Model of Cloud Computing Realisation on the Base of Infrastructure IaaS

Peter PENIAK[1], *Maria FRANEKOVA*[1], *Iveta ZOLOTOVA*[2]

[1]Department of Control and Information Systems, Faculty of Electrical Engineering, University of Zilina, Univerzitna 1, 010 26 Zilina, Slovakia
[2]Department of Technical Cybernetics and Artificial Intelligent, Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Letna 9, 042 00 Kosice, Slovakia

peter.peniak@fel.uniza.sk,maria.franekova@fel.uniza.sk,iveta.zolotva@tuke.sk

DOI: 10.15598/aeee.v14i2.1565

Abstract. *The paper deals with the problems of cloud computing applied for industrial applicationson the ground of practical experiences in certain manufacturing corporation. The main part of paperis orientated to proposal of the numerical model on the base of Infrastructure as a Service (IaaS) and its mathematical description. In addition the model has been extended to include the requirements of mission critical systems with real time behaviour and fail-safe features. The models were realised via virtualisation software Hypervisor which creates a group of available virtual resources through physical infrastructure, which can be offered to customers. Proposal solution enables to create a proper sizeof cloud infrastructure for hardware provisioning according to customer requirements.*

Keywords

Cloud computing, control and information systems, host up, hypervisor, IaaS, Industrial applications, model, SaaS, virtual machine.

1. Introduction

Cloud computing is a modern term applied to large hosted datacenters that offer various computational services on a "utility" basis [1], [2]. The general understanding of cloud computing is related to an on-demand service model by which various different resources (hardware, software, and services) are delivered over the network. This networkcould be the intranet of a company or the internet when the service is ordered from an external provider [2], [3]. The resources involved in cloud computing primarily are computational resources (e.g. server, storage, network, software), and they are primarily provided as services for the users according to defined rules [4], [5]. Currently, there are three possible cloud service layers that can be used in combination to build a full end-to-end cloud.

- Software as a Service (SaaS) - offers an application on demand over the network.
- Platform as a Service (PaaS) - providers a complete development platform including the necessary built-in services, such the databases and the middle-ware on demand over the network.
- Infrastructure as a Service (IaaS) - a service which is offering the hardware and the software components, such as computers and the storage systems.

A cloud may be hosted by an enterprise or a service provider and to provide services to the clients represented by PCs, tablets andsmartphones. Datacenters can be distributed geographically to achieve higher redundancy and service independency on physical location of physical host systems. There are three deployment models defined for cloud computing [6]:

- Public Cloud - that is available to clients from a third party service provider via the Internet.
- Private Cloud - implemented internallywithin the company or organization and its private network.
- Hybrid Cloud - a combination of public and private cloud.

2. Approach on the Base of IaaS Service Layer

This service layer delivers computer infrastructure (typically hardware, storage, servers and data center space or even network components to customer). It may include basic software within provided infrastructure. IaaS enables users to self-provision these resources in order to run own platforms and applications. It is also known as Hardware as a Service (HaaS). However more advanced services can be provided on the same infrastructure as well such as database and WEB provisioning (PaaS) or even selected applications (SaaS) as it is shown in Fig. 1. Generally there are several options for providing IT infrastructure via Cloud. The most commonly used method is hardware virtualization. Virtualization is based on a common network infrastructure and physical servers (hosts) that can be installed in various locations and facilities. This infrastructurewith help of virtualization software called hypervisor creates a group of available virtual resources which can be used through so called virtual machines (VM). It is up to a customer which an operation system and applications are installed on provided VM.

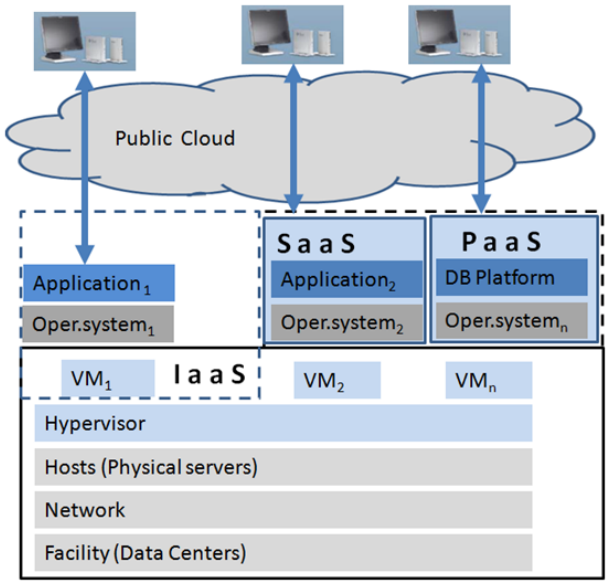

Fig. 1: Solution of service layers of cloud computing.

3. Model Realisation and its Mathematical Description

Virtual Machine can represent a logical server or desktop object which simulatesa logical hardware object thatbehaves exactly like a physical computer or a storage. It has its own "Virtual" Processor unit (VP), Virtual memory RAM (VRAM), Virtual Network Interface Card (VNIC) and virtual hard disk (VFS) and runs as an isolated operating system installation on shared hardware, called a host. The process of virtualization works by inserting "hypervisor", directly into the computer hardware. Remote workstations (PCs, thin-clients, workstations) can access to virtual machines via terminal software with selected protocol for example Remote Desktop Protocol (RDP) [7].The proposed model of IaaS is shown in Fig. 2. Host servers provide their physical hardware resources which are mapped by Hypervisor to virtual resources. Virtual resources are assigned to defined virtual machines VMs consisting one or more virtual processors, requested amount of RAM, disk space and virtual network card. The hardware sizing for needed VM resources must be in a correlation with available physical resources of the host computers. The correct configuration of host servers is a prerequisite for implementation of server virtualization in IaaS. Based on the provided model we can define the following parameters:

- N_{hosts} - number of physical servers,
- $N_{\text{hosts}_{\text{VP}}}$ - number of physical servers as to VProcessors,
- $N_{\text{hosts}_{\text{VRAM}}}$ - number of physical servers as to VRAM,
- $N_{\mu\text{P}}$ - number of processors in host (sockets),
- N_{Cores} - number of cores per processor,
- VP_{ij} - virtual processor of VM,
- RAM_{host} - RAM size in Hosts (GB),
- RAM - RAM size needed for host itself (GB),
- n - number of VMs,
- m - number of virtual processors per VM,
- VFS_i - size of requested virtual file systems,
- FS - requested file system for physical host itself (GB) or overall size of file systems connected to hosts (GB).

The number of physical host servers is most significant design parameter for IaaS. It can be defined by:

$$N_{\text{hosts}_{\text{VP}}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i} \text{VP}_{ij}}{N_{\text{Cores}} \cdot N_{\mu\text{P}}}, \tag{1}$$

or amount of virtual memory by all requested virtual machines,

$$N_{\text{hosts}_{\text{VRAM}}} = \frac{\sum_{i=1}^{n} \text{VRAM}_i}{\text{RAM}_{\text{host}} - \Delta\text{RAM}}. \tag{2}$$

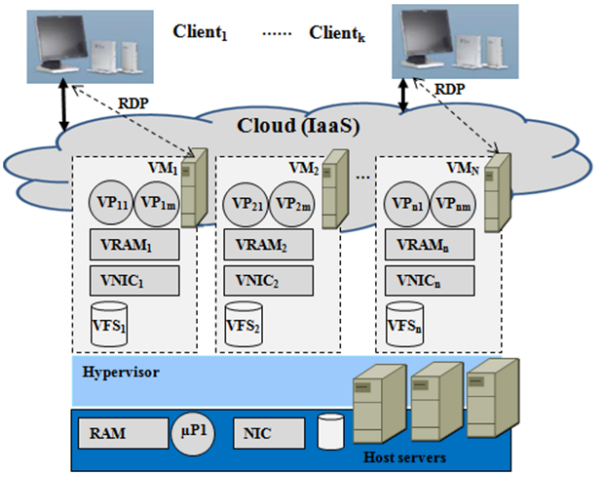

Fig. 2: Proposed model of cloud computing on the base of IaaS.

The final result is calculated as a maximum from parameters $N_{\text{hosts}_{\text{VP}}}$ and $N_{\text{hosts}_{\text{VRAM}}}$ but has to be rounded to the nearest higher value:

$$N_{\text{hosts}} => \max\left\{N_{\text{hosts}_{\text{VP}}}, N_{\text{hosts}_{\text{VRAM}}}\right\}. \qquad (3)$$

In addition to the requested amount of the physical host servers a size of file system is also needed for proper setting of overall IaaS infrastructure. It has to be calculated according to:

$$\text{FS} = \sum_{i=1}^{n} \text{VFS}_i + \Delta\text{FS} \times N_{\text{hosts}}. \qquad (4)$$

The proposed model can be used for calculation of the number of requested host computers with associated amount of file space. This calculation enables configuration of IaaS service provisioning for generic applications.

4. Extended Model Realisation for Fail-Safe and Real Time Support

In case of need to support mission critical systems in cloud environment,generally, there are two critical requirements that have to be considered supported for hardware provisioning via Cloud Computing and IaaS:

- Support of Fail-Safe operation (for safety-related applications) [8], [9].
- Real time features.

4.1. Proposal of Assurance for Fail-Safe Support

In case of a failure of host computer all the hosted virtual machines are affected in the same time. It would mean that the service is significantly affected and required hardware is not provided to cloud subscribers until host is recovered and reachable again. This behavior cannot be accepted by any mission critical system [10]. Therefore it is a nature feature of the Cloud Computing that the service delivery is independent on the physical position of a hardware component within the cloud. Switching of the virtual machines between different host systems is allowed. In case of a failure of a host computer all the virtual machines have to be automatically moved and restarted on a different physical host server. The switching of VM machines between the host systems is managed by dedicated virtualization software. However, the common data store with all VM server snapshots has to be available to enable process with fail-over procedure. The described feature is illustrated in Fig. 3, where virtual machine VM1 has been broken and virtualization management software has initiated the fail-over procedure with restart of VM1 server on different physical server. This situation would remain till the original server is again brought to production state and VM1 can migrate back to the former host. This approach shall be reflected in previously proposed model. It means that number of the physical servers must be increased with an additional host server, so that VMs can be redirected to the spare server in the case of failure of any of host computers. Therefore amount of needed physical servers can be defined by:

$$N_{\text{hosts}} => \max\left\{N_{\text{hosts}_{\text{VP}}}, N_{\text{hosts}_{\text{VRAM}}}\right\} + 1. \qquad (5)$$

Another option is to provide fail-over operation between hosts in two independent server rooms. In addition to the standard switching, the virtualization has to deal with situation when the data-stores are located in the different storage systems.

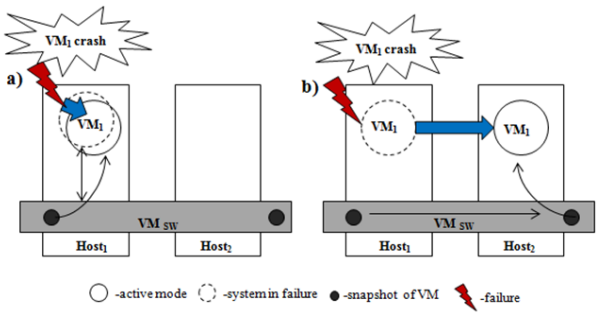

Fig. 3: Extended model for VM machines with Fail-over procedure.

In such case, there is a need to replicate data stores between data centers with using an additional tech-

nology, which has to perform switching of data stores and application data between control rooms in case of failure in one control room.

4.2. Proposal of Assurance for Real Time Conditions

The hypervisor layer presents multiple sets of "virtual hardware" to defined virtual machines. For hardware virtualization, an additionalvirtualization layer has to be added between physical hardware and a virtual machine with dedicated operation system and requested control application software. The new layer might cause not only an additional latency, but could significantly influence overall response time by compromising requested deterministic framework. Therefore, we can propose an approach based on a real-time capable Hypervisor (RTH), as it is illustrated in Fig. 4. RT hypervisor would have to control assignment of physical resources (RAM, NIC, μPC) to virtual machines (VM1,...,VMN) equally in precise time slots, emulating token based approach among virtual machines. It means that host resources are assigned to each VM cyclically. This approach enables deterministic response time and real-time behavior. According to proposed solution, a numerical model can be created to evaluate basic interaction of Hypervisor module with particular VM servers. Having taken into consideration the request to support real-time properties by Hypervisor, let us assume that the model can be represented by finite population queuing model that is often used for interactive systems [11]. Hypervisor subsystem consists of queue for centrally managed hardware resources. Each VM server (VMi) interacts with Hypervisor (HV1) and can be in either an idle status, without request for resources, or requesting status, or receiving status. In order to achieve real time properties, a time-sharing model is suggested. Hypervisor resources are assigned equally among all requesting VM servers. After solving the first assigned interaction, Hypervisor serves another interaction in queue.

The performance can be evaluated by service rate (S_r), which is defined as a performance of Hypervisor (HV) instantly assigned among all interactions for given physical resource, as defined by:

$$S_r = \frac{\mu}{n}, \tag{6}$$

where:

- S_r - Hypervisor service rate,
- n - number of interaction among Hypervisor and VMs,
- μ - Hypervisor performance per resource (MIPS).

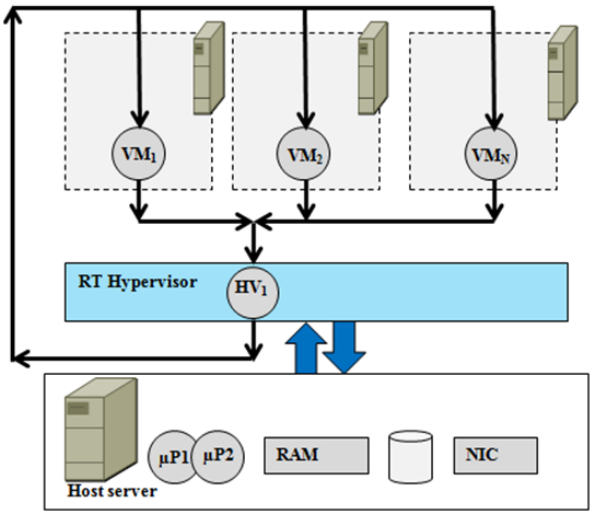

Fig. 4: Extended model for RT hypervisor.

The proposed numerical model of finite population queuing system can be expressed by the following equations:

$$W = \frac{N \cdot E(s)}{\rho} - E(t), \tag{7}$$

$$\rho = 1 - p_0, \tag{8}$$

$$\lambda_T = \frac{\rho}{E(s)}, \tag{9}$$

$$p_0 = \frac{1}{\sum_{n=0}^{N} \frac{N!}{(N-n)!} \cdot \left(\frac{E[s]}{E[t]}\right)^n}, \tag{10}$$

where:

- W - response time of Hypervisor,
- $E(s)$ - service time per VM interaction,
- $E(t)$ - idle time between finished and new interaction,
- N - amount of VM servers,
- ϱ - Hypervisor utilization,
- p_0 - probability that Hypervisor is idle,
- λ_T - throughput in interactions per VM unit time.

The most important parameter, which is used to describe performance, is overall response time of Hypervisor, represented by Eq. (6). It is calculated from assigned service time per each interaction $E(s)$, which is multiplied by number of all VM hosts (N) and divided by Hypervisor utilization and time to next request for interaction $E(t)$. Hypervisor utilization is related to probability that hypervisor is not idle and there are requirements for interactions (n) from possible VM hosts

(N), based on Eq. (8). Another useful parameter that represents capabilities of hypervisor is its throughput (λ_T), in interactions per assigned service time for virtual machine.

5. Service Contract for Provisioning of IT Infrastructure via IaaS

The demands of customers with respect to the number of requested virtual machines (client, servers), versions of operation systems and detailed system parameters (number of processors, RAM size,...), have to be specified in form of "Service contract", which is valid for provided IaaS services of Cloud Computing. Having defined the models for provisioning of IT infrastructure from previous chapters, we can apply these results also fora service contract specification.The proposed service contract is based on Service scope and Service Level Agreement (SLA), as shown by Fig. 5.

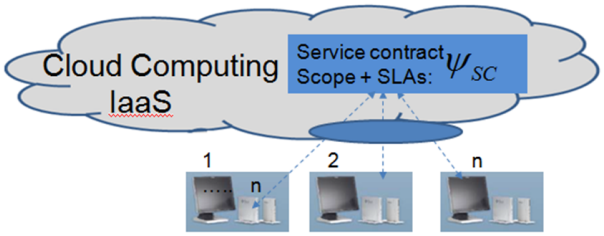

Fig. 5: Specification of service contract for IaaS.

The service scope is defined by provided infrastructure objects and their parameters as expressed by equation Eq. (11). Quality of the service is specified by key-performance indicators (KPI), which are to be delivered and kept during service provision. The system response time (W) and throughput (λ_{Ti}), which is represented by number of interactions (I/0) per requested virtual machine (VM), are proposed as typical KPIs of SLA.

$$\psi_{SC} = \{\mathrm{VM}_i, \mathrm{KPI}_i\}, i = 1, \ldots, n, \tag{11}$$

$$\mathrm{VM}_i = \{\mathrm{VP}_{ij}, \mathrm{VRAM}, \mathrm{VFS}_i\}, \tag{12}$$

$$\mathrm{KPI}_i = \{W_i, \lambda_{Ti}\}, \tag{13}$$

where:

- ψ_{SC} - service contract of requested service,
- KPI_i service level agreement per service.

6. Obtained Results

The proposed model was applied for verification of alredy implemented VM infrastructure without considerng the required disk space and service level agreement with KPIs. The simplified model is represented according to equation Eq. (14).

$$\psi_{SC} = \{\mathrm{VM}_i\}, \; i = 1, \ldots, n, \tag{14}$$

and we can define a service contract just as number of defined virtual machines (VM = 54) with required virtual processors (VP = 2) and virtual RAM (VRAM = 16 GB). The host server infrastructure (e.g. HP BL 390) is represented by the following parameters:

- $N_{\mu\mathrm{P}} = 2$,
- $N_{\mathrm{Cores}} = 8$,
- $\mathrm{RAM}_{\mathrm{host}} = 196$,
- $\Delta\mathrm{RAM} = 4$.

Next parameters were determined according to Eq. (15) to Eq. (17):

$$N_{\mathrm{hosts}_{\mathrm{VP}}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathrm{VP}_{ij}}{N_{\mathrm{Cores}} \times N_{\mu\mathrm{P}}} = \frac{54 \cdot 2}{8 \cdot 2} = 7, \tag{15}$$

$$N_{\mathrm{hosts}_{\mathrm{VRAM}}} = \frac{\sum_{i=1}^{n}\mathrm{VRAM}_i}{\mathrm{RAM}_{\mathrm{host}} - \Delta\mathrm{RAM}} = \frac{54 \cdot 16}{196 - 4} = 5, \tag{16}$$

$$N_{\mathrm{hosts}_{\mathrm{VRAM}}} = \max\{N_{\mathrm{hosts}_{\mathrm{VP}}}, N_{\mathrm{hosts}_{\mathrm{VRAM}}}\} + 1 = 8. \tag{17}$$

After applying the proposed numerical model, provision of IaaS with 54 virtual machinescan be achieved by 8 host servers, based on given hardware parameters of host servers and used fail-safe approach (+1 redundant server). In contrast to applied model, the real implemented host infrastructure, for given specification, was based on 10 host servers (5 plus 5) that were installed in two data processing centers, having a free unused capacity that might be used for future service extension.

7. Conclusion

The paper dealt with cloud computing and possibilities to use Infrastructure as a Service for hardware provisioning through hardware virtualization. The model is based on so called Virtual Machine (VM) representing logical server object or desktop. The main focus was paid on creation of the numerical model which is proposed as a tool for a calculation and sizing of physical infrastructure. The proposed model has been extended so that the requirements of mission critical systems such as fail-safe behavior and real-time features could

be supported as well. This model therefore provides the required parameters for hardware sizing including additional hardware components for VM move procedure. Moreover it includes the finite population queuing model to concentrate on overall system throughput and response time of Virtual machines. The both models can be used for more detailed analysis of virtualization and its influence to mission critical systems. However, this is just the first step that is to be extended with service level agreement model and KPIs. The simplified model was succesfully verified on existing host infrastructure where the number of required physical host servers is very close to the real number of used servers considering practical reserve for next extensions.

Acknowledgment

This work has been supported by the Educational Grant Agency of the Slovak Republic - Projects KEGA Number: 008ZU-4/2015: Innovation of HW and SW tools and methods of laboratory education focused on safety aspects of ICT within safety critical applications of processes control (50%) and 001TUKE-4/2015: CyberLabTrainSystem demonstrationaland training of informationcontrol systems-innovation (50%).

References

[1] PENIAK, P. *Cloud computing and integration manufacturing information systems with process control.* Zilina, 2014. Inaugural dissertation work. University of Zilina.

[2] HOFMANN, P. and D. WOODS. Cloud computing: The limits of public clouds for business applications. *IEEE Internet Computing.* 2010, vol. 11, iss. 6, pp. 90–93. ISSN 1089-7801. DOI: 10.1109/MIC.2010.136.

[3] BERNSTEIN, D., E. LUDVIGSON, K. SANKAR and S. DIAMOND. Blueprint for the Inter-cloud. Protocols and Formats for Cloud Computing. In: *The Fourth International Conference Internet and Web Applications and Services.* Venice/Mestre: IEEE, 2009, pp. 328–336. ISBN 978-1-4244-3851-8. DOI: 10.1109/ICIW.2009.55.

[4] GIRIRAJ, M. and S. MUTHU. A Cloud Computing Methodology for Industrial Automation and Manufacturing Execution System. *Journal of Theoretical & Applied Information Technology.* 2013, vol. 52, iss. 3, pp. 301–307. ISSN 1992-8645.

[5] PAVLIK, M., R. MIHAL, L. LACINAK and I. ZOLOTOVA. Supervisory control and data acquisition systems in virtual architecture built via VMware vSphare platform. In: *The 16th WSEAS International Conference on Circuits.* Kos Island: WSEAS, 2012, pp. 389–393. ISBN 978-1-61804-108-1.

[6] LOJKA, T. and I. ZOLOTOVA. Improvement of Human-Plant Interactivity via Industrial Cloud-Based Supervisory Control and Data Acquisition System. In: *Advances in Production Management Systems: APMS 2014.* Ajaccio: Springer, 2014, pp. 83–90. ISBN 978-3-662-44732-1. DOI: 10.1007/978-3-662-44733-8_11.

[7] PENIAK, P. and M. FRANEKOVA. Visualization of MES systems integrated within distributed control systems. *ATP Journal Plus.* 2013, vol. 2013, no. 2, pp. 74–77. ISSN 1336-5010.

[8] RASTOCNY, K., M. FRANEKOVA, I. ZOLOTOVA and K. RASTOCNY. Quantitative assessment of safety integrity level of message transmission between safety-related equipment. *The journal Computing and Informatics.* 2014, vol. 33, no. 2, pp. 334–368. ISSN 1335-9150.

[9] ILAVSKY, J., K. RASTOCNY and J. ZDANSKY. Common-cause failures as major issue in safety of control systems. *Advances in Electrical and Electronic Engineering.* 2013, vol. 11, no. 2, pp. 86–93. ISSN 1804-3119. DOI: 10.15598/aeee.v11i2.748.

[10] FRANEKOVA, M. *Safety communications of industrial networks.* Zilina: EDIS, 2007.

[11] ALLEN, A. O. Queueing Models of Computer Systems. *Computer.* 1990, vol. 13, no. 4, pp. 7–30. ISSN 0018-9162.

About Authors

Peter PENIAK was born in Povazska Bystrica, (Slovakia) in 1968. He received his Assoc. Prof. degree in 2014 in University of Zilina in the field of Automation with orientation to "Cloud computing and integration manufacturing information systems with process control". His research interests include manufacturing MES systems and its implementation within industry applications. He is head of IT division in Continental Matador Truck Tires s.r.o.

Maria FRANEKOVA was born in Brezno, (Slovakia) in 1961. She received her Prof. in 2011 in the field of Automation with orientation to "Safety-related Control and Communication Systems" in University of Zilina, Slovakia. Her research interests

include analysis of safety communications, methods of safety assessment of safety-related communication systemsfor safety-critical applications.

Iveta ZOLOTOVA was born in Michalovce, (Slovakia) in 1959. She received her Prof. in 2010 in the field of Cybernetics with orientation to "Selected aspects of visualization as a support in the management and decision-making" in Technical University of Kosice, Slovakia. Her scientific research is focused on industrial internet of thinks, cloud manufacturing, networked control, data acquisition, and human machine interface and web labs.