REMLABNET – BaaS, Backup as a Service in Remote Laboratories and Increase Sciences and Research’s Data Security Precautions against Ransomware

Pavel Beňo¹, Sandra Šprinková², František Schauer³
Trnava University in Trnava¹,
Hornopotočná 23,
91843 Trnava, Slovakia,
Trnava University in Trnava²
Faculty of Education, Priemyselná 4,
91843 Trnava, Slovakia,
Tomas Bata University in Zlin³
Faculty of Applied Information,
Czech Republic,
e-mail: pavel.beno@truni.sk¹, sandra.sprinkova@tvu.sk², fschauer@fai.utb.cz³

DOI: https://doi.org/10.15414/isd2022.s5-2.02

Abstract
The article focuses on the importance of the REMLABNET system, which deals with the backup of scientific research and student data. The system helps scientists keep their data safe from destruction or theft. The article also discusses the introduction of the REMLABNET Cloud computing service called Backup as a Service (BaaS). The service provides data backup for researchers, teachers, and students. Every part of the university will find applications for the security of its data here. The service contains a large number of battles with Ransomware, right at the level of disk arrays. Like everything in CC-REMLABNET, we provide these services through the list of services listed on our website. Here we present the Beta version of our new portal for access to our laboratories and also the services that REMLABNET provides.

Keywords: remote laboratories, virtual laboratories, backup, services, cloud computing, education, REMLABNET, science

JEL Classification: L86, D85, L63

1. Introduction
We have been dealing with the topic of remote laboratories (RLs) for more than a decade, and during our career we have already researched and created a lot in this area. This is one of the reasons why it probably doesn't make sense to describe the issue itself, as it has become an ordinary part of our lives. We meet her especially in the distance education of students, which exacerbated the pandemic covid19 and thus pointed out the need to conduct lessons (lessons) in a different way than the school was used until recently. Today, RLs are dealt with by more and more institutions around the world. The traditional way of teaching process had to be changed from the ground up, and as teaching methods changed, so did RLs. They are constantly evolving and are gaining more and more popularity.

Our new structure has changed radically compared to the flood design. We have moved from the usual RLs solution (described here (Schauer, 2014)) to a fully cloud solution, which is a priority for us. This is because the cloud solution allows the migration of individual RLs across connected data centres (DTCs). The floods aimed at integrating all RLs into the cloud, called Cloud computing REMLABNET © (CC-REMLABNET ©), are coming true and becoming a reality. Figure 1 shows our proposed timeline for integration and creation of a new solution.
When we look at the timeline of what was built, we can see the progress in RLs. In the first place, it is obvious that the main functionalities of individual RLs are not changing, but some functionalities, ensured by the cloud, are added. First, of all, REMLABNET (without cloud) provided functionalities, summarized below (Panel/Block 1). In 2013, we started installing cloud computing and placed our REMLABNET in a virtual cloud (Panel/Block 2), providing basic IaaS, PaaS and SaaS services (Panel/Block 3). Later, we begun creating additional services and options that our cloud might provide, and we built many extended options. According to our experience, the real interest is only in three, the most commonly used: Remote Laboratory as a Service (RLaaS), Simulation as a Service (SIaaS) and Storage as a Service (STaaS). In this paper, we will focus primarily on STaaS technology and its modification for Backup as a Service (BaaS). The virtualization and cloud solutions are the main topic of IT technologies nowadays, when servers, networks, disk arrays, individual Luns on the storages are being virtualized. The trend touches also the field of RLs (Panel/Block 4). We have to conclude that at present, the approaches in RLs world community, takes similar trajectory of building RLs, to provide RLs as service (Tawfik, 2014) (Garcia-Loro, 2019).
No matter how the structure changes, it is necessary that the main functions and advantages of using RLs remain. For that, summarizing, the functions of REMLABNET are:

- To embed individual RLs and provide their functioning,
- Enable Diagnostics of RLs,
- Enable communication in the virtual classes,
- Enable connectivity with another RLMS,
- Provide connectivity of clients from arbitrary locality,
- Provide unification of interfaces,
- Provide booking system,
- Ensure storing of measured data.

When we focus on integrating the RLs itself into the cloud, we must realize that there is no longer a service computer for each experiment, but all the software components run in a virtualized cloud storage. An example of this is in Figure 2, where you can see that there are really only HW devices such as Camera and converter, respectively LAN connection interface, in the experiment itself.

![Figure 11 - Basic scheme of the server-client communication in RL embedded in cloud CC - REMLABNET©](source)

A description of the connection between CC-REMLABNET © and the expert himself is beyond the scope of this publication. In short, we will at least state that the main cloud blocks for this purpose are shown in Figure 3. On the fig 3, we see how the user comes to the CC-REMLABNET © portal, which serves as a user interface. Subsequently, using the FlexConnect block, it is redirected to selected RLs that no longer need service PCs, but the control software is located on separate virtual machines (VMs). Currently Remlabnet (https://remlabnet.eu) is undergoing reconstruction and its new version is located on the domain https://remlabnet.truni.sk, where the individual RLs are divided into logical blocks:
1. Virtual laboratories - where there are completely virtual or simulation tools that help in the teaching process. For example, simulation of a rocket flight with the possibility of adjusting the mass and fuel, where the student sees how to model the model in the gravitational field of the earth (flight height), Mendelian periodic table of elements showing the atomic structure for each element (number of electrons in the envelope,…), Solar system, where real data and orbiting planet orbits are used with the ability to track the difference between the orbits of the two planets, and more.

2. Physical laboratories - where RLs are located on the basis of ISES (Schauer, 2008) but also the newly supported Poseidon, Arduino or Rasbery PI systems. Examples are electromagnetic induction, measuring weather conditions in the city, measuring heat transfer on various materials, and more.

3. Experiments according to the curriculum of the Ministry of Education - where we have created and are constantly supplementing the individual RLs needed for teaching physics and science at primary and secondary schools in the Slovak Republic and the Czech Republic. This section is governed by approved curricula for each class of individual schools.

Fig 3 - FlexConnect in block diagram of the CC-REMLABNET©

The whole description of the cloud solution design is also described here (Schauer, 2020). In this paper, we want to primarily deal with the service that the cloud provides us automatically for REMLABNET clients - Backup as a Service (BaaS). It is no coincidence that we have worked on this issue. In the past, we have created both the possibility of establishing research activities and also the possibility of providing disk space for clients of the REMLABNET, like advanced technology in our cloud. Advanced technology has also invited some severe security threats, like the cyber threat. Moving your data to the cloud is considered to achieve their maximum safety (Ismail, 2018). Cloud storage refers to the online space where data can be stored. This online space resides across multiple remote servers housed inside massive DTCs, built on virtualization technology, which means that storage space spans across different servers and possibly even multiple and geographically different locations.

The Storage as a Service (STaaS) takes care about storing all generated data both by measurements, calculation and modelling on one side and control traffic data on the other. The
added value is a possibility to distribute the cloud storage locally to even very remote localities and thus suppress vulnerability of both storage HW and data stored.

Every day, internet users create trillions of bytes of information, an inconceivable mass of ones and zeros, that are then stored in the cloud computing farms run by the world’s biggest companies (Decker, 2018). In the REMLABNET is situation “little bit” better, but our clients are producing data every day in measurements, or they are sharing the data. In addition to their processing, this data must also be stored and, of course, backed up, which has a great impact on storage capacity.

2. Purpose
The proposed solution has a very wide application. First of all, we need to realize where the RLs are going, and that is education and, of course, research. As far as school activities are concerned, it is, of course, necessary to back up the various works of students. Within the RLs, these are mainly measurement protocols and submitted laboratory work - measurement outputs. Due to the legislation, it is necessary to maintain these parts of teaching for at least 5 years. It is useless for such work to be accessible on fast disk arrays, taking up space for applications and more up-to-date data. For this reason, the system is designed so that the current data is located on fast disks together with applications running under the university DTC - Fig. 4. The same is true for research activities, research data and development work not only in the development and development of REMLABNET, but also within the whole university with running applications. As part of the University of Trnava in Trnava, all research teams use our proposed concept.

The data collected by the application and REMLABNET itself are stored and worked on fast SAS disk arrays, which are then backed up to slower SATA disk arrays. Backup as well is unnecessarily described. They are clearly among the basic issues of IT issues. An incredibly large amount of time and energy is devoted to him. But why is that so? The answer is always clear. The need to go back to the data in time. There may be a number of phenomena that affect our stored data. On the one hand, there is a possibility that we will damage them ourselves, whether by careless handling or simply deleting something that we find useful. For example, the measurement of electromagnetic induction from 2019 (Beňo, 2019). The second reason for having a backup is disaster recovery. This is an external intervention, where current data can be damaged by equipment failure or hacker attacks. Both options are very common.

Fig 4 - Distribution of the disk arrays for REMLABNET applications
Source: Author
In general, we can say, that data backup is a process of duplicating data to allow retrieval of the duplicate set after a data loss event. Backups have two different purposes. The primary goal is to recover data after it is deleted or damaged. The secondary goal of the backup is to restore the data from the previous period in accordance with the determination of the data retention policy that is normally configured in the long-term backup application for how copies of the data are needed. Although backups are a necessary form of disaster recovery and should be part of the REMLABNET disaster recovery plan.

In forming the scheme of the storage we stick to the rule „2+1= Data backup best practice“, that means for critical data of all activities we should make two full copies, maintained on separate physical devices. Of course, this only applies to data marked with priority “1”. Our systems adhere to the university’s security guidelines, which specify the priority of systems and their relevance. In addition, a third copy should be kept offline, preferably situated at another location.

Let us look at how we archive the data of REMLABNET. Data archiving is the process of moving data that is no longer actively used on a separate long-term storage device. Archival data consists of legacy data that remains relevant to the organization or must be retained for future use or for regulatory reasons. Data archives are indexed and have search capabilities, so needed files can be located and retrieved (Rouse, 2019).

At this point, we need to mention one important thing and the crucial thing that is going on. Ransomware. Ransomware knows about everyone. It is quite possible that he met him in person when he found the encrypted data on his PC in the morning. Ransomware is a type of malicious software that is used by cybercriminals to block the work of an infected system. It attacks both local PCs and large systems. We have already encountered encrypted data found in the HyperV environment. However, this publication is not aimed at ransomware and how to defend against it. A number of publications have already been published about this, and some of them are very good and, in addition, offer some solutions to problems. For example (Eset, 2022) (Kharraz, 2018). We also want to point out how the backup system can protect our data from Ransomware.

3. Methods
We have described and presented ways of connecting and creating backup snapshots in our publication ROAD TO STRENGTHEN OF VIRTUAL INFRASTRUCTURE AND SECURITY OF REMOTE LABORATORIES ON TRNAVA UNIVERSITY IN TRNAVA (Schauer, 2020). In short, we can say that the snapshots that we create are set up so that they save a huge space on disk arrays and their creation and restore are much faster, because it does not copy such a huge amount of data.

In the first phase, our NetApp FAS2700 series data storage completely replaced three old data storages from various vendors. Because it natively allows the combination of block and file protocols on a single disk space, it can be very flexibly integrated into the infrastructure. Flexibility plays a huge role in creating data backup and archiving. Scientific data needs to be backed up at the lowest level. We are helping our scientists not to have to think that it is necessary to solve some backup, thus relieving them of unnecessary effort and so that they can devote themselves fully to their work. It is the same with student data, as students back up data in this way throughout the semester. Thanks to the ability to create a Storage Virtual Machine (SVM), we can provide each application with the necessary protocol, dedicated logical interface (LIF), resources and define Quality of Service (QoS). Subsequently, simple SVM /
LIF transfer between individual controllers or between multiple disk arrays within a cluster allows us to implement maintenance and future technological upgrades without the need to interrupt service provision. Moving data to faster layers of the cloud is seamless and transparent to applications. The integrated Non-Volatile Memory Express (NVMe) cache on the controllers acts as an excellent accelerator for IO operations, which were insufficient for the original data storage and slowed down the operation of applications.

Thanks to the new ONTAP® storage operating system, we can automatically create images with virtually no restrictions and without negatively affecting performance (1023 images per volume). A copy of a snapshot is an image of a file system at a specific time. Low-overhead image copying is possible thanks to the unique features of Write Anywhere File Layout (WAFL®) storage virtualization technology, which is part of Data ONTAP®. Like a database, WAFL uses pointers to actual blocks of data on disk, but unlike a database, WAFL does not overwrite existing blocks; writes the updated data to a new block and changes the pointer. The NetApp Snapshot copy simply manipulates block pointers and creates a "frozen" read-only view of the WAFL volume, which allows applications to access older versions of files, directory hierarchies, and / or logical unit numbers (LUNs) without special programming. Because the actual blocks of data are not copied, copies of the slides are extremely efficient in terms of the time required to create them and the storage space (NetApp, 2011).

The creation of the snapshot itself is in Figure 5. The snapshot is captured in Figure 5. In Figure 6, the changed data is written to the new block and the pointer is updated, but the snapshot pointer still points to the old block, giving you a live view of the data and a historical view. The next snapshot will be created in Figure 7 and you now have access to 3 generations of your data without taking up disk space, which would require 3 unique copies: live image, snapshot 2 and snapshot 1 in order of age.

![Figure 5 - Take a snapshot (resource NetApp)](resource NetApp)

*Source: NetApp Snapshot Technology, Datasheet DS-2477-1111, 2011*

![Figure 6 - Write data to a new block (resource NetApp)](resource NetApp)

*Source: NetApp Snapshot Technology, Datasheet DS-2477-1111, 2011*
Figure 7 - Create another snapshot (resource NetApp) 

Snapshot technology forms the basis of a unique ecosystem of high-availability, disaster-tolerant, and data protection solutions. The great advantage of setting up our snapshot system is that each user can manage their snapshots and their backups.

Ransomware protection is also a very important part of this technology. We have been dealing with the fight against Ransomware since the very beginnings of this technology. Much of our research focuses on cloud services, as we mentioned above. One service that we tried a long time ago was also Ransomware as a Service (RWaaS), but we also cancelled this technology after the deployment and creation of the service, because we do not consider it moral to offer another such tool. There is more scope and technology to combat Ransomware, but we are already working at the disk array to achieve data security. Gartner also has an interesting look at this:

Gartner’s requirements of next generation cyber storage: “By 2025, 40% of all enterprises will require storage products to have integrated ransomware defence mechanisms, up from 10% in 2021”.

The systems we use currently have basic anti-Ransomware features:

- Real-time detection using file system analytics and file activity entropy calculator
- Upon ransomware detection:
  - Alerts administrators to attack
  - Triggers automatic snapshot, to minimize damage and simplify recovery
- Included in new Anti-Ransomware Suite

Gartner also defined requirements of next generation cyber storage, which consist of mainly (figure 8):

Identify

- Assess your data protection and security posture
- Classify type of data, location, and permissions

Protect

- Block malicious data from being written to disk
- Create granular, immutable copies to thwart infection
- Prevent data deletion with indelible data copies
Detect

• Monitor user behavior for suspicious activity
• Detect storage behavior anomalies

Respond

• Initiate NetApp® Snapshot™ copies if an attack is identified
• Block malicious user accounts

Recover

• Restore data in minutes and bring applications back online
• Apply intelligent forensics to identify the source of the threat

Figure 8 - Gartner’s requirements of next generation cyber storage (Identify, Protect, Detect, Respond, Recover)
Source: Author

4. Conclusion

Our idea of using cloud computing has been confirmed and discussed with experts in this research section. The way we work is good and has made great progress. We can provide the new Remote Labs as a Service (RLaaS) service in our cloud system. Our clients are primary school teachers, students and the brainwashing of universities and high schools, but access is possible for all consumers via the internet. This shows how heavily the university network is in terms of communication and transport. This states that the network and all parts of the IT structure must be free of failure and latency. And stay safe for research data management and protection.

Snapshots allow us to efficiently back up, go back in time with unwanted changes with extremely high granularity, and last but not least, we have a tool to easily recover from a possible ransomware attack (SnapShots are ransomware resistant), because NetApp Copy Snapshot is a read-only, static and fixed copy. Using SnapManager, we can set up an automatic backup cycle, application runtime, and retention policy. The high performance, scalability and stability of NetApp Snapshot technology means it provides the ideal online backup for user-driven recovery. Other solutions allow you to copy backups to disk offline or to tape and archive them.

This is really the most reliable ransomware protection we’ve found in REMLABNET and universities. Our Consortium want to continue with research of the increase data protection and bring absolutely or best security of data.

Acknowledgement

The support of the Eunis Slovakia is acknowledged.
References

[1] Beňo P., Cloud computing solutions and security of EU remote interactive laboratories network, Univerzita Tomáše Bati ve Zlíně, Zlín, 2019


