UDC 621.39

RAICS AS ADVANCED CLOUD BACKUP TECHNOLOGY IN TELECOMMUNICATION NETWORKS

Andriy Luntovskyy[1], Volodymyr Vasyutynskyy[2], Josef Spillner[2]

[1]BA Dresden University of Cooperative Education, Dresden, Germany
[2]Dresden University of Technology, Dresden, Germany

Data crashes can cause unpredictable and even hard-out effects for an enterprise or authority. Backup strategies as antidote unify a complex of organizational and technical measures that are necessary for data restoring, processing and transfer as well as for data security and defense against its loss, crash and tampering. High-performance modern Internet allows delivery of backup functions and is complemented by attractive (mobile) services with a Quality of Service comparable to that in Local Area Networks. One of the most efficient backup strategies acts the delegation of this functionality to an external provider, an online or Cloud Storage system. This article argues for a consideration of intelligently distributed backup over multiple storage providers in addition to the use of local resources. Some examples of Cloud Computing deployment in the USA, the European Union as well as in Ukraine and the Russian Federation are introduced to identify the benefits and challenges of distributed backup with Cloud Storage.

Introduction

Up-to-date network technologies aimed at backup and restore technologies of critical enterprise/authority data are discussed. A comparative analysis of existing complex solutions and standalone tools is represented. Essential advantages in restore technologies for critical enterprise or authority data can be offered via a newly developed original Cloud Backup concept [1—3] in comparison with the traditional data-centric backups [4]. But the complex constellation of international law and multilateral data safety requirements limit in some way the development of Network Technologies for Cloud Backup.

One of the possible ways for solving the mentioned problems offers intelligent combination of well-known commercial storage clouds with the use of efficient cryptographic methods and stripes/parity dispersal functionality for authenticated, transparently encrypted and reliable data backups. This approach has obtained a patent in the USA with the name RAIC (Redundant Arrays of Independent Clouds) [5]. Yet, from both a scientific and a practical perspective, there are shortcomings in conventional RAICs when e.g. dismissing the cost and trust characteristics of the associated storage services.

Backup as important component of informational safety

Disruption of critical data has unforeseen and heavy consequences for companies or organizations. It may have different reasons, but the main result remains always the same: a significant risk of losing data or access to it. This may lead to impediments in reaching the goals of companies or organizations, errors in documents, malfunctions of tools and machines, losing reputation on the side of partners. Very often the risks of losing data are caused by natural phenomena as shown in Table 1 where they are presented along with statistical probabilities and human factors.

Table 1. Causes and probabilities of losing critical data due to natural and human factors.

Cause of losing data	Statistical probability
Natural phenomena	
Hurricanes	1 %
Fire	6 %
Water	8 %
Short-circuit	16 %
Lightning stroke	17 %
Other natural phenomena	17 %
Human factor	
Usage faults	25 %
Stealing	10 %

The next problems of the company or organization are significant costs for the recovery of critical data and compensation of damages. For these reasons, backup technologies are a very practical task and a relevant part of securing data and assuring information safety of the company or organization.

The purpose of data backup is the regular creation of copies of files, databases, applications, settings on external backup systems, which in most cases are storage units managed by a backup application. Modern network/off-site backup systems support this process with separation of locality for reasons of saving and recovering the data and prevent the risks of data loss in a company or organization that may appear because of: hardware malfunction because of voltage jumps or devastating natural disasters, such as fire, water; attacks of malicious software, like computer viruses, Trojans; system errors during data storage; stealing the data.

Backup includes organizational and technical measures for storing, processing and transferring back important data and guarantees their protection from loss, destruction or disruption. The main distinctive features of modern network backup systems are the target devices (smartphone, tablet, PC, rack server form factors) along with the target storage media (magnetic disks or tapes, electronic flash memory and optical disks), delay of data access (ms), maximal time of safe data storage (months, years), error rate, GB costs. An example of a combined backup system for a small or medium-sized company or organization is shown in Fig. 1.

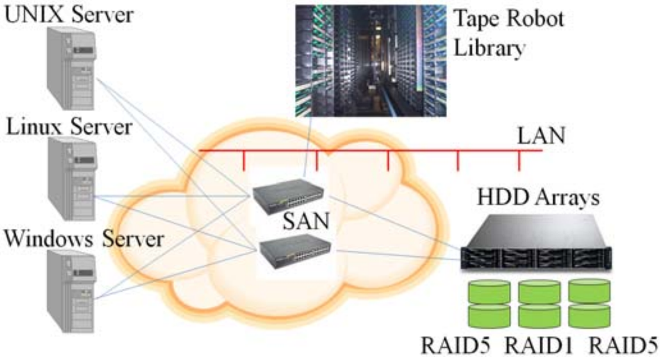


Fig. 1. Example of backup system structure.

The main components of the system are: an optical network (ATM, 10GbE), SAN (Storage Area Network), Tape Library, RAID file server systems (Redundant Array of Independent Disks).

According to Table 2 [1—5], the main criteria for the choice of suitable backup media and networking technologies include high-speed connections (~1GB/s over LAN), very large data volumes of overall storage (from 100 Petabytes up to Exabytes), long guaranteed usage duration (months, years), all when at the same time having a low probability of errors and costs per data unit. This list is not conclusive; good handling of small files and backup schemes are further factors.

As it can be seen from Table 2, the streamer tools (Streamers SLR, DLT, DAT/DDS, LTO, VXA) guarantee a low probability of errors and costs per data unit, long guaranteed duration and large data volumes, as well as a good price/value ratio. But a non-linear restore operation from such media is a time-consuming task, leading to the requirement of balanced choices.

The RAID technology (Redundant Array of Independent/Inexpensive Disks) is based on the creation of a redundant array of independent (multiple vendors) and inexpensive (consumer SATA instead of SAS) hard disc drives (HDDs), which work in one system to improve selectively both speed and reliability of I/O operations. The array of HDDs is controlled by a special RAID controller (hardware or software array controller), which provides the functionality of storing and retrieving data in the array as well as creating and checking the checksums. This allows making the underlying system transparent to the external users and presenting it as one logical I/O channel. Thanks to parallel runs of read/write operations on several discs, the disc array provides a higher speed of data exchange compared to one large disc.

Table 2. Overview of backup media.

Media for backup	Max data volume	Cost per 1GB	Guaranteed usage duration	Probability of failures
DVD	4.7—8.5 GB	0.05 €	small, ~1 year	high
USB flash	2—256 GB	0.97 €	very small	medium
USB-HDD	0.5—4 TB	0.04 €	very small	medium
Streamer LTO	0.2—3 TB	0.06 €	30 years	low
Streamer DLT	0.16—1.6 TB	0.17 €	30 years	low
Systems of redundant discs, RAID	max 10 TB	multiple of HDD costs	several years	low

The RAID technology was created first in 1988 by D. A. Patterson, G. Gibson, R. H. Katz, researchers of University of California, Berkeley. For regular backups, different variants of underlying storage types exist: streamers connected via local network (method 1); backup via LAN (method 2); backup via SAN (method 3); backup via NAS (Network-Attached Storage, method 4); backup via external backup provider (data center or cloud system) (method 5).

For choosing the right backup method for a company or organization, different methods and factors should be considered including: size of the company or organization, structure of available networks, number of users (a small enterprise with 20 users or a big company with more than 1,000 users), costs of backup, requirements on data safety and security as well as administration efforts.

In recent years, network technologies made a great progress in QoS (due to WdM, 10GbE), mobility (HSDPA, LTE) and easy access to computing centers [6—8]. In fact, the “Internet of Services” ensuring the application based on service-oriented architecture (SOA) principles has been created. High-speed Internet enables providing functionality and services with the same quality as known from local networks, and hence makes the shift of formerly relatively local functions such as backup into the network feasible. The new IT paradigm of delegating the services to external providers is known as Cloud Computing.

One of the most effective backup strategies is thus the delegation of the entire backup process to an external provider by interfacing up-to-date cloud systems [1—8]. This is achieved by placing the backup services into a public cloud offered by a capable and trustworthy cloud provider. Cloud computing is becoming more and more popular when several companies transfer their IT infrastructure (completely or partly) into clouds. This may lead to a lack of transparency of data access (who, when, where, why and what) and cloud reliability and raises the risk of loss of all critical data if the cloud provider leaves the market. To mitigate these risks to some extent, the deployment model of private clouds (method 6) under operational control from the client may be used. Furthermore, intelligent client-side techniques can further reduce the risks. Below, a very precise, adopted from the NIST and Amazon definition of the concept of Cloud Computing is given [9, 10]: “Cloud Computing is the on-demand and pay-per-use application of virtualized IT services over the Internet. The Clouds can offer: on-demand self-service; broadband network access; resource pooling; measured and optimized service; rapid elasticity”.

The adoption of Cloud Computing provides the following advantages [1—3]: relative reliability and security while giving up physical possession; staying in control when demand changes, the control can be exerted through vertical and horizontal scaling and migration to other providers; availability of attractive multi-layer services from infrastructure to software applications, efficient platforms/stacks and convenient client integration (Table 3).

The broad range of platforms and choices in functionality leads to a discussion of the most important domain-specific criteria for Cloud Backup. These criteria based on those for general Backup and those for general Cloud Computing are: Quality of Service parameters such as throughput, data rate, delays and reaction time; convenience (comfort, suitability, effectiveness); user control; trustworthiness, security and privacy; price per data extent and time.

Fig. 2 depicts a comparison of the mentioned criteria: convenience for the use regarding to user control for certain well-known systems. There are some disadvantages accompanying the clouds deployment also as follows [1—3]. Performance and convenience of offered clouds are questionable since cloud provider trustworthiness must be discussed. The failures of provider services can follow despite of existing SLA (Service Level Agreement) conditions.

Table 3. Well-known Cloud Platforms.

Platform	Provider
Amazon EC2	Amazon Web Services (AWS) for Elastic Compute Cloud (EC2)
Cloud Computing Yahoo!	Cloud services from Yahoo Platforms
Cloud Computing Resource Kit	Cloud services from Oracle/Sun
Eucalyptus	IaaS stack which reimplements the Amazon APIs
Sales.Force	Cloud services from Force.com, mostly on the SaaS level
Google App Engine	Google (a PaaS model)
Google Docs	Google (a SaaS model)
Google Compute Engine	Google (an IaaS model)
iCloud	A virtual OS on a Cloud basis
meebox	Online file management in the frame of a SaaS model
MS Windows Azure	Multiple Cloud Services in the frame of the Win Azure Platform (Microsoft)
Nimbula	A private/hybrid cloud technology of former AWS-collaborators
OnLive	An interactive Games-on-Demand-Platform with compression methods for computer graphics and videogames
Open Cirrus	Open Cloud Computing Research Testbed from opencirrus.org
OpenStack.org	Open Cloud from Rackspace, Citrix, NASA, Dell
OpenNebula	Commercialized European research project for data center virtualization and service markets
OpenShift	PaaS from Red Hat
T-Systems Dynamic Services	A private Cloud-system for dynamic deployment of SAP-applications from SAP GmbH
Verpura	Online-Cloud for Enterprise Resource Planning in SME
VMware vSphere	A virtual OS on the Cloud-Basis of VMWare

The next position might be the organizational reliability (trustworthiness of a cloud provider) because a provider can disappear from horizon unexpectedly, for instance, due to own economic, legal or political reasons. Data security is required since the risks of data losses and compromises by provider maintenance via third parties are still unreasonably high.

Regular backup software

Backup software is the basis for realization of the backup strategy in the company or organization which allows automation of the backup tasks. The software triggers the backup process in a certain point of time, provides the full or incremental backup of the selected data and informs the IT administrator.

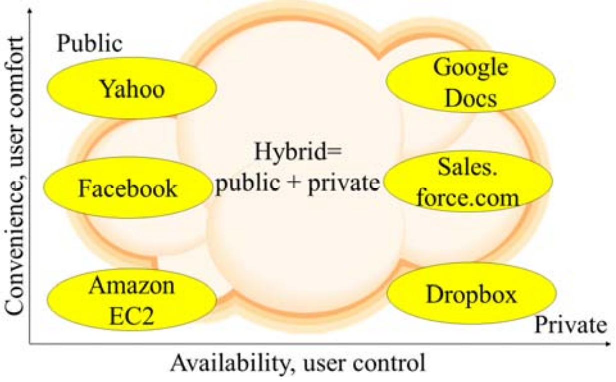


Fig. 2. Criteria comparison for well-known systems: convenience-to-control.

The choice of backup software may include cost-free open source software, which may in some cases miss the required functionality, effectiveness and reliability [1—3], but typically offer a tremendous flexibility for customization. On the other side, the commercial complex backup solutions ensure high reliability, high data transfer effectiveness, advanced configuration settings and additional features. The commercial solutions are, however, more expensive and may lead to a backup software lock-in which should be avoided similar to a storage provider lock-in. That is why in all backup planning projects a compromise should be made between the costs and added value of the backup solution (functionality, effectiveness and reliability), cf. Table 4.

Delegation of network backup functionality to Cloud providers

The functionality of a Cloud delivers services by accessing the virtualized resources which internal structure is unknown to the users, providing certain common operations, resource-intensive tasks, consolidation and distribution of resources and integration of applications in IT systems of companies [6—8]. Distribution of responsibilities in different cloud technologies compared to a traditional IT is shown in Table 5. Providers of "Internet of services" deliver the services at different hierarchical levels. The functionality of the computers and further interaction devices [6] as thin clients of end users in the Cloud is limited to providing a graphical or multi-modal interface (service frontend), caching the data, selection of and access to external network services. We see a resurrection of this host-node computing model in the increased use of consumption-oriented notebooks, netbooks, smartphones, tablets etc. Access to network resources can be provided by using the standardized web service protocols XMPP (Extensible Messaging and Presence Protocol) and SOAP (Simple Object Access Protocol), including a range of extensions to both, for permanent sessions and request-response models, respectively.

Table 4. Selected backup software.

Software	Description	Costs
DAR (Disk Archive)	Uses an own archive compression format, distributes the backup copies into different fragments and discs, supports common encryption methods.	Freeware
Rsnapshot	Creates hard links between different stored routes that requires the storage media support of the hard links. When a file changes, not only the change difference is backed up, but the whole file.	Freeware
Duplicity	Creates backup copies in encrypted format GPG (PGP) and archived in GZIP. Backup copies can be made practically for all types of operation systems; supports upload of backup copies over FTP, systems SSG, Rsync, WebDAV, HSi, and Amazon S3.	Freeware
Acronis Backup & Recovery Advanced Server	Popular but expensive software for MS Windows, allows creating image and file backups, is oriented on using HDD, tape libraries, cloud technologies.	about 1,100€
Drive Backup Server	Provide different backup functions, e.g. storage on internal and external media, CD/DVD/BR discs, NAS systems, FTP with support of virtual machines VMWare.	about 500€
Symantec Backup Exec 2012	Similar to Drive Backup Server	about 900€
Rsync	Allows scripts for configuration of shell, copying files and their parts. The special feature of Rsync is effective synchronization of file tree over network.	GNU General Public License /Unix-Distributions
Cron-Daemon	System process of Unix for timer-based triggering of processes like backup. The backup tasks can be triggered periodically according to "crontabs" tables and are called "cronjobs". They create backups on specified servers.	Unix-Distributions
Bup	A combination of Rsync and Git (version control) concepts. It offers Par2 redundancy	GNU LGPL v2
Bacula	Client-server based network backup application for individual computers up to large networks	GNU AGPL v3
Amanda	Advanced Marayland Automatic Network Disc Archiver with support for tape drives, disks and optical media, with native Windows client	BSD-style

Access to these resources can be also ensured via RESTful (Representational State Transfer) methods, a session-less paradigm which transfers state by modifying resources on the server.

Table 5. Load balancing and functionality distribution between Cloud Computing and conventional IT.

Conventional IT	IaaS	PaaS	SaaS
Applications	+	+	Applications*
Data	+	+	Data*
Runtime	+	Runtime*	Runtime*
Middleware	+	Middleware*	Middleware*
Web Services	+	Web Services*	Web Services*
OS	OS*	OS*	OS*
Virtual Resources	Virtual Resources*	Virtual Resources*	Virtual Resources*
Server	Server*	Server*	Server*
Storage	Storage*	Storage*	Storage*
Network	Network*	Network*	Network*
+	for self-responsibility		
*	delivered from the cloud		

The processing and archiving tasks, database querying, calling and encapsulation of further internal function calls are delegated to the Cloud provider. There are closed (private), public and hybrid clouds, which include file servers, databases, archiving backup systems, high-performance computers, computer grids and multiprocessor clusters. Peer-to-peer Clouds are not yet widely used but they are considered as a future trend in research, in particular for trustworthy mutual Backup. Service Level Agreements between cloud providers and end users guarantee a certain QoS, and aim to achieve a high level of users' satisfaction called quality-of-experience (QoE). Cloud computing provides following functionality: outsourcing of IT infrastructure to the cloud provider which may be less expensive than maintaining the own one, hosting of services saving costs for administration and maintaining the IT infrastructure, outsourcing of data archives and applications (mail servers, file servers, data bases, backup services etc.), cost-saving by using of high-performance computer cluster/grids.

The main Cloud models given by the NIST and Microsoft definitions [9, 11] are presented in Fig. 3. Software as a service (SaaS) is the most simple model using service-oriented web applications providing the access to resources in the cloud via frontends. Platform as a service (PaaS) provides an integrated platform for developing and testing web applications (testbed) and eventually running them. Infrastructure as a Service (IaaS) provides services of virtual networks by using remote servers, systems of networked hard disc drives (SAN/NAS), virtual machines (VMs) with network management exploiting the SNMP protocol and upcoming OCCI interfaces. The IaaS layer can be further subdivided into compute, storage and communication resources.

Communication is an implicit prerequisite for the other two so that they can be used over the network. For Cloud backup systems, the main interest is in storage resources which are accessed through network resources. In practice, these resources are not universally described.

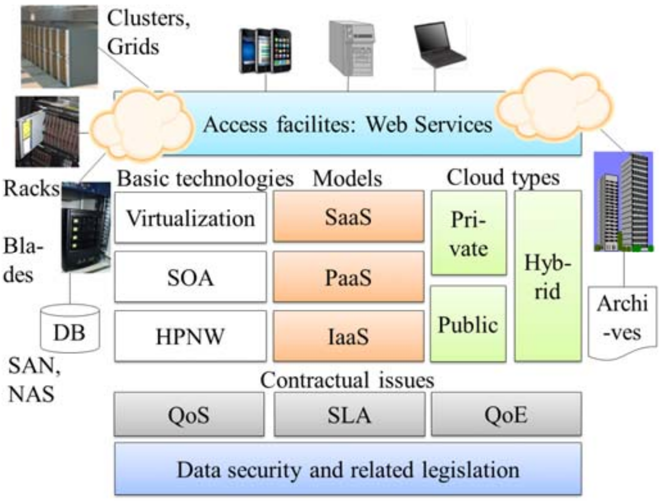


Fig. 3. Architecture of cloud computing: (HPNW) denotes High-Performance Network.

When creating, commissioning and maintaining Cloud services, a lot of questions of IT security still remain open limiting the further spread of cloud technology. This could be addressed by the creation of a non-profit cloud security alliance aiming to collect the best practices of effectiveness, legal compliance and IT security. Researchers already started an outreach into this direction through surveys [12].

These abstract challenges shall now be demonstrated with examples from a selection of countries with a varying level of development and Cloud adoption rates. With regards to Cloud Computing, legal acts of Ukraine regulate in general the operations in the area of IT security and related fields (intellectual property, telecommunications, cyber-crime, television). In general, it can be evaluated as systematic and complete thanks to the consideration of existing international best practices [13].

The providers of Cloud computing in Ukraine and Russia mainly provide now installation and later maintenance of server and communication hardware of clients [2, 3, 13]. Physically, the hardware is installed into 19-inch racks. The performance and reliability of communication channels of data centers influence directly the quality of service and quality of experience. One of the main criteria is the boot-up-time, i.e. the minimal time when the virtual service or server gets ready to work.

The oldest and largest providers in Ukraine are ColoCall and Hosting.ua, and in Russia Selektel, Stack Group, ISG, WideXS, Telehouse Caravan, IBS DataFort, KiaeHouse, DataDome, Filanko, DataLine, StoreData, KROK, PTKOMM (a group of companies of Rostelecom) [2, 3].

One current scientific task is the optimization of the service characteristics of these providers regarding QoS and QoE. Great importance is given to the uptake of mobile services based on LTE and 4G networks with access through modern mobile devices running on iOS, iPhoneOS, Windows Phone 8 or Android OS, and the newer challengers FirefoxOS, Ubuntu Phone and Sailfish, all equipped with web browsers and personal data vaults.

The development of these technologies is widely supported by governments of developed countries, since it allows a significant resource saving, but requires coordination of providers in areas of efficiency, legal issues and IT security of Clouds [13]. Hence, for designing optimal Cloud Backup systems, the non-functional properties of the storage media, storage access nodes, the network connections and the client integration around the backup software need to be considered and evaluated.

Modern systems for Cloud backup

One of the most promising backup strategies is to delegate backup to an external provider, e.g. to a Cloud Backup system. A short overview of Cloud storage providers suitable for backup is given in Table 6. Online cloud resource brokers and marketplaces are updated periodically for an up-to-date view on the choices based on rich provider descriptions [14] which facilitate the exchange of the information through open markets. A comfortable access to the cloud backup systems is possible through dynamic and non-intrusive service selection even with mobile devices like tablets or smartphones.

If the company or organization does not trust the cloud provider, it could use the technology of private clouds, which limits the access to the cloud for external users and lets the data within the company, which underlines the benefits of cloud computing. Hybrid clouds combine placing a part of the data into a public cloud and processing the other part of data in an own private cloud.

An example of a cloud backup system is the Amazon Web Services provisioning platform (AWS), which also includes the Amazon Elastic Compute Cloud (Amazon EC2) and consequently follows the service-oriented architecture principles [10].

The Amazon Web Services platform provides access to a large number of different further services like application access, virtual machines, backup of files, databases, processing queues, online-memory (see an overview in Fig. 4 and Fig. 5).

Table 6. Overview of cloud backup platforms.

Name of Cloud Backup system	Region of storage	Max volume of cost-free storage	Max volume of paid storage	Platform
Amazon Cloud Drive	USA	5 GB	No limits	Win, Mac, Linux, iOS, Android, Windows Phone
Dropbox	USA	2 GB	No limits	Win, Mac, Linux, iOS, Android, Blackberry
Windows Live Skydrive	Ireland	25 GB	100 GB	Win, Mac, Windows Phone, iOS, Android
Strato HiDrive	Germany	-	5000 GB	Win, Mac, Android, WP7, Chrome, Synology
Google Drive	USA	5 GB	16000 GB	Win, Mac, iOS, Android, Linux
HighSecurity Backup	Germany	10 GB (up to 30 days)	No limits	Win, Linux, Mac, DBs, Exchange, Lotus, VMWare, Hyper-V
Ubuntu One	Isle of Man	5 GB	50 GB	Win, Linux, Android, iOS
SafeSync	Japan	500GB (up to 30 days)	No limits	Win, Mac, iOS, Android
F-Secure	Finland	-	No limits	Win, Mac
Daten-Safe	Austria	-	No limits	Win, Linux, Mac, DBs, Exchange, Lotus, VMWare
Four-Shared	USA	10 GB	No limits	Win, Linux, Mac, Blackberry etc.

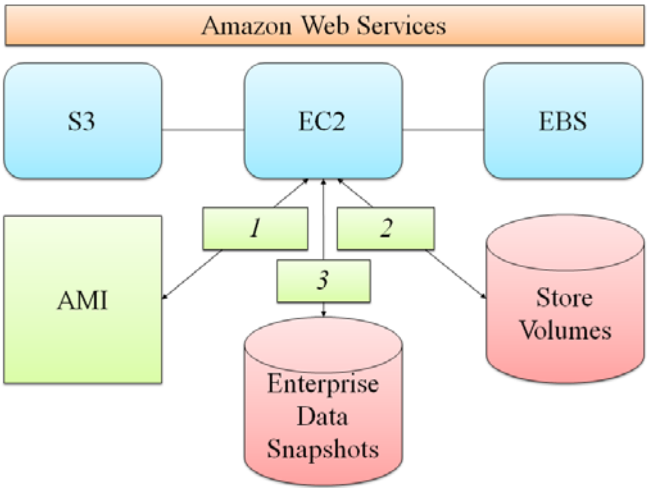
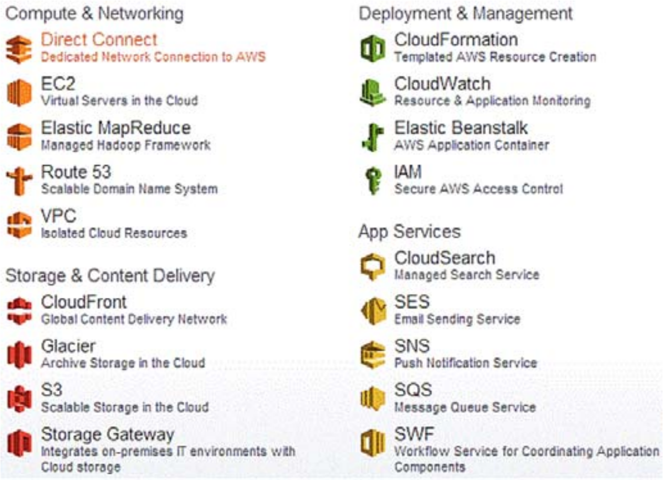


Fig. 4. Structure and components of Amazon Web Services: (AMI) Amazon Machine Images; (EC2) Amazon Elastic Compute Cloud; (S3) Amazon Simple Storage Service; (EBS) Amazon Elastic Block Store; Operations: (*1*) Deploy, (*2*) Attach, (*3*) Backup.



Compute & Networking
Direct Connect
Dedicated Network Connection to AWS
EC2
Virtual Servers in the Cloud
Elastic MapReduce
Managed Hadoop Framework
Route 53
Scalable Domain Name System
VPC
Isolated Cloud Resources

Storage & Content Delivery
CloudFront
Global Content Delivery Network
Glacier
Archive Storage in the Cloud
S3
Scalable Storage in the Cloud
Storage Gateway
Integrates on-premises IT environments with Cloud storage

Deployment & Management
CloudFormation
Templated AWS Resource Creation
CloudWatch
Resource & Application Monitoring
Elastic Beanstalk
AWS Application Container
IAM
Secure AWS Access Control

App Services
CloudSearch
Managed Search Service
SES
Email Sending Service
SNS
Push Notification Service
SQS
Message Queue Service
SWF
Workflow Service for Coordinating Application Components

Fig. 5. The screenshot of main panel of Amazon Web Services.

Hybrid Cloud backup concept

The modern challenges and solution ideas responding to the listed problems have to be elaborated on. One of them is the deployment of the hybrid clouds as a combination of private and public clouds in certain topologies. The combined hybrid clouds with additional cryptographic protection functionality and management layer (so called “cockpit features”) at the customer side is often an appropriate solution [1—3, 5, 12, 14]. Taken to the extreme, such setups can include peripheral devices such as USB sticks for a four-eye principle in access control.

Further key points of a hybrid Cloud Backup concept under the given circumstances are as follows: use of effective and transparent encryption methods (for instance, RSA+AES+PKI, see Fig. 6) for increased security; deployment of a stripe and parity based dispersion (analogically to the known RAID techniques, see Fig. 7) for increased safety.

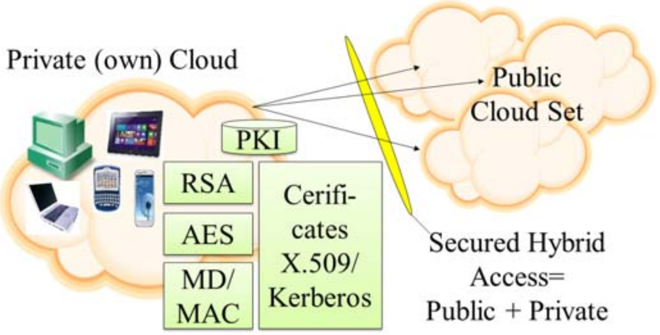

Private (own) Cloud
Public Cloud Set
PKI
RSA
AES
MD/ MAC
Cerifi- cates X.509/ Kerberos
Secured Hybrid Access= Public + Private

Fig. 6. Cloud Backup and Transparent Encryption: (MD) Message Digest; (MAC) Message Identification Code; (AES) Advanced Encryption Standard; (RSA) Rivest, Shamir and Adleman Encryption; (PKI) Public Key Infrastructure (X.509/ Kerberos).

The notion of transparent encryption for Cloud Backup encompasses the following features (Fig. 7): efficient cryptography methods such as AES, RSA, MD/MAC; X.509/Kerberos private key certificates, PKI deployment; document classification and demarcation; analysis of structured, unstructured data and context information; user authentication and respective keys granting. In local backup setups, the most popular systems are the RAIDs numbered as 0, 1, and 5, respectively, with two or four disks of which zero or one are redundant.

The functionality of RAIDs is based on stripes and parity dispersal routines [15, 16]. For a RAID5, a representation is depicted in Fig. 7*c*. With different colors the partition in the usual disks array is given: firstly for the data (the so called “Stripe Set”, e.g. A1 or C3) and then the distribution of the parity sums (“Parity Set”, e. g. BP or DP) through the four disks Disk 1 … Disk 4. In the given case, the common available volume *V* for the data backup can be calculated by the formula (Fig. 7*c*):

$$V = (n-1)\cdot V_{\min} \quad (1)$$

where *n* denotes the number of used HDDs; $V_{\min}$ is the minimal available HDD volume in the array. The redundancy is self-evident preconditioned via the parity set.

Let us consider the example with four arrays each of 500 GByte:

$$V = (4-1)\cdot 500\text{GByte} = \quad (2)$$

1500 GB pure for data backup as well as
500 GB for the parity control (see Fig. 7*c*).

Therefore a next constructive idea is the deployment of redundant cloud arrays (stripe and parity based dispersion). There are naturally a lot of further RAID concepts optimized for minimum access time, minimum failure probability, maximum volumes, minimum costs [15, 16].

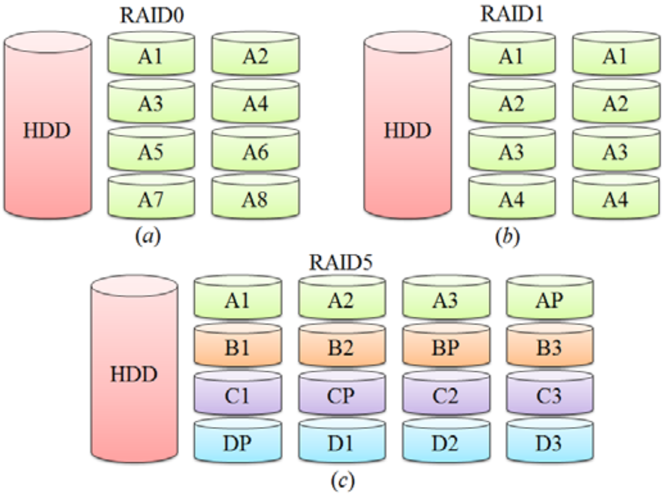


Fig. 7. The most used systems: RAID 0, 1, 5: (RAID) Redundant Array of Independent Disks; (HDD) Hard Disk Drives (Disk 1 … Disk 4).

Practically, these multiple concepts can be continued and mapped into clouds with the old-new slogan "From RAID toward RAIC!". There are already numerous RAICs, or Redundant Arrays of Independent Clouds [5, 12, 14, 16]. The possible variations to the concept are also: Redundant Array of Independent Networked Storages (RAINS) as well as Random Array of Independent Data Centers (RAIDC) or Redundant Array of Optimal Clouds, an extension to RAIC which emphasizes an enforcement of user requirements on the selection and maintenance of storage service arrays (RAOC).

The software architecture suitable for the realization of RAIC is depicted in Fig. 8. The predominant client-side software for RAICs consists of the following three layers with the related functionality: 1. integration layer (with logical partition and interface to the backup application); 2. pre-processing layer (with stripes and parity dispersal routine, encryption and other modifications); 3. transport layer (with block transfer operations). The clients obtain the possibility of the reliable and efficient access to an array of HDD storage media with added organizational and spatial independence. This software considers the state-of-the-art.

The advanced software architecture realizes a new layered RAIC concept and includes the following already known components but with the extended functionality.

Firstly, the advanced integration layer (1) includes multiple network file system protocols like NFS, CIFS/SMB, WebDAV or, alternatively, a local virtual file system interface or a Web Services interface. Additionally, CVS/SVN/Git (version control subsystems) and synchronization overlays are integrated. On the other hand, an advanced pre-processing layer (2) consists of necessary codecs aimed to classification of document types and its efficient coding (text files, MPEG, PDF). Then the policies on the data storage subjects and paths are included here as well as the routines for stripes and parity dispersion, authentication with MD/RSA/PKI and encryption with AES/RSA/PKI.

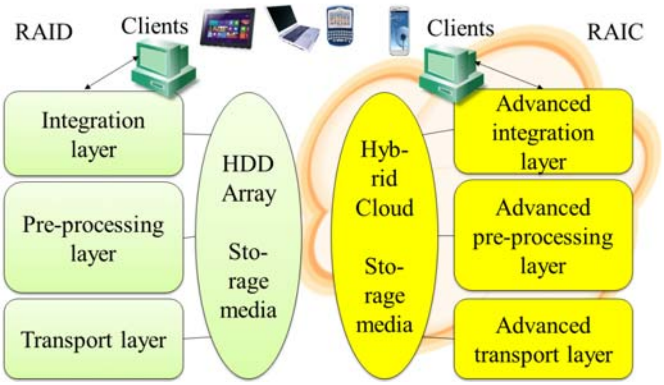


Fig. 8. Offered software architecture to realization of a RAIC: (CVS) Concurrent Versions System; (RAIC) Redundant Arrays of Independent Clouds.

Finally, the advanced transport layer (3) integrates the parallel and block-wise streaming, caching and local persistence procedures as well as includes the adapters for multiple provider APIs. The multi-modal cloud clients (desktops, tablets, and smartphones) enjoy with the reliable and efficient resource access to the set of the hybrid (private-public) Cloud Storage Media, namely to the RAIC.

An example of implementation

At this point an advanced example of an implementation for the RAIC and RAOC concepts can be mentioned. It is rooted in the FlexCloud@TUD (Dresden University of Technology) young researcher group, led by Alexander Schill, Josef Spillner and Stephan Gross. The project is funded by the European Union through ESF. Its goals are oriented towards a user-controllable and secure cloud life cycle. The basic project aims are [12, 14, 16]: avoiding uninformed Cloud provider selections through formal descriptions of resource, data and software properties; avoiding the Cloud provider lock-in effect through Multi-Cloud scenarios and migration paths; towards inter-connected personal Clouds, under the control of the user, which can be federated into a powerful network of Clouds; finally, means to exert the control with an appropriate management UI representing a personal Cloud cockpit.

The Cloud storage controller designed and developed within FlexCloud is called Nubisave (from Latin "Nubes" meaning "Cloud"). It sets up an aggregated view across multiple Cloud storage providers and enables higher-level storage tasks such as policy-enforcing data gateways, adaptive synchronization between devices, backup and collaborative sharing. Nubisave exports a virtual file system through FUSE which can be used as an underlay target media of backup software. All write accesses received by Nubisave are multiplexed onto the configured Cloud storage providers, and all read accesses reassemble the data. Encryption and versioning can entirely be performed on the client side. In case of failures, affected storage providers can be replaced by others and a replication of data from the remaining ones takes place automatically. Nubisave is available as open source software which has been demonstrated and discussed at both commercial events in Hannover and academic events in Stuttgart.

Conclusion

This paper can be generally characterized as a Work-In-Process (WIP). The next problems to be solved can be listed as follows:

1. Analysis of integration options for existing mature backup tools and emerging cloud backup services.

2. Optimization of the innovative RAIC techniques, development of a software controller based on Web Services for management and cryptographic protection of a RAICs, for instance, RAIC5. In the FlexCloud@TUD project this option is called π -Cloud with a π -Cockpit.

3. Development and securing a notation (for instance, based on WSDL or ontologies) for the meta-data aimed to RAIC description and management.

4. Development of acceptable conditions for the enterprise access with offer of the increased QoS (transfer rate, advanced security, data control and comfort for the users).

5. Further development of collaboration scenarios with file sharing, access by external entities, CVS and group working, as well as automatic classification of data.

6. Improving performance via scheduling, caching and parallelization algorithms.

The obtained results can be widely applied for efficient, automated and secured backup of critical enterprise data as well as for speed data access via up-to-date mobile or fixed networks.

References

1. Cloud computing / C. Baun, M. Kunze, J. Nimis, S. Tai // Web-based dynamic IT-Services. — Springer Verlag, 2010.

2. Luntovskyy A., Guetter D., Melnyk I. Planung und Optimierung von Rechnernetzen: Methoden, Modelle, Tools für Entwurf, Diagnose und Management im Lebenszyklus von drahtgebundenen und drahtlosen Rechnernetzen // Handbook for German universities. — Vieweg + Teubner Verlag Wiesbaden, Springer Fachmedien Wiesbaden GmbH, 2011. — 411 p. (in German).

3. Luntovskyy A., Klymash M., Semenko A. Distributed services for telecommunication networks: Ubiquitous computing and cloud technologies. — Lviv: Lvivska Politechnika, 2012. — 368 p. (in Ukrainian).

4. Ordinary backup technologies (Online, in German): http://www.tecchannel.de/storage/backup.

5. Decasper D., Samuels A., Stone J. RAIC – Redundant Array of Independent Clouds // Patent USA 2012: Reg. No.: 12/860, 810, Publishing No.: US 2012/0047339 A1.

6. Luntovskyy A. Programming technologies of distributed applications. — Kiev: DUIKT University of Telecommunications, 2010. — 474 p. (in Ukrainian).

7. Luntovskyy A., Klymash M. The service-oriented Internet // Proceedings of IEEE 11th TCSET 2012 Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science. — Lviv — Slavsk, February 2012, Lviv Polytechnic National University. — P. 256 (IEEE Xplore).

8. Luntovskyy A., Guetter D. A concept for a modern virtual telecommunication engineering office // Telecommunication Sciences. — 2012. — Vol. 3, Issue 1. — P. 15—21.

9. Mell P., Grace T. The NIST definition of cloud computing // NIST Special Publication. — 800-145. — September 2011.

10. Amazon Web Services (Online): http://aws.amazon.com.

11. Kommalapati H. Azure Platform (Online): http://msdn.microsoft.com/en-us/magazine/ee309870.aspx/.

12. Spillner J., Schill A. A versatile and scalable everything-as-a-service registry and discovery // Proceedings of 3rd International Conference on Cloud Computing and Service Science (CLOSER), Aachen, Germany, May 2012. — P. 23—27.

13. Ukrainian legislation regarding to data security (Online): http://zakon.rada.gov.ua/.

14. Schill A., Spillner J., Gross S. FlexCloud@TUD project/ Dresden University of Technology TUD (Online): http://www.rn.inf.tu-dresden.de/.

15. Metadata efficiency in versioning file systems / C. A. N. Soules, G. R. Goodson, J. D. Strunk, G. R. Ganger // Proceedings of the Third USENIX Conference on File and Storage Technologies, March 31 — April 2, 2003, San Francisco, California, USA (Online): http://static.usenix.org/ publications/.

16. Seiger R., Gross S., and Schill A. A secure cloud storage integrator for enterprises // International Workshop on Clouds for Enterprises, Luxemburg, September 2011. — P. 12—15.

Received in final form October 24, 2012