UDC 004.94+681.5

V. Kolesnik, O. Rolik

THE APPROACH OF ONTOLOGICAL STRATIFICATION IN IT INFRASTRUCTURE FOR QUALITY OF IT SERVICE MANAGEMENT

Abstract: The common service-based approach to distinguishing hierarchical layers for service level management in IT infrastructure covers only discrete scopes of IT infrastructure and is very general. As a result, it can be applied to any kind of IT infrastructure. However, coordination between scopes can miss some parameters which may be crossed among the layers. Moreover, such approach does not consider peculiarities of specific domains.

In order to fix this problem, it is suggested to introduce the ontology of domain area and to extend service-based stratification with domain-specific definitions. Also, in accordance to domain each general-based service such as infrastructure or platform as a service can be extended to cover interlinked dependencies between layers. For the purpose of quality of service management an approach of statistical process control is used.

Keywords: IT infrastructure, IT service management, quality of service, QoS, hierarchical stratification, QoS management, theory of automatic control, statistical process control.

Introduction

The cornerstone of modern IT infrastructures is the distribution and provisioning of IT services which has become possible as a result of service-oriented computing [1] paradigm applied to development of complex distributed systems inside a data center. Eventually, all connections and dependencies between IT services have to be measured, evaluated, planned, forecasted where possible and controlled. Therefore, a quality of service indicators must be introduced. In accordance with these indicators appropriate control actions should be made to maintain service quality and reduce negative impact of periods when quality could drop slightly.

Nowadays, IT services imply not only software or it’s piece which provides some utility to the end user, like calculus, but also can encompass quite complex utility scenarios, like web site hosting, library search indexation, complicated data querying, etc. In fact, as for today almost all of the IT services provides such complex utilities and consequently requires special theory to provide formalization basis for further object research. From this perspective, IT infrastructure which aids IT services to provide the utility to users can be formalized with the help of hierarchical multilevel system theory and system theory in general.

As for now, a lot of attempts were made to decompose IT infrastructure into layered hierarchy, and the results are widely used. The decomposition-compensation approach enlightens a way for distinguishing levels of service management which could provide a mechanism for moving from the bottom level of hardware and to the topmost level which is related to business processes and therefore to coordinate work between levels in order to compensate negative impact of inner and outer disturbances. The service-oriented computing paradigm created a base for cloud provisioning models which are also can be considered as abstract levels, namely another point of view of decomposition.

Ontological Stratification of IT infrastructure

Some applications of hierarchy theory imply the decomposition of any system into a few layers each of which can be considered as a submissive to upper layer, and the top-most layer will not have any layer above, same as bottom-most layer will not have any layer under it. Such approach was used during application of decomposition approach to hierarchical management in cloud [2]. In the mentioned research discrete decomposition was applied as a division of IT infrastructure layers into three most widespread provisioning models, which are software, platform and infrastructure as a service (SaaS, PaaS, IaaS) or in general it can be described as everything as a service (XaaS) [3]. After distinction of those layers a penalty function is applied in accordance to the negative impact on final quality of service from a specific layer. Despite that, a viral criterion for quality of service management was used, which is minimization of deviation of calculated SLA parameters from required ones.

Historically, the approach of decomposing complex system into different hierarchical layers was introduced in [4] as a stratification operation. Each layer of abstraction in terms of stratification is represented by stratum, which handles domain-specific inputs and produces related outputs. With this approach it is possible to separate physical processes from mathematical operations and higher abstraction activities, like business flows, economics interventions, power-related issues, etc. In general case the results of stratification have something in common with results obtained from the application of XaaS approach. However, hierarchical management can go far beyond than distinguishing XaaS levels. This can be explained by the roots of each approach. Everything as a service has emerged from service-oriented computing paradigm, which “uses services to support the development of rapid, low-cost, interoperable, evolvable, and massively distributed applications.” [1]. In fact, logical stratification is also applied in SOC in order to distinguish three planes: foundation, composition and management. Described planes can be considered as a roadmap for any service-based architecture in order to comply with client’s requirements and adapt to any changes in a flexible manner.

According to [5] there are two different forms of hierarchy: compositional and subsumptive. From the methodological point of view, compositional hierarchy can be divided into parts. In turn, subsumptive hierarchy implies division into discrete stages which are based on antecedent or ancestral conditions of the system or information about the system in time. From the ontological perspective, the compositional hierarchy indicates that the structure consists of parts and size of the structure is increasing or reducing depending on the point of consideration. On the contrary, subsumptive hierarchy is denoted by development from earlier conditions which means it might be part of evolutionary trajectory.

The term “ontology” in the area of information systems has a wide range of definitions. Each definition can be reduced to common description, which is described as “explicit formal conceptualization of a shared understanding of the domain of interest” [6] with incorporation of terms, semantics and pragmatics of latter. The most common IT infrastructure-related approach is to take one specific ontology and obtain results in the scope of considered area. However, decomposition of IT infrastructure components can go beyond the chosen ontology, which means that it is possible to consider the same IT infrastructure on different detailing level, which can include or exclude some parts of overall ontology definitions. Such decomposition is related to stratification approach in hierarchical system theory [4], however ontology-based stratification deals with the same inputs for each stratum, whereas original stratification recognizes different inputs for each layer which are interrelated. Therefore, the ontological stratification is introduced as an approach for quality of IT service management in IT infrastructure.

An example of the ontological stratification is depicted on Fig. 1. An IT infrastructure can be studied from different perspectives, depending on the context of considered objective. Also, ontology of IT infrastructure is a domain-specific if generalize it to information system. But, as long as object of the research is IT infrastructure it is suggested to think of the ontology at the topmost layer. Then, a domain ontology with task and problem-solving ontology will also depend on the context, for example consideration of constraints for critical operations or energy awareness, or cost saving, etc. Each constraint will introduce its own additional pragmatics for already existing semantics and terms.

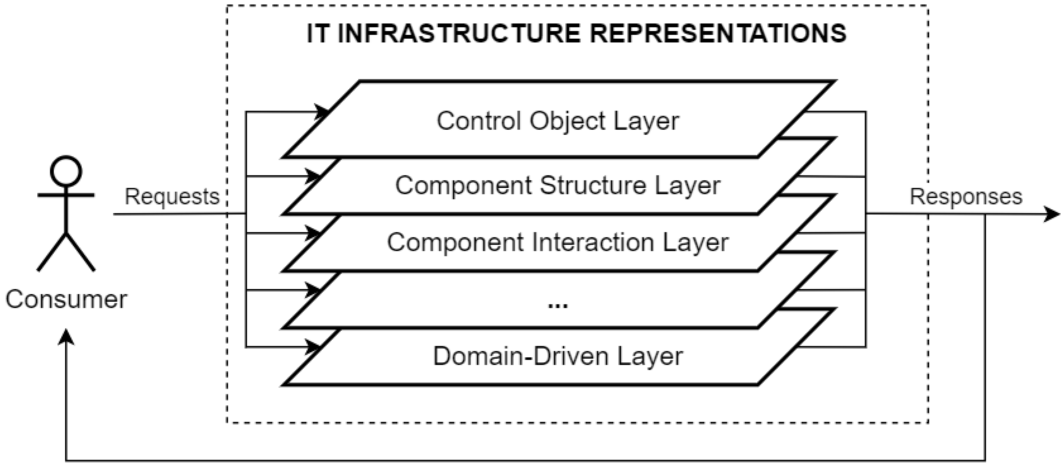
IT INFRASTRUCTURE REPRESENTATIONS
Control Object Layer
Component Structure Layer
Component Interaction Layer
...
Domain-Driven Layer
Requests
Responses
Consumer

Figure 1. Ontological representations of IT infrastructure

The highest stratum is Control Object Layer. On this level an IT infrastructure is considered as a black-box, which receives requests and provide responses. This layer can be used for analysis phase. It allows to get statistical or behavioral properties of input objects and to map their characteristics to output objects. On this layer it is possible to perform some correlation and covariance tests between those objects. Also, this layer can be utilized for application of different system consideration strategies: terminal or goal-seeking. According to [7] these two approaches can be absolutely different and moreover a system can be defined only in terms of goal-seeking approach but might not have any description based on terminal approach. Any of these resolves the problem of constructive specification. If the system is presented by system object sets $X = \{X_1,...,X_n\}$ then the terminal approach is related to mapping input which is represented by a family of sets $X = X_1 \times ... \times X_m$ to output $Y = X_{m+1} \times ... \times X_n$ with auxiliary functions, and can be explained as simple as providing functional over inputs in order to obtain outputs (1):

$$S \subset X \times Y . \tag{1}$$

Should be mentioned that product operations denote Cartesian product and each set indicates object [7].

On the next layer of Component Structure, the core entities of IT infrastructure can be defined according to an ontology. Among those entities are servers, containers, redirection entities (load balancer), etc. However, this stratum doesn’t imply specification of properties of those entities, so IT infrastructure is no longer a black-box, but those entities are black-boxes on this layer. The redirection entities can impact vector, which take part in optimization problem because this entity is controllable by IT infrastructure holder or researcher.

In turn, Component Interaction Layer is responsible for defining relations and connections between components introduced on the previous layer or even extended entities can be introduced here. This layer denotes a network in IT infrastructure and some properties of network entities. After this layer additional layers may be defined, which denote some domain-specific behaviors or extend existing framework, defined on previous layers.

The most bottom layer is a domain layer. On this layer domain-specific tasks are described and specific goals to be achieved. For example, this stratum can contain a description of tasks related to the behavior of provided services or applications, for example health-care services require more examinations comparing to other products because they introduce additional risks regarding people lives and health.

Control loops of Information Control System

When stratification is completed it is necessary to outline information system which will be responsible for management operations in IT infrastructure. According to [8] a high-level Information Control System was introduced, as described on Fig. 2.

Starting from the top there is the outer loop which is responsible for handling top-level business processes. This loop manages design, implementation and fulfillment of business processes and their quality. Therefore, information control system (ICS) is essential for efficient business functioning.

The inner loop, which is depicted on Fig. 2 in the middle and affects IT infrastructure and business management staff, stands for establishing service level agreements and continuously controlling service quality on agreed level. Also, each process in IT infrastructure (either business or production process) uses criteria for business efficacy. With this in mind ICS should be considered as a system, which “combines business management staff and IT infrastructure” [9].

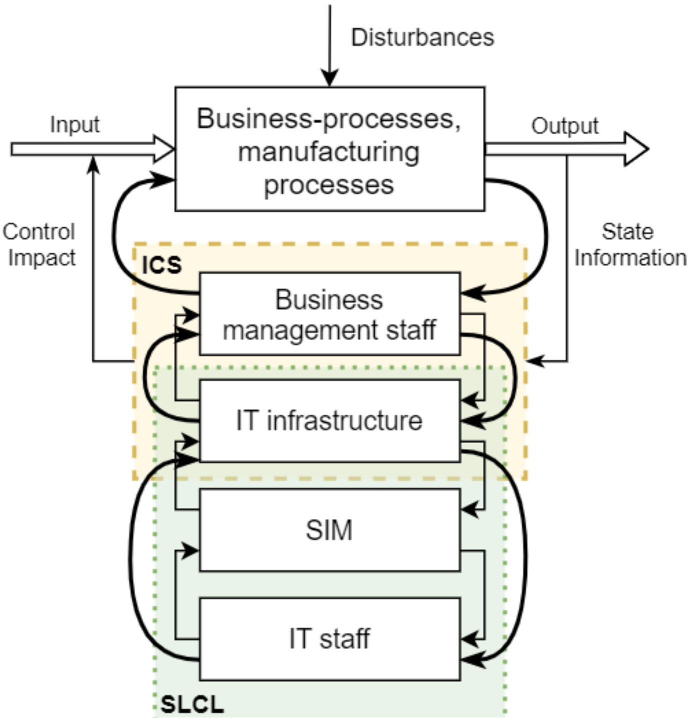

Figure 2. Control loops of Information Control System

On the bottom there is the service level control loop (SLCL). It combines IT infrastructure, supporting office and maintenance tools (software and hardware) which according to [9] can be generalized to “system of IT infrastructure management (SIM)”. This loop is responsible for ensuring quality of IT services on agreed level.

Suggested approach for Quality of IT service control

For the topmost levels of consideration where connections and appropriate weights are unknown it is possible to use statistical process control method which conforms with industry standard ISO 9000 for quality management. Despite the fact that initially the standard was developed for manufacturing, later it has been developed to cover generic approach to any business process. So far, IT infrastructure also contains a lot of processes which have direct or indirect impact on the business value. For the purposes of statistical control, it is important to “determine control limits that would apply to future data obtained from a process.” [10]

Afterwards, the objective of such control reduces to determination of the fact “that the process is being maintained in a state of statistical control.” [10]

In fact, described activities are possible only after definition of suggested distribution for consecutive application on a target process. In the result, common statistical properties of the process are being analyzed and appropriate actions are taken in case if process is going out of statistical control. It means, that application of statistical control enables signals of non-conformities, which can give an idea of adjustment of process parameters. However, the decision about performing some action should be put on an actor, either person or a controller.

Additionally, according to [10], the utilization of statistical control methods gives better results when applied to process variables instead of product variables. Despite that product variables are not taken into consideration for statistical control they still are going to be connected with the processes in terms of provided specification of required quality. For the area of IT infrastructures one of such specification is service level agreement (shortly SLA), which describes a list of quality values guaranteed by the provider to a client.

As was mentioned before, statistical control method just provides indicators which define whether the process is being observed under statistical control or not. However, adjustment of control parameters should be made by controller. In order to decrease human factor during control it makes sense to apply automatic control theory in combination with statistical process control. Indeed, automatic control theory is also considered as the best approach for technical systems, however in light of recent developments in informational technology (IT) field a software system can also be considered as the technical system to some extent. The only difference between pure technical system and software one is in units used in control loop, for example sensors; physical sensors are real devices and sensors in a software system are also software-based.

The automatic control theory is not very popular option for optimizing IT systems, yet some tries have been made in this area in [11-14] and therefore it proves that automatic control theory can be applied in IT systems likewise.

According to the described approaches ontology-based stratum can be elaborated in more details. Sample structure of Control Object Layer is depicted on Fig. 3.

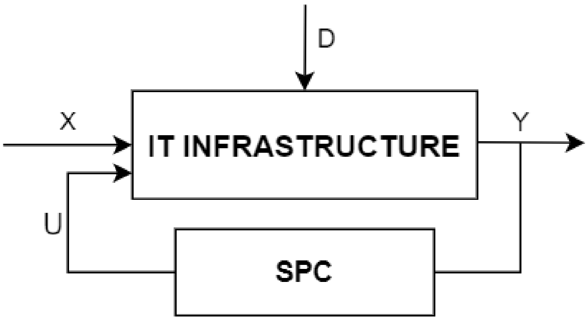

Figure 3. Control Object Layer

From the figure can be seen a set of inputs and outputs. A vector X denotes to consumer requests and vector Y defines responses to the input layer, in turn vector D represents disturbances in IT infrastructure. As far as internal structure is not taken into consideration on this layer, statistical process control unit is responsible for a feed-back connection in a control system and therefore provides a vector U of control impacts.

A noteworthy fact that general intention of theory of automatic control was concentrated on the processes of a real world. A real world is continuous itself and therefore describes behavior with introduced lags. Basically, it means that impact from provided control input can be observed only after some time. On the contrary, a computer program sometimes can change its state instantly, for example changing size of program’s cache from one value to another is not continuous. In the result, “the dynamics of computer applications tend to be discontinuous and "hard"” [15].

Afterall, knowledge about computer program’s behavior leads to understanding of any IT system’s behavior. One of such systems is IT infrastructure.

Conclusion

Consideration of different ontologies during modelling of IT infrastructure can decrease complexity of a model, but introduce different levels of consideration. Such approach allows to create basic set of levels which can be used without relation to any domain area. On the other hand, the domain-specific layers can give insights in relation to basic layers which can help in coordination of work or even find out some derived information which can lead to unexpected behaviors of IT infrastructure for cases of inconsistencies between layers.

In addition, the usage of statistical process control approach on each layer except analytical layers can empirically optimize functioning of IT infrastructure based on statistical properties of entities, which are being enriched during IT service provisioning. It allows to control, which entities can go out of scope of consideration, and which should remain in the consideration for a longer time.

REFERENCES

1. Papazoglou M.P. Service-Oriented Computing: State of the Art and Research Challenges / M.P. Papazoglou, P. Traverso, S. Dustdarand, F. Leymann //Computer, vol. 40, no. 11, Nov. 2007, pp. 38–45. doi: 10.1109/MC.2007.400

2. Zharikov E. Decomposition Approach to Hierarchical Management of Cloud Data Center Services / E. Zharikov, O. Rolik, S. Telenyk // 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), 11-14 Sept. 2018, Lviv, Ukraine. – IEEE, 2018, vol. 1. – pp. 42–47. doi: 10.1109/STC-CSIT.2018.8526764

3. Robison S. The next wave: Everything as a service / S. Robison // Executive Viewpoint, 2008, [online] Available: www.hp.com.

4. Mesarovic M.D. Theory of hierarchical, multilevel, systems / M.D. Mesarovic, D. Macko, Y. Takahara. – Academic Press, New York, 1970. – 294 pp.

5. Salthe S.N. Hierarchical Structures / S.N. Salthe // Axiomathes, 2012, 22(3), pp. 355–383. doi:10.1007/s10516-012-9185-0

6. Kabilan V. Ontology for information systems (04IS) design methodology: conceptualizing, designing and representing domain ontologies / V. Kabilan // Stockholm: KTH , 2007, p. xxii, 332.

7. Mesarovic M. General Systems Theory: Mathematical Foundations (Mathematics in Science and Engineering) / M. Mesarovic, Y. Takahara. – Moscow: Mir, 1978. – 312 p.

8. Rolik A.I. Upravlenie korporativnoy IT-infrastrukturoy / A.I. Rolik, S.F. Telenyk, M.V. Yasochka. – Kiev: Naukova dumka, 2018. – 576 p. ISBN 098-966-00-1665-1

9. Rolik O. Decomposition-Compensation Method for IT Service Management / Rolik O., Kolesnik V., Halushko D. // Kulczycki P. (eds) Information Technology and Computational Physics. CITCEP 2016. Advances in Intelligent Systems and Computing, vol 462, Springer, Cham, 2017, pp. 89–107. doi: 10.1007/978-3-319-44260-0_6

10. Ryan T.P. Statistical methods for quality improvement / T.P. Ryan. – John Wiley & Sons, 2011. – 666 p. doi:10.1002/9781118058114

11. Stahl E. Towards a control-theory approach for minimizing unused grid resources / E. Stahl, A.G. Yabo, O. Richard, B. Bzeznik, B. Robu, E. Rutten // Proceedings of the 1st International Workshop on Autonomous Infrastructure for Science. – 2018, June, pp. 1–8.

12. Yabo A.G. A control-theory approach for cluster autonomic management: maximizing usage while avoiding overload / A.G. Yabo, B. Robu, O. Richard, B. Bzeznik, E.Rutten // 2019 IEEE Conference on Control Technology and Applications (CCTA), 2019, August, pp. 189–195.

13. Berekmeri M. Feedback autonomic provisioning for guaranteeing performance in mapreduce systems / M. Berekmeri, D. Serrano, S. Bouchenak, N. Marchand, B. Robu// IEEE Transactions on Cloud Computing, 2016, 6(4), pp. 1004–1016.

14. Cerf S. Adaptive feedforward and feedback control for cloud services / S. Cerf, M. Berekmeri, B. Robu, N. Marchand, S. Bouchenak, I.D. Landau // IFAC-PapersOnLine, 2017, 50(1), pp. 5504–5509.

15. Janert P.K. Feedback control for computer systems: introducing control theory to enterprise programmers / P.K. Janert. – O'Reilly Media, Inc. 2013. – 330 p.