IT Security in research support: The MyDisk platform at UPC

Cyber Security Month, September 2022

IT security concerns are nowadays well-known for everybody. The Internet network has provided in a few years an unprecedented boost of possibilities, but also a myriad of dangers and threats in this new digital world.

In research support, national or European projects as well as technology transfer activities, this issue is usually feared and sometimes, at least, partially ignored because of their complexity: GDPR[3], Open research and FAIR principles, Data Management Plan or Ethical committees. And also, Digital sovereignty concerns: vendor locking vs. free software, independence, trustworthiness, or the public-nonprofit transparency society commitments.

In this text we will share our experience in research support at the /rdlab – UPC, hoping it will be useful.

Gabriel Verdejo Alvarez
(gabriel@cs.upc.edu)

is the IT manager at the Research and Development Lab (*/rdlab)* since 2010 and full-time staff at UPC (*Universitat Politècnica de Catalunya*), Computer Science Dept (*CS*) after 2003. Previously, worked as a senior IT consultant for innovation and database processes.

The /rdlab provides specific IT support for the UPC research groups, fellow universities and research centers in their national and European projects in order to foster their technology transfer initiatives.

1. Researchers and the research process in a public institution

In order to understand our proposal, the MyDisk Platform[1][2], we need some context about where we come from. Research, unforeseeable by its nature, it is not a standard process and requires a high degree of freedom and flexibility. Unlike other regular processes and activities (student enrollment or exam qualification), the research needs are not predictable and will evolve with the project progress.

Furthermore, every researcher/group/partner has its own needs and "work methodology", so, a one-size-fit-all solution or a too rigid environment will deliver negative effects in the project outcomes. For example, some researchers can work overseas (different timezone), or in a restricted environment (China), or has strong security policies (hospitals, military research centers or hi-tech companies).

With these (and some more) considerations in mind, we designed a solution for our research groups that ensures that researchers can focus their time and efforts in the research process itself, and should not spend time looking or configuring external IT infrastructures, or complementary services not related to the project itself.

As a public institution, the UPC[4] is strongly committed to transparency and public service values, encouraging the usage of free software in all their areas following the guidelines of European initiatives like FOSSEPS[5]/EU-FOSSA-2[6]. In addition, free software allows access to the source code which provides additional benefits and possibilities to enhance security and transparency, for example[7][8]:

- Source code can be audited in order to understand and confirm what is really doing
- Source code can be checked and tested in order to find bugs
- Updates and security patches are always available to everybody
- Source code lets you contribute/adapt/modify/fork the project according to your needs

2. The MyDisk platform

The MyDisk service provides a safe and easy-to-use web-desk environment where researchers can work standalone or in a collaborative mode with their partners using just a browser (see figure-1). You can store, access, edit and share your data worldwide. Access our source code Gitlab service, safe real-time chat and videoconference meetings, project tracking (Agile Kanban) or manage your digital secrets.

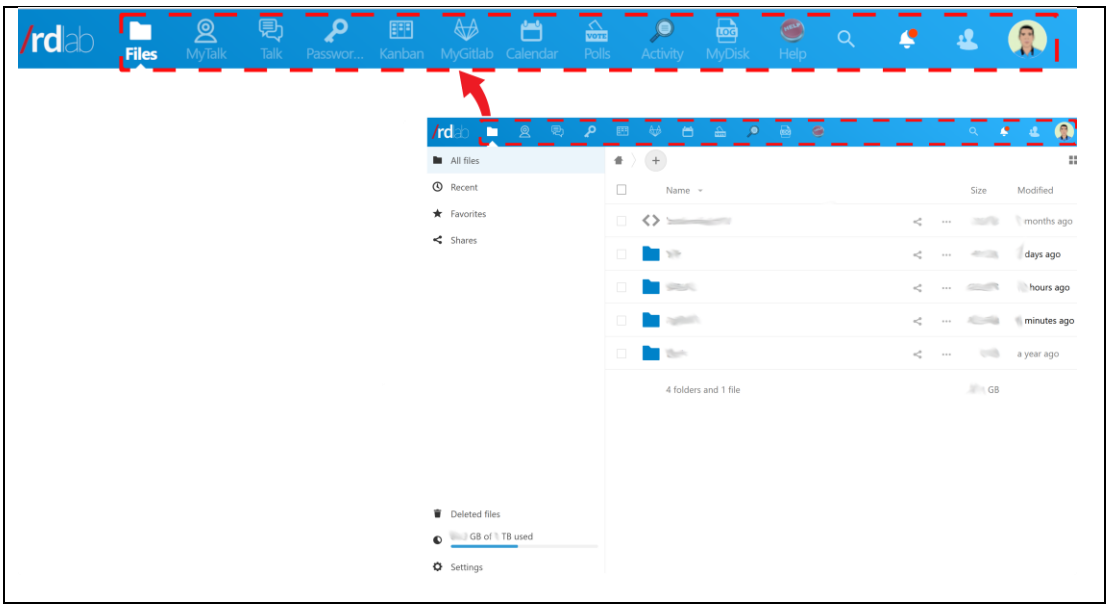

Figure-1: The MyDisk web desktop dashboard

All our servers and services are located in Barcelona, at the main UPC Datacenter facilities[9]. A 250m^2 TIER II+ ANSI/TIA-942 certified space, with a dual cooling system, redundant power generator and a 24x7 monitoring service provided by UPCnet.

The MyDisk platform integrates several free software projects in a single web dashboard, offering a virtual desktop environment to work autonomous regardless your conditions. We proudly use these free software projects in our systems and services:

- *GitLab* [10]: Source code repository
- *Linux*[12]: Operating System
- *Jitsi*[14]: Videoconferences
- *NextCloud*[16]: Web access and file management
- *OpenNebula*[18]: Cloud services management
- *HAProxy*[11]: HA and load balancing (web)
- *Lustre/ZFS*[13]: Storage management
- *MariaDB*/Galera Cluster[15]: Database services
- *OnlyOffice*[17]: Online collaborative web editing
- *ProxySQL*[19]: HA and load balancing (databases)

3. Security in MyDisk

A centralized service makes easier to deploy common security policies for users and services, simplifying the protection model. Nevertheless, you have a single point of failure[20] who will receive most of the attacks and security problems (e.g. brute force[22] or denial of service[21]). Moreover, in research we deal with sensitive data, therefore, we have to enforce data Confidentiality, Integrity and Availability (CIA triad[23]).

In order to decrease threats, and to strengthen our platform as much as possible, we mainly address IT security issues at two levels:

A. Architecture layer: This level involves the design area, defining how the hardware and the services are connected and how will operate (see Figure-2).

Our network is segmented by service areas, isolating every function/service from the rest and forbidding access outside their scope. A perimeter firewall filters the Internet requests and a local firewall at each server ensures minimum damage policies.

Every service is executed in a private cloud environment, making good use of virtualization capabilities: isolation, scalability, availability, redundancy and flexibility. Our storage service is connected only to the system hosts, using a local bypass system VirtioFS[24][25] to ensure that every VM can access only their required "slice" and keeping a good security-performance balance.

Our disk system relies on a parallel high performance network filesystem (Lustre[13]) with a ZFS[26][27] backend. Unlike most filesystems, ZFS stores a checksum[28] for every datablock guaranteeing the data integrity and self-healing capabilities. In addition, our system keeps 3 exact copies of every block for quorum and performance sake using our triad system[29][30].

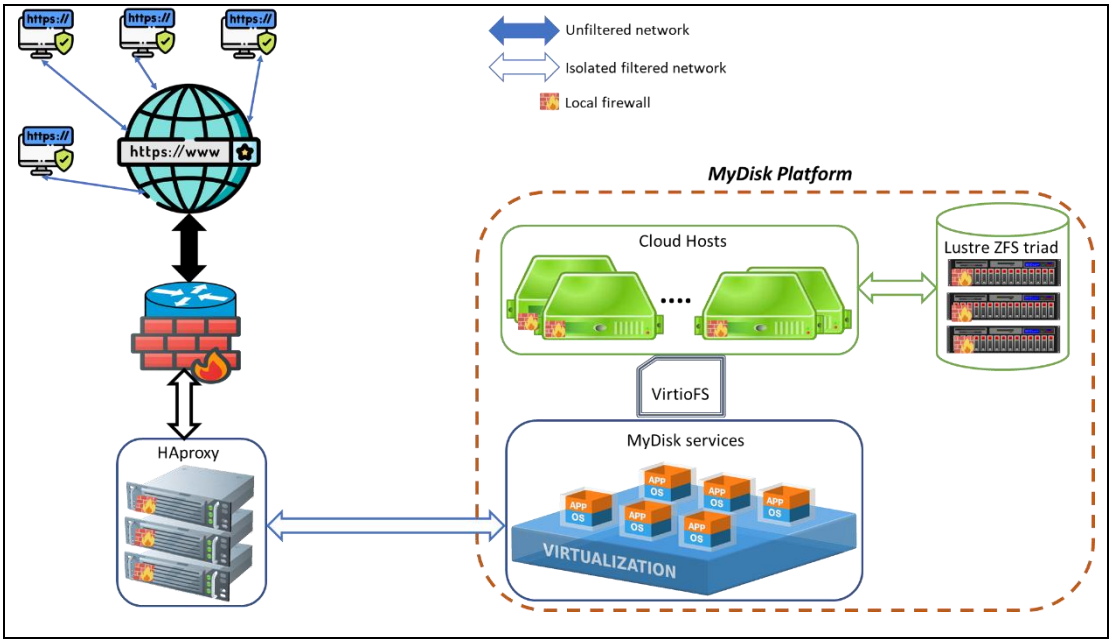

Figure-2: The MyDisk architecture layer

B. Service layer: This level provides a secure user access and enforces good security behaviors inside the platform (see Figure-1).

MyDisk services are only available through a secure web access (https), providing universal connection and an open device policy. Researchers can work with just a browser avoiding extra complexity like VPN setups and preventing the use of insecure alternatives like SMB, NFS or VNC.

Our web service has a self-defense protection system against common attacks, like brute-force or random scans[31]. Any DOS attack is minimized by our load balancing service spreading the requests among several VM instances and through the countermeasures provided by our ISP (UPCnet, CSUC and RedIRIS).

The user access service supports MultiFactor/2FA[32][33] authentication through email pin, TOTP applications, physical token devices and user's one-use pre-generated code. Besides, the user authentication can be federated with renowned projects (SIR, UNIFICAT, eduGAIN…).

In order to promote good security behaviors and increase the overall platform security, MyDisk also offers a password and digital secrets (certificates, tokens…) management service:

- A trustworthy centralized place to control all your authentication credentials and a safe new password generator. Additionally, automatically checks for weak or compromised credentials[34] warning the user if necessary.

- A safely way to share passwords among team members avoiding unsafe practices like sending clear text passwords by email or using chat messaging apps. MyDisk forbids group password sharing because group members could change over time, concealing who had access to a specific credential and worsening security leak tracking.

We want to stress that this service always uses a safe random per-user generated unique key to encrypt data credentials[35] (even in the database tables!) guaranteeing privacy against database dumps or malicious system administrators.

For security's sake, direct data sharing capabilities are restricted only between team members (group A members cannot "see" group B members unless they share a common group). For external shared data, we enforce security using date expiration, password protected, read or write only access options.

Mydisk allows private internal communication for group members using an instant messaging service and videoconferences with external users using just a browser or specific app clients available for any OS.

Regarding users and data access accountability, we log every interaction for a full month due the large log file size. Every researcher can review their own actions and the activity on accessible shares, defining to be noticed on specific events. For special audit requirements, we can also provide a lifetime monitor log service (see activity and MyDisk log, Figure-1).

A regular daily backup is made every night, but in addition, MyDisk offers a built-in file control version management system that automatically keeps versions of every modified file using user's quota free space[36]. This way, every user can recover by itself accidental modified files or malicious ransomware attacks.

4. The future of MyDisk

Our platform was designed to provide a safe, flexible and useful workspace for researchers, therefore, its evolution evolves along the researcher's needs and requirements. Immediate future plans include performance improvement and increasing storage space. Additionally, user authentication and data shared federation with other fellow research services has been successfully tested, opening a path for future European collaborations.

Do not hesitate in contacting us for further details, we will be glad to hear from you!

References:

[1] https://rdlab.cs.upc.edu/MyDisk/
[2] https://rdlab.cs.upc.edu/wp-content/uploads/2020/04/Tecniris-Teletrabajo-coronavirus.pdf
[3] https://gdpr.eu
[4] https://www.upc.edu/en/the-upc/the-institution
[5] https://joinup.ec.europa.eu/collection/fosseps/about
[6] https://joinup.ec.europa.eu/collection/eu-fossa-2
[7] https://www.gnu.org/philosophy/free-sw.en.html
[8] https://rdlab.cs.upc.edu/free-software/
[9] https://rdlab.cs.upc.edu/our-facilities/
[10] https://about.gitlab.com
[11] http://www.haproxy.org
[12] https://en.wikipedia.org/wiki/Linux
[13] https://www.lustre.org
[14] https://jitsi.org
[15] https://mariadb.com/kb/en/what-is-mariadb-galera-cluster
[16] https://nextcloud.com
[17] https://onlyoffice.com
[18] https://opennebula.io
[19] https://proxysql.com
[20] https://en.wikipedia.org/wiki/Single_point_of_failure
[21] https://en.wikipedia.org/wiki/Denial-of-service_attack
[22] https://en.wikipedia.org/wiki/Brute-force_attack
[23] https://blog.itgovernanceusa.com/blog/how-nist-can-protect-the-cia-triad-including-the-often-overlooked-i-integrity
[24] https://virtio-fs.gitlab.io
[25] https://archive.fosdem.org/2020/schedule/event/vai_virtio_fs/
[26] https://zfsonlinux.org
[27] https://en.wikipedia.org/wiki/ZFS
[28] https://openzfs.github.io/openzfs-docs/Basic%20Concepts/Checksums.html
[29] https://rdlab.cs.upc.edu/rdlab-at-the-lustre-user-group-conference-lug2021/
[30] https://rdlab.cs.upc.edu/wp-content/uploads/2021/05/Abstract-TriadBased-architecture-rdlab.pdf
[31] https://www.fail2ban.org/wiki/index.php/Main_Page
[32] https://en.wikipedia.org/wiki/Multi-factor_authentication
[33] https://rdlab.cs.upc.edu/two-factor-authentication-2fa-on-mydisk
[34] https://haveibeenpwned.com
[35] https://rdlab.cs.upc.edu/wp-content/uploads/2022/09/PM_MyDisk.pdf
[36] https://docs.nextcloud.com/server/latest/user_manual/en/files/version_control.html