A Systematic Mapping Study on Real-time Cloud Services

Jakob Danielsson, Nandinbaatar Tsog, Ashalatha Kunnappilly
Mälardalen University, Västerås, Sweden
{jakob.danielsson, nandinbaatar.tsog, ashalatha.kunnappilly}@mdh.se

***Abstract*—Cloud computing is relatively a new technique to host and use the services and applications from the internet. Although it offers a multitude of advantages like scalability, low operating cost, easy access and maintainability, etc., often they are not utilized to the fullest due to the lack of timeliness property associated with the cloud. The cloud services are mainly designed to maximize throughput and utilization of resources and hence incorporating predictable execution properties in the cloud is arduous. Nevertheless, cloud still remains a highly attractive platform for hosting real-time applications and services owing to features like elasticity, multi-tenancy, ability to survive hardware failures, virtualization support and abstraction layer support which provides flexibility and portability. In order for real-time safety critical applications to exploit the potential of cloud computing, it is very essential to ensure the predictable real-time behavior of cloud services. So in this paper, we perform a detailed systematic mapping study on real-time cloud services to identify the current research directions and potential research gaps. Our study focuses on analyzing the current architectures and software techniques that are available at present to incorporate real-time property of the cloud services. We also aim at investigating the current challenges involved in realizing a predictable real-time behavior of cloud services.**

***Index Terms*—Cloud computing, real-time, safety-critical applications, systematic mapping study.**

I. Introduction

The recent years have witnessed a tremendous increase in the usage of cloud computing technologies by the industries and this has been accounted due to the fact that cloud services can deliver high performance solutions, with high flexibility. Cloud computing services can typically be used in any system which require computational power, e.g., autonomous vehicle systems and big data centers. The characteristics of a cloud based service is that most of its computational power is located in one specific location and when it receives a computational request from a node, it performs the necessary computation and then distribute the results to the requesting node. A cloud system uses three types of services - Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) [1].

There is a tremendous potential for the cloud computing services in the present scenario, but most of it is not being utilized and this is mainly due to the lack of real-time property of cloud computing solutions. For instance, systems such as airbag systems and braking systems often contain strict real-time constraints to avoid disastrous consequences. In order to avoid such catastrophic consequences, all tasks inside a system should be executed within a predictable time interval, thus creating a real-time task schedule. Since a cloud computing service contains many layers which are dependent on each other, implementing real-time systems using cloud services can be very complex, e.g., a real-time task at the top SaaS layer can be dependent on the lower layers to achieve real-time performance. In addition, there are also other major challenges in achieving a real-time cloud such as dealing with scheduling algorithms which are not time-based, non-support for a real-time clock as an internal reference, delay in provisioning resources and virtual machines, lack of predictability of execution of tasks, etc. In this paper, we conduct a systematic mapping study on 'Real-time cloud services' with the aim of analyzing the techniques for achieving predictable real-time behavior in cloud computing scenario and also identifying papers which have addressed the challenges listed above [2], [3].

The rest of the paper is organized as follows: Section 2 describes the process followed for our systematic mapping study, including screening methods, classification scheme and initial search results. In Section 3, we describe in detail our results and analysis generated from the mapping study. Section 4 concludes the paper and give some directions for future work.

II. Mapping Study Process

A systematic mapping study aims at providing a deep overview of researches conducted in a particular field and aids in structuring the research results in an efficient manner to identify research gaps [2]. In order to perform a systematic mapping study on our topic of interest "Real-time cloud services", the authors have followed the mapping study approach detailed by Petersen et.al [2]. The study approach is divided in to 5 steps which starts with the definition of research questions, followed by conducting the search, screening the papers, key wording and data extraction and finally the mapping process. The steps are illustrated in Figure 1. In the following

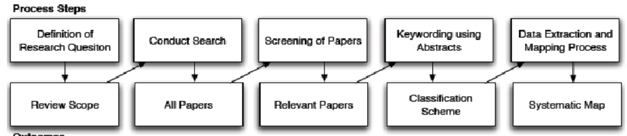

Figure 1: Systematic mapping process.

subsections, we detail each of these steps with respect to our research focus area, which is “Real-time cloud services”.

A. Research Questions

The first step of mapping study comprises of framing the research questions to identify the scope of the research. The main goal of our study is to analyze how predictable real-time behavior can be achieved within a cloud environment and to identify the hindrances, if any, to achieve this behavior. This bigger goal is subdivided into 3 research questions to ease our analysis. The research questions are detailed below:

Research question 1: Which software techniques are most commonly used for achieving real-time behavior in a cloud environment?

The question aims at answering all software related queries and techniques, like the type of scheduling algorithms and other software processes developed for achieving real-time behavior for cloud systems.

Research question 2: What are the major cloud architectures developed for achieving predictable real-time cloud services and how can we modify the existing cloud architectures to achieve the desired behavior?

Under this question, we would like to explore how the existing cloud platforms like Saas, PaaS and IaaS are being used for supporting real-time applications and also identify some new frameworks, if any, that can support real-time behavior. Since there exist different layers for a cloud system, each one of them may have to be altered in order to achieve a predictable behavior of the cloud system.

Research question 3: Are there any challenges identified for achieving a predictable real-time behavior in cloud?

There are many challenges that are accounted even in the simplest of real-time systems such as scheduling algorithms and cache design. Therefore the main challenges for ensuring predictability in complex systems such as cloud must not be left without investigation. There might also be some unknown challenges that expose itself only when we start experimenting with various real-time cloud solutions.

After the definition of research questions, we proceed to Step 2 of our mapping study, which is conducting the search in various databases.

B. Conducting the search

In order to conduct the search, we first define a set of keywords that can be inserted into the search string for retrieving publications from various databases. In our study, we restrict the databases to IEEE, Scopus, Springer, ACM and Science direct. We have not used Google Scholar in our study as it retrieved us with a lot of non-peer reviewed articles and we wanted to restrict the scope of this study to peer reviewed articles. The search queries we defined are based on our field of study, which is a combination of real-time and cloud. With the motive of getting all the papers related to real-time cloud services, we arranged our search query in such a way to contain either the key phrases ”cloud computing” or ”cloud services”, thus the full query we used is: ”real-time” AND (”cloud computing” OR ”cloud services”). Table I show the initial search results generated.

Database	Initial hits	By Title	By Abstract	Full text
IEEE	3628	70	34	23
ACM	255	14	12	7
Springer	1,595,798	121	12	9
Scopus	11,221	88	9	3
Science direct	3,754	6	2	2

Table I: Search query results.

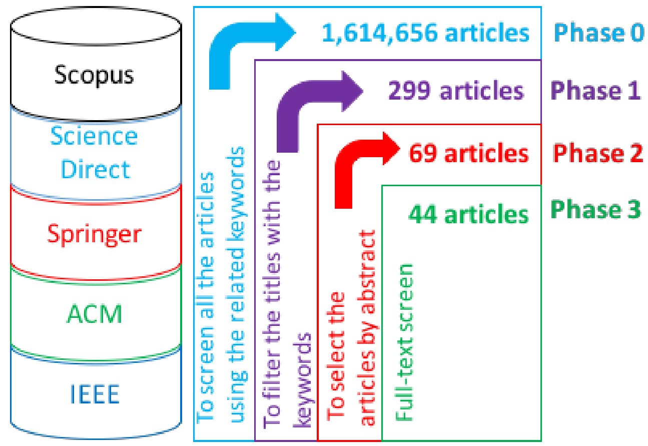

Figure 2: Selection phases.

It was found that our initial search with the above search string extracted 3628 articles from IEEE, 255 from ACM, 1,595,798 from Springer, 11,221 from Scopus and 3754 from Science direct. These numbers were huge to deal with and hence the works falling outside the research focus area should be eliminated and hence we proceed with the third step, screening the papers.

C. Screening the papers

As mentioned, a screening procedure was required to effectively select the relevant papers within the area of our study. The screening mechanism which we employed in our study is based on several inclusion-exclusion criteria. Our inclusion-exclusion criteria is explained in detail below.

1) **Inclusion Criteria**
   a) The papers must be peer- reviewed.
   b) Paper should be based on techniques to achieve real-time cloud services, be it software related or architecture related.
   c) Papers must report challenges in incorporating timeliness property to cloud services.
   d) Paper must address the cloud from “cloud computing” and “cloud services” perspectives.
2) **Exclusion Criteria**
   a) Paper must report on ”cloud” related to “real-time”.
   b) All the non peer reviewed papers and articles in the form of abstracts, editorials or keynotes are excluded from the mapping study.

c) Papers which are not in English language are excluded.
d) Only the papers in the year range of 2009-2016 were selected to include the most recent developments.

The inclusion-exclusion criteria which we defined could successfully eliminate papers which were outside the scope of the current study. The screening procedure based on the above inclusion-exclusion criteria was conducted in 4 phases. The diagram showing different phases of study is described in Figure 2 and the results are tabularised in Table I. The phases of study are described in detail below.

1) **Phase 0:** In this phase, we have screened the papers based on the application of our search string to various database sources. The results of Phase 0 are already obtained in Step 2 and it yielded us with a total of 1,614,656 articles from different database sources.
2) **Phase 1:** In Phase 1, we consider all the papers from Phase 0, and do a 'Title-based selection' and this could efficiently bring down the paper results when compared to Phase 0. By performing a title based search, we could successfully bring down the papers to 299 of which IEEE, ACM, Springer, Scopus and Science direct constitute 70,14,121,88 and 6 articles respectively.
3) **Phase 2:** After screening of papers based on title, we proceeded to Phase 3, which is an abstract based selection. In this phase, we also eliminated the duplicate papers which brought down the paper count to 69 articles, of which IEEE has a contribution of 34, ACM 12, Springer 12, Scopus 3 and Science direct 2 articles.
4) **Phase 3:** In some of the papers, reading the abstract was insufficient to capture the necessary information and hence we proceeded by reading the full-text, mainly the introduction and conclusion sections. This could further narrow down our search to 44 papers, which we consider further for our mapping study.

D. Key-wording technique and generating the classification scheme

After performing the phase-wise screening of papers based on inclusion and exclusion criteria, we assigned keywords to each paper. Based on the research questions defined earlier, our motive was to identify the papers that were related to software techniques and architecture frameworks developed for achieving real-time cloud services. At the same time, we were also keen in analyzing the challenges for achieving predictable performance in cloud services, which further helped us to identify the research gaps in the area. We found that reading the abstracts was not enough to generate efficient keywords and build a classification approach as there were domain overlaps and hence we proceeded by reading the introduction and conclusion as well, and then utilized this information for key-wording. The keywords which we assigned are 'software-techniques', 'architecture based approaches' and 'challenges'. After analyzing the contribution and context of research from the selected papers, a higher level view of the research was identified, which helped us to generate a more detailed classification scheme. The classification scheme is illustrated in Figure 3.

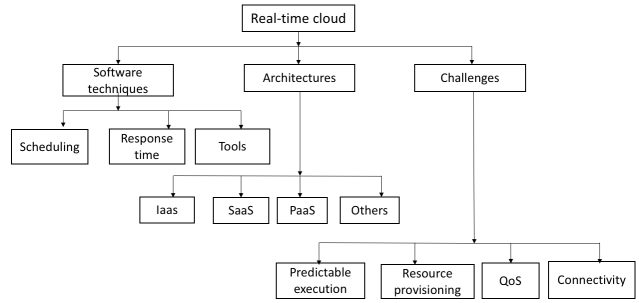

Figure 3: Classification Scheme.

The software techniques for achieving real-time cloud services were again categorized as scheduling based, response-time based and tool-based. Under scheduling based papers, we had papers discussing about task scheduling, resource allocation and minimizing the total cost. Apart from that, there were also some papers dealing with the response-time analysis and some others reasoning about various methodologies/tools for achieving predictable cloud services, e.g, an improved cache based system and therefore we could classify them into categories named response- time based and tool-based papers. Under the category of architecture frameworks for achieving real-time services, we accomplished a detailed classification based on the major cloud platforms used like Saas, PaaS and IaaS. The remaining real-time cloud architecture categories were classified into the category 'others', which included Control as a Service (CaaS) architectures, community cloud and graph based architectures. Identifying the challenges in real-time cloud computing was the most trickiest part of the lot as each of the papers we analyzed presented its own limitations. However, we restricted our focus on identifying major challenges in achieving a predictable real-time environment and hence the challenges were further classified into papers mainly describing the hurdles to achieve predictable execution in cloud computing environment, challenges in successful resource provisioning for achieving real-time cloud services and also papers describing QoS and connectivity hindrances. The results are presented in detail in the next section.

III. MAPPING STUDY RESULTS

The previous sections unveiled the detailed mapping study approach which we followed for structuring the research results in the field of 'Real-time cloud'. In this section, we will discuss the results obtained by this mapping study. The results are categorized into 3 sub-sections. In the first subsection, we categorize our primary results based on research focus area, publication type and contribution type and also by the detailed classification scheme which we explained earlier. In the second subsection, we present our extensive mapping study results based on 44 papers we selected for performing the mapping study. We did a comparison-based approach by detailing the research contributions, approaches, scope of the works, limitations and results [3]. And finally in the third subsection, we discuss the research gaps identified from our study.

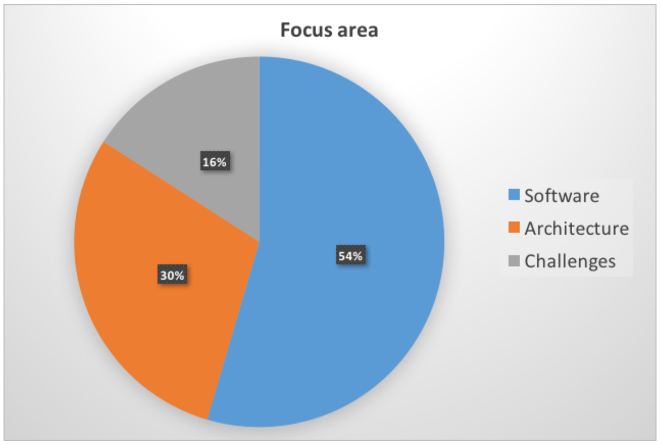

Figure 4: Distribution of primary studies by research focus area.

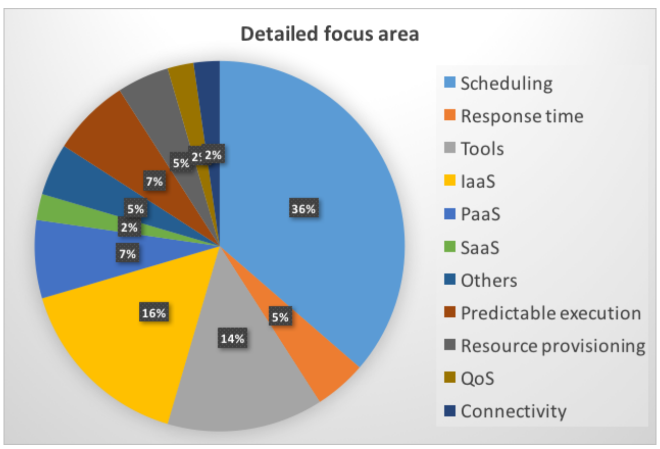

Figure 5: Distribution of primary studies by detailed focus areas.

A. Primary results

Under this subsection, we will illustrate the primary results from our systematic mapping study, generated from 44 articles, which comprised of various journal articles, conference and workshop proceedings and book chapters from 2009-2016. We have developed detailed classification schemes for the distribution of our primary studies based on publication type, research focus area and contribution type. The distribution of papers by publication type demonstrated that out of 44 papers we selected, 24 papers were conference proceedings, 9 were journal articles, 6 were book chapters and 5 were workshop proceedings. The results are tabularised in Table II. Another classification we performed was based on research focus areas, which aimed at finding out whether the paper

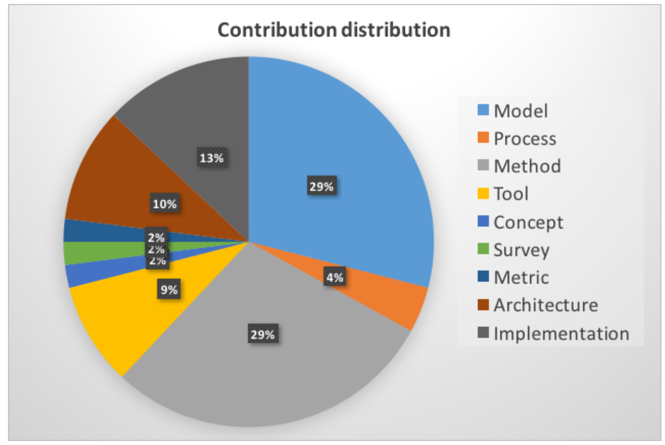

Figure 6: Distribution of primary studies by contribution type.

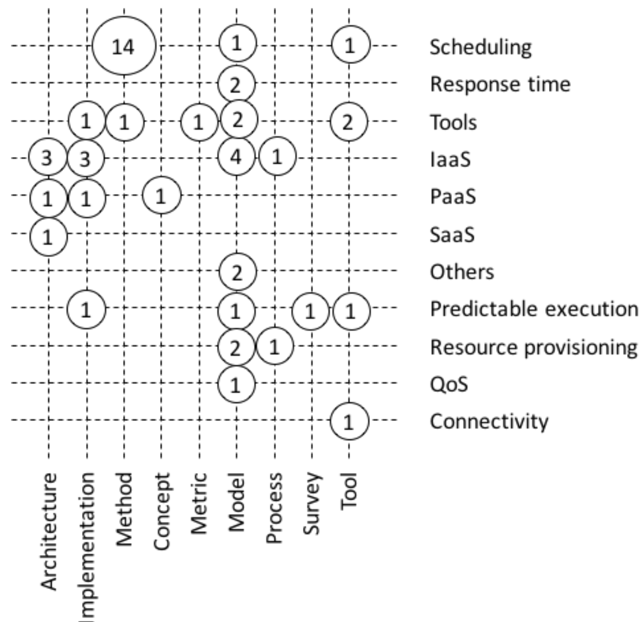

Figure 7: Map of focus areas in research on real-time cloud services by contribution type.

falls under the category of software techniques, architectures or challenges tackled. Figure 4 show a pie-graph, depicting that 54% of the papers were software techniques related, 30% were architecture related and another 16% addressing the challenges. Further we proceeded to identify the detailed division of focus areas based on the classification scheme which we detailed earlier. The results are depicted in Table III and also the percentages are depicted via Pie-graph, shown in Figure 5. By analyzing the results, we could identify that papers on software techniques for achieving real-time cloud were further classified as papers based on scheduling (16 papers, 36%), based on response time analysis (2 papers, 5%) and on tools (6 papers, 14%). Architecture-based papers are also sub classified as IaaS (7 papers, 16 %), PaaS (3 papers,7 %), SaaS (1 paper,2 %), and other categories (2 papers, 5%). And

Publication type	2009	2010	2011	2012	2013	2014	2015	2016	Total	%
Conference proceedings	-	3	4	3	5	6	1	2	24	55
Journal article	-	-	1	-	-	3	4	1	9	20
Book chapter	-	1	-	-	-	2	1	2	6	14
Workshop proceedings	1	-	-	1	-	-	2	1	5	11
Total	1	4	5	4	5	12	8	6	44	100

Table II: Distribution of primary studies by publication type.

Focus area	2009	2010	2011	2012	2013	2014	2015	2016	Total	%
Scheduling	-	2	2	-	2	6	1	3	16	36
Response time	-	-	-	-	1	-	1	-	2	5
Tools	-	1	1	3	-	-	1	-	6	14
IaaS	-	-	-	-	1	4	1	1	7	16
PaaS	-	1	-	1	1	-	-	-	3	7
SaaS	-	-	1	-	-	-	-	-	1	2
Others	-	-	-	-	-	-	2	-	2	5
Predictable execution	-	-	1	-	-	1	1	-	3	7
Resource provisioning	1	-	-	-	-	-	-	1	2	5
QoS	-	-	-	-	-	-	1	-	1	2
Connectivity	-	-	-	-	-	-	-	1	1	2
Total	1	4	5	4	5	12	8	6	44	100

Table III: Distribution of primary studies by focus area.

Contribution type	2009	2010	2011	2012	2013	2014	2015	2016	Total	%
Model	1	1	2	1	2	3	4	1	15	29
Process	1	-	-	-	-	1	-	-	2	4
Method	-	1	2	1	2	5	1	3	15	29
Tool	-	1	1	-	-	1	1	1	5	10
Concept-based	-	-	-	1	-	-	-	-	1	2
Survey	-	-	-	-	-	1	-	-	1	2
Metric	-	-	-	-	-	-	1	-	1	2
Architecture	-	1	1	-	1	1	-	1	5	10
Implementation	-	-	1	1	1	1	1	1	6	13
Total	2	4	7	4	6	14	8	7	51	100

Table IV: Distribution of primary studies by contribution type.

also, the papers reporting the challenges were also categorized into papers reporting on predictable execution challenges (3 papers,7%), resource provisioning challenges (2 papers, 5%), QoS issues (1 paper, 2%) and connectivity problems (1 paper, 2%).

We have also created a grouping of papers based on contribution type. The results are detailed in Table IV and the percentage wise distribution depicted by a pie-graph as shown in Figure 6. The tables and pie-chart distribution depicts 29% of papers based on model (15 papers), 4% on process (2 papers), 29% on methodology (15 papers), 10% on tool (5 papers), 2% concept-based (1 paper), 2% survey-based (1 paper), 2% metric-based (1 paper), 10% based on architecture (5 papers) and another 13% based on implementation (6 papers). We have also generated a map of focus areas in research on real-time cloud services as shown in Figure 7.

B. Extensive mapping study results and comparisons

The scope of this section is to illustrate our extensive mapping study results. The results are tabulated and presented in Table V. We have done a deeper analysis on 44 papers which we have selected for the mapping study. Our approach consisted of identifying the major contributions of the papers and then classifying it based on our classification scheme, scope of research and approaches. We also proceeded to analyze the results of the articles and then checked whether the results are experimentally validated or not. In each of the paper, we also tried to identify the limitations / constraints encountered while experimenting the particular approach. This approach helped us to efficiently structure the research results within the area of real-time cloud computing and assisted us further in identifying the research gaps in the area. The research gaps identified will definitely pave the way for future researches in this area. The major research gaps which we identified are discussed in the next section.

C. Discussion

In this section, we detail about the potential research gaps we have identified based on our mapping study and they are detailed below.

1) *Lack of unified methods and integrated architecture support for achieving real-time cloud services.*
   The existing research works in the area of real-time cloud architectures and methodologies lack a unified approach or architecture which can be used to reliably host real-time applications. Though most of the existing works are aimed at altering the existing IaaS platform to suit the real-time functionality, we could not find an approach integrating the various platforms like IaaS,

PaaS, SaaS which we believe is very essential to further exploit the advantages of real-time cloud.

Author	RQ#	Focus area	Problem addressed	Approach	Scope of approach	Limitations/Constraints	Validated	Results of validation
Hasan et al. [4]	1		Designed an architecture framework called 'Control as a Service' (Caas) to check the extent of cloud control possible in modern day car for realizing real-time services like braking, acceleration etc. with WLAN/CAN gateway and conducted a feasibility study using cloud based throttle limitation scenario.	Model	Delivers an architecture solution to validate the control CaaS framework in automotive domain.	Handling real-time varying network delay, rt-performance, connectivity etc are questions to be analyzed before it can be implemented and utilized in modern cars.	Yes	Simulation results for validating proof of concept of Control as Service in throttle limitation scenario in modern cars.
Mohammad Aazam, Eui-Nam Huh [5]	3	QoS	Developed a reimburse framework for a given up service in a highly dynamic environment when using cloud services for achiveing real time communication leading to an effective reimbursement considering quality degradation	Model	Developed a framework to effectively model QoS degradation and underutilized service using Netflix and SD streams and their pricing in model through which fairness and customer satisfaction of cloud consumers can be met	Quality of Experience from different users was not considered while developing the model and the extend to which it fairness is ensured	yes	Degradation factor was modeled in the form of an equation and effective reimbursement policy was calculated.
Dominik Meilnder, Sergei Gorlatch [6]	3	Resource provisioning	Developed a generic scalability model for ROIA on clouds that monitors the application performance at runtime is proposed and it predicts the load-balancing decisions and also whether and how often to redistribute workload or add/remove cloud resources.	Model	Scalability model for predicting the maximum supported numberof users per server in shooter online gaming scenario.	Load balancing is a critical challenge for ROIA on Clouds.	yes	Scalability is modeled w.r.t. to two kinds of resources - computation (CPU)and communication (network) and combining them and experimentally validated.
Kyoungho An et al. [7]	1	Tools	A publish/subscribe middleware based on Data Distribution Service (DDS) called Scalable and QoS-enabled virtual resource monitoring system for Real-Time applications in Clouds(SQRT-C) for dependable and real-time resource monitoring in the cloud was developed.	Model	Developed a middleware based on publish/subscribe paradigms, which can effectively allocate resources to ensure real-time property of the cloud	Constraints: We require resource information be obtained as a service rather than directly by accessing physical/virtual machines. Limitations: did not test QoS settings of DDS, only tested with OpenNebula, and experiment did not show the optimal amount of publishers	yes	Experimental results to show that SQRT-C outperforms the REST-ful monitoring approach in terms of (a) latency, (b) scala-bility, and (c) jitter
Kyong Hoon Kim et al. [8]	3	Resource provisioning	1) Provided real-time Cloud service framework for requesting a virtual platform 2) Investigated various power-aware VM provisioning schemes based on DVFS (Dynamic Voltage Frequency Scaling) schemes	Model and Process	Developed a framework for power-aware provisioning of virtual machines for real-time services.	1) Impact on cooling systems not analysed 2) Need analysis across many approaches like bin packing or linear programming etc. to strongly prove the results.	yes	1) Developed the real-time cloud service framework based on real-time virtual machine model. 2) Evaluate simulations of power-aware real-time services using the CloudSim toolkit with additional development of power-awareness capability.
Lars Lundberg, Sogand Shirinbab [9]	1	Response time	How to apply the existing real-time related theory to cloud-based virtualized environments	Model	Described a technique for calculating a period and a WCET for a VM containing a RT application with hard deadlines		no	
Cong-Thinh Huynh et al. [10]	1	Scheduling	How to tradeoff QoS, response time, against cost efficency	Method	Proposed the Cost efficient Real-time Application Scheduling (CERAS) algorithm		Yes	The results show that how efficient the CERAS algorithm can guarantee application's deadline while achieving the optimal resources need and cost

Author	RQ#	Focus area	Problem addresses	Approach	Scope of approach	Limitations/Constraints	Validated	Results of validation
Marcelo Dutra Ös, Graça Bressan [11]	2	IaaS	Proposed an IaaS based community cloud architecture(decentralized architecture) for a category of real time application, called 'trading application'. It also analyzes the schedulability using scheduling algorithm based on the heuristic function to ensure real-time property.	Model and Process	Developed a community cloud architecture, which is a highly decentralised architecture based on Iaas to use it for real-time finiacial application and analysed the schedulability of resources	Need to analyze the results on numerous other cloud environments. Delay involved in the provisioning of virtual machines in cloud environments was a challenge and paper has assumer preprovisoning of required resources.	Yes	Different simulation scenarios was used in order to validate the architecture and also with the different values the heuristic function for scheduling.
Laura Belli et al. [12]	2	Others	Proposed a a Graph-Based Cloud Architecture relying on the concepts of data listener and data-oriented processing graph in order to implement a scalable, highly configurable, and dynamic chain of computations on incoming real-time big streams and to dispatch data with a push-based approach, thus providing the lowest delay between the generation of information and its consumption.	Model	Developed a novel Cloud architecture for the managementof Big Stream applications in IoT scenarios.	Validation of the architecture and testing with RT data in various scenarios are not yet done.	No	The architecture is prposed and it has to be implemenetd and tested for various RT scenarios.
Jing Zhang, Qianmu Li·Wei Zho [13]	3	Connectivity	Proposed a distributed layered cache system for real-time cloud services, HDCache, built on the top of Hadoop Distributed File System (HDFS).	Tool	Developed a novel cache system for ensuring simultaneous access of file data to support real-time functionality of cloud services.	The HDCache system is realtively simple,based on the simple sparse-write-intensive-read data access model in the cloud. However, many real-time services have more complex and dynamic data access models. Hence the study fails to address this concern. There are many other techniques for which the technique needs to be analysed for like NoSQL database, MapReduce framework and distributed data warehouse.	Yes	Experimental results proved that HDCache can store files with a wide range of sizes and has the access performance in a millisecond level under highly concurrent environments. Hit ratio obtained was higher than 95%.
Meng Xu et al. [14]	3	Predictable exe-cu-tion	Proposed a framework for cache-aware compositional analysis technique to be used for timing guarantees of components scheduled on a multicore virtualization platform. Developed three different approaches,baseline,task-centric and model-centric.	Tool	Developed a novel framework framework that can provide timing guarantees for multiple components running on a shared multicore virtualization platform based on multicore compositional analysis.	Cache-aware compositional analysis for multicore virtualization platforms is chal-lenging because virtualization introduces additional overhead that is difficult to predict. It is hard to account for cache misses that can occur when a VCPU finishes its budget and stops its execution.	Yes	Experimental results on varying workload (using randomly generated workloads) showed a significant improvement in reducing resource bandwidth and improves resource savings and offer better predictability by reducing miss rates.
Fei Teng et al. [15]	1	Scheduling; task schedul-ing	To propose the Paused Rate Monotonic algorithm for scheduling hard real-time tasks on a MapReduce-based cloud	Method	Formalizing the scheduling problem by specifying the characteristics of a task, cluster, and algorithm. Also proposes the PRM algorithm. Finally proves a utilization bound for a set of real-time tasks fully utilizing the MapReduce-based cluster	Analyzed theoretically	yes	The PRM algorithm outperforms traditional real-time algorithms by improving the probability that a real-time task set can be scheduled on a MapReduce-based cloud

Author	RQ#	Focus area	Problem addresses	Approach	Scope of approach	Limitations/Constraints	Validated	Results of validation
Seyedmehdi Hosseini-motlagh et al. [16]	1	Scheduling min-i-mize en-ergy con-sump-tion	How to minimize energy consumption of the host while keeping certain number of instruction executions	Method	Introduced an optimal utilization level of a host to execute a cetain number of instructions to minimize energy consumption of the host. Also proposed a virtual machine scheduling algorithm		yes	The simulation result shows that the proposed algorithm not only reduces total energy consumption of a Cloud by 60%, but also has a profound impact on turnaround times of RT tasks by 94%.
Huangning Chen, Wenzhong Guo [17]	1	Scheduling; task schedul-ing	How to perform task scheduling in large-scale distributed systems such as cloud systems.	Method	Proposed a soft realtime task scheduling algorithm based on particle swarm optimization approach for cloud computing.		yes	The proposed algorithm can effectively minimize deadline missing ratio, maximize the profit of cloud service provider and achieve better load balancing compared with baseline algorithms.
Yi Zhang et al. [18]	1	Scheduling; task schedul-ing	How to achieve energy reductuion for cloud providers and payment saving for cloud users, and at the same time, without introducing VM migration overhead and without compromising deadline guarantees for user tasks.	Method	Proposed a heuristic task scheduling algorithm called Energy and Deadline Aware with Non-Migration Scheduling algorithm.		yes	The algorithm performs better than other existing algorithms on achieving energy efficiency without introducing VM migration overhead and without compromising deadline guarantees.
Song Wu et al. [19]	1	Scheduling; task schedul-ing	How to guarantee real-time constraints in virtualization and cloud computing.	Tool	Proposed a mechanism called multi-core dynamic partitioning to divide physical CPUs into two pools dynamically in a real-time scheduling framework in virtualized environment.	Xen hypervisor	yes	The experimental results show that the real-time scheduling framework (Risa) supports real-time applications well, reduces operation expense, and improves CPU utilization.
Higinio Mora Mora et al. [20]	1	Response time	How embedded and mobile systems can take advantage of the remote computing resources to meet with real time constraints.	Model	Developed a simple application examplewhich shows a real scenario.		yes	The simulation result shows the simplicity of the proposed model and ease to design scheduling tasks.
Kyong Hoon Kim et al. [21]	1	Tools	Energy-efficient management of computing resources via Virtual Machine (VM) provisioning in Cloud computing environments.	Model and Tool	Developed a real-time cloud service framework to investigate the topic.	It needs to use CloudSim toolkit	yes	The results show that data centers can reduce power consumption and increase the profit with their method.
Joe Hoffert et al. [22]	1	Tools	How machine learning can effectively perform autonomic configuration for fog.	Tool	Implemented adaptive middleware and network transports platform (ADAMANT)	by configuring Emulab	yes	The results show how computing hardware environments affect QoS for these systems and how ADAMANT configures the system based on the computing resources
Marisol García-Valls et al. [23]	3	Predictable exe-cu-tion	To identify the technical challenges in supporting RT applications. To survey recent advancement in RT virtualization and cloud computing technology. To offer research directions to enable cloud-based real-time applications in the future	Survey	Survey for real-time application in cloud computing		no	

Author	RQ#	Focus area	Problem addresses	Approach	Scope of approach	Limitations/Constraints	Validated	Results of validation
Huangke Chena et al. [24]	1	Scheduling; task scheduling	To describe the uncertainty of the computing environment and a scheduling architecture to mitigate the impact of uncertainty on the task scheduling quality for a cloud data center.	Method	PRS methods for scheduling real-time, aperiodic, independent tasks.	An experimental cloud environment using Apache CloudStack 4.2.0.	yes	PRS performs better than those algorithms, and can effectively improve the performance of a cloud data center.
Khalid Alhamazani et al. [25]	1	Tools	To introduce application on cloud that it requires extensive monitoring and benchmarking mechanisms to ensure run-time QoS.	Metric	Developed CLAMBS (Cross-Layer Multi-Cloud Application Monitoring and Benchmarking as a Service.)		yes	CLAMBS is flexible, scalable and resource efficient and can be used to monitor and benchmark several applications and cloud resources distributed across multiple clouds.
Shuo Liu et al. [26]	1	Scheduling; task scheduling	To propose the algorithm for two different TUFs that represent profit and penalty.	Method	Non-preemptive scheduling algorithm.		yes	”The results show that the proposed algorithm outperform the EDF, the Generic Benefit, Scheduling, and the Profit Penalty aware scheduling.
Karthik Kumar et al. [27]	1	Scheduling; task scheduling	How to apply the early deadline first until zero laxity algorithm to the cloud service scheduling.	Method	On-line scheduling algorithm.		yes	EDZL scheduling alogorithm is suitable to be employed on the cloud platform.
R. Santhosh, T. Ravichandran [28]	1	Scheduling; task scheduling	To present a preemptive online scheduling algorithm of real time services with task migration to minimize the response time and to improve the efficiency.	Method	Migrate the tasks when it misses its deadline.		yes	The simulation result shows that the proposed algorithm can significantly outperform the EDF and Non Preemptive scheduling algorithm.
Seyedmehdi Hosseinimotlagh et al. [29]	1	Scheduling; task scheduling	To propose a cooperative two-tier task scheduling approach to benefit both cloud providers and their customers.	Method	Proposed two levels of scheduling techniques; local and global levels		yes	The simulation results show that the proposed task scheduling approach not only reduces the total energy consumption of a cloud by 41%, but also has profound impacts on turnaround times of real-time tasks by 85%.
Sonika P Reddy, Chandan H K S [30]	1	Scheduling; task-scheduling	To present a system that handles real-time and non real-time tasks in an energy efficient method without compromising much on neither reliability nor performance.	Method	EDF and EDL for real-time tasks and FCFS for non real-time tasks		yes	The simulation results show significant energy savings compared to the existing Stand-by Sparing for Periodic Tasks (SSPT) for a few execution scenarios.
Yuhuan Du, Gustavo de Veciana [31]	1	Scheduling and resource-allocation	Resource allocation for a computing system with multiple resources supporting heterogeneous soft real-time applications subject to Quality of Service (QoS) constraints on failures to meet processing deadlines.	Method	To develop a general outer bound on the feasible QoS region for nonclairvoyant resource allocation policies, and an inner bound for a natural class of policies based on dynamically prioritizing applications' tasks by favoring those with the largest (QoS) deficits.		yes	Analysis and simulations show substantial resource savings for such policies over reservation-based designs.
Worachat Chawarut, Lilakiatsakun Woraphon [32]	2	Scheduling	To propose a CPU re-allocation algorithm to improve efficiency of energy management and adaptation scheme on real-time service.	Method	The algorithm works as three characteristics; Longest Completion Time, Highest Utilization and Lowest Utilization.		yes	The result shows the reduction of energy consumption and execution time.

Author	RQ#	Focus area	Problem addresses	Approach	Scope of approach	Limitations/Constraints	Validated	Results of validation
Tsai et al. [33]	1	Scheduling and resource-allocation	Adresses and discusses the entire problem with using real-time oriented cloud computing. Furthermore discusses the problems which can occur both on client-side and server-side.	Model	Using a model which suggests a paritioning scheme of the databases to achieve full isolation and therefore avoiding race-conditions.	Very theoretical, does not validate results and only considers partitioning of databases .	No	Model of paritioning scheme
Ozono et al. [34]	2	PaaS	Implementation real-time framework in PaaS application performance at runtime is proposed and it predicts the load-balancing decisions and also whether and how often to redistributeworkload or add/remove Cloud resources when the numberof users changes in shooter online gaming scenario .	Method	Re-implementation of WFE for RTCE. The authors chose to implement WFE inside the google app engine. Main problem is to implement real-time services inside the GAE. The maximum supported numberof users per server in shooter online gaming scenario	Limited to googles environment.	Yes	Real-time measurements which compares old WFE to the newly implemented WFE-S environment with promising results.
Lin et al. [35]	1	Tools	Suggest framework for real-time carpool services.	Method	Suggest using 2 frameworks: a mobile client (MC) and cloud global carpool services (CGCS) for achieving fast send and reception of carpool services. Includes control systems and especially control of and distribution of data from one client to another client	Does not account for probabilistic systems, but instead uses real-time as a speed performance parameter.	Yes	Improvement of algorithm speed
D. von Söhsten and S. Murilo [36]	2	PaaS	Uses cloud technologies for optimizing the face recognition time.	Method	Presents 4 different models which can be used for face recognition. By using windows Azure and Emgu CV for image processing, the work tries to create a way for faster face recognition.	No real-time contraints defined	Yes	Successful results, revealing that the implemented architecture outperforms other solutions. Presents valid comparsion measurements.
C. McGregor [37]	3	Predictable execution-time	This paper tries to find out how it is possible to obtain real-time data from patients in critical care units.	Model	Tests 3rd party software for health care applications. This paper uses the Artemis cloudsuite, which is a platform that enables the multidimensional real-time analysis of physiological data. The artemis cloud service can be used with the inosphere streams runtime which enables real-time performance. Furthermore, a model for how the data extraction should be done clinically is presented.	No proof of concept which validates the theory of real-time performance, also has dependency upon 3rd party software in which real-time constraints are not explained	No	
Wang et al. [38]	2	SaaS	How to create real-time networks in vehicular systems.	Architecture	Presents a theory of architecture. Creating a real-time behaviour in the SaaS layer using a 3-tier V-cloud architecture which includes hardware/actuator units, communication media and server media. Furthermore discussion about sub tier issues in the V-cloud architecture such as body area sensor networks.	No proof of concept, also complex	No	

Author	RQ#	Focus area	Problem addresses	Approach	Scope of approach	Limitations/Constraints	Validated	Results of validation
Xi et al. [39]	2	IaaS	Addresses the problems where the virtual machines within the cloud cannot achieve real-time compatibility in bare-metal hypervisors.	Implementation	Implementation of a resource management system - RT-openstack (an extension of openstack) which can be used for co-hosting real-time and regular virtual machines.	Only comparing one hypervisor	Yes	Successful in revealing that an RT-open stack resouce management system successfully increases the real-time capabilities of a VM cloud system by a lot.
Krishnappa et al. [40]	1	Tools	The problem inside this paper addresses how to achieve real-time behaviour of weather forecasting applications using an IaaS model.	Method	Since the weather forcasting applications often transfers large amounts of data, different forecasting algorithms are used and compared. Furthermore, the validity of real-time within different weather forecasting algorithms is investigated.		Yes	Conclusion that the cloud algorithm ExoGENI performs the best. It is also concluded that all weather forecasts are feasible for real-time behaviour.
S. Deniziak and S. Bak [41]	2	IaaS	How cloud can be used in real-time and how it can be used to satisfy all user needs	Architecute	Finding a distrubted architecture which can satisfy all user requirements and furthermore develop an iterative algorithm which can be used for real-time cloud constraints.	No proof of concept by implementation	Yes	Results prooven by theoretical scheduling, no results showing the response time of tasks.
Kumar et al. [42]	2	IaaS	Fault tolerance in real-time systems	Model	Apply model which can be used for providing high fault tolerance using MIPS virutal machines	Usage of MIPS, the simplest processor architecture	Yes	Implementation of HAFTRC which was proven a reliable option for fault tolerance in applications
Cordeschi et al. [43]	2	IaaS	How to achieve hard real-time using cloud	Model	Mathematical model of a Virtual machine approach		Yes	Results confirm that aquiring real-time requires a dynamic tradeoff
Tseng-Yi Chen et al. [44]	1	Scheduling; task scheduling	How to apply the early deadline first until zero laxity algorithm to the cloud service scheduling	Method	On-line scheduling algorithm		yes	EDZL scheduling alogorithm is suitable to be employed on the cloud platform
Mohammed et al. [45]	2	IaaS	How to use fault tolerence in the cloud	Architecture	Developing a new, optimized infrastructure for IaaS cloud platforms by optimizing the successrate of virtual nodes and virutal machines	No real-life scenario	Yes	Successful as faults are repaired before a deadline miss occur
Boniface et al. [46]	2	PaaS	This paper addresses the problem with cloud based PaaS platforms in order to achieve real-time in interactive multimedia.	Architecture	To be able to create interactive multimedia, this paper suggest using an architecture which includes real-time key features such as: QoS specification, Event prediction, Dynamic SLA navigation, etc.	No real results are are presented in this work, just a model without validation.	No	Theory results, discussion why the solution would provide real-time QoS
Piyare et al. [47]	2	IaaS	This paper tries to find an architecture in which it is possible to easily implement wireless sensor networks.	Architecture	The approach is to suggest an architecture which uses the cloud based sensor data platform open.sen.	No Real-time constraints	Yes	Shows comparative measurements of battery lifetime depending on which event is triggered.

Table V: Mapping Study Results

2) *Problems hindering the predictable execution of tasks in cloud platform.*
Cloud platforms offer a lot of challenges while integrating it with real-time computing environment. Practically, all real-time safety critical applications demand high performance computing where performance sustainability, resource guarantees and timely guarantee for results are highly essential, and cloud environment is simply not designed to handle all of these, therefore the usage of cloud for hosting hard real-time applications is still beyond realities. These issues become more troublesome in multi-core environment, where it is very complex to handle vrtualizations and multicore timing.
In addition to these issues, cloud has a sub-optimal physical topology intended to accommodate a large number of applications, which give rise to performance penalties, which cannot be accommodated in real-time computing. It is also complex to account for the delays in provisioning of resources and virtual machines to ensure predictability. Though some works have proceeded in this direction, they need to mature a lot to use it for practical applications.
3) *Lack of efficient methods for identifying QoS based degradation in real-time cloud services.*
Yet another important concern is maintaining the desired QoS for real-time cloud services. There are lot of QoS issues that needs to be handled in a real-time scenario. For e.g, ensuring fairness when a service goes down, especially when you use it as a pay as you go manner, is very difficult and is a serious problem for all soft real-time applications in cloud. Moreover achieving scalability of real-time online interactions in cloud is again tricky. There are also many other issues like effectively managing power in the systems in data-centers while they are used for real-time applications, improving performance of file access inorder to simultaneously retrieve a large amount of data for achieving real-time cloud services etc., which needs to be investigated beyond the scopes of present works, before which the potential of cloud -computing cannot be fully utilized to host real-time applications.

IV. Conclusion

In this paper, we have conducted a systematic mapping study of real-time cloud services to effectively structure the research results and establish a detailed state-of the-art research. Our mapping study could identify some potential research gaps in the research works published within this area. The major issue which we found is the considerable lack of efficient works in this area that can serve as a base for future researches. Almost all of the existing works are fragmented, with no support for an integrated architecture for achieving predictability of cloud services. We earnestly hope that the future works in this area will definitely proceed in this directions and address the existing challenges.

References

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica *et al.*, "Above the clouds: A berkeley view of cloud computing." Technical Report UCB/EECS-2009-28, EECS Department, University of California, Berkeley, 2009.

[2] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," in *12th international conference on evaluation and assessment in software engineering*, vol. 17, no. 1. sn, 2008.

[3] A. Abdelmaboud, D. N. Jawawi, I. Ghani, A. Elsafi, and B. Kitchenham, "Quality of service approaches in cloud computing: A systematic mapping study," in *The Journal of Systems and Software 101*, 2015, p. 159–179.

[4] H. Esen, M. Adachi, D. Bernardini, A. Bemporad, D. Rost, and J. Knodel, "Control as a service (caas): cloud-based software architecture for automotive control applications," in *Proceedings of the Second International Workshop on the Swarm at the Edge of the Cloud*. ACM, 2015, pp. 13–18.

[5] M. Aazam and E.-N. Huh, "Qos degradation based reimbursement for real-time cloud communication," in *Proceedings of the 1st Workshop on All-Web Real-Time Systems*. ACM, 2015, p. 6.

[6] D. Meiländer and S. Gorlatch, "Modelling the scalability of real-time online interactive applications on clouds," in *Proceedings of the Third International Workshop on Adaptive Resource Management and Scheduling for Cloud Computing*. ACM, 2016, pp. 14–20.

[7] K. An, S. Pradhan, F. Caglar, and A. Gokhale, "A publish/subscribe middleware for dependable and real-time resource monitoring in the cloud," in *Proceedings of the Workshop on Secure and Dependable Middleware for Cloud Monitoring and Management*. ACM, 2012, p. 3.

[8] K. H. Kim, A. Beloglazov, and R. Buyya, "Power-aware provisioning of cloud resources for real-time services," in *Proceedings of the 7th International Workshop on Middleware for Grids, Clouds and e-Science*. ACM, 2009, p. 1.

[9] L. Lundberg and S. Shirinbab, "Real-time scheduling in cloud-based virtualized software systems," in *Proceedings of the Second Nordic Symposium on Cloud Computing & Internet Technologies*. ACM, 2013, pp. 54–58.

[10] C.-T. Huynh, T.-D. Nguyen, H.-Q. Nguyen, and E.-N. Huh, "Cost efficient real-time applications scheduling in mobile cloud computing," in *Proceedings of the Fifth Symposium on Information and Communication Technology*. ACM, 2014, pp. 248–255.

[11] M. D. Ös and G. Bressan, "A community cloud for a real-time financial application-requirements, architecture and mechanisms," in *International Conference on Algorithms and Architectures for Parallel Processing*. Springer, 2014, pp. 364–377.

[12] L. Belli, S. Cirani, G. Ferrari, L. Melegari, and M. Picone, "A graph-based cloud architecture for big stream real-time applications in the internet of things," in *European Conference on Service-Oriented and Cloud Computing*. Springer, 2014, pp. 91–105.

[13] J. Zhang, Q. Li, and W. Zhou, "Hdcache: A distributed cache system for real-time cloud services," *Journal of Grid Computing*, pp. 1–22, 2016.

[14] M. Xu, L. T. X. Phan, O. Sokolsky, S. Xi, C. Lu, C. Gill, and I. Lee, "Cache-aware compositional analysis of real-time multicore virtualization platforms," *Real-Time Systems*, vol. 51, no. 6, pp. 675–723, 2015.

[15] F. Teng, F. Magoulès, L. Yu, and T. Li, "A novel real-time scheduling algorithm and performance analysis of a mapreduce-based cloud," *The Journal of Supercomputing*, vol. 69, no. 2, pp. 739–765, 2014.

[16] S. Hosseinimotlagh, F. Khunjush, and R. Samadzadeh, "Seats: smart energy-aware task scheduling in real-time cloud computing," *The Journal of Supercomputing*, vol. 71, no. 1, pp. 45–66, 2015.

[17] H. Chen and W. Guo, "Real-time task scheduling algorithm for cloud computing based on particle swarm optimization," in *International Conference on Cloud Computing and Big Data in Asia*. Springer, 2015, pp. 141–152.

[18] Y. Zhang, L. Chen, H. Shen, and X. Cheng, "An energy-efficient task scheduling heuristic algorithm without virtual machine migration in real-time cloud environments," in *International Conference on Network and System Security*. Springer, 2016, pp. 80–97.

[19] S. Wu, L. Zhou, D. Fu, H. Jin, and X. Shi, "A real-time scheduling framework based on multi-core dynamic partitioning in virtualized

environment," in *IFIP International Conference on Network and Parallel Computing*. Springer, 2014, pp. 195–207.

[20] H. Mora Mora, D. Gil, J. F. Colom López, and M. T. Signes Pont, "Flexible framework for real-time embedded systems based on mobile cloud computing paradigm," *Mobile Information Systems*, vol. 2015, 2015.

[21] K. H. Kim, A. Beloglazov, and R. Buyya, "Power-aware provisioning of virtual machines for real-time cloud services," *Concurrency and Computation: Practice and Experience*, vol. 23, no. 13, pp. 1491–1505, 2011.

[22] J. Hoffert, D. C. Schmidt, and A. Gokhale, "Adapting distributed real-time and embedded pub/sub middleware for cloud computing environments," in *Proceedings of the ACM/IFIP/USENIX 11th International Conference on Middleware*. Springer-Verlag, 2010, pp. 21–41.

[23] M. García-Valls, T. Cucinotta, and C. Lu, "Challenges in real-time virtualization and predictable cloud computing," *Journal of Systems Architecture*, vol. 60, no. 9, pp. 726–740, 2014.

[24] H. Chen, X. Zhu, H. Guo, J. Zhu, X. Qin, and J. Wu, "Towards energy-efficient scheduling for real-time tasks under uncertain cloud computing environment," *Journal of Systems and Software*, vol. 99, pp. 20–35, 2015.

[25] K. Alhamazani, R. Ranjan, P. P. Jayaraman, K. Mitra, F. Rabhi, D. Georgakopoulos, and L. Wang, "Cross-layer multi-cloud real-time application qos monitoring and benchmarking as-a-service framework," 2015.

[26] S. Liu, G. Quan, and S. Ren, "On-line scheduling of real-time services for cloud computing," in *2010 6th World Congress on Services*. IEEE, 2010, pp. 459–464.

[27] K. Kumar, J. Feng, Y. Nimmagadda, and Y.-H. Lu, "Resource allocation for real-time tasks using cloud computing," in *Computer Communications and Networks (ICCCN), 2011 Proceedings of 20th International Conference on*. IEEE, 2011, pp. 1–7.

[28] R. Santhosh and T. Ravichandran, "Pre-emptive scheduling of on-line real time services with task migration for cloud computing," in *Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on*. IEEE, 2013, pp. 271–276.

[29] S. Hosseinimotlagh, F. Khunjush, and S. Hosseinimotlagh, "A cooperative two-tier energy-aware scheduling for real-time tasks in computing clouds," in *2014 22nd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing*. IEEE, 2014, pp. 178–182.

[30] S. P. Reddy and H. Chandan, "Energy aware scheduling of real-time and non real-time tasks on cloud processors (green cloud computing)," in *Information Communication and Embedded Systems (ICICES), 2014 International Conference on*. IEEE, 2014, pp. 1–5.

[31] Y. Du and G. de Veciana, "Scheduling for cloud-based computing systems to support soft real-time applications," *arXiv preprint arXiv:1601.06333*, 2016.

[32] W. Chawarut and L. Woraphon, "Energy-aware and real-time service management in cloud computing," in *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2013 10th International Conference on*. IEEE, 2013, pp. 1–5.

[33] W. T. Tsai, Q. Shao, X. Sun, and J. Elston, "Real-time service-oriented cloud computing," in *2010 6th World Congress on Services*, July 2010, pp. 473–478.

[34] T. Ozono, R. M. E. Swezey, S. Shiramatsu, T. Shintani, R. Inoue, Y. Kato, and T. Goda, "A real-time collaborative web page editing system wfe-s based on cloud computing environment," in *Advanced Applied Informatics (IIAIAAI), 2012 IIAI International Conference on*, Sept 2012, pp. 224–229.

[35] C.-H. Lin, M.-K. Jiau, and S.-C. Huang, "A cloud computing framework for real-time carpooling services," in *Information Science and Service Science and Data Mining (ISSDM), 2012 6th International Conference on New Trends in*, Oct 2012, pp. 266–271.

[36] D. von Söhsten and S. Murilo, "Multiple face recognition in real-time using cloud computing, emgu cv and windows azure," in *2013 13th International Conference on Intellient Systems Design and Applications*, Dec 2013, pp. 137–140.

[37] C. McGregor, "A cloud computing framework for real-time rural and remote service of critical care," in *Computer-Based Medical Systems (CBMS), 2011 24th International Symposium on*, June 2011, pp. 1–6.

[38] J. Wang, J. Cho, S. Lee, and T. Ma, "Real time services for future cloud computing enabled vehicle networks," in *Wireless Communications and Signal Processing (WCSP), 2011 International Conference on*, Nov 2011, pp. 1–5.

[39] S. Xi, C. Li, C. Lu, C. D. Gill, M. Xu, L. T. X. Phan, I. Lee, and O. Sokolsky, "Rt-open stack: Cpu resource management for real-time cloud computing," in *2015 IEEE 8th International Conference on Cloud Computing*, June 2015, pp. 179–186.

[40] D. K. Krishnappa, E. Lyons, D. Irwin, and M. Zink, "Network capabilities of cloud services for a real time scientific application," in *Local Computer Networks (LCN), 2012 IEEE 37th Conference on*, Oct 2012, pp. 487–495.

[41] S. Deniziak and S. Bak, "Synthesis of real time distributed applications for cloud computing," in *Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on*. IEEE, 2014, pp. 743–752.

[42] P. Kumar, G. Raj, and A. K. Rai, "A novel high adaptive fault tolerance model in real time cloud computing," in *Confluence The Next Generation Information Technology Summit (Confluence), 2014 5th International Conference -*, Sept 2014, pp. 138–143.

[43] N. Cordeschi, D. Amendola, F. D. Rango, and E. Baccarelli, "Networking-computing resource allocation for hard real-time green cloud applications," in *2014 IFIP Wireless Days (WD)*, Nov 2014, pp. 1–4.

[44] T.-Y. Chen, H.-W. Wei, J.-S. Leu, and W.-K. Shih, "Edzl scheduling for large-scale cyber service on real-time cloud," in *2011 IEEE International Conference on Service-Oriented Computing and Applications (SOCA)*. IEEE, 2011, pp. 1–3.

[45] B. Mohammed, M. Kiran, I. U. Awan, and K. M. Maiyama, "Optimising fault tolerance in real-time cloud computing iaas environment," in *2016 IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud)*, Aug 2016, pp. 363–370.

[46] M. Boniface, B. Nasser, J. Papay, S. C. Phillips, A. Servin, X. Yang, Z. Zlatev, S. V. Gogouvitis, G. Katsaros, K. Konstanteli, G. Kousiouris, A. Menychtas, and D. Kyriazis, "Internet and web applications and services," in *Internet and Web Applications and Services (ICIW), 2010 Fifth International Conference on*, May 2010, pp. 155–160.

[47] R. Piyare, S. Park, S. Y. Maeng, S. H. Park, S. C. Oh, S. G. Choi, H. S. Choi, and S. R. Lee, "Integrating wireless sensor network into cloud services for real-time data collection," in *2013 International Conference on ICT Convergence (ICTC)*, Oct 2013, pp. 752–756.