The LAIA Database – A Secure and Decentralized Platform for Environmental Data

Center for Social Ecology and Free Technology
SEFT
seft@riseup.net

***Abstract*—The current centralized storage of data in private servers exposes the whole society to the threats of surveillance, censorship and malicious attacks coming from different sources. Under this context, the environmental and climate data are also a target of these threats, therefore a shift on the topology of data storage and sharing is needed. The present article proposes the LAIA Database, a decentralized and encrypted platform focused on environmental and climate data that provides a total control over the data by the users. LAIA is a friend-to-friend network (F2F), which only allows access to the data by trusted nodes on the network. The project aims to improve the research collaboration between groups as it allows to share important information between trusted parts, while it protects both research groups and their data against censorship, persecution and other forms of political retaliation.**

Introduction

The concepts of privacy, anonymity and digital self-defense are not only restricted to activists and groups directly involved with digital rights and other political movements. Since the surveillance apparatuses of the state were revealed in detail by Edward Snowden, many social movements started to deal with technological devices in a different way. Back then, the massive storage of data in a centralized database under US jurisdiction came out to the public spot.The issue of mass surveillance not only affected citizens but also other nation-states depending on USA-based infrastructures, which eventually turned out into an international crisis (Sargsyan 2016).

Recently there has been a rise of critical voices against investments in environmental and climate research programs. For some of them, environmental policies are seen as a threat to industries and the development of the country, this is something on which many politician base their rhetoric (Trump, for instance, as a climate-change-denier based his campaign on this issue with the slogan "make America great again"). Once in power conservative politicians supported by climate-change-deniers groups has moved from a mere rhetoric into political actions across the globe (De Pryck and Gemenne 2017). This new political scenario resulted in a strong reaction from researchers and activists that started to take some measures in order to protect the data from being altered, deleted or removed (an example of this measures was the mass-download of US government climate data before Donald Trump's administration took over). Nonetheless, what it is seen is a shift of data storage from US-based servers to servers under different jurisdiction but owned by private groups and big corporations keeping high centralization of the knowledge on the Web in a few large providers (Halpin and Monnin 2016). Hence, although this action intends to rescue the data, its strategy is very ineffective as the political scenario is constantly changing in accordance to centralized interests.

A more efficient and low cost way to protect the data is based on a real paradigmatic shift on data storage, from a centralized and vulnerable structure to a decentralized and serverless architecture. This is a path to recover the autonomy and control over environmental data by reducing the risks of data loss, the single point of failure and the exploitation of centralized data controlled by corporations and governments. This distributed structure is also social, where the autonomy is associated to the strength of the social bonds and the flux of information. Even though there is already some decentralized projects available (e.g BitTorrent, git, etc), none of them was developed focusing on climate and environmental data.

The current improvement of environmental data generation, from universities and research institutes to citizen scientific labs and collectives, contrasts with the lack of an integrated platform infrastructure where the data is published and the communication between different groups involved with the environmental monitoring is enhanced. Concerning to the aforementioned limitations and vulnerabilities of the mainstream database structures, this project aims to develop an accessible and decentralized database based on trustworthy networks, privacy and anonymity. Under this structure a central service is not required and if a node is under attack, the other parts of the network can work without interruption.

Serverless means that each node is server&client at the same time, since no external service is needed for the network to function. Thus, it's based on the principle of a Network of Participation, which means that the data will never disappear as long as there is a network of people interested on keep it available in their computer. Moreover, with the multi-source sharing (swarming) and the multi-hop a small number of hops among nodes is expected to access files resulting on faster downloads (low latency).

The Database Architecture

The LAIA Database is a cross-platform database focused on the user experience in order to build up a Web of Trust between peers. It integrates a strong encrypted protocol with usability and gives a total control over the data by users. In the decentralized structure of LAIA database the data is hosted in swarms, where a user who join the swarm become a peer and can download pieces of the data from other node since establish connection to at least one trusted peers in the swarm. In return, the user uploads pieces of data they already have to trusted peers who need it. Once the user have all of the data, they can choose to remain in the swarm and continue sharing with other trusted peers, which makes them a seed. Accordingly, in this network of participation the structure the more popular the data the bigger the swarm and the faster the downloads.

LAIA database a private peer-to-peer (P2P) in which each network node exchanges data only with a designated list of "friend" nodes. Also known as Friend-to-Friend (F2F), this network topology builds its overlay on top of pre-existent trust relationships among its users. A digital life is a dimension of real-life and this is a very important base of LAIA's principles where only trusted nodes establish connection among them, resulting in a significantly more secure network. With F2F it is possible to ensure a cryptographically secure connection established between nodes that had met each other in a past experience outside the platform. Thus, unlike P2P network, with F2F it is possible to control who accesses each others data through encrypted connections between trusted parts that had already shared secrets needed to establish these secure connections.

Related to the search for files, it is possible for nodes to look for files accessing a large network even when still only connected to their trusted nodes. The trusted node's search works by propagating search and file transfers recursively to all friends of friends to a certain degree of separation. This topology ends up in a non-homogeneous structure with regard to the numbers of nodes and bandwidth (Fig. 1), where the swarming structure ensure either technical or legal means to overcome traditional ways for limiting the access to the Web (Fig. 2).

The LAIA database uses the reputation scores to give a general overview of the peers' level of participation in the network. The role of this tool is to protect nodes against the problems of P2P network with spams and malicious nodes, preventing disturbing data spreading without control. Reputation score is accounted for each node considering negative score as bad, positive as good and zero as neutral. If the score is too low, the identity is flagged as bad. During the bootstrapping process for joining the P2P network, peers can potentially use reputations to decide who to directly connect to in the overlay topology.

Security and Privacy

The creation of a profile is actually the generation of a PGP key, where a passphrase is used to to validate the profile. At the login, the PGP key is used to decrypt the encrypted SSL certificate and its passphrase is read. The SSL passphrase is chosen randomly when creating the location and encrypted using the users' PGP key.

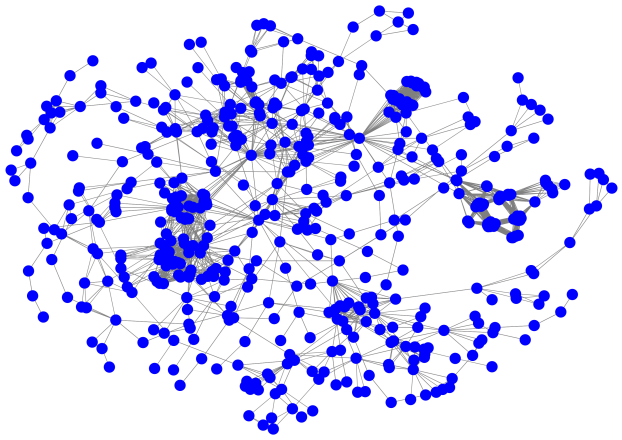

Fig. 1. Friend-to-Friend: The network of no-homogeneous networks. As the swarming is built up based on trust between nodes, the networks might vary in matter of bandwidth and size

The data is identified by their specific fingerprint, and unlike other decentralized projects (e.g. IPFS) it is also possible to look for files by their cryptographic hashes, names or even by their sizes. The security relies on cryptographic algorithms in which the connection between nodes are encrypted using SSL, while the identity of node is represented by its PGP key. Once the software is installed and started for the first time, a profile and a certificate are automatically generated. That certificate, in turn, is used to authenticate and connect safely with the trusted node to include in the network.

Once a folder is added to be shared, all the files present inside it will be hashed (i.e. there will be created a fingerprint for each file) before being shared by the node. This folder can be shared as "browsable", in which only trusted nodes can see and download the shared files, "network wide", where friends of friend can download the files via anonymous tunnels. In this case, the friends of friends nodes are not capable to see which files are being shared and the node that is the source of these files, but they can find the files using the search.

a) Using Multiples Accounts For a Group: LAIA allows to use an existing PGP key for signing multiple identities, and thus makes team work easy. As detailed before, the PGP keys will be used for encrypting the SSL passphrase on disk and to sign the location's SSL certificate.

b) Files Hashing: A way LAIA stores files is versioning them through the 40-character secure hash algorithm SHA-1. In 2017, some concerns came out since researchers had announced a hash collision of the SHA-1. A hash collision occurs when two separate inputs produce the same hash result output.

Although a has collision is unlikely to occur , there is an important difference between using cryptographic hash for

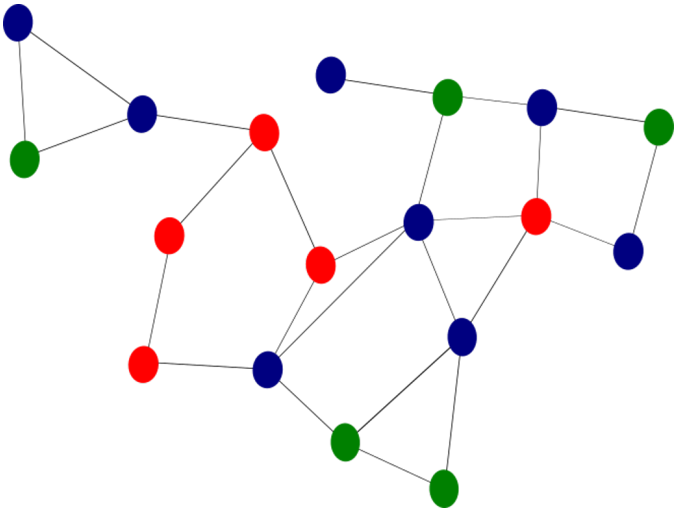

Fig. 2. Privacy and distribution of data throughout the network: Only trusted nodes (blue dots) can access the data from the source (green dots), while nodes "friend-of-friend" (red dots) can't see the data

security signing, and using it for generating a file content identifier. In the case of a hash used for security it is used to trust users in order to protect people using a given platform from malicious users. On the other hand, in content-addressable systems, like in LAIA, the hash for content identifier is not used to trust users but, as previously said, for file content identification.

c) Estabilishing a secure connection: LAIA's privacy is protected with anonymous tunnels. The swarms often consist of computers distributed around the world (in which national laws cannot actually achieve the censorship), and its structure is independent from a centralized server controlled by corporations or nation-states. Thus, there is no single entity to sue or pressure financially, and the data sharing overcomes the infrastructure jurisdiction and implication in giving users data to intelligence agencies.

The secure connection of two nodes is done using the PGP signature that authenticates SSL links between them and only the trusted nodes can access details like each other's IP addresses, which files are being shared, etc. One way to improve the privacy and anonymity is using the Retro-Tor tool that make it possible to run LAIA over the Tor network, where IP is not visible even to connected nodes. An advantage for users is that it creates a hidden node with one click away that enhances the secure data sharing, and overcomes digital embargo or censorship. This is an important tool for activities that require a secure and private contact with a trusted persons.

CONCLUSION

The LAIA database proposes a new efficient, strong and secure alternative to the centralized structure of environmental data storage with many advantage comparing to both centralized and decentralized platforms. The motivations of the project, based on autonomy, social participation, and freedom of speech are align to the essential principles for the improvement of researches and the social role of science.

The reputation system, as already pointed out by other projects, is an effective system of self-management, since the majority of the nodes in the networks are interested to maintain the system stable and free from trolls and distractions. Furthermore it is faster, safer and ensures that the data will never disappear as there is at least one node hosting it. The data is private until a node allows a trusted one to access to its data. It is also easier to overcome the problem with human friendly names for files and enhances privacy using Tor network.

The F2F network used by the project is a very strong concept which allows private sharing of data, and protects either users hosting data or accessing it. Thus, in LAIA Database the control of access to the file is not used to restrict the access of information, but to ensure that researchers groups, collectives connected to the network, as well as the knowledge itself are safe. It allows data sharing between trusted nodes by using a free and open source network with high scalability gathering universities, citizen scientists and stakeholders promoting their participation and cooperation in the network.

References

De Pryck, Kari, and Franois Gemenne. 2017. "The Denier-in-Chief: Climate Change, Science and the Election of Donald J. Trump." *Law and Critique* 28 (2): 119–26. https://doi.org/10.1007/s10978-017-9207-6.

Halpin, Harry, and Alexandre Monnin. 2016. "The Decentralization of Knowledge: How Carnap and Heidegger Influenced the Web." *First Monday* 21 (12). https://doi.org/10.5210/fm.v21i12.7109.

Sargsyan, Tatevik. 2016. "Data Localization and the Role of Infrastructure for Surveillance, Privacy, and Security." *International Journal of Communication* 10 (0): 17. https://ijoc.org/index.php/ijoc/article/view/3854.