Experience and quantitative results from the deployment of an open source production-grade virtualization platform

EKT

Panagiotis Stathopoulos

Alexandros Soumplis

Nikos Houssos

email: {pstath;soumplis;nhoussos}@ekt.gr, tel: +30 210 7273 996

National Documentation Center /N.H.R.F.

Εθνικό Κέντρο Τεκμηρίωσης/Ε.Ι.Ε.

- **Introducing NDC's requirements**
- **Description of NDC virtualization platform**
- **Performance metrics and experiences drawn**
- **Benefits devired**
- **Virtualization and FLOSS - conclusions**

National Documentation Center/Εθνικό Κέντρο Τεκμηρίωσης:

- **Uses Information Technology for providing information services for Research, Science and Technology in Greece during the last 25 years**

- **«*The National Documentation Centre (EKT) is the backbone organisation of the national infrastructure for scientific documentation, online information and support services on research, science and technology.*»**

- **Part of the National Hellenic Research Foundation (*Εθνικό Ίδρυμα Ερευνών*)**
  - NHRF: 50 years of life
  - NDC: 25 years

A multitude of services, targeted at users of different requirements and demands:

–Hellenic Dissertations Database – 2.500.000 scanned pages – *http://www.ekt.gr*

–Union Catallogues (academic libraries journals, public libraries, etc.), http://www.ekt.gr

–Develops the AVEKT Library Automation System

–Provides a Z39.50 gateway, *http://argo.ekt.gr/*

–Digitilization services

–3rd party application hosting, *http://wok.ekt.gr*

–The Greek gateway for Research, Technology and Innovation, *http://www.ekt.gr/content/*

–Variety of web sites for Research and Technology, Cultural related etc.

–Information services provided locally to end users

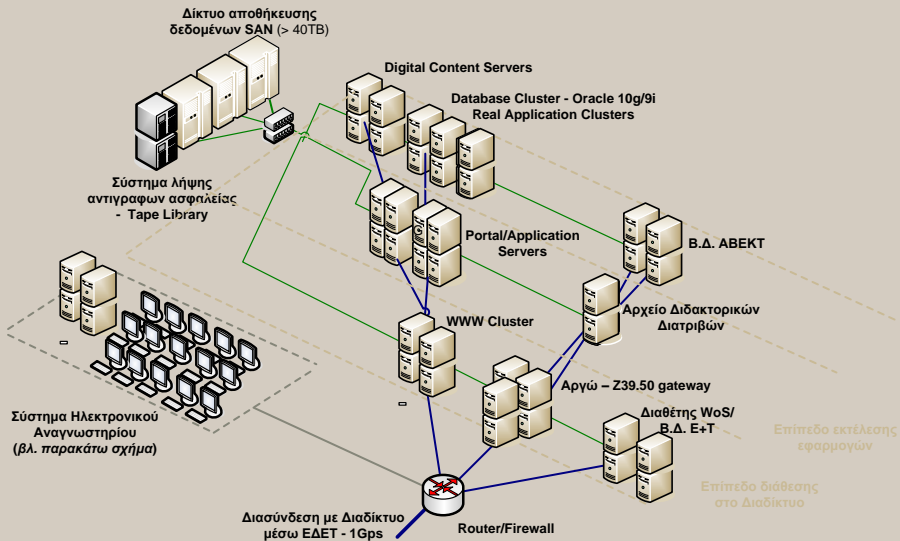–*Digital Library, manages the wireless hotspot of NHRF, http://www.broadbandcity.gr*

Activities under development:

–Open access services: NHRF institutional repository under development, open access journals, *http://www.byzsym.org* & *http://www.openaccess.gr*

- IT infrastructure supports NDC's services. **Main characteristic**:
  - Many different categories of end user services
  - In order to provide information services for S&T a whole range of **diverse** of applications and services is needed, e.g. :
    - Internet connectivity and services (www,email, ftp, etc), in house development,3rd party application hosting (e.g. WOS/WOK), digitalization, metadata production, digital library, electronic reading room.
  - Each of them poses **different requirements**,
  - Frequently facing **rapidly changing hw and sw requirements**
    - especially in hosting 3rd party applications
  - Moreover, services and applications **evolve** over time
    - Different parts are made at different time periods, many times with different underlying technologies and assumptions
  - **A single size does not fit all**
    - Different technologies for bibliographical systems or small web sites,e.g ruby on rails or enterprise databasese.g. J2EE + Oracle or Postgress
  - The aforementioned characteristics are frequently overlooked as major IT issues however they are common enough in a variety environments.

- Agile development under pressing deadlines
- Heterogeneity in platforms and technologies used
- "Heavy" customization and extensions on open source software
- Frequent evaluations of software (mostly open source)
  - During the last 12 months:
    - CMS/Portal platforms: dotCMS, Liferay, OpenCMS, Jahia, Alfresco, Plone, eXo, Jetspeed, JBoss Portal, OpenPortal, Magnolia
    - Digital repositories: EPrints, DSpace, Fedora, Fez
    - Journal publishing platforms: OJS, DPubs, Topaz
    - Federated search: dbWiz, LibraryFind
    - Wikis: MediaWiki, DekiWiki
  - Various technologies: Java/JEE, PHP, Ruby on Rails, Python, .NET, Perl, …

NDC infrastructure before virtualization

ΣΥΣΤΗΜΑ ΔΙΑΘΕΣΗΣ ΔΙΑΔΙΚΤΥΑΚΩΝ ΥΠΗΡΕΣΙΩΝ ΕΚΤ

Δίκτυο αποθήκευσης δεδομένων SAN (> 40TB)

Digital Content Servers

Database Cluster - Oracle 10g/9i
Real Application Clusters

Σύστημα λήψης αντιγραφων ασφαλείας - Tape Library

Portal/Application Servers

Β.Δ. ΑΒΕΚΤ

WWW Cluster

Αρχείο Διδακτορικών Διατριβών

Σύστημα Ηλεκτρονικού Αναγνωστηρίου (βλ. παρακάτω σχήμα)

Αργώ – Z39.50 gateway

Διαθέτης WoS/ Β.Δ. E+T

Επίπεδο εκτέλεσης εφαρμογών

Επίπεδο διάθεσης στο Διαδίκτυο

Διασύνδεση με Διαδίκτυο μέσω ΕΔΕΤ - 1Gps

Router/Firewall

Διασύνδεση Fibre Channel 2Gps

Διασύνδεση Gigabit Ethernet 1Gpps

Εξυπηρετητές σε διαμόρφωση Υψηλής Διαθεσιμότητας (Cluster)

Εξυπηρετητής Υψηλής Απόδοσης

Storage Area Network

Router/Firewall

Τερματικό Χρηστών Η.Α.

Smart Card Reader

Services offered: Web site development και hosting,,3rd party applications hosting, e.g. WoS database, Internal user support και Internet Applications (network, helpdesk, etc), Union Catalogues, Databases development, etc., Electronic reading room, smart cards, thin clients,κ.α. Wireless hotspot, Conference rooms.

- **Three tier architecture**:
  - Web, application, database layer
  - Bibliographic applications
  - Applications and database servers could vary depending on the service needs.
  - Fully redundant

- **Satisfactory slowly changing (static) homogeneous environment** - limited flexibility at:
  - Allocating resources
  - Porting applications to different hardware platforms
  - Dynamically dimensioning services
  - New functional requirements and demands

- NDC equipment at NDC/NHRF's 100m^2 computer room: 46KVA UPS supply, 150.000BTU cooling supply.

- Hosts also the HellasGrid/EKT node operated from the HellasGrid team and featuring with independent power and cooling supplies.

- Wikipedia:
  - "**virtualization** is … refers to the <u>abstraction</u> of computer resources... hides the physical characteristics of computing resources. This includes making a single physical resource (such as a <u>server</u>, an <u>operating system</u>, an application, or <u>storage device</u>) appear to function as multiple virtual resources; " (http://www.wikipedia.org")
- Features
  - **Independency** of applications and OSes from particular hardware platform installed
    - Capability for the transparent movement of computing systems from host machine to host machine
  - Greatly increases **flexibility** and overall availability
  - Long term acquisition and maintenance **cost reduction**
  - **Transparent management** of raw computing resources
  - **Application porting and preservation** is highly simplified

- Practical requirements:
    - System heterogeneity:
        - A realistic goal is not to eliminate different applications, services tailored for each application but to manage this heterogeneity effectively
        - Each application and framework has a **purpose**
    - Empirical experience from legacy applications:
        - Significant dependency of end services to the whole hardware/OS/application software stack. Layers of OS patches, application patches and modifications, especially for closed source ones, frequently bind the application to a particular piece of hardware.
    - Application development requirements:
        - Requires vigorous testing and different development environments especially when time constraints are crucial
    - Limited annual budgets
        - Need for flexibility, and for resolving needs that can not be predicted or dimensioned beforehand.
    - Data Center Infrastructure
        - Beyond simple Computer Rooms and towards information processing factories.
        - Cooling and power facilities are significant, expensive and complex in their installation.
            - Frequently surpassing in complexity and issues arising the pure IT issues faced.
    - **A decision was made to evaluate and if possible employ open source virtualization**

Virtual Platform Migration Plan

A' Testing Period (1 month)
- Install CentOS 5 + Virtualization
- Install VMs with CentOS 5
- Hardware and Software compatibility check.
- Performance Metrics.

B' Testing Period (3 months)
- Install/Test administration server for management and monitoring (Nagios, Cacti)
- Install/Test application server for testing applications development.
- Install/Test LAMP staging server for testing of production web sites.

A' Production Period (3 months)
- Migration of development servers from physical machines (Mongrel, Tomcat, Jboss, MySQL, Mono application Server)
- Testing of Internal Mail Server (cloning of production server, working in parallel on the backend zone)
- Migration of internal support applications (SVN, Mantis Bug Tracker, Wiki)

B' Production Period (2 months)
- Migrate Internal Mail Server
- Install production LAMP Servers for migration of production web sites (www.broadbandcity.gr, academyofathens.ekt.gr, etc)
- Install production application servers (Tomcat and Mongrel) to support production web sites (thesis.ekt.gr, coins.ekt.gr, etc)

C' Production Period (2 months)
- Migration to full virtualized 3-tier architecture by installing Database servers (MySQL, Postgres, Oracle).
- Integrate Mail Server with existing Directory infrastructure.
- Install Backup Server

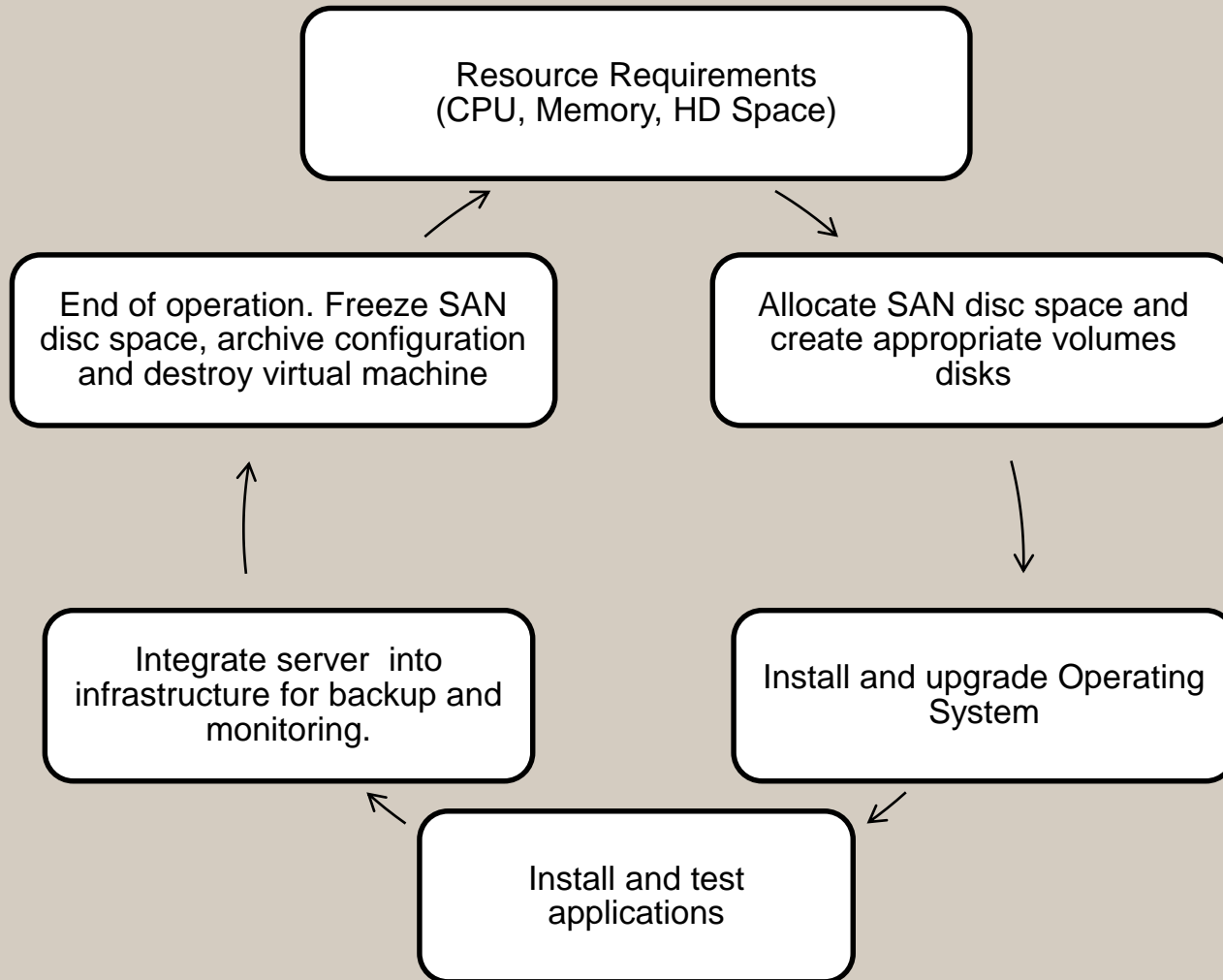# Virtual Platform Migration Plan

- **A' Testing Period**
  - Duration: 1 month
  - Evaluation of different Virtualization solutions.
    - We evaluated Xen, VMware and MS Virtual Server.
    - Election Criteria: Performance, Licensing, Support (Commercial and Community)
    - We chose Xen on CentOS 5. Full Open Source Software.

- **B' Testing Period**
  - Duration: 3 months
  - We have installed Virtual Machines in order to test and evaluate the usual usage scenarios related to NDC.
  - We gathered performance metrics and we also evaluated administration and maintenance cost.

Virtual Platform Migration Plan

- **A' Production Period**
  - Duration: 3 months
  - We brought out of production several physical machines (Development Servers) due to problems regarding stability, reliability, performance and security.
  - Internal support software consolidation and upgrade (svn, bugtracker).

- **B' Production Period**
  - Duration: 2 months
  - Migration of critical production servers (ex. internal mail server).

- **C' Production Period**
  - Duration: 2 months
  - Completeness of a full virtualized 3-tier architecture.
  - Separation of Databases into different virtual machines. Different servers for legacy , Oracle applications, MySQL and Postgres.

```
                    ┌────────────────────────────┐
                    │  Resource Requirements     │
                    │  (CPU, Memory, HD Space)   │
                    └────────────────────────────┘
```

Resource Requirements (CPU, Memory, HD Space)

End of operation. Freeze SAN disc space, archive configuration and destroy virtual machine

Allocate SAN disc space and create appropriate volumes disks

Integrate server into infrastructure for backup and monitoring.

Install and upgrade Operating System

Install and test applications

Virtual Machine Life Cycle

- No difference from the Life Cycle of a physical server.

- **Resource management flexibility**
  We have a minimalistic approach. We allocate the minimum resources possible and we increase them dynamically depending on demand.

- **Install through Templates.**
  No need to install OS from scratch. Ready templates and preinstalled images due to common virtual hardware.

- **Safe upgrade and patching**
  We have the ability to revert back to a snapshot backup almost immediately in case of failure or incompatibility.

- **Freeze Virtual Machine.**
  When we no longer need a virtual machine we do not have to delete it. We freeze the SAN disc space it uses and save its configuration. In case we need it again in the future we can bring it up in minutes.

Performance metrics - Some Numbers

- **2 physical servers host 16 virtual machines for development and production:**
  - 3 Database Servers (Oracle 9i, MySQL)
  - 5 Application Servers (Tomcat, Mono)
  - 4 LAMP Servers
  - 1 Internal Mail Server (Sendmail, Dovecot)
  - 1 NMS Server (Nagios, Cacti for more than 1500 performance metrics)
  - Backup Server (Bacula)
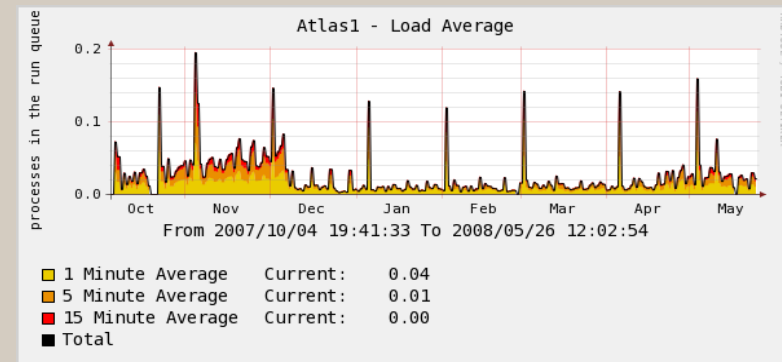 - Log Server (collect and parse logs for 40 services and 20 web sites)
- **Server Specifications:**
  - 4 x 2,2GHz AMD Dual Core CPU
  - 16GB Ram
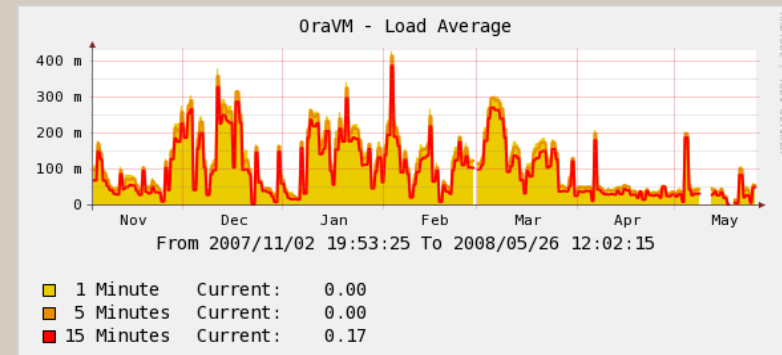 - 3,5 TB SAN space (shared)
  - 2 FC Emulex PCI Controllers.
- **Very Low CPU Usage overall.**
  - 10% Average CPU usage on the physical server.
  - 20% Average CPU usage on the virtual server.
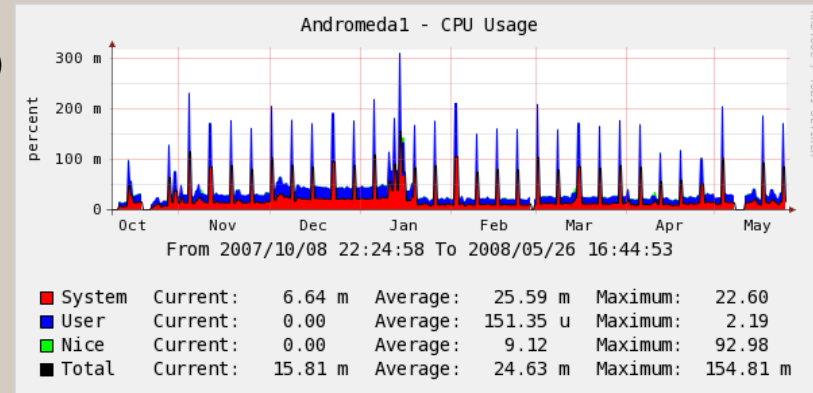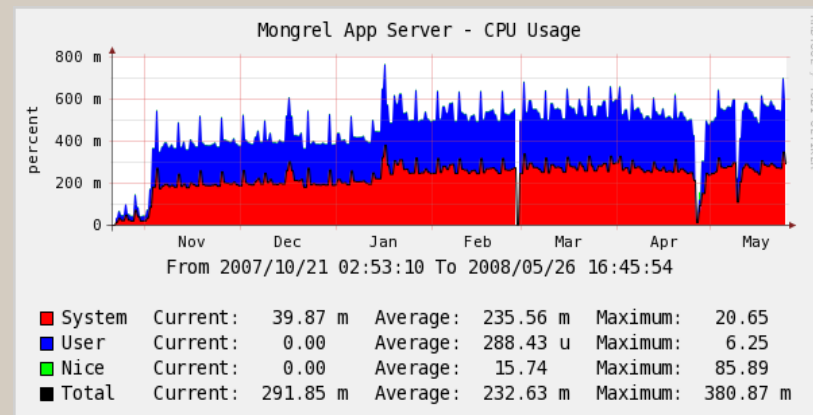
Physical Server

Virtual Server

Performance metrics - Some More Numbers

- **2 physical servers host 4 production servers:**
  - 2 LAMP / SVN Server
  - 1 Application Server (Tomcat)
  - 1 Application Server for Ruby on Rails (Mongrel)
- **Server Specifications:**
  - 2 x 2,2GHz AMD Dual Core CPU
  - 4GB Ram
  - 3,5 TB SAN space (shared)
  - 2 FC Emulex PCI Controllers.
- **Almost no CPU usage (less than 1%)!!!**

Physical Server

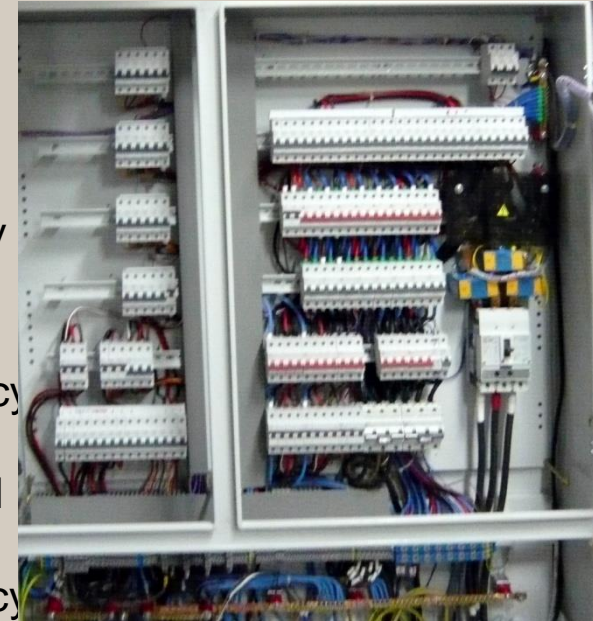Andromeda1 - CPU Usage

From 2007/10/08 22:24:58 To 2008/05/26 16:44:53

■ System	Current:	6.64 m	Average:	25.59 m	Maximum:	22.60
■ User	Current:	0.00	Average:	151.35 u	Maximum:	2.19
■ Nice	Current:	0.00	Average:	9.12	Maximum:	92.98
■ Total	Current:	15.81 m	Average:	24.63 m	Maximum:	154.81 m

Virtual Server

Mongrel App Server - CPU Usage

From 2007/10/21 02:53:10 To 2008/05/26 16:45:54

■ System	Current:	39.87 m	Average:	235.56 m	Maximum:	20.65
■ User	Current:	0.00	Average:	288.43 u	Maximum:	6.25
■ Nice	Current:	0.00	Average:	15.74	Maximum:	85.89
■ Total	Current:	291.85 m	Average:	232.63 m	Maximum:	380.87 m

- Lack of GUI applications to manage virtual machines.

- Steep learning curve.

- Lack of authorization control on host machines. We cannot authorize operators to power on and off only certain virtual machines across infrastructure.

- Xen does not support MS Windows if hardware does not support VT extensions.

- Performance penalties exist comparing to physical servers but in most cases are not noticeable by end users and applications.

- Not intended to give an absolute value but an estimation of benefits with a verifiable manner.

- Assumptions:
  - Compare:
    - *Current system*: virtualization applied and legacy systems retired

    to:

    - A) *Initial System*: no virtualization including legacy systems
    - B) *Realistic alternative scenario*: fully functional equivalent system without virtualization
    - C) *Worst case scenario*: no virtualization + legacy systems

- Results:
  - Nominal power consumption,
  - PDU sample metered power consumption.
    - Include cooling power requirements

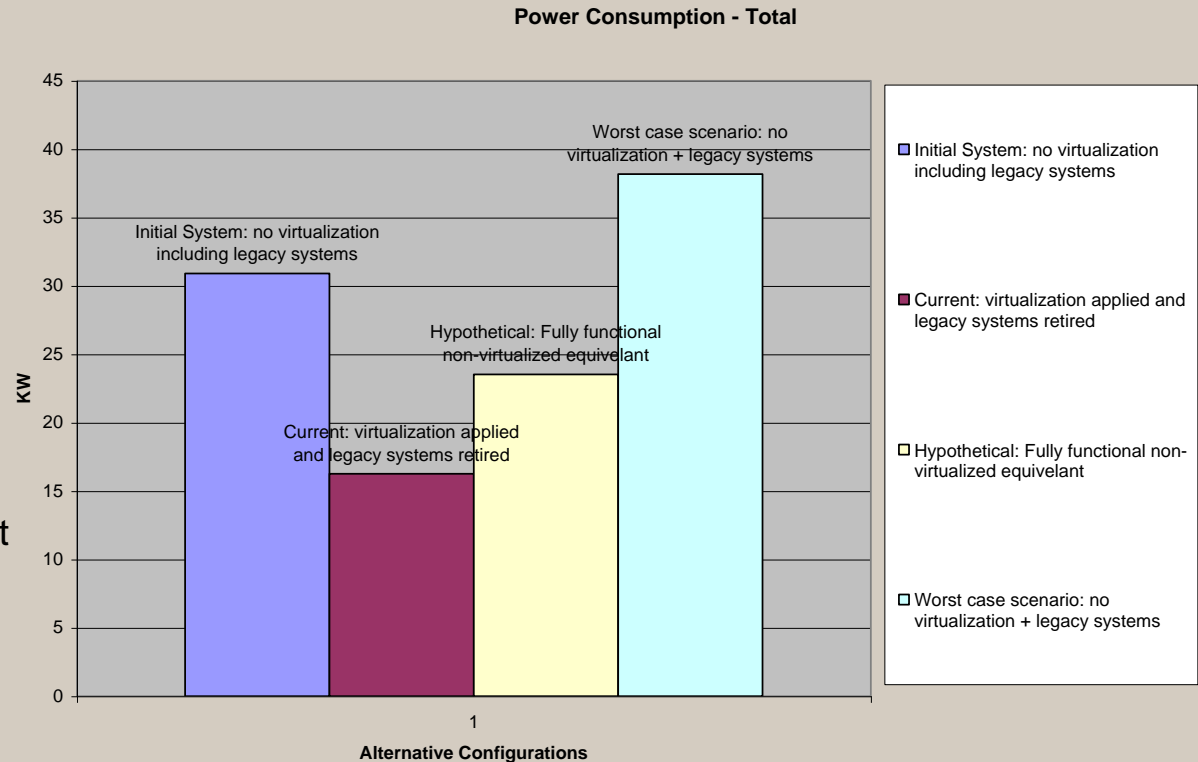- **Figures from specifications:**
- **8-core servers:**
    - Demand on cooling:
      850 BTU/hr per server
    - Demand on power for cooling:
      249W per server
    - Nominal power supply: 760W
- **4-core Servers**
    - Demand on cooling:
      850 BTU/hr per server
    - Demand on power for cooling:
      249W per server
    - Nominal power supply: 500W
- 350W are calculated for equivalent virtual servers – benefits already **UNDERESTIMATED**

Power Consumption - Total

Legend
Initial System: no virtualization including legacy systems
Current: virtualization applied and legacy systems retired
Hypothetical: Fully functional non-virtualized equivelant
Worst case scenario: no virtualization + legacy systems

Current/Initial: 47%, Current/Equivelant: 30%, Current Worst case: 67%

Performance metrics - Measured Power Consumption Savings

Power Consumption

- **Measurements from sample PDUs (*Power Distribution Units*)**

- **In idle state both type of sample servers consumed %50 of the nominal power.**
    - **1,6A for 8-core server**
    - **0,8A for 4-core server**

- **Assumption: similarly for the cooling power required.**

- **Percentage relationships remain**

Maximum (Nominal) Power Consumption - Total
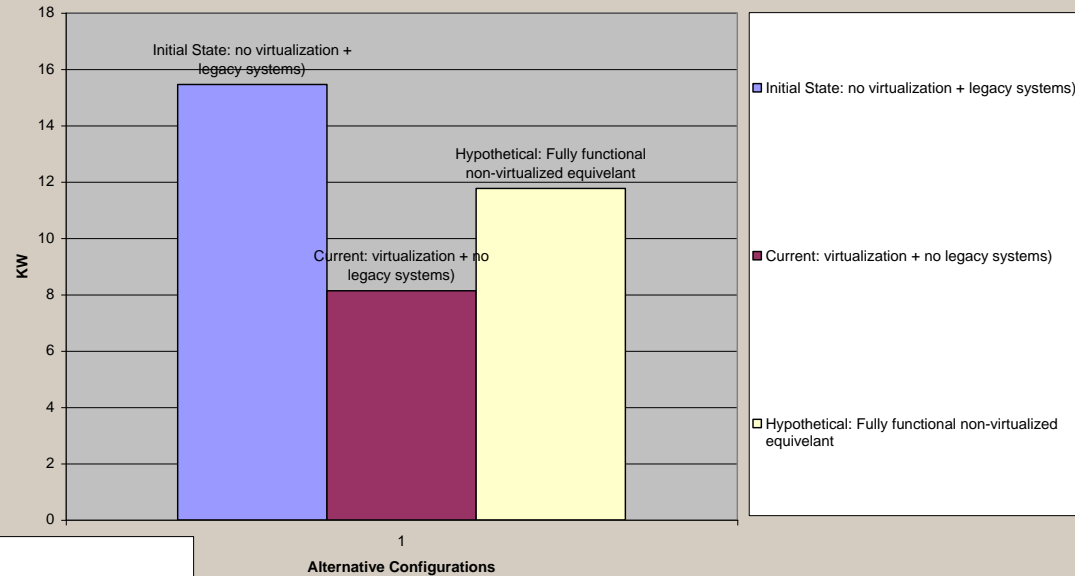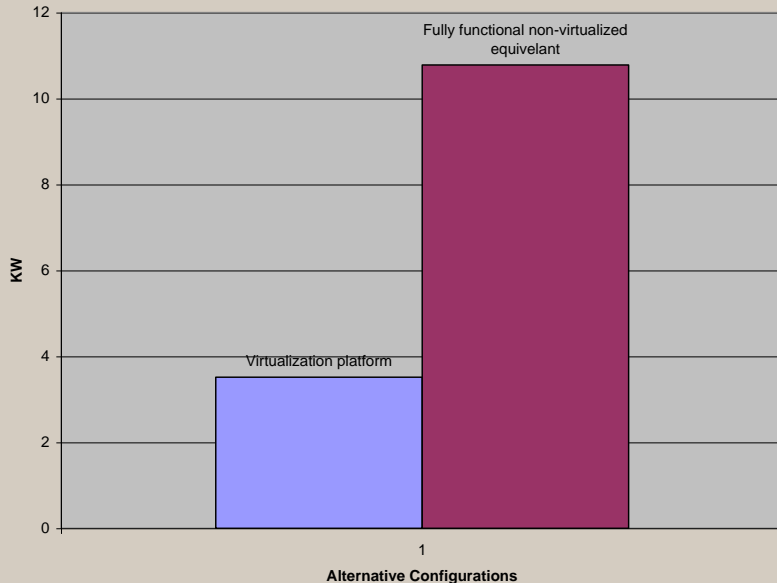
- **When counting only the virtualization platform actual power consumption estimated at 3,5KW instead of 11KW for a fully equivalent platform**

Is the nominal (maximum) power supply figures an overestimation?

- **NO – Datacenter design depends on this figure for dimensioning power and cooling supplies.**

- **Actual consumption although significant – since operating savings are dependent on this - is not definitive for specifying datacenter design parameters and costs.**

- Dedicated virtual machines for specific application servers
  - Avoid installation delays due to conflicts / incompatibilities
  - Less constraints in resource allocation (e.g., memory)
  - No chance of accidental problems due to different teams working on the same machine at the same time
- Flexibility in:
  - Production installations -> increase/decrease resource allocation (e.g., memory, disk) on-demand to reflect changing scalability requirements
  - Testing configurations
    - Example: stress test app servers with different memory sizes
    - Easier to setup test environments for distributed systems (e.g., SOAs)

Virtualization Benefits

- **Server Provisioning**
    - Server Deployment based on templates.
    - Stream line OS upgrades with a common upgrade procedure for all virtual servers.
    - Common hardware (Virtual Hardware).
    - Common network tuning and configuration.
- **Cut down maintenance costs.**
    - Maintain and support hardware only for 4 servers instead of 20!
    - Minimize downtime and troubleshooting time regarding hardware related issues.
- **Manageability**
    - Dynamic resource allocation. Ability to respond to burst traffic.
    - Virtual machine Live Migration --- No downtime in case of physical server error or preventive maintenance and update.
    - Manage and administer storage out of the virtual machines (file system integrity check, alter partition layout, increase size of volumes and file systems).

- **Virtualization software + virtualized OSes.**
  - **Total software cost of adding virtual machines == 0.**
    - **OS license**
    - **Backup license**
    - **Antivirus license**
    - **Etc.**
  - **This cost analysis drives a new paradigm for virtual machines roll out**
    - **Software devices instead of servers**
    - **The middle ground between overly fragmentation and flexibility should be found**
- **Suitability for the public sector:**
  - **Managerial difficulties in obtaining new equipment**
  - **Limited yearly budgets**
  - **Development of organizations through mid term projects, e.g. Κοινωνία της Πληροφορίας, where precise needs are difficult to be accurately predicted.**
    - **Virtualization could be a key solution for realizing such mid term and long term projects without over- or under-estimating hw needs.**
  - **IT managers could surely appreciate the reduced need for procurement procedures and negotiations.**
- **Overall conclusion: FLOSS virtualization is mature for production-grade environments.**

Thank you for your attention!

http://www.ekt.gr
http://www.openaccess.gr