A Risk Assessment Framework for Cloud Computing

Karim Djemame, *Member, IEEE,* Django Armstrong, Jordi Guitart, and Mario Macias

Abstract—Cloud service providers offer access to their resources through formal Service Level Agreements (SLA), and need well-balanced infrastructures so that they can maximise the Quality of Service (QoS) they offer and minimise the number of SLA violations. This paper focuses on a specific aspect of risk assesment as applied in cloud computing: methods within a framework that can be used by cloud service providers and service consumers to assess risk during service deployment and operation. It describes the various stages in the service lifecycle wheres risk assessment takes place, and the corresponding risk models that have been designed and implemented. The impact of risk on architectural components, with special emphasis on holistic management support at service operation, is also described. The risk assessor is shown to be effective through the experimental evaluation of the implementation, and is already integrated in a cloud computing toolkit.

Index Terms—Cloud computing, risk assessment, risk modelling, holistic management

1 Introduction

ADVANCES in cloud computing research have in recent years resulted in considerable commercial interest in utilising cloud infrastructures to support commercial applications and services. However, significant developments in the areas of risk and dependability are necessary before widespread commercial adoption can become a reality. Specifically, risk management mechanisms need to be incorporated into cloud infrastructures, in order to move beyond the best-effort approach to service provision that current cloud infrastructures follow [1].

The importance of risk management in cloud computing is a consequence of the need to support various parties involved in making informed decisions regarding contractual agreements. The lack of adequate confidence in a cloud service in terms of the uncertainties associated with its level of quality may prevent a cloud service consumer from adopting cloud technologies. Although the provision of a zero-risk service is not practical, if not impossible, an effective and efficient risk assessment of service provision and consumption, together with the corresponding mitigation mechanisms, may at least provide a technological insurance that will lead to high confidence of cloud service consumers on one side and a cost-effective and reliable productivity of cloud service providers resources on the other side.

Consider an end-user (a service provider or a broker acting on their behalf) who is a participant from the broader public approaching the cloud in order to perform a task comprising of one or more services. The end-user must indicate the task and associated requirements formally within a Service Level Agreement (SLA) template. Based on this information, the end-user wishes to negotiate access with Infrastructure Providers (IPs) offering these services, in order that the task is completed.

IPs offer access to resources and services through formal SLAs specifying risk, price and penalty. Interactions between IPs and end-users can then be governed through a contract defining the IP's obligations, the price the end-user must pay and the penalty the IP needs to pay in the event that it fails to fulfill its obligations. The use of SLAs to govern such interactions in cloud computing is gaining momentum [1]. Moreover, IPs need well-balanced infrastructures, so they can maximise the Quality of Service (QoS) and minimise the number of SLA violations. Such an approach increases the economic benefit and motivation of end-users to outsource their IT tasks. A prerequisite to this is the IP's trustworthiness and their ability to successfully deliver an agreed SLA.

Risk assessment is considered in all phases of the service lifecycle for these stakeholders: end-users during service deployment and operation, and IPs during service admission control and internal operations.

In service deployment, risk assessment is considered in the following context: 1) before sending an SLA request to IPs, what is the risk of dealing with them, and which IP is less risky? 2) Once an IP receives an SLA request, what is the risk of dealing with the end-user from which the request came from? 3) In the admission control the IP performs, what is the risk of accepting the SLA request? and 4) Once an end-user receives an SLA offer, what is the the risk associated with deploying a service in an IP i.e. entering an SLA with the IP? Risk

- *K. Djemame and D. Armstrong are with the School of Computing, University of Leeds, UK, LS2 9JT.*
E-mail: {K.Djemame,een4dja}@leeds.ac.uk
- *J. Guitart and M. Macias are with Barcelona Supercomputing Center and Universitat Politecnica de Catalunya - Barcelona Tech., Spain.*
Email: {jguitart,mario}@ac.upc.edu

assessment allows the IP to selectively choose which SLA requests to accept (and consequently which to monitor and fulfil at service operation). On the other hand, end-users must make informed, risk-aware decisions on the SLA quotes they receive from the IPs so that the decision is acceptable and balances cost, time and risk. They clearly benefit from an evaluation of the risk of an SLA violation, since it allows them to determine the economic implications of agreeing to a particular SLA offer. This is also where risk assesment can play a key role by evaluating the reliability of an IP's own risk assessment.

In service operation, risk assessment helps support the following: 1) from the end-user perspective, what is the risk of failure of the SLA? 2) similarly from the IP perspective what is the risk of failure of a specific SLA? of the cloud infrastructure? Here, IPs perform continuous risk assessment at service operation, monitoring low-level events from the infrastructure such as risk of failure of physical hosts/VMs, security, legal, and data management risk. On the other hand, end-users also perform continuous risk assessment, monitoring service level non-functional Quality of Service (QoS) metrics such as the availability of VMs.

Risk assessment has been introduced into utility computing such as Grids and clouds either as a general methodology [2], [3], [4] or focusing on a specific type of risk, such as security and SLA fulfilment [5], [6]. However, the aim of this paper is to propose a risk assessment framework for cloud service provision, in terms of assessing and improving the reliability and productivity of fulfilling an SLA in a cloud environment. Based on this framework, a software tool is designed and implemented as a risk assessment related module, which can be integrated into other high level cloud management and control software systems for both end-users and IPs. This paper builds on our existing research on Risk Assessment within the context of Cloud Computing. The nature of SLAs in clouds and their associated workloads are different to those in other paradigms and thus effects the event prediction and the associated risks that must be considered within any given framework. The life-cycle of a Cloud, orientated towards long lived services, also differs from that of Grids and a risk assessment framework must therefore consider its semantics to service deployment and operation.

The main contributions of this paper are:

- A risk assessment framework for cloud computing. Risk assessment is supported at service deployment and operation, and benefit both end-users as well as infrastructure providers.
- A model for infrastruture providers to assess at service operation the risk of failure of 1) physical nodes; 2) VMs; 3) SLAs, and 4) entire cloud infrastructure.
- An evaluation of the risk model on a cloud infrastructure and through simulation.

The remainder of the paper is organised as follows. Section 2 introduces the risk management discipline. Section 3 explains the vision of risk in cloud computing. Section 4 presents the proposed risk model, section 5 its implementation, and Section 6 its evaluation on a cloud infrastructure. Section 7 presents some related work. In conclusion, section 8 provides a summary of the research.

2 Risk Management

Risk management plays an important role in a wide range of fields, including statistics, economics, systems analysis, biology and operations research. Risk is defined as the possibility of a hazardous event occurring that will have an impact on the achievement of objectives. Risk is measured in terms of *consequence (or impact)* and *likelihood* of the event [7]. Qualitatively, risk is considered proportional to the expected losses which can be caused by an event and to the probability of this event. Quantitatively, it is the product of probability of hazardous event and the consequences.

The most central concepts in risk management are the following: an *asset* is something to which a party assigns value and hence for which the party requires protection. An *unwanted incident* is an event that harms or reduces the value of an asset. A *threat* is a potential cause of an unwanted incident whereas a *vulnerability* is a weakness, flaw or deficiency that opens for, or may be exploited by, a threat to cause harm to or reduce the value of an asset. Finally, *risk* is the likelihood of an unwanted incident and its consequence for a specific asset, and *risk level* is the level or value of a risk derived from its likelihood and consequence. For example, a server is an asset, a threat may be a computer virus, the vulnerability a virus protection not up to date, which leads to an unwanted incident: a hacker getting access to this server. The likelihood of the virus creating a back door to the server may be medium, but the integrity of the server (consequence in terms of harm) may be high.

A fundamental issue in the characterisation and representation of risk is to properly and appropriately carry out the following steps:

- Analyse the triggering events of the risk, and by breaking down those events formulate adequately their accurate structure.
- Estimate the losses associated with each event in case of its realisation.
- Forecast the probabilities or the possibilities of the events by using either statistical methods with probabilistic assessments, or subjective judgements with approximate reasoning.

After the possible risks have been identified, they are assessed in terms of their potential severity of loss and probability or possibility of occurrence. This process is called *Risk Assessment (RA)*. The input quantities for Risk Assessment can range from simple to measurable (when estimating the value of a lost asset or contracted penalty associated with non-delivery) to impossible to know for certain (when trying to quantify the probability of a very unlikely event). *Risk Management (RM)* is the process of measuring or assessing risk and on the basis of the results developing strategies to manage that risk and control its implications. Managing a type of risk includes the issues of determining whether an action or a set of actions - is required, and if so finding the optimal strategy of actions to deal with the risk. The actions applied in a comprehensive strategy consist of an appropriate combination of the following measures:

- Transferring the risk to another party.
- Avoiding the risk.
- Reducing the negative effects of the risk, and
- Accepting or absorbing some or all of the consequences of a particular risk.

In *Quantitative Risk Assessment (QRA)* a numerical estimate is made of the probability that a defined harm will result from the occurrence of a particular event. Quantitative Risk Analysis is performed on risks that have been prioritized. The effects on those risk events are analysed and a numerical rating to those risks are assigned. A risk level can then be represented using for example a 7-point rating scale [8]: 1: trivial, 2: minor (-), 3: minor (+), 4: significant (-), 5: significant (+), 6: major, and 7: catastrophic.

This paper focuses on a specific aspect of risk management as applied to cloud computing: methods that can be used by a cloud provider to evaluate risk through the service lifecycle: construction, deployment, and operation. In this context, assets include physical nodes, virtual machines, SLAs etc. Considering a physical node as an asset, a threat may be a loss of its connectivity, the vulnerability a fault in hardware, which leads to an unwanted incident: the failure of the resource.

3 Risk Aware Cloud Computing - The Framework

The overall vision is the provision of a framework allowing individuals to negotiate and consume cloud resources using Service Level Agreements (SLA). This embraces an extended approach to the utility computing business model, which fits in an open market business model (for example for access to infrastructure as a service) as used in sectors such as finance, automotive, and energy. This section presents the main actors (service provider and infrastructure provider), and the various stages in the service lifecycle where risk assessment takes place.

3.1 Actors

The main actors are Service Providers (SPs) and Infrastructure Providers (IPs):

- Service providers offer economically efficient services using hardware resources provisioned by infrastructure providers. SPs participate in all phases of the service lifecycle, by implementing the service, deploying it, and overseeing its operation.
- Infrastructure providers offer physical infrastructure resources required for hosting services. Their goal is typically to maximize their profit by making efficient use of the infrastructure and by possibly outsourcing partial workloads to partner IPs. The application logic of a service is transparent to the IPs, which instead uses VMs as black-box management units.

The element of focus is the *service*. A service can provide any sort of functionality to an end-user. This functionality is delivered by one or more VMs, which each run applications that deliver the service logic. Each service is associated with a number of functional and non-functional requirements specified in a *service manifest*. The functional requirements include a requested performance profile, both in terms of hardware configuration, e.g., amount of physical memory and CPU characteristics for VMs as well as application-level Key Performance Indicators (KPIs) such as transaction rate, number of concurrent users, response time, which are used for elastic auto-scaling. The Service Provider can if needs be over come uncertainty associated with the performance profile through acquired historic knowledge of other applications it has deployed. Thus this issue can be abstracted away from and no longer becomes a concern of the underlying risk model. The non-functional requirements include, e.g., service and security level policies, and ecological profile. All of these requirements in the manifest configure the service for deployment and operation. The service manifest is thus the principal mean of interaction between the SP and IP.

3.2 Risk Assessment and Service Lifecyle

Consider the situation where an IP wishes to offer use of its resources as a pay-per-use service to potential SPs, and where the use of SLAs govern the interaction between them. An IP may need to implement an effective risk assessment procedure prior to making an SLA offer. If the SP's service requirements can be satisfied, the IP makes an SLA offer. The SP either commits to the SLA or rejects it. Risk is then considered in all phases of the service lifecycle for the two stakeholders: SP during service construction, deployment, and operation, and IP during service admission control and internal operations.

In the risk assessment process specific issues imposed by law or regulations, as well as operational risks inherent to the use of Cloud systems, either local or external assets, aer found. These risks can have a great impact on the operation of SPs and IPs, making it inconsistent with their respective business strategies, represented by means of Business Level Objectives (BLOs) and/or constraints. Therefore, a risk assessment methodology of cloud service provision is needed in order to assess and improve the reliability and productivity of fulfilling an SLA in a cloud environment, and the impact of some traditional risks must be re-evaluated in clouds. Figure

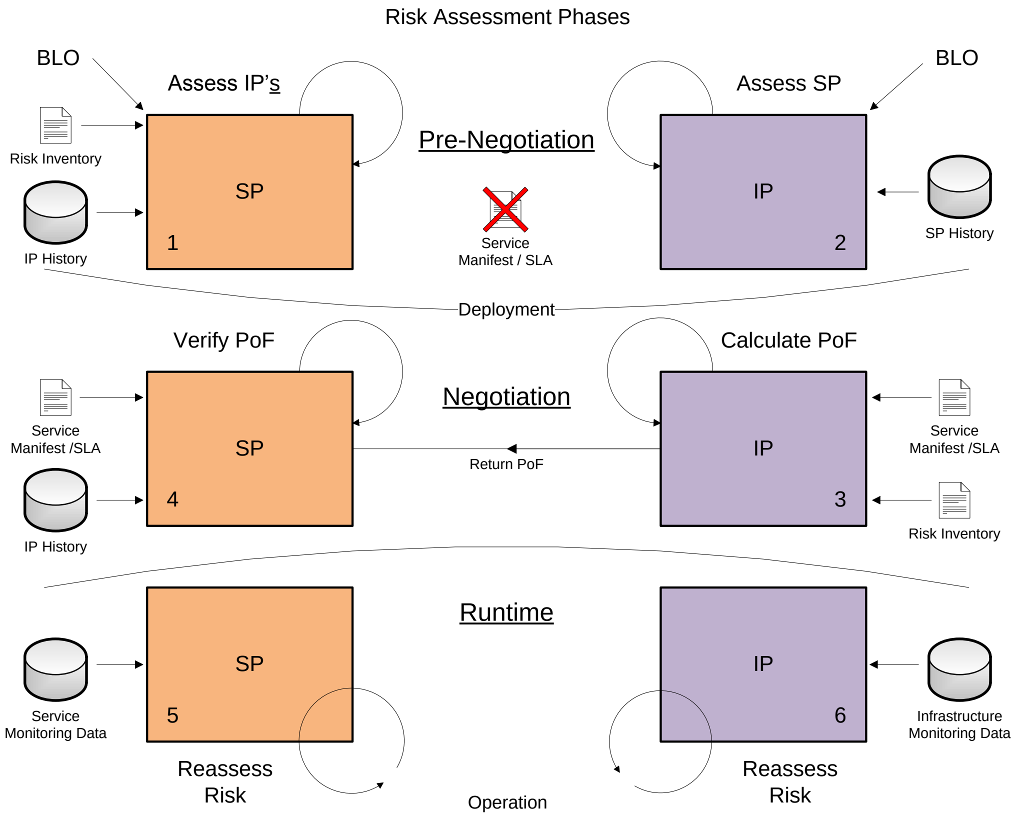


Fig. 1. Risk Assessment Stages

1 shows an overview of the stages where risk assessment takes place.

3.2.1 Service Deployment

An overview of the SP-IP interaction at service deployment when the SLA negotiation takes place, and the corresponding risk assessments are explained next.

Infrastructure Provider Risk Assessment

One of the objectives of the SP is to assess the IP before committing to an SLA. To support its strategic and business planning, the SP requires protection of specific assets, namely the actual SLAs as well as its reputation. This is called in risk assessment an *indirect asset*, which is an asset that, with respect to the target and scope of the risk analysis, is harmed only via harm to other assets. The harm relation expresses that an asset can be harmed through harm to another asset. Therefore, the SP, before sending an SLA request, assesses the risk of dealing with all known IPs (see Figure 1, stage 1).

Such provider assessment builds on the research in [6] which addresses the problem of dealing with missing provider information to produce a ranking of the providers based on specified criteria. The assessment of an IP by an SP is based on seven criteria, which are based on information collected from Cloud providers willing to share this information with the general public, or proposed guidelines by the National Institute for Standards and Technology (NIST) [9] and the European Network and Information Security Agency (ENISA)[10]. The assessment criteria are:

1) Past SLA Performance: this includes the number of past successful SLAs.
2) Geography Information: Geographic threat level, stability level, jurisdiction transparency level, jurisdiction overlapping level.
3) Certifications and Standards Compliance: facility related certification level, operation related certification level, and industry standard compliance level.
4) Business Stability: this includes for example the business history, the number of employees, and the number of customers.
5) General Infrastructure Practice: Available compute resources, available spare resources, average node availability, storage backup frequency.
6) General Security Practice: facility security level.
7) General Privacy Practice: facility and data access control level, personal data protection level.

A value between 0 and 1 is computed for each of the criterion by evaluating an IP with respect to a number of sub-criteria. These values are used as the basis for the evaluation. There are two important features of the evaluation system. Firstly, it takes into account SP's preferences. Different SPs are likely to value the criteria differently. Therefore SPs are able to specify the importance of each of the criteria, e.g. on a scale of 0 to 10. These are then translated into criteria weights that encapsulate the amount of influence a particular criterion should have and incorporated into the provider evaluation. Secondly, it is able to handle missing data. Some IPs may be unwilling to share all of the information necessary to compute the criteria values. Alternatively, data may have been corrupted.

This is achieved through an implementation of Dempster-Shafer Analytical Hierarchy Process (DS-AHP) [11], whereby each decision alternative (in this case each IP) is mapped onto a belief and plausibility interval. Consider a set of providers as corresponding to the proposition that the providers in that set are preferable to all other providers considered in the evaluation but not to each other. The SP preference weights, w_i, are computed for each criterion, $i = 1..N$. Pair-wise comparison of decision alternatives (for providers) are used to derive weights for the criteria, $r_j^{(i)}$ for the i^{th} criterion and j^{th} provider. A weight or Basic Probability Assignment (BPA) is computed for each provider, and providers which are indistinguishable with respect to a criterion are grouped together as a single proposition (that the providers in this group are the best alternative). This results in the bpas in the form $m_i(s)$ where s is a set of one or more providers. Criteria bpas are combined using Dempster's rule of combination [11]. A set of intervals [Belief, Plausibility] for each single IP are then used to compute the order of preference for the IPs. Finally, a preference value for provider A relative to provider B is computed, and if $P(A > B) > 0.5$ then provider A is preferred. Details on the risk assessment model are found in [1], [12], [6].

Service Provider Risk Assessment

An IP receives an SLA request and assesses the risk of dealing with the SP from which the request came from (see Figure 1, stage 2). Similarly, to support their strategic and business planning, IPs require protection of their assets, which include their infrastructure (physical hosts, Virtual Machines, networks, disk storage) as well as indirect assets, in this case their reputation. The DS-AHP based assessment (see Section 3.2.1) of the SP is done using the following criteria based on proposed guidelines by NIST [9] and ENISA[10].:

1) Past SLA Performance: this includes the number of past successful SLAs deployed by the SP.
2) Business Stability: this includes for example the business history, the number of employees, and the number of customers.
3) General Security Practice: Facility Security Level.

Details on the risk assessment model are found in [1], [12], [6].

SLA Request Risk Assessment

The IP assesses the risk of accepting the SLA request from the SP (see Figure 1, stage 3). To do so: 1) the IP's Admission Controller (AC) determines whether or not to accept a new service request by handling the tradeoff between increased revenue and increased workload as well as the impact on running services, and 2) the IP carries out a risk assessment based on the requirements specified in the service manifest, which may include physical nodes, security, and legal requirements to name a few.

The assessment of physical nodes' risk of failure within cloud infrastructures builds on the research in [13] which is based on Semi-Markov models. However, the risk of failure is assessed considering the cloud resource historical information stored in the IP's Historical Database and the Cumulative Distribution Function (CDF) of the Time To Fail (TTF). This enables an IP to identify infrastructure bottlenecks, estimate the likelihood of physical host failures and identify possible infrastructure bottlenecks.

Security is a priority concern for cloud service consumers; many of them will make buying choices on the basis of the reputation for confidentiality, integrity and availability of, and the security services offered by, a cloud provider. Therefore, this is a strong driver for an IP to improve security practices.. The IP faces security challenges posed by the transparency of distribution, abstraction of configuration and automation of services and needs to perform a detailed threat analysis across different deployment scenarios (private, bursting, federation or multi-clouds). Our research in [14] presents a systematic approach for threat analysis based on standard threats for distributed systems, adopted in cloud computing. The proposed methodology uses the CORAS risk modeling methodology [15] coupled with Information Risk Analysis Methodology (IRAM), tailored for specific cloud computing security risk assessment.

The IP may also require to carry out a legal risk assessment for the benefit of legal compliance. Details on the legal issues surrounding risk assessment in cloud computing are found in [16]. Specifically, risk is analysed regarding data protection and security, and the requirements of an inherent risk inventory are presented.

Once the various risks in relation to the requirements in the service manifest are assessed, an aggregated risk (failure of physical node, security, and legal) of accepting the SLA request is estimated, possibly allocating weights to the various risks depending on the service requirements. The IP then sends a response back to the SP, either in the form of an SLA *offer* or *rejection*.

SLA Offer Risk Assessment

The SP assesses the risk associated with deploying a service in an IP (entering an SLA with the IP) (see Figure 1, stage 4).

Since an IP may not want to share detailed data about their own infrastructure or their risk assessment methods, it is difficult for the SP (and possibly cloud brokers) to verify the reliability of IP's risk assessment. This problem may be exacerbated if the SP wishes to choose the best (according to their own criteria) IP for their services and have little or no past dealings with many providers. When making an SLA offer, an IP usually presents an associated availability to the entity (SP) it is negotiating with. Most of commercial cloud providers such as Amazon "guarantee" a service availability of 99%. This metric can be translated into a Probability of Failure (PoF), therefore a risk of service unavailability. However, the SP will not necessary trust this IP's metric. Even if the IP utilises sophisticated and accurate risk assessment techniques, they might wish to convince the SP that the PoF is lower than it is in reality, in order to increase the likelihood that the SLA negotiation is successful [17]. If the IP's offered PoFs were consistently lower than the actual PoF by a considerable margin then its reputation may be damaged and the SP may become aware of this. However, if the difference is not too large (while still being statistically significant and therefore potentially having an impact on the SP's own assessment) it may require a significant amount of historical data to enable the IP's inaccuracy or dishonesty to be identified. Hence it would clearly be of value if the SP could obtain additional information to provide some indication of the reliability of an IP's risk assessment prior to accepting an SLA offer. This information is a measure of the IP's reliability and the SP's own PoF estimate (which differs from the provider's if it is considered unreliable).

A risk model is used for evaluating IP's reliability, with respect to systematic errors and is based on research in [6]. Here, systematic errors refer to provider errors whereby their risk assessments exhibit a typical trend in the sense that they tend to overestimate the PoF or tend to underestimate the PoF. The scenario from the SPs perspective is as follows:

1) An IP makes an SLA offer and includes a PoF estimate (e.g. 99%) associated with that SLA.
2) Each time an offer is accepted, the details are stored in the SP's historical database, including the final status (Success/Fail) and the offered PoF.

With this information, the problem is now to determine whether risk assessment based on IPs past data can be considered reliable. The reliability assessment model is based on: 1) a reliability estimate for systematic errors; 2) scenarios where the IP is under or over estimating the PoF, and 3) Accounting for static/dynamic provider behaviour. The IP dynamic behaviour may change as a consequence of a variety of factors, e.g.

- The IPs infrastructure is updated. This may have an effect on the reliability of subsequent risk assessments.
- The IPs risk assessment methodology or model parameterisation may change.
- The IPs policy may change, for example due to economic considerations. For example, they may decide that they can make more profit if they start to give overoptimistic estimates to an SP/broker, in order to persuade them to agree offers.

3.2.2 Service Operation

Once the service is deployed and its operation is underway, the SP and IP need a combination of monitoring and assessment in order to create a self-managed cloud infrastructure driven on one hand by their Business-Level Objectives (BLO), and by the SLA fulfillment on the other. For an IP BLOs represent how to manage the infrastructure, e.g. reducing costs of operation at the expense of risk, whereas for the SP they represent how to manage the service, e.g. scaling in/out, elasticity, and VM migration to another IP.

SP Dynamic Risk Assessment

The SP requires protection of its assets (running services) and therefore performs continuous risk assessment at service operation, monitoring service level non-functional Quality of Service (QoS) metrics such as availability of VMs (see Figure 1, stage 5). Service management involves a wide range of tasks, including the provision of an information repository for services deployed on the VMs of IPs, service deployment and undeployment, and monitoring the execution of services to ensure SLA fulfillment.

Models are needed to assist the SP in continuously assess risk as service operation. This will very much depend on the data available for the SP at service operation, data the IP is willing to share with the SP, and the SP's own historical data.

IP Dynamic Risk Assessment

To protect its assets and as part of SLA fulfilment the IP performs continual risk assessment at service operation, monitoring low-level events from the infrastructure such as risk of failure of physical hosts and VMs, as well as risk in relation to security, legal, and data management (see Figure 1, stage 6).

3.3 Summary

This section has presented the risk assessment framework allowing the actors (SPs and IPs) to negotiate and consume cloud resources using SLAs, as well as the various risk models these actors make use of at various stages in the service lifecycle:

- When an SLA negotiation takes place, an SP is able to assess the risk of dealing with IPs. It is useful for the SP to know which IP is less risky, as this brings the confidence that the SLA once signed has a good chance to be fulfilled.
- When an SLA negotiation takes place, an IP is able to assess the risk of dealing with an SP. Similarly, this is useful for the IP as it brings the confidence

that once the SLA is signed the IP's assets are protected ensuring the SLA fulfilment.

- Prior to committing to an SLA, the IP carries out a risk assessment based on the service requirements. This is SLA dependent and may include assessing a security risk, a legal risk, the risk of failure of physical hosts, or the risk of failure of VMs. The risk of accepting an SLA request is seen as an aggregate of various risks.
- Prior to committing to an SLA, the SP carries out a risk of service unavailability by evaluating the IP's reliability with respect to past behaviour. By doing so, the SP has the confidence that the SLA can be fulfilled by a reliable IP.
- Once a service is in operation, the SP constantly monitors the service execution and continuously performs risk assessment as part of the SLA fulfilment.
- Similarly, the IP constantly monitors its infrastructure in service operation and continuously performs risk assessment as part of the SLA fulfilment.

The remainder of this paper will focus on a model that can be used by an IP at service operation to assess the risk of failure of 1) physical nodes; 2) VMs; 3) SLAs, and 4) entire cloud infrastructure.

4 Risk Assessment Modelling - IP Service Operation

In order to assess the risk associated with cloud resources based on (real-time) data at service operation, it is essential to know what data is required for such risk assessment, and how it is going to be analysed to estimate the actual risk. For this purpose, a *risk inventory* is used and is presented next.

4.1 Elements of Risk

The risk inventory is populated with:

- **Assets**: Virtual Machine (VM), physical host, SLA, with a description of their characteristics. Risk events are assessed in terms of these.
- **Incidents / Risk Scenarios**: aim to describe any event, condition or their combination that has the potential to reduce the capacity or availability of an asset. Incidents are composed of:
- **Vulnerabilities**: describe inherent weaknesses of the asset (e.g. a faulty hardware) and their impact reflects the possibility of a risk incident, e.g. violations of the Quality of Service (QoS), and SLA indicators, inherent to the assets.
- **Threats**: represent the other side of the risk which depends on factors independent to the asset, e.g. loss of connectivity of a physical host.
- **Adaptive capacity**: description of the mitigating strategies in place for the specific asset, e.g. server replication.
- **Impact/Consequence** of a risk incident, e.g. failure of a physical host, and is defined using as degraded performance, loss of data, or unavailability. The evaluation is performed according to the indicators selected to describe the asset as well as associated costs, e.g. of not meeting predefined service levels.

4.2 Process

A quantitative risk assessment approach which makes use of measurable, objective data to determine the likelihood of events and associated risk is then applied to estimate the level of risk attached to VMs, Physical Hosts, and SLAs thanks to the data gathered by the Cloud Monitoring Infrastructure (CMI). Therefore, an identification of the elements of risk in the risk inventory for VMs, Physical Hosts, and SLAs becomes important. It should be noted that the nature of risks may differ thus, the quantitative risk estimation too.

Figure 2 shows the risk assessment methodology, which divides the risk assessment process into the following stages:

Risk Inventory. At this stage, requirements analysis is performed to identify how the risk inventory is populated.

Vulnerability identification. A vulnerability is considered as a weakness or flaw in system procedures, design or internal, management controls that can be accidentally triggered or intentionally exploited. Let each vulnerability be represented as a single bit in the vulnerability vector:

$\vec{V} = \{V_i\}$ = 1,0 $\forall i$, i=1,2, . . . n

where V_i represents an individual vulnerability. The value 1 indicates the presence of this vulnerability in the system under assessment, otherwise 0.

Threat identification. During a threat analysis process potential threat sources and actions that may exploit system vulnerabilities are identified. Information about threats can be gathered from experts and log files. Let each threat be represented as a single bit in the threat vector:

$\vec{T} = \{T_j\}$ = 1,0 $\forall j$, i=1,2, . . . m

where T_j represents an individual threat. The value 1 indicates the presence of this threat in the system, otherwise 0.

Data Monitoring. At this stage, the data requirements that need support are identified. In service operation it is expected that a cloud monitoring infrastructure will provide such data.

Event Analysis. An event can be defined as a pair: a vulnerability and it matching threat. Events can be identified from facts, which take place in a specific context. In order to identify the possibility of an event occurring, the likelihood should be estimated considering factors of threat-source motivation and capability, and nature of the vulnerability. Therefore, the likelihood of threat acting over vulnerability is defined as : $L_{ji} = \langle T_j, V_i \rangle$

Quantitative Risk Analysis. Risk is defined as the likelihood of an event and its consequence. After potential events and their likelihood have been identified

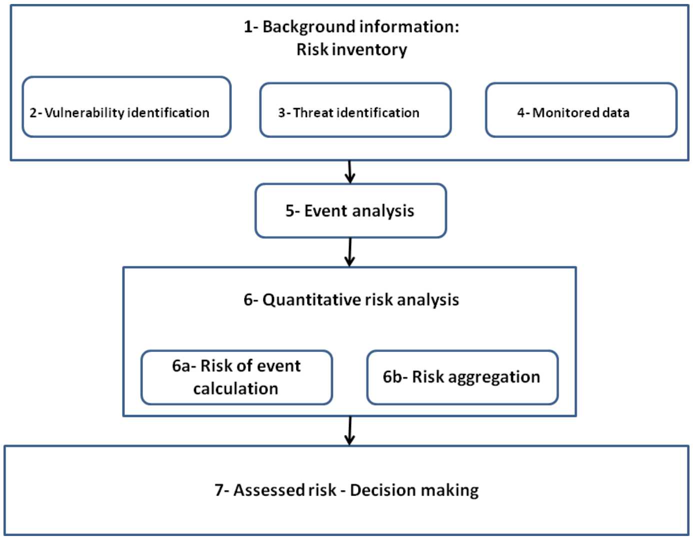

1- Background information:
Risk inventory
2- Vulnerability identification
3- Threat identification
4- Monitored data
5- Event analysis
6- Quantitative risk analysis
6a- Risk of event calculation
6b- Risk aggregation
7- Assessed risk - Decision making

Fig. 2. Risk Assessment Methodology Steps

the quantitative risk assessment approach is applied to estimate a level of risk for VMs, Physical Hosts, SLAs and the IP.

4.3 Model

Individual risks associated with each event (vulnerability, threat) are first calculated and then an aggregated risk for enhancing knowledge based on these individual risks is estimated. Within the general risk assessment model, several elements of risk are identified:

$$R_{j,i} = L_{ji} \cdot I_i \tag{1}$$

Thereafter, the risk for an individual element within an asset under a context specific environment can be calculated as follows:

$$R_E = 1 - \prod_{j=1}^{m}(1 - R_{ji}) \tag{2}$$

where E = 1,2, ... is an individual element of risk within the asset. The formula only applies in cases when an element has threats and vulnerabilities associated.

The aggregated risk consists of all individual risks within an asset and is defined as:

$$R_{agg} = 1 - (R_{E_1} \cdot R_{E_2} \cdot \ldots R_{E_k}) \tag{3}$$

5 Implementation

The risk assessment framework has been implemented in the context of the OPTIMIS project [1]. This section provides the framework implementation details as well as the components involved in the risk assessment process at service operation.

5.1 Context

OPTIMIS innovations can be summarized as a combination of technologies to create a dependable ecosystem of cloud providers and consumers that will be the foundation of efficient operations of services and infrastructures [1]. This includes innovations for optimizing the whole service lifecycle. The main result of OPTIMIS is the OPTIMIS Toolkit [18] (see Figure 3), a set of software tools for cloud providers and users to support various multi-cloud architectures, optimize the cloud service lifecycle, and simplify cloud self-management.

The OPTIMIS toolkit simplifies for SPs the development of new services, makes informed deployment decisions for these, and monitors their execution. Similarly, IPs can use OPTIMIS tools to decide whether to accept additional services and optimize provisioning for the already hosted ones.

SPs and IPs decision making are based not only on low-level functional properties, but also on non-functional factors relating to trust (including reputation), risk (likelihood-consequence analysis for valuable assets), eco-efficiency (power consumption and ecological factors), as well as cost (economical models for service and infrastructure provisioning expenses).

The OPTIMIS Toolkit enables dynamic and pro-active management of cloud infrastructures and services. This ensures adaptability, reliability, and scalability of infrastructures that handle predicted and unforeseen changes in services and continuously optimize operation.

5.2 Architectural Components and Interaction

As stated in section 4.2 the data requirements for risk assessment are provided by the CMI (see Figure 4). The CMI provides aggregation of data from multiple Information Providers, management of that data and a Data Model designed for flexible post-processing of that monitoring data. It effectively stores, manages and

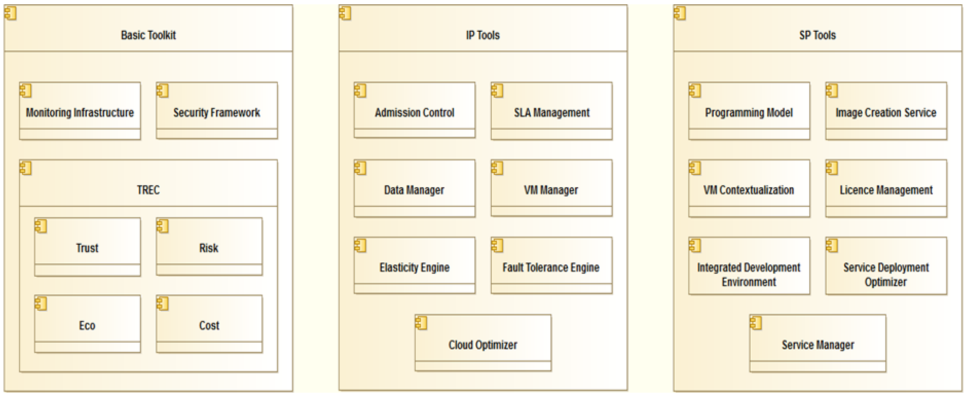


Fig. 3. OPTIMIS Toolkit Overview

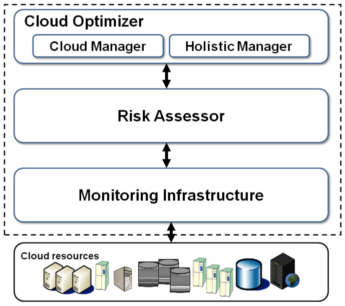


Fig. 4. Components Interaction: Monitoring Infrastructure, Risk Assesor, and Cloud Optimizer/Holistic Manager

offers data to the risk assessor component (see Figure 4), addressing the challenges in relation to the use of virtualization technologies, the distribution of the infrastructure, and the scalability of the applications [19]. It must be noted that the discloser of internal infrastructure level monitoring data may be limited if it comes from a third party. This availability of what monitoring data can be access would in turn have an impact of the effectiveness of the risk assessment and could be a possible extra source of revenue for a provider if the required data is deemed to be essential to mitigating one specific risk. This would not have an impact on other monitoring metrics sources such as those coming internally from a VM.

The Cloud Optimizer (CO) links directly to the OPTIMIS cloud toolkit and components including cloud service optimisation and risk assessment services. It combines the monitoring and assessment tools in the OPTIMIS toolkit with various management engines in order to create a self-managed cloud infrastructure driven by IPs BLOs. The CO consists of two main components: the Holistic Manager (HM) and the Cloud Manager (CM).

The Holistic Manager ensures *reactive* and *proactive* risk assessment linked to service level goals. The HM is responsible for harmonizing the operation of the IP risk management in order to fulfill high-level BLOs which represent how an IP provider wants to manage the infrastructure, i.e. set the overall IP risk values. To do so, the HM translates BLOs to risk-based objective functions and constraints.

The Cloud Manager is responsible for the actual deployment/release of the different VMs based on the recommendations given by the risk assessor. In addition, the CM arbitrates between private versus external deployment of the VMs arbitrating between different cloud deployment models. The CM receives a notification from the risk assessor when a risk level goes over/below a given threshold. When this occurs, the CM forwards this notification to the low-level managers to try to solve the issue. If the CM receives subsequent notifications, it initiates more drastic actions, such as canceling a running service, or elastic VMs within services. To this end, the CM repeats the process above but calculating the *forecast* of risk for the IP when cancelling the corresponding VMs.

5.3 Risk Inventory

Considering physical nodes, VMs, SLAs, and entire infrastructure as IP's main assets, the risk model introduced in 4.3 is implemented considering 1) the vulnerabilities and threats associated with physical nodes, VMs, SLAs, and entire infrastructure, and 2) data provided by the CMI.

As part of implementation of the risk assessor, the risk inventory introduced in Section 4 plays a critical role in the output of the risk model and what assets, threats, vulnerabilities and impacts are considered. The following defines the elements of the risk inventory used as input to the risk model in its current implementation.

5.3.1 Assets

The physical resources of an IP relate directly to its financial performance. Thus the following items list the high level assets included in the current implementation of the risk assessor:

- Computational Resources
- Storage Resources (both volatile and non-volatile)
- Networking Resources

5.3.2 Threats

Currently three external threats have been considered in the current implementation of the risk assessor and are defined in relation to the previously defined IP assets:

- Denial of Service: A malicious attacker causes a spike in resource usage and service requests are dropped.
- Flash Crowd: A sharp increase in the number of service users cause resource capacity to be exceeded and service requests to be dropped.
- Poor Quality Hardware Vendor: A vendor provides an IP with hardware that exhibits higher than normal failure rates. The hardware failures prevent requests from being serviced.

5.3.3 Vulnerabilities

The internal vulnerabilities of the IP are defined as a shortfall in the following resources due either to exceeding resource capacity or hardware failure at the physical, virtual or infrastructure level, when trying to meet a service's defined QoS:

- CPU
- Memory
- Network Bandwidth
- Disk Storage

5.3.4 Impacts

The impacts of both a lack of resource capacity and failure of a resource are as follows:

- Physical Resource:
  - Capacity Exceeded: Medium
  - Failure: Medium
- Virtual Resource:
  - Capacity Exceeded: Low
  - Failure: Low
- IP Resource:
  - Capacity Exceeded: High
  - Failure: High

It should be noted however that these lists of elements are not meant to be exhaustive as there are many possible extensions, which are touched on in Section 6.3. In the subsequent section these elements of the Risk Inventory are used in experimentation.

6 Evaluation

To evaluate the IP Service Operation risk model, two experiments have been conducted. The first is performed in the context of a real cloud testbed using the prototype risk assessor. The second, takes a black box approach and fabricates input for the risk assessor enabling greater control over experimental variables.

A numerical rating from 1 to 7 is assigned to the various risks to be assessed (physical node, VM, SLA, and infrastructure) as introduced in section 2. This setting is carried out by the HM in a proactive mode to control the operation of the IP risk assessment in order to fulfill high-level BLOs and is directly linked to existing service level goals.

Recall that risk is defined as the likelihood of an event and its impact. For simplicity the internal vulnerabilities of the cloud infrastructure are defined as a shortfall in physical and virtual resources due exceeding resource capacity (CPU, memory, network bandwidth, and disk storage). The risk of SLA failure is translated as the aggregate of the risk of failure of physical hosts and VMs the SLA makes use of.

As stated in section 4.2 the data requirements for risk assessment are provided by the CMI (see Figure 4).

In service operation the risk assessor is continuously running and its data requirements in relation to vulnerabilities are provided by the CMI. Once a particular risk level is reached the risk assessor initiates an automatic notification to the CM. An activation threshold prevents the CM from initiating corrective actions unnecessarily in the presence of abrupt but short-lived risk level peaks. The activation threshold indicates the maximum number of accumulated times (based on the monitoring rate of physical hosts and VMs) that a risk level can be above before attempting corrective actions.

For clarity, all illustrated results show time series graphs, expressed as Time since the Unix epoch on the x-axis with Risk-Level on the y-axis. The objectives, experimental setup and results of each are discussed in detail in the following sections.

6.1 Prototype Evaluation

This section discusses the outcome of testing the prototype risk assessor and its accompanying real time risk visualization tool.

6.1.1 Objective

The objective of this experiment was to ascertain that the risk model when monitoring a real service on an OPTIMIS cloud testbed can provide an end user with real time feedback on current service risk levels.

6.1.2 Experimental Setup

The experiment was performed on an cloud testbed comprised of two physical machines each in turn comprised of a Dual AMD Opteron 6234 (16 Core) Processors with 64GB of RAM. The risk assessor was deployed as part of the wider OPTIMIS toolkit. An OPTIMIS service manifest was deployed that comprised of 8 VMs each providing 2 virtual CPU cores and 2GB of memory. Each VM executed an idle Tomcat container on deployment. The risk assessor was invoked on deployment and took monitoring metrics from the OPTIMIS monitoring infrastructure.

6.1.3 Results

Figure 5 shows the outcome of the evaluation from the perspective of the real-time risk visualization tool. The user interface illustrates the four levels of risk associated with the deployment of this service in the IP's infrastructure. From the figure it can be seen that there are two initial periods of high risk. The first is due to the resource utilization of the service VM's operating system during the boot-up sequence. The second spike is caused by the resource utilization of the tomcat container starting up and deploying its services. The risk levels then later subside as resource utilization stabilizes. If a *conservative* approach towards risk was adopted by the HM by considering a risk level of 2 and above as *significant* then the risk assessor would have initiated an automatic notification to the CM.

6.2 Functional Evaluation

The aim of this functional evaluation of the risk assessor is to ascertain the effect of cloud environment and risk inventory input on the output over time of the risk assessment model for IP service operation via the use of fabricated monitoring metrics. To set the scene, the experimental scenario is motivated via a hypothetical IPs desire to maintain profitability and meet financial obligations to its stakeholders. To achieve this goal, an IP running numerous services uses risk assessment to prevent or mitigate the cost of breaking end-user agreed Quality of Service (QoS) through the protection of its assets.

6.2.1 Objective

The objective of this evaluation is to expose the relationship and correlation between the four risk levels of the model. Additionally, the experiment will uncover the effects of a different Cloud environment on the risk models output and disclose the ramifications of running the risk assessments on different types of service with varying attributes.

6.2.2 Experimental Setup

This evaluation considers a hypothetical IP that has ten physical machines with the following characteristics: 8 CPU Cores, 4GB RAM, 1Gbit NIC, 4TB HDD, 0.001% Porbability of Failure. Through the course of the evaluation three different service profiles with the following virtual machine characteristics are used:

TABLE 1
Virtual Machine Characteristics

Service Profile	CPU Cores	RAM	Network	Disk Storage	Chance of Failure
1	1	512	100Mbit	100GB	0.002%
2	2	1024	100Mbit	200GB	0.002%
3	4	2048	100Mbit	300GB	0.002%

These service profiles are representative of the typical small, medium and large virtual machine instances provisioned by real cloud providers such as Amazon. In addition to these service profiles two service workloads were used. The first, a low variance workload with between 25-50% resource utilization, the second a high variance workload with between 25% and 100% resource utilization, generated using a uniform normal distribution. Virtual machines are allocated to physical resources using a round robin approach to maximise resource utilization and the assets of the IP.

To evaluate the risk assessor three experiments have been devised. The first experiment considers the impact of running 1 to 10 concurrent homogeneous services comprised of 5 VMs using a low variance workload. A new service is added every 10 time steps. This experiment is repeated three times, each time using a different service profile. This experiment will ascertain the impact of increasing resource utilization over time on the four risk levels. The second experiment considers 2 services using Service Profile One with a low variance workload and Service Profile Three with a high variance work load. Each service is composed of 10 VMs. This experiment will ascertain the impact of workload on risk level between concurrently executing services. The final experiment executes 3 concurrent services each of a different service profile composed of 5 VMs. This experiment will confirm that the utilization of an IP's resources only impacts a service's risk levels when the service is actually making use of these resources.

6.2.3 Results

The following section discuss the results of the three experiments using the previously outlined profiles and experimental setup. Figures 6, 7 and 8 show the results of running one to ten concurrent services using Service Profile One with 5 VMs per service.

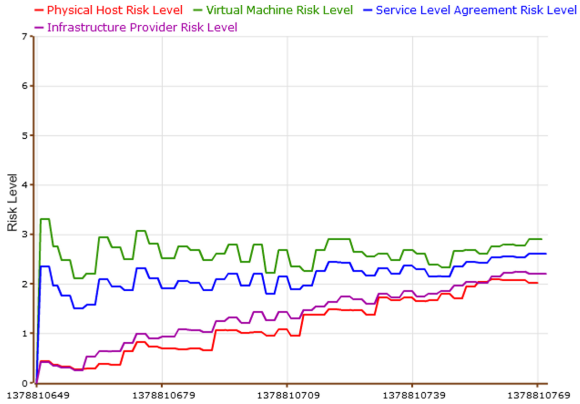
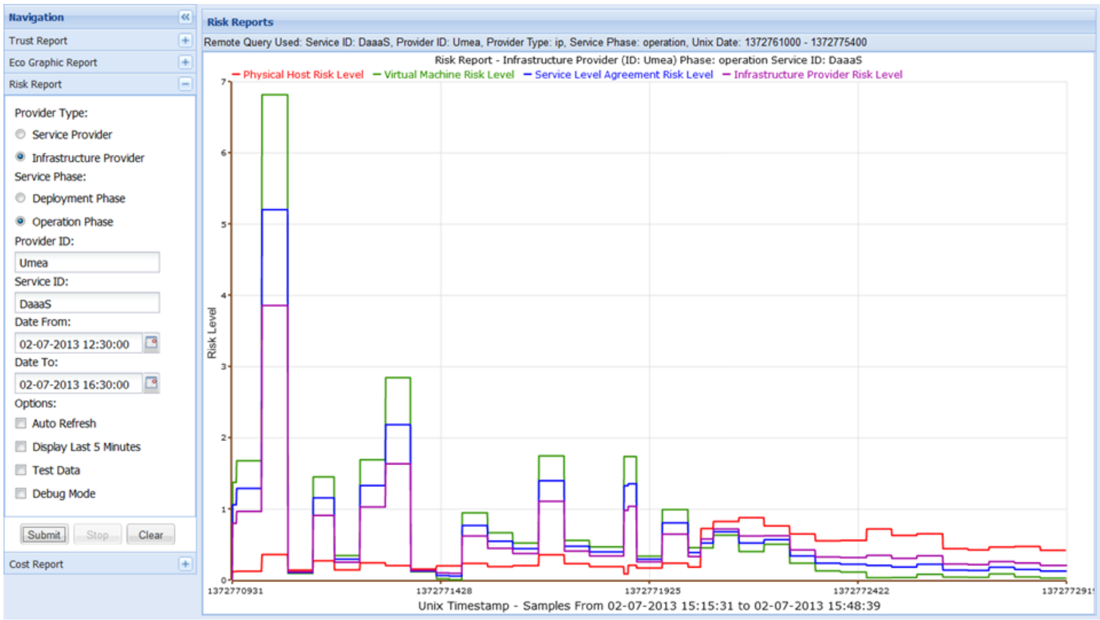


Fig. 6. 1-10 Homogeneous Concurrent Services, Service Profile One, 5 VM Per Service

The figures show that as the capacity of the cloud increases there is an associated increase across all Risk levels. This is more pronounced in both the Physical Host Risk Level and the Infrastructure Provider Risk Level where physical resource capacity plays a larger role in these risk level calculation. As would be expected, when many services with larger resource requirements



Navigation

Trust Report

Eco Graphic Report

Risk Report

Provider Type:

○ Service Provider

● Infrastructure Provider

Service Phase:

○ Deployment Phase

● Operation Phase

Provider ID: Umea

Service ID: DaaaS

Date From: 02-07-2013 12:30:00

Date To: 02-07-2013 16:30:00

Options:

□ Auto Refresh

□ Display Last 5 Minutes

□ Test Data

□ Debug Mode

Risk Reports

Remote Query Used: Service ID: DaaaS, Provider ID: Umea, Provider Type: ip, Service Phase: operation, Unix Date: 1372761000 - 1372775400

Submit
Stop
Clear
Cost Report
Risk Report - Infrastructure Provider (ID: Umea) Phase: operation Service ID: DaaaS
Physical Host Risk Level
Virtual Machine Risk Level
Service Level Agreement Risk Level
Infrastructure Provider Risk Level
Risk Level
0
1
2
3
4
5
6
7
1372770931
1372771428
1372771925
1372772422
137277291
Unix Timestamp - Samples From 02-07-2013 15:15:31 to 02-07-2013 15:48:39

Fig. 5. Four Levels Risk Assessment at Service Operation

are deployed in an infrastructure provider, the impact is seen more quickly across all Risk Levels. These figures also illustrate that the risks associated with physical resource capacity do not have an impact on the risks associated with virtual resource capacity. Additionally Figure 8 shows the relationship between the risk levels when the capacity of an IP is filled. From this it can be seen that the Physical Host Risk Level of a service is greater than the Infrastructure Provider Risk Level. The Infrastructure Provider Risk Level adjusts capacity related risks with the risks to an IP across all currently deployed services. Thus if a IP has many services running and is reaching its capacity but the service of interest is consuming a smaller proportion of the total resources available from the IP, this is perceived to be a lower risk to the service.

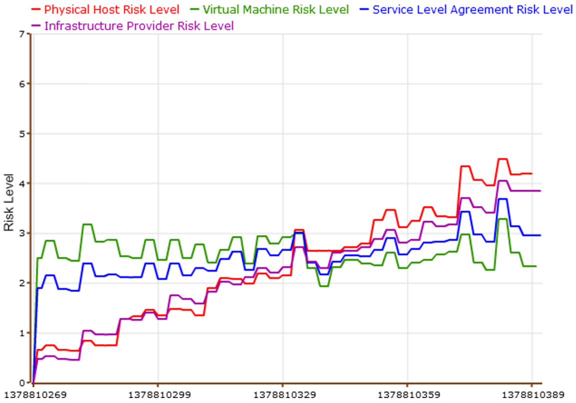


Fig. 7. 1-10 Homogeneous Concurrent Services, Service Profile Two, 5 VM Per Service

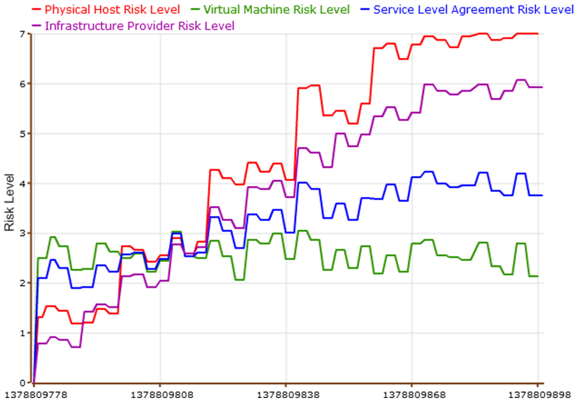


Fig. 8. 1-10 Homogeneous Concurrent Services, Service Profile Three, 5 VM Per Service

In addition, these figures show the relationship between Physical Host Risk Level and Virtual Machine Risk Level, where both resource capacity risks are accounted for in the SLA Risk Level. Since the Virtual Machine Risk Level remains relatively static as the number of services running on the IP increases, the rate at which the SLA Risk Level increases is less. Finally, the figures show that when the physical resources used by a service are shared and multi-tenant to other services, the Physical Host Risk Level to the service of interests increases due to the increasing perceived impact this has on the performance of the service.

Figures 9 shows the results of running two services concurrently with different workload characteristics. The left graph shows that the high variance workload of the larger service (Service Profile Three) shown on the right has an impact on both the Physical Resource Risk Level and the Infrastructure Provider Risk Level of the smaller service (Service Profile One). Finally, the experiment shown in Figure 10 shows how the Physical Host Risk

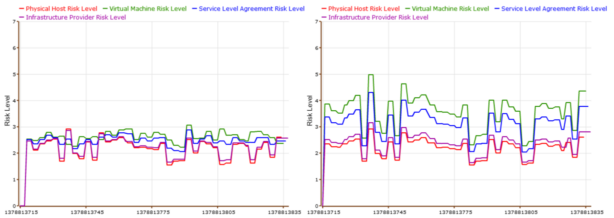


Fig. 9. 2 Concurrent Heterogeneous Services, Left: Service Profile One (Low Variance), Right: Service Profile Two (High Variance), 10 VMs,

Level accounts for only the relevant capacity related risks to a service. Thus this risk level only considers the actual physical resources usage of a service within an IP and not those used by other services. The graphs left and center within this figure, share the same resources while the graph right uses a different subset of resources within the IP due to the round robin nature of physical resource allocation to virtual resources. It can be seen that the Physical Host Risk Level of the service shown on the right graph is dissimilar to the others as expected.

6.3 Extensions

There are many possible extensions to the risk assessor. For example, the risk to a service at operation could be extended to provide risk forecasts by extrapolating the input data currently passed to the risk model. This would include a validation to provide a comparison of real versus forecast values. In addition to this, evaluating how the failure rates affect the risk model will provide additional insights into its effectiveness. Furthermore, the current implementation of the risk assessor supports a number of scenarios that are considered by the OPTIMIS project: (a) cloud federation, (b) private cloud, (c) multi-cloud, (d) broker, and (e) cloud bursting, which are imperative for proactive operation of a provider. These scenarios would benefit from the addition of further elements in the risk inventory that consider the additional vulnerabilities and threats associated not only with failures of physical hosts and VMs, but also with security, legal, and data management aspects. Further work is planned to validate the output of the risk assessor and perform further evaluations of the risk model on different cloud testbeds with services that exhibit a wider variety of varying attributes, including the consideration of the size of the hypothetical IP.

7 Related Work

Risk assessment in distributed computing has been extensively studied in the literature, either as a general methodology [2], [3], [4] or focusing on a specific type of risk, such as security and SLA fulfilment [5], [6]. These projects have included objectives ranging from information protection, evaluation /prediction of QoS and probability of SLA failures [2], [20].

The objective of the Consequence project [21] is to provide an information protection framework and to thereby identify the security risk in sharing data in a distributed environment. The risk items are used as a checklist of items to be addressed in the Consequence architecture, without any assessment of the probability and the negative impact of a risk item. The SLA@SOI project [22] does not explicitly address risk assessment, although it does propose the utilisation of a prediction service for estimating the probability of software and network failures, as well as hardware availability in an attempt to evaluate QoS.

The AssessGrid project [23] proposes a model to estimate the probability of SLA failures in Grid environments, and considers the probability of n resources failing for the scheduled duration of a task as well as the probability that m reserved resources are available for that duration. The probability of node failure is calculated by assuming that the node failures represent a Poisson process, which is non-homogenous in time. The resource provider risk assessment techniques enable the identification of infrastructure bottlenecks, evaluate the likelihood of an SLA violation and, where appropriate, mitigate potential risk, in some cases by identifying fault-tolerance mechanisms such as job migration to prevent SLA violations [24], [25]. The AssessGrid broker acts as a matchmaker between end-users and providers, furnishing a risk optimised assignment of SLA requests to SLA quotes [12] by evaluating the provider reliability with respect to systematic errors. Here, systematic errors refer to provider errors whereby their risk assessments exhibit a typical trend in the sense that they tend to overestimate/underestimate the risk of failure.

Risk as a basis for proactive or reactive service management in Clouds can be seen within the domain of information security / privacy [26]. Information security is suited to the management of risk in clouds as risk can be defined and linked to existing ways of expressing security policies [27]. In these cases an organisation or user can associate events expressed in policy which can be measured using risk assessment, proactive and reactive action can then added to the process to act upon the risk [28]. Risk as a management concept has a significant background in the concept of systems auditing and third party insight into systems [29].

A recent study focusing on clouds analyzes the risks of overbooking resources and proposes a threshold-based overbooking scheme [30]. The trade-off between overbooking and performance degradation is closely related to SLA management, studied e.g. [31] which proposes to extend standard availability SLAs to also include probability of successfully launching additional VMs (model based on CPU usage). They present an algorithmic framework that uses cloud effective demand to estimate the total physical capacity required for performing the overbooking. In order to influence risk assessment along any business/third party grounds the management and the setting of the risk assessment is taken outside of the

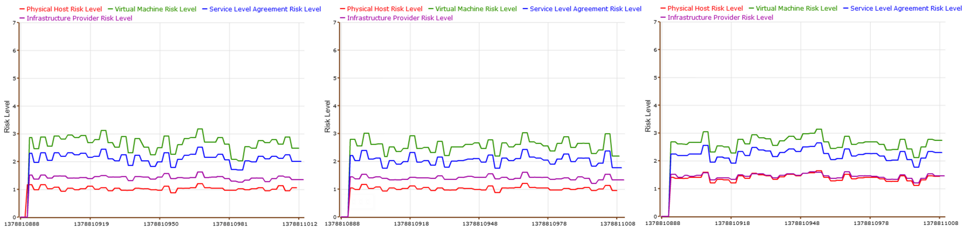


Fig. 10. 3 Concurrent Heterogeneous Services, 5 VMs

cloud fabric and turned it into a service level [4]. In terms of wider resource failure a wide range of studies exist for distributed computing environments [32], [33]. Risk from a third party service as an extension to risk assessment mechanisms has also been explored in cloud environments [34].

In the model proposed in this paper risk is monitored in terms of computing resource behaviour within the domain of the cloud infrastructure provider and interfaces are presented for third party auditing and control. The focus is with respect to internal threats to service execution based on analysis of historical and live data from the cloud infrastructure. This is important for the support of a holistic solution including complementary proactive-reactive risk assessment approaches to tolerate both physical, VM, SLA and infrastructure failures.

8 Conclusion

This paper has proposed a risk assessment framework for cloud computing. The framework is beneficial for end-users and service providers approaching the cloud to deploy and run services, as well as infrastructure providers to deploy and operate those services. These benefits include supporting various parties for making informed decisions regarding contractual agreements, as highlighted in Section 5.

Motivating scenarios for the need for a risk assessment framework in cloud computing were first presented. The importance of risk assessment at various stages of the service life cycle was identified, and the models for assesssing the risk pre-SLA negotiation, SLA negotiation, and service runtime were presented. The need for monitored historical data is showcased, along with the data collection process at service runtime.

The risk model the IP uses at service operation to perform continual risk assessment was developed. The model relies on monitoring low-level events from the infrastructure to support the risk of failure in relation to four assets (physical hosts, VMs, SLAs, and IP infrastructure), and considers vulnerabilities and threats associated with them. The risk model was implemented and evaluated in the context of a real cloud testbed and through fabricated monitoring data that enabled fine-grained control over experimental variables. The software prototype is accompanied with a visualization tool, which provides real-time feedback on current service levels. It is integrated in the risk assessment framework and is a viable contender to enable an IP to identify infrastructure bottlenecks and mitigate potential risks, in some cases by identifying fault-tolerance mechanisms to prevent SLA violations.

The paper has demonstrated the flexibility of the risk model, which has been illustrated to be easily extended in several ways, e.g. for the provision of risk forecasts by extrapolating the input data from the CMI, and by considering additional vulnerabilities and threats associated not only with failures of physical hosts and VMs, but also with security, legal, and data management aspects. The risk assessment model did not consider the intensity of the workload running on a cloud resource. However, there is evidence of a correlation between the type and intensity of the workload and the failure rate of the resource [33]. More importantly, extending the model to cater for this information will provide a more accurate risk estimation.

The risk assessment framework is fully integrated in the OPTIMIS toolkit [18], which simplifies cloud self-management, optimizes the cloud service lifecycle, and supports various cloud architectures. The holistic management ensures risk assessment is linked to service level goals, which combines risk avoidance, reduction in associated negative effects, or the absorption of some or all of its consequences. However, in the current incarnation of the risk assessment framework the SP dynamic risk assessment is limited due to the lack of support for service consumer's side monitoring tools, in addition to the limited availability of shared monitored data from IPs. This will be the subject of future research. The deployment of the OPTIMIS toolkit on a production cloud would be beneficial to further evaluate the risk assessment framework's overall performance and usability.

Acknowledgments

The authors would like to thank the European Commission for supporting this work under FP7-ICT-2009-5 contract 257115 Optimized Infrastructure Services (OPTIMIS), the Ministry of Science and Technology of Spain under contract TIN2007-60625, and the Generalitat de Catalunya under contract 2009-SGR-980.

References

[1] A. J. Ferrer, F. Hernandez, J. Tordsson, E. Elmroth, A. Ali-Eldin, C. Zsigri, R. Sirvent, J. Guitart, R. M. Badia, K. Djemame,

W. Ziegler, T. Dimitrakos, S. K. Nair, G. Kousiouris, K. Konstanteli, T. Varvarigou, B. Hudzia, A. Kipp, S. Wesner, M. Corrales, N. Forgo, T. Sharif, and C. Sheridan, "Optimis: A holistic approach to cloud service provisioning," *Future Generation Computer Systems*, vol. 28, no. 1, pp. 66 – 77, 2012.

[2] K. Djemame, I. Gourlay, J. Padgett, K. Voss, and O. Kao, *Market-Oriented Grid Computing, R. Buyya and K. Bubendorfer (Eds.)*. Wiley, 2009, ch. Risk Management in Grids.

[3] X. Zhang, N. Wuwong, H. Li, and X. Zhang, "Information security risk management framework for the cloud computing environments," in *Proceedings of the 2010 10th IEEE International Conference on Computer and Information Technology*, ser. CIT '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 1328–1334.

[4] J. O. Fitó and J. Guitart, "Business-driven management of infrastructure-level risks in cloud providers," *Future Generation Computer Systems*, vol. 32, pp. 41–53, Mar. 2014.

[5] P. Saripalli and B. Walters, "Quirc: A quantitative impact and risk assessment framework for cloud security," in *Proceedings of the 2010 IEEE 3rd International Conference on Cloud Computing*, ser. CLOUD '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 280–288.

[6] K. Djemame, J. Padgett, I. Gourlay, and D. Armstrong, "Brokering of risk-aware service level agreements in grids," *Concurrency and Computation: Practice and Experience*, vol. 23, no. 7, 2011.

[7] K. Misra, "Risk analysis and management: An introduction," in *Handbook of Performability Engineering*, K. Misra, Ed. Springer London, 2008, pp. 667–681.

[8] "Iso 31000:2009 risk management - principles and guidelines," December 2013, http://www.iso.org/iso/catalogue_detail?csnumber=43170.

[9] W. Jansen and T. Grance, "Guidelines on security and privacy in public cloud computing," National Institute of Standards & Technology, Gaithersburg, MD, United States, Tech. Rep. SP 800-144, 2011.

[10] E. Network and I. S. Agency, "Cloud computing security risk assessment," 2009.

[11] Z. Hua, B. Gong, and X. Xu, "A DS-AHP Approach for Multi-attribute Decision Making Problem with Incomplete Information," *Journal of Expert Systems with Applications,*, vol. 34, pp. 2221–2227, 2008.

[12] K. Djemame, D. Armstrong, M. Kiran, and M. Jiang, "A risk assessment framework and software toolkit for cloud service ecosystems," in *Proceedings of the Second International Conference on Cloud Computing, GRIDs, and Virtualization*, Rome, Italy, Sep. 2011.

[13] K. Djemame and R. Alsoghayer, "Resource failures risk assessment modelling in distributed environments," *Journal of Systems and Software*, 2014, elsevier.

[14] A. Khan, M. Oriol, M. Kiran, M. Jiang, and K. Djemame, "Security risks and their management in cloud computing," in *Cloud Computing Technology and Science (CloudCom), 2012 IEEE 4th International Conference on*, 2012, pp. 121–128.

[15] M. Lund, B. Solhaug, and K. Stolen, *Model-Driven Risk Analysis - The CORAS Approach*. Springer, 2011.

[16] K. Djemame, B. Barnitzke, M. Corrales, M. Kiran, M. Jiang, D. Armstrong, N. Forg, and I. Nwankwo, "Legal issues in clouds: towards a risk inventory," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1983, 2013.

[17] M. Macías and J. Guitart, "Cheat-proof trust model for cloud computing markets," in *GECON'2012*, ser. Lecture Notes in Computer Science, K. Vanmechelen, J. Altmann, and O. F. Rana, Eds., vol. 7714. Springer, 2012, pp. 154–168.

[18] "Optimis toolkit," 2013, http://optimistoolkit.com/.

[19] G. Katsaros, G. Gallizo, R. Kübert, T. Wang, J. O. Fitó, and D. Henriksson, "A multi-level architecture for collecting and managing monitoring information in cloud environments," in *Proceedings of the 1st International Conference on Cloud Computing and Services Science (CLOSER'2011)*, Noordwijkerhout, The Netherlands, May 2011, pp. 232–239.

[20] A. Morali and R. Wieringa, "Risk-based confidentiality requirements specification for outsourced it systems," *Requirements Engineering, IEEE International Conference on*, pp. 199–208, 2010.

[21] "Consequence: the context-aware data-centric information sharing," 2011, http://www.consequence-project.eu.

[22] "Sla@soi: Slas empowering a dependable service economy," 2011, http://sla-at-soi.eu.

[23] K. Djemame, I. Gourlay, J. Padgett, G. Birkenheuer, M. Hovestadt, O. Kao, and K. Vo, "Introducing Risk Management into the Grid," in *Proceedings of the 2nd IEEE International Conference on e-Science and Grid Computing (eScience2006)*. Amsterdam, Netherlands: IEEE Computer Society, 2006.

[24] C. Carlsson, "Risk Assessment for Grid Computing with Predictive Probabilities and Possibilistic Models," in *Proceedings of the 5th International Workshop on Preferences and Decisions*, Trento, Italy, Apr 2009.

[25] D. Battré, G. Birkenheuer, M. Hovestadt, O. Kao, and K. Voss, "Applying Risk Management to Support SLA Provisioning," in *The 8th Cracow Grid Workshop*. Academic Computer Center CYFRONET AGH, 2008.

[26] P. Srivastava, S. Singh, A. A. Pinto, S. Verma, V. K. Chaurasiya, and R. Gupta, "An architecture based on proactive model for security in cloud computing," in *Recent Trends in Information Technology (ICRTIT), 2011 International Conference on*. IEEE, 2011, pp. 661–666.

[27] C. Chen, W. Han, and J. Yong, "Specify and enforce the policies of quantified risk adaptive access control," in *Computer Supported Cooperative Work in Design (CSCWD), 2010 14th International Conference on*. IEEE, 2010, pp. 110–115.

[28] S. Pearson, "Toward accountability in the cloud," *Internet Computing, IEEE*, vol. 15, no. 4, pp. 64–69, 2011.

[29] L. Willcocks and H. Margetts, "Risk assessment and information systems," *European Journal of Information Systems*, vol. 3, pp. 127–127, 1994.

[30] R. Ghosh and V. K. Naik, "Biting off safely more than you can chew: Predictive analytics for resource over-commit in iaas cloud," *2012 IEEE Fifth International Conference on Cloud Computing*, vol. 0, pp. 25–32, 2012.

[31] D. Breitgand, Z. Dubitzky, A. Epstein, A. Glikson, and I. Shapira, "Sla-aware resource over-commit in an iaas cloud," in *CNSM*. IEEE, 2012, pp. 73–81.

[32] T. J. Hacker, F. Romero, and C. D. Carothers, "An analysis of clustered failures on large supercomputing systems," *Journal of Parallel and Distributed Computing*, vol. 69, no. 7, pp. 652–665, Jul 2009.

[33] B. Schroeder and G. A. Gibson, "A large-scale study of failures in high-performance computing systems," *Dependable and Secure Computing, IEEE Transactions on*, vol. 7, no. 4, pp. 337–350, 2010.

[34] B. S. Kaliski Jr and W. Pauley, "Toward risk assessment as a service in cloud environments," in *Proceedings of the 2nd USENIX conference on Hot topics in cloud computing*. USENIX Association, 2010, pp. 13–13.

Karim Djemame Karim Djemame was awarded a Ph.D. at the University of Glasgow, UK, in 1999, and is currently holding a Senior Lecturer position at the School of Computing, University of Leeds. He sits on a number of international programme committees for cloud middleware, computer networks and performance evaluation. He was the investigator of various e-Science/Grid projects including DAME, BROADEN, and AssesGrid. He is currently involved in various research projects including OPTIMIS, ISQoS, and STRAPP. His main research areas focus on Grid/Cloud computing, including system architectures, resource management, and risk assessment. Dr. Djemame is a member of the IEEE.

Django Armstrong Django Armstrong was awarded a Ph.D. at the University of Leeds, UK, in 2013, and is currently a research associate at the School of Computing at the University of Leeds. He has experience working in a number of EC-funded projects including AssessGrid and OPTIMIS. His research is in the field of Distributed Systems and the complementary paradigms of Grid and Cloud Computing, with a specific interest in quality of service, performance evaluation and virtualization.

Jordi Guitart Jordi Guitart received the MS and Ph.D. Degrees in Computer Science at the Technical University of Catalonia (UPC), in 1999 and 2005, respectively. Currently, he is an associate professor at the Computer Architecture Department of the UPC and an associate researcher at Barcelona Supercomputing Center (B.SC.) within the Autonomic Systems and eBusiness Platforms research line. His research interests are oriented toward innovative resource management approaches for modern distributed computing systems. He is involved in a number of European projects.

Mario Macias Mario Macias obtained his Master Degree in Computer Science at the Universitat Autonoma de Barcelona, Spain. After a few years working in the IT industry, he joined Barcelona Supercomputing Center (BSC) in 2006 to work in European Research Projects and carry out his PhD studies at Technical University of Catalonia (UPC). He combines his main job at BSC with his lectures at UPC.