Specifications for Centralized DataCenter serving the educational cloud for Bulgaria

Radoslav Yoshinov
Laboratory of Telematics
Bulgarian Academy of Sciences - BAS
Sofia, Bulgaria
yoshinov@cc.bas.bg

Daniela Pavlova
State University of Library Studies and Information Technologies
Sofia, Bulgaria
d.pavlova@unibit.bg

Monka Kotseva
Laboratory of Telematics
Bulgarian Academy of Sciences - BAS
Sofia, Bulgaria
mkotseva@cc.bas.bg

***Abstract* — Electronic Education in Bulgaria, according to the country's National ICT Strategy in Education is based on the idea of developing one Integrated Information Resource Infrastructure for Education and Research. A very important part of this integrated infrastructure is to establish a Centralized DataCenter, where the three main components of Electronic Education – Storage Infrastructure, E-learning Platform, and Digital Content, can be focused (concentrated) and technologically refined. In the article below, we discuss the Basic advantages of implementation of Centralized Datacenter in Bulgarian e-education, serving its educational cloud. Security issues of the Datacenter are discussed. Bulgarian National Educational Claud is defined as integrated structure of all described components in their interoperability.**

Keywords— Storage,e-Learning Platform, Datacenter Security, Cloud Data Managemen, DDos attacks, datacenter, Content Delivery Network, Technological Pedagogical Content Knowledge

I. INTRODUCTION

Electronic Education in Bulgaria, according to the country's National Strategy is based on the idea of developing one integrated Information –resource Infrastructure for Education and Research. A very important part of this integrated infrastructure is to establish a Centralized Datacenter, where the three main components of Electronic Education – Storage Infrastructure, E-learning Platform, and Digital Content, can be delivered in a more focused, and technologically refined manner. In this article the authors will try to answer key questions about the technological pillars of e-Education and their integration, forming the National educational cloud.

1) Which are the technological pillars of the Bulgarian e-Education

2) How to formulate the basic and extended principles to be implemented in the Centralized Datacenter for Bulgarian e-Education;

3) How to choose the appropriate storage and devices for the Centralized datacenter to make it match for educational & research demanded technologies;

4) How to be up to date with the contemporary tendencies in Datacenter security;

5) How to choose the appropriate Equipment for creating, operating and managing centralized storage as main part of a centralized datacenter for the Bulgarian e-Education;

II. FORMING THE EDUCATIONAL CLOUD FOR BULGARIA

The electronic education in Bulgaria, according to the National Strategy is based on main pillars. Applied to e-Learning, we evolve from Shulman's Pedagogical Content Knowledge framework (PCK, 1987) [1] to Technological Pedagogical Content Knowledge framework of Mishra and Koehler (2006) and Koehler and Mishra (2008) [4], reflecting the general scheme of e-learning using learner centered pedagogy.

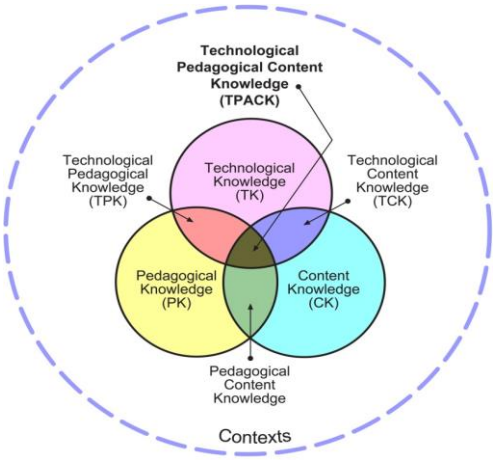

Fig. 1: Technological Pedagogical Content Knowledge framework

If we stress on technological knowledge (TK) from the TPCK framework, it concerns Educational technology giving the effective use of technological tools in learning. As a concept, it concerns an array of tools, such as media, machines and networking hardware, as well as considering underlying theoretical perspectives for their effective application. Acquiring TK in this manner enables a person to accomplish a variety of different tasks using information technology and to develop different ways of accomplishing a given task.

So we derive the four main technological pillars of electronic education in Bulgaria, according to the National ICT Strategy for e-learning to be:

- Network infrastructure

Containing National backbone with "Last mile" to the institution (regional inspectorate of education, university, research institute, research laboratory, college, school etc.), extended with the local network of the institution.

- Storage infrastructure.

A federation of repositories shall be implemented to make decentralized resources available through a single point of access. This federation has to respect the autonomy of each repository but at the same time allow the cross-querying and discovery of relevant assets stored in different places. A common schema for assets description that allows precise and rich description of repositories content is a key enabler of seamless data exchange. The approach is to semantically interlink the federated repositories, preserving their autonomy and decouple while at the same time avoiding imposing specific technologies.

- E-learning platform.

Basic advantages of implementation of e-learning platform are: Effectiveness of spending; Support of all type of educational activities and mixed educational programs; Rich functionalities; High level of security; Multilingualism; Easy to use and user friendly; Easy maintenance and management; Interoperability with existing institution (regional inspectorate of education, university, research institute, research laboratory, college, school etc.) management systems; Support of third party digital content

- Digital educational content.

Rich digital content is a powerful way of providing today's students with high quality, relevant and up-to-date instructional materials. Rich digital content can take many forms. It can be provided in standards-based packages that build upon textbooks, with teacher's guides, assessments and multimedia content all included and aligned to standards. It can be created collaboratively, in open source format, by a variety of experts. Or it can be drawn from multiple sources — subscriptions, free online resources and other digitized material.

Common features of digital content include: Multimedia elements such as still images and graphics, video virtual reality, animations, simulations, audio, music, interactive, and gaming elements; Embedded tools (survey, calculator, spreadsheet, etc.) to facilitate student highlighting, annotating, calculations and more; Additional tools (wikis, video/graphics editors, academic networking tools) to support collaboration and creation; A variety of languages; Adaptive and assistive technology designed to meet special needs; Embedded links to external sources and access to remote experts and mentors; Technologies that evaluate student responses, provide customized content and redirect students to data-indicated areas of need; A seamless continuum of instruction and assessment; The ability to be updated and enriched continuously and seamlessly; Site licenses or subscriptions that ensure a dependable supply of "perfect" copies; The ability for teachers to search, sort and select by standards based needs and queries; Options for exporting, reformatting and combining text and other content so it can be used beyond the original package for presentation and dissemination in various ways.

Remark: It is presumed by the National ICT Strategy for e-learning to have one integrated network infrastructure for Education and Research for Bulgaria.

Of course the main thing is accepting and applying standards for every one of the technological pillars.

Network standards facilitate the interoperability of network technologies. All networking standards to be used are "open" standards, administered by a standards organization or industry group. Technical standards applied for the network infrastructure assure interoperability of optic, cable, Wi-Fi (Bluetooth, infrared …) infrastructure, network devices, addressing and identification, crosslayer design, network virtualization, routing and of the traffic engineering, dynamic switching of optical circuits, decoupling of control and data, service discovery and composition, and of course network management.

Storage standards as the Storage Management Initiative-Specification (SMI-S), eXtensible Access Method (XAM), encryption key management and the Fabric Application Interface Standard (FAIS), ISO 35.220: Data storage devices, promise to make it much easier to manage storage and data.

Standards in Digital educational content assure interoperability of e-Learning content and Learning Management Systems. The primary benefit of these standards is that electronic content developed to be standards-conformant can be easily integrated into LMSs compliant with these standards. Here are some examples.

Standards for search, discovery and retrieval - Dublin Core, IEEE LOM, IMS Digital Repositories Interoperability, Vocabulary Definition Exchange, Resource List Interoperability, Open Digital Rights Language, Metadata Encoding and Transmission Standard etc.

Standards for planning of education scenarios, formulating and sharing of educational and pedagogical content and tests - IMS Learning Design, Simple Sequencing, IMS Content Packaging, METS, IMS Sharable State Persistence, ADL SCORM, IMS Question and Test Interoperability, IMS Learner Information Packages, IMS Reusable Definition of Competencies or Educational Objectives, IMS Accessibility for LIP, W3CWAI, etc.

The four pillars, taken in their integrity and interoperability form the educational cloud for Bulgaria.

Important part of this integrated infrastructure for Education and Research for Bulgaria is a Centralized Datacenter, where we can concentrate and technologically refine three of the four pillars of the Bulgarian electronic education. These are the storage infrastructure, the E-learning platform with the digital educational content. Important part is the switch of the datacenter to the National Network infrastructure [9, 10].

III. BASIC ADVANTAGES OF IMPLEMENTATION OF CENTRALIZED DATACENTER IN BULGARIAN E-EDUCATION.

Datacenters, consist of high capacity storage resources, hardware resources, memory and network resources. There are several fundamental types of data centers. Traditional "systems of record" data centers, the extension and globalized verson of it is the Monolithic data center. The basic requirements are reliability, availability and serviceability *(RAS)* together with fast data access combined with other often conflicting requirements for retention, legal discovery, longevity, integrity, accessibility, security, disaster recovery and so on. [2] Data centers intended to support cloud-oriented operations beside the basic RAS requirements have additional requirements for switching and virtual storage infrastructure. Engineering data centers, which incorporate a mix of cloud and traditional models. Datacenters for smaller businesses (SMBs) typically using less robust grade of storage because of economic considerations. [5, 9, 10] For now the base principles applied in the centralized datacenter for Bulgarian e-education are:

Educational content stored and retrieved "on one counter"

- Storage and maintenance of data on one place;
- Effectiveness in using resources;
- Easier IP right implementation on content;
- Effectiveness of spending;
- Support of all type of educational activities and mixed educational programs;
- Flexibility and rich functionalities;
- High level of security;
- Easier to implement multilingualism;
- Advantage in maintenance and management;
- Interoperability with existing institution (regional inspectorate of education, university, research institute, research laboratory, college, school etc.) management systems;
- Being flexible and adaptable, allowing students to learn at their own pace, and in their own style;
- Offering teachers and administrators the power to select and modify content as desired;
- Connecting students with outside resources as well as experts and mentors to support their learning;
- Providing a seamless continuum of instruction and assessment, thus providing data to inform teacher practice and improve student performance;
- Offering opportunities for students to share ideas and collaborate with one another through such tools as wikis or social/academic networks;
- Challenging and motivating students to create their own meaning in the form of blogs, multimedia presentations or other original content that builds on what they have learned and is delivered to an authentic audience.

We tend to fulfill also these extended principles for handling applications on the centralized datacenter for Bulgarian e-education:

• Applications are easily and highly parallelizable
• All applications can be hosted on the same architecture
• The applications are at incredible scale (global)
• The applications are purpose built
• Storage and computation are advantageously collocated
• Small amounts of data loss are excusable

Assumption: The Education Data Center provides resources only to legitimate clients.

Specification of running Environments in the Centralized Datacenter [5]

- Development
- Testing
- Staging
- Production
- Running day to day storage support

Specification of the basic category of the host's functions (roles) of the servers in the datacenter:

- Application Server (non-web)
- Database Server
- SFTP server
- Mail Server
- Name Server
- Configuration Management (puppet/ansible/etc.)
- Monitoring Server (nagios, sensu, etc.)
- Proxy/Load Balancer (software)
- SSH Jump/Bastion Host
- Storage Server
- Version Control Software Server (Git/SVN/CVS/etc.)
- Virtual Machine Manager
- Storage arrays with management
- Web Server

Power supply and cooling of Datacenter: [5]

- Power Usage Effectiveness of less than 1.3
- Power connection of over 1 megawatt, Electricity connection of over 2 MW, connected to 10 kV power supply
- Redundant transformers in separate rooms
- 230 V and 400 V circuits available

- Electricity consumption graphics can be downloaded in real time
- Up to 30 kW for high-performance racks
- Redundant, environmentally friendly USV facilities (Redundant, environmentally friendly UPS systems)
- Redundant transformers in separate rooms
- High-performance diesel generator sets for up to 96 hours of autonomous operation
- 104 kW cooling performance per conditioning cabinet
- Cold/hot aisles with complete rack housing for large servers
- 50 cm double floor height for optimal ventilation

Tier of Datacenter: The Datacenter for education is to be minimum Tier 2 (better Tier 3) Datacenter, according to the four tier model of Datacenters. [2]

REMOTE HANDS AND EYES for the datacenter - engineers and technicians, ready to perform basic tasks on customer equipment under on-line customer supervision

DISASTER RECOVERY SPACE - Private or shared office space for Disaster Recovery with direct cable connections to redundant IT equipment on site.

IP rights preserving scheme for the stored content

Encryption: concepts and roles if required encryption of data.

- Encrypted servers
- Encrypted VoIP systems
- Encrypted VPN connections
- Encrypted end devices
- Encrypted site connectivity

Using Content Delivery Network (CDN) concept for Datacenter architecture.

A CDN (Content Delivery Network) consists of a source server and several cache servers, to which users' queries are directed. If someone requests a particular file, the cache server checks to see if it is already available in the local cache. If so, it delivers it immediately. If not, it checks whether it is available in the global cache. If it is not available, it requests the file from the source server

After analyzing the situation, an individual CDN concept to be developed, which takes the geographical distribution of the visitors to be taken into account.

IV. SECURITY:

The Datacenter hosts critical data and contains core assets, including user's information, intellectual property, and other critical data. With emerging trends such as Big Data, bring-your-own-device (BYOD) mobility, and global online collaboration sparking an explosion of data, the data center will only become more important and will extending be the target of advanced malware and other cyber attacks. [7] What is needed to be done to secure a Datacenter is:

- To shield data centers from advanced persistent threats (APTs) and sophisticated malware found in content stores, web and application servers, and common file shares
- To stop attacks entering organizations via mobile devices and portable storage
- To receive on-target analysis to pinpoint possible gaps that need addressing
- To protect key assets and prevent attacks with products and services that work together and share and threat intelligence
- To prevent attacks with a nimble, adaptive cyber security strategy
- To safeguard Datacenter from attacks that use web servers and other data center infrastructure to host malware
- To detect threats quickly to reduce lag time before resolution
- To get reliable, fast malware analysis with agentless network-based threat detection and protection engine
- To provide continuous, dynamic, non-disruptive resolution to incidents

One of the main threats is the denial of services, which puts down functionally the Datacenter. The goal here is to protect Datacenter from Distributed Denial of Service (DDoS) attacks and assure its functionality to the clients.

Let's define what we understand as DDos attack. A DDoS (Distributed Denial of Service) attack uses more than one Zombie computer to launch a synchronized attack against the victim (it can be one or more target). Because of the client-server technology, attacker or executor is able to increase the success of the Denial of Service drastically by utilizing the resources of multiple unaware zombie systems which serve as an attack enhancer. [3] Types of DDoS attack:

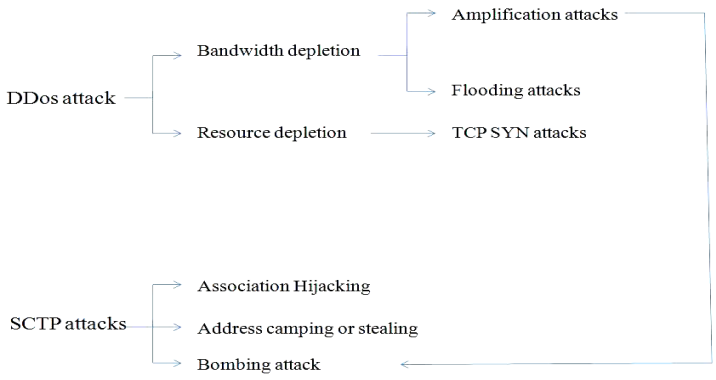

Fig. 2 Types of DDoS attack

There are numerous models for DDoS detection and DDoS scenarios in real-time. One of the most advanced is neural classifier architecture having four stages for detecting DDoS attacks, which are Data collection, preprocessing, classification, response. DDoS attackers employ botnets and zombie servers to launch DDoS to deplete server resources. Classification of attacks is true positive, true negative, false positive and false negative.

Proposed solution for defense is framework of cooperative intrusion detection system (IDS), an agent based system and network treat detection system which enhances the system by distributing the IDS nodes across the network. Host IDS collects audit data from operating system. Network IDS collect data from network packets. When any malicious intrusion is detected, system generates reports and alerts. It prevents the system from single point of failure attack. [6]

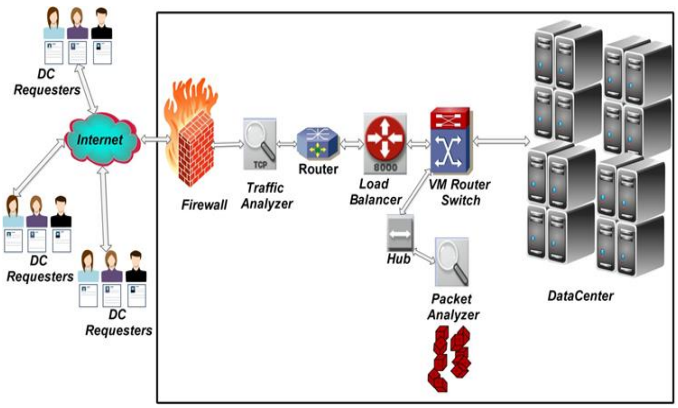

Fig. 3 Proposed technological scheme for protection

This scheme also identifies the aggressive legitimate user and prevents them entering into the firewall until the session expiry. The proposed scheme is efficient in terms of serving the clients by caring the time-sensitiveness, a characteristic of cloud computing. The detection capability can be increased with parallelizing the required hardware that is responsible for improved detection accuracy even at increased traffic. The main goal of this concept is to restrict the number of requests to a value that can be handled by the server efficiently [8]

Cheaper alternative is the p-Filter concept, based on the assumption that every server can serve a limited number of requests at a time and if the number of requests (invalid or valid) is within that range the server can work efficiently and can serve those requests. The number of requests can be calculated on the basis of the available resources and the bandwidth allocated for particular server. (Fig. 4)

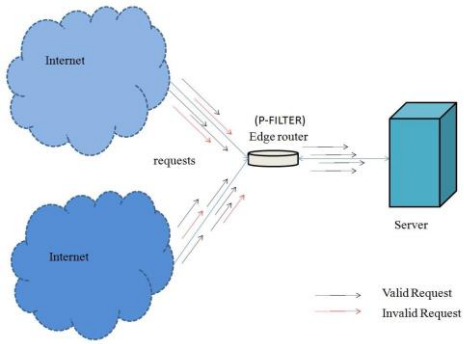

Fig. 4 The p-Filter concept

The main purpose of this system is to balance the load of the system, which is the ultimate goal of each and every detection and prevention techniques.

V. STORAGE: AS IMPORTANT PART OF A DATACENTER. [2, 5]

The single most important consideration related to storage when planning a data center is the intended use of the data center. Storage is an increasingly important and complex component in modern data centers, especially taking into account the recent move toward cloud infrastructures. There are numerous storage technologies that impact data center design, following principles of consolidation of storage into a central location, removal of the storage burden from host OSes, and so on. Storage Area Networks (SANs) evolved in an environment dominated by structured data, aggregating several drives together and giving the capability of carving out partitions, called Logical Units (LUNs). Contemporary SANs use several types of virtualization to allow sophisticated data management techniques such as growing and shrinking of LUNs, mirroring of data, including to remote sites, using Fibre Channel (FC), allowing switching and deployment at much larger distances, etc. In SANs the host is responsible for mounting and maintaining any filesystems it might need on the raw block storage offered by the SAN. Network Attached Storage (NAS) has many of the same goals as SAN, using Fibre Channel fabric evolving in an environment dominated by unstructured data. NAS systems use a file sharing protocol such as NFS or CIFS to create, read, write and delete files. Both SAN and NAS installations require a significant switching infrastructure and splitof datacenter equipment into computing, networking and storage components. All of the capacity optimization technologies like parity RAID, delta snapshots, thin provisioning etc. are applicable to storage of both structured and unstructured data. Used devices are of SSD, SATA, SAS or FC types. A traditional mixed-use data center means two data fabrics: a switched Ethernet infrastructure and a switched Fibre Channel infrastructure ususlly of a “tree”-like or “star”-like architecture, with small switches at the rack level and director-class switches routing traffic between the smaller switches.

Storage management:

The term storage management encompasses the technologies and processes organizations use to maximize or improve the performance of their data storage resources. It's a broad category that includes virtualization, replication, mirroring, security, compression, traffic analysis, process automation, storage provisioning and related techniques. Storage management techniques can be applied to primary, backup or archived storage. Deployment and implementation procedures will vary widely depending on the type of storage management selected and the vendor. In addition, the skills and training of storage administrators and other personnel add another level to an organization's storage management capabilities.

Storage Management gives many benefits:

Technologies, like storage virtualization, deduplication and compression, allow better utilization of existing storage, resulting in lower costs for operating and maintain storage devices. They simplify the management of storage networks and devices, reducing overall storage operating costs.

The appropriate storage management improves data center's resilience, performance, reliability and availability, agility and resilience.

Replication, mirroring and security are often particularly important for backup and archive storage.

The primary organization involved in establishing storage management standards is the Storage Networking Industry Association (SNIA). It has put forth several important storage specifications, including the Storage Management Initiative Specification (SMI-S) and the Cloud Data Management Interface (CDMI). Storage management is also closely associated with networked storage solutions, such as storage area networks (SANs) and network-attached storage (NAS) systems.

VI. CONCLUSIONS AND RECOMMENDATIONS

To develop the educational cloud for Bulgaria based on the four pillars, taken in their integrity and interoperability using common standards.

Recommendation to use Content Delivery Network concept for the Datacenter architecture.

Distributed Denial of Service is one of the serious security threats that challenge the availability of the DC resources to the intended clients.

Recommended is the Adaptive Defense approach to cyber security delivers technology, expertise, and intelligence in a unified, nimble framework, which demands the adaptation of the security architecture to prevent today's cyber attacks and avert their worst effects.

Servers to be built exclusively for the data center, designing them so they don't include unnecessary hardware or software—reducing the number of potential vulnerabilities.

Recommendations about storage as a complex and demanding piece of the datacenter infrastructure are:

- Consolidation of storage as much as possible.
- Maximum use of green technologies, Parity RAID, thin provisioning, and tiering or caching with SATA drives
- Using centralized backup and archive resources.
- Planning for increasing levels of virtualization
- Emphasizing flexibility especially in the fabric and switching layout,
- Investing in storage systems that can be easily expanded in capacity, I/O capabilities and storage virtualization.

VII. BIBLIOGRAPHY

[1] Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. Educational Researcher, Vol. 15, No. 2. (Feb., 1986), pp. 4-14.

[2] Burkey, Roxanne E.; Breakfield, Charles V. (2000). Designing a total data solution: technology, implementation and deployment. Auerbach Best Practices. CRC Press. p. 24.

[3] Stephen Specht Ruby Lee, "Taxonomies of Distributed Denial of Service Networks, Attacks, Tools, and Countermeasures", Technical Report CE-L2003-03, May 16, 2003

[4] Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. Teachers College Record, 108(6), 1017-1054.

[5] Cosmano, Joe (2009), Choosing a Data Center (PDF), Disaster Recovery Journal,

[6] Chun-ChiehHuang, Ku, J.,A Cooperative Intrusion Detection System Framework for Cloud Computing Networks, 39th International Conference on Parallel Processing Workshops, (2010) 280 – 284.

[7] Jayadivya S K, Jaya Nirmala S, Mary SairaBhanu S, Fault tolerant workflow scheduling based on replication and resubmission of tasks in Cloud Computing, International Journal on Computer Science and Engineering, 4 (2012) 996-1006.

[8] Ricardo Lent, Evaluating a migration-based response to DoS attacks in a system of distributed auctions, Computers & amp Security, 31 (2012) 327-343.

[9] Стратегия за ефективно прилагане на информационни и комуникационни технологии в образованието и науката на Република България (2014-2020 г.) https://www.mon.bg/?go=page&pageId=74&subpageId=143

[10] [План за реализация на Стратегията за ефективно внедряване на ИКТ в образованието и науката (2014-2020 г.) Фаза 1: е-обучение - 2015-2017

https://www.mon.bg/?h=downloadFile&fileId=7173