Multi-Criteria Virtual Machine Placement in Cloud Computing Environments: A Literature Review

Wissal Attaoui
NEST Research Group, ENSEM
IN & Mobile Charging Department
INWI Corporate, Casablanca 20190, Morocco
Email: w.attaoui@ensem.ac.ma

Essaid Sabir
TÉLUQ, Department of Science and Technology, University of Quebec
Montreal, H2S 3L4, Canada
Email: essaid.sabir@teluq.ca

Abstract—Cloud computing is a set of innovative and powerful technologies that have completely rethought networks' design and operation. It provides an excellent resiliency level since virtual machines (VMs) run workloads elastically on physical hosts. The placement of VMs in cloud systems is a significant issue that has been thoroughly investigated, although not yet entirely resolved. This paper presents a systematic literature review of VM placement in cloud environment using SPAR 4 SLR protocol. We use VOSViewer tool for bibliometric performance and intellectual structure (i.e., thematic performance). As a result, nine clusters were identified based on co-occurrence keyword performance. These clusters shed light on the various algorithms and techniques employed to address VM placement challenges. These techniques are employed to achieve diverse optimization objectives, including maximizing performance, minimizing energy consumption, maximizing resource utilization, minimizing cost, enhancing security, and ensuring quality of service (QoS).

Index Terms—Virtual Machine Placement; Cloud computing; SPAR-4-SLR.

I. INTRODUCTION

Nowadays, with the tremendous growth of mobile service demands, operators have to provide low latencies and high throughput, hence the need to rethink traditional physical architectures. Therefore, 5G tend to be virtualized, eventually, the 5G network core, i.e. the set of transmission and switching media in which the most crucial part of the traffic is processed, will no longer be carried by physical equipment as in 4G architectures, but it will be supported by the software. This virtualization is based on software-defined networking (SDN) and network function virtualization (NFV) technologies. It offers many advantages and helps providing personalized connectivity services through network slicing technology. Thus, the operator will be able to agilely activate the functions required for each service, adapting the network sizing and topology according to customer needs and cloud properties: low or high throughput, low latency, high reliability, more or less distributed architecture, etc.

The 5G deployment will be gradual (i.e., initially 4G and 5G will coexist seamlessly as was the case for 3G and 4G). In addition, pooling physical infrastructure through virtualization techniques opens the way to create a universal 5G network core that is totally agnostic to the type of access (wireless, wireline, etc.). Therefore, this will ensure homogeneous management of operator’s network. In this way, network slicing will enable mobile network operators to manage different virtual networks on the same physical network infrastructure.

In the digital age, the proliferation of cloud computing has revolutionized how businesses and individuals consume and deliver IT services. The virtual machine (VM) placement concept stands at the heart of this paradigm shift, where resource optimization, performance enhancement, and cost-effectiveness intersect within cloud infrastructures. As cloud environments continue to evolve in terms of scale and complexity, the importance of effective VM placement remains paramount, driving the need for in-depth research and thinking in this area. The rapid expansion of cloud computing ecosystems, fuelled by exponential data growth and the demand for scalable, agile IT solutions, underlines the crucial importance of optimizing resource utilization and performance. Cloud computing enables users to consume on-demand computing resources, such as storage, servers, applications, and networks, as instances (VMs) instead of building physical infrastructure. These resources may be swiftly provided and managed with minimal effort by cloud providers. Cloud computing provides numerous exciting advantages for companies and ultimate consumers like on-demand provisioning of virtual resources, self-service ability, high elasticity, flexibility, scalability, broad network access, and resource pooling. Virtualization technologies allow users to box their required computing resources into VMs. A virtual machine placement algorithm is used to define the locations of these VMs at suitable computing centers. Many issues should be considered in VM placement (e.g., power consumption, traffic, resource wastage, security, QoS, cost, etc.). For each matter, various solutions are proposed through different surveys and research papers in which each one addresses the VM placement problem in a specific manner by suggesting various methods based on different types of algorithms such as heuristic, meta-heuristic, approximate, deterministic, and AI/learning algorithms. As already mentioned in our previous literature review [4] where we have used a classic method to classify papers based on objective functions in optimization problems.

In contrast to conventional methods of literature review, where extensive reading, selection, and subsequent classification are the norm, our approach signifies a departure towards innovation. Inspired by the groundbreaking work of Paul et
al. in 2021 [29], we embrace the SPAR 4 SLR methodology as a transformative tool for systematic literature review. This methodological shift allows for a more structured and efficient analysis, streamlining the identification and synthesis of relevant literature. By employing SPAR 4 SLR, we transcend the limitations of traditional approaches, ensuring a more comprehensive and nuanced understanding of the research landscape surrounding VM placement in cloud computing. Our adoption of this novel methodology not only enhances the rigor of our review but also exemplifies our commitment to staying at the forefront of methodological advancements in academic inquiry.

In this literature review, we present a pioneering analysis of virtual machine (VM) placement within cloud computing environments. The outstanding feature of our study is the innovative use of the SPAR 4 SLR protocol and VOSViewer to dissect performance trends in this crucial area. To the best of our knowledge, this is the first comprehensive examination employing these methodologies. By leveraging the SPAR 4 SLR protocol, we ensure a systematic and rigorous approach to collecting and synthesizing relevant literature. Furthermore, the integration of VOSViewer empowers us to visually map and interpret the intricate connections between various studies, shedding light on the evolving landscape of VM placement in cloud computing.

The remainder of the paper is structured as follows: Section II presents a brief overview of VMP in cloud environment. Next, section III provides the methods and procedures employed in our systematic review of VMP. Subsequently, section IV provides the results of bibliometric and intellectual analysis of the VMP review process. Finally, a synthesis of the findings is presented along with concluding remarks.

II. THEORETICAL BACKGROUND

The virtual machine placement (VMP) concerns the mapping between VMs and Physical Machines (PMs) to optimize the use of available resources [23]. VMP is the process of selecting a suitable host where VMs can be implemented. VM migration enables different VMP strategies to be explored based on Service Level Agreement (SLA) constraints in order to address various workloads [9]. There are several restrictions that can hinder the efficient placement of VMs, including high cost, power consumption, wasted resources, degraded performance, and security vulnerabilities (see Figure 1).

A. High Cost and Energy consumption

One of the foremost challenges in VM placement within cloud computing environments revolves around the dual concerns of energy consumption and cost optimization. As the scale and complexity of cloud infrastructures continue to expand, the energy footprint associated with powering and cooling data centers has become a significant operational expense and environmental concern. Suboptimal VM placement strategies can exacerbate this issue by leading to inefficient resource utilization, resulting in underutilized servers and increased energy consumption [3]. Energy consumption emerges as a multifaceted challenge with various dimensions affecting operational efficiency and sustainability. Firstly, the idle power consumption of physical machines (PMs) hosting underutilized VMs contributes significantly to energy wastage. Studies such as [33] have highlighted the impact of VM consolidation techniques in mitigating this issue by dynamically reallocating VMs to reduce the number of active PMs. Additionally, the phenomenon of “stranded power” arises when VM placement strategies lead to uneven distribution of workload across PMs, resulting in suboptimal utilization and increased energy consumption. This issue underscores the importance of load-balancing mechanisms in achieving energy efficiency within cloud infrastructures [5]. Additionally, the unnecessary VM migrations lead to additional management costs [16] and high energy consumption [22] due to the re-engineering, deletion, and creation of VMs.

Large-scale cloud systems entail a high cost to cloud partners regarding energy dissipation, whereby data centers remain the main contributor to global CO2 emissions [25]. In addition, the wasteful use of computing resources can lead to high consumption of energy.

B. Resource Wastage

Data centers in the cloud are influenced by spatiotemporal alterations of customer requirements and limited existing resources. The arisen question is how data centers can meet the high number of customer demands while respecting the constraints of computing resources and link capacities [27] [21]. In addition, the uneven utilization of physical hosts leads to computing resource wastage, which has a significant impact on VM placement [1] [35] [17].

C. Performance degradation

The performance degradation generated by intensive virtual consolidation is another problem to be addressed; Therefore, a balance has to be achieved between the utilization of the acquired resources and the eventual performance degradation [14]. In this vein, a body of literature [10], [37] focuses on the trade-off between ensuring high performance and reducing energy consumption while maintaining low SLA violation.
D. Network Scalability

The VMP can be affected by traffic congestion, connection disruption, data transfer time, and network links [15] [30] [36]. Moreover, numerous applications require high network resources, such as intense bandwidth requirements [13]. Therefore, many concerns should be taken into account for the scalability of the underlying network architecture.

E. Security

Security has become a significant problem for VM placement in cloud infrastructure. Placing a new VM with known vulnerabilities in a physical host may cause security risks due to the mapping relationship between VMs and PMs, which generates a rapid spread of vulnerabilities across the entire Infrastructure as a Service (IaaS) [32]. Different VM placement approaches are proposed to reduce the security risks for both VMs and PMs [24]. A security assessment can depend on four aspects: VM vulnerabilities, hypervisor vulnerabilities, co-residence, and networking [20]. Security problems related to co-location and resource utilization are tackled in [8] and [2], where results prove that combining security and resource utilization drives powerful VM placement that mitigates co-location attacks. On the other hand, instead of including resource optimization, the placement strategy based on the tradeoff between co-location and startup time can enormously reduce the co-location attacks within a short time [28].

III. METHODOLOGY

This study applied a Systematic Literature Review (SLR) to observe the existing relevant literature on virtual machine placement in order to analyze the related body of literature. To achieve this goal, we use the proposed protocol by Paul et al. [29], namely SPAR-4 SLR, aiming to conduct the study in three stages (i.e., assembling, arranging, and assessing) and six sub-stages (i.e., identification, acquisition, organization, purification, evaluation, and reporting).

A. Assembling

The first stage of the SPAR-4-SLR protocol is assembling, which involves the identification (i.e., review domain, research questions, source type, and source quality) and acquisition (i.e., search mechanism and material acquisition, search period, search keywords) of articles for review. It includes the identification and acquisition of literature that has not been synthesized [29].

In the sub-stage of identification, the domain, research questions, source type, and source quality are determined. The domain and the research questions, which guided this review are presented in detail in Figure 2. The material of the study consisted of peer-reviewed research articles either published or in a state of “in press” in international peer-reviewed journals. Scopus journal quality list was used to evaluate the source quality and identification of journals. Furthermore, the quality of articles was assessed with journal impact factors and article citations. The impact factor of the publishing journal (Journal Citation Reports, 2024) was considered as a proxy of an article’s quality, as in other systematic literature reviews. Additionally, the publishing journals were checked to screen out any predatory journals.

In order to conduct the search, we chose (1) “Virtual Machine Placement”, and (2) “Cloud Computing”, as exact keywords for the search because of the centrality of these concepts to our review. We checked the “title of the article” option in “advanced search” in order to produce finer-grained search results. The keywords were consistent in both American and British English, and thus, alternative spellings were not considered.

In the second sub-stage, acquisition, the research articles published or in a state of “in press” between 2009 and 2024 were gathered in April 2024 in Scopus database using the search words “Virtual Machine Placement” for titles, abstracts, and keywords. Research terms generated 1957 studies in Scopus. In addition, the reference lists of the articles found through systematic searches were scrutinized to capture all the relevant material in the background of this review.

B. Arranging

The second stage of the SPAR-4-SLR protocol is referred to as arranging, which involves the organization of the literature by employing organizing codes, and purification of the material [29]. In the organization step, articles from the English language were filtered, the document type selected was all research and review articles, and the source type was selected as journals only. After the initial stage of limiting the search to English language and peer-reviewed articles, 159 papers were excluded and duplicates were removed (114). The total number of articles returned after refinement was 1684, which was downloaded in bib text format. In terms of purification, 273 articles were not directly related to the context of customer engagement, and some of them were duplicated. So after purification and data cleaning, a total of 1001 documents were included for review purposes. The full texts were reviewed to assess the relevance of the research focus to our study.

C. Assessing

The last stage of assessing includes the evaluation and reporting of the reviewed literature [29]. In this literature review, the material was analyzed with content analysis and thematic analysis. Findings were reported through the presentation of figures, tables, and textual explanations. Limitations of the study include reliance on a single database, namely Scopus, and a focus limited to bibliometric analysis and content cluster analysis using keyword co-occurrence networks. Notably, no financial support was required as the review was based on secondary data sources.

IV. RESULTS

A. Bibliometric performance

Bibliometric analysis serves as a crucial tool in evaluating the scholarly landscape of a particular research domain. By quantifying and analyzing publication patterns, citation networks, and other bibliographic data, researchers can gain
Figure 2: The SPAR-4-SLR protocol.
valuable insights into the evolution, impact, and trends within a field. In the realm of cloud computing, where the dynamic nature of technology drives rapid innovation, understanding the bibliometric landscape of key topics such as virtual machine (VM) placement is paramount.

In this systematic literature review (SLR), we employ bibliometric analysis to comprehensively examine the scholarly output related to virtual machine placement in cloud computing. By leveraging established protocols such as the Systematic Review Protocol for Software Engineering (SPAR 4 SLR), we aim to provide a rigorous and structured analysis that sheds light on the key contributors, seminal works, and emerging trends within this critical domain.

We seek to address fundamental questions surrounding the publication output, citation impact, collaboration networks, and geographical distribution of research in virtual machine placement. By doing so, we not only provide a snapshot of the current state of the field but also identify avenues for future research and collaboration, thereby fostering continued advancement in cloud computing technologies.

1) Corpus performance: The analysis of corpus performance reveals a dynamic trajectory in research on virtual machine placement within cloud computing, influenced by various factors, including the global COVID-19 pandemic. In the early years, from 2009 to 2012, limited publications may have mirrored the nascent stage of cloud technology adoption, coupled with the need for foundational research in virtualization and resource allocation. However, from 2013 to 2018, a notable surge in publications occurred, driven by advancements in virtualization technologies, growing industry investments, and escalating demand for efficient resource management solutions. The subsequent stabilization of research output from 2019 to 2021 suggests a potential saturation of certain research avenues or a consolidation of existing knowledge, as the field reached a level of maturity and equilibrium. Conversely, the decline observed from 2022 onwards may be attributed in part to the disruptive impact of the COVID-19 pandemic, which catalyzed an accelerated adoption of cloud technologies and virtual machine placement solutions. As businesses worldwide rapidly scaled up their cloud infrastructure to support remote work and digital transformation initiatives, there was heightened urgency in addressing emerging challenges and optimizing resource utilization in distributed environments. Notably, the peak observed in 2022 followed by a decline in 2023 hint at the dynamic interplay of internal and external factors shaping research agendas, including technological advancements, funding dynamics, and possibly the disruptive impacts of global events. However, it’s essential to acknowledge that the data for 2024 is preliminary, as the year is ongoing, and the final publication count may vary as more research results are published and made available. Therefore, any analysis of corpus performance for 2024 should be approached with a degree of caution and may require revisiting and updating as more data becomes available. Figure 3 displays the number of publications per year.

2) Article performance: A co-citation analysis was conducted to identify the top VMP articles that have been cited the most by other VMP articles (local citations), and thus, enables a more accurate evaluation of the citation performance as seen in Figure 4. Table 1 presents the 10 most-cited VMP articles by 1001 articles in the CE corpus retrieved from Scopus for the period between 2009 and 2024.

The landscape of virtual machine placement in cloud computing is enriched by seminal contributions from various researchers, each offering unique insights and solutions to address key challenges within the domain. Kristenpart’s groundbreaking work in 2009 stands out as a cornerstone, with its innovative approaches to data storage security garnering significant attention and citations. Beloglazov’s research in 2010 and 2012 delves into energy-efficient management of virtual machines and provides a comprehensive taxonomy of energy-efficient data centers and cloud computing systems, respectively, shaping discourse on sustainability in cloud infrastructure. Gao’s 2013 paper on energy-efficient task scheduling algorithms and Gu’s 2017 contribution on predictive analytics for virtual machine provisioning advance optimization techniques, enhancing resource utilization and provisioning efficiency. Meanwhile, Chaisiri’s works in 2009 and 2012 offer insights into network-aware virtual machine placement and minimizing network traffic, crucial for optimizing performance in cloud environments. Lastly, Liu’s 2018 survey comprehensively catalogs virtual machine placement techniques, providing a roadmap for navigating the evolving landscape of cloud computing research. Together, these seminal papers constitute foundational pillars in the field, guiding research endeavors and fostering innovation to meet the evolving demands of cloud computing infrastructure.

3) Author performance: For author performance, our focus shifts to identifying authors who have made significant contributions to the field, employing a selection criterion of a minimum of four papers authored by each individual. This rigorous criterion ensures that we capture the insights and expertise of authors with a substantive body of work in virtual machine placement. Through meticulous analysis, we aim to uncover key trends, seminal works, and influential authors who have shaped the discourse and advancement of virtual machine placement within cloud computing.

In this context, Table 2 provides an insightful overview of author performance within the field of virtual machine placement.
placement in cloud computing, highlighting the number of documents authored by each researcher, their corresponding citation counts, and total link strength. Rajkumar Buyya emerges as the most prolific and influential author, with 18 documents and a remarkable citation count of 2656, contributing significantly to the scholarly discourse. Other notable contributors include Jie Wu, Ashutosh Kumar Singh, and Ashok Kumar Turuk, each with a substantial number of documents and citations, indicating their substantial impact on the field. Additionally, authors such as Sourav Kanti Addya, Yu-Chu Tian, and Benjamín Barán exhibit high total link strength, suggesting strong collaboration networks and influential partnerships within the research community. The diverse array of authors and their respective contributions underscores the interdisciplinary nature of virtual machine placement research, with researchers from various backgrounds contributing to the advancement of the field through their expertise and collaborations.

4) Country performance: The data, obtained from VOSViewer, reveals significant variations in research output and impact across different countries. Notably, China emerges as a dominant player in the field, with a staggering 446 documents and 9020 citations, reflecting its substantial investment and contributions to cloud computing research. India also exhibits strong performance, with 506 documents and 4557 citations, indicative of its growing influence in the field.

Among Western countries, the United States and the United Kingdom stand out with 294 and 81 documents respectively, reflecting their longstanding leadership and robust research ecosystems in cloud computing. Canada also demonstrates notable performance, with 80 documents and 1933 citations, underscoring its active engagement in the field. Furthermore, several emerging economies such as Brazil, Iran, and Turkey showcase significant research activity, highlighting the global nature of virtual machine placement research and the diverse range of contributors shaping its development. It’s important to note that while some countries exhibit high document counts, their citation and total link strength metrics vary, indicating differences in the impact and collaboration networks within...
Table II: Twenty-five most-cited VMP authors according to h-index and total citations.

| Author               | Docs | Citations | Total link strength |
|----------------------|------|-----------|---------------------|
| Buyya, Rajkumar      | 18   | 2656      | 9129                |
| Wu, Jie              | 15   | 305       | 3600                |
| Singh, Ashutosh Kumar| 13   | 124       | 3997                |
| Turuk, Ashok Kumar   | 13   | 241       | 5123                |
| Addya, Sourav Kanti  | 12   | 160       | 5194                |
| Sahoo, Bibhudatta    | 12   | 237       | 3130                |
| Tian, Yu-Chu         | 12   | 290       | 7107                |
| Barán, Benjamín      | 11   | 161       | 8198                |
| Zomaya, Albert y.    | 11   | 366       | 3453                |
| Mann, Zoltán Adám    | 10   | 391       | 3453                |
| Tang, Maolin         | 10   | 391       | 8383                |
| Li, Xin              | 9    | 412       | 4448                |
| López-Pires, Fabio   | 9    | 299       | 1554                |
| Toroghi Haghighat, Abolfazl | 9 | 118 | 1267 |
| Wang, Shangguang     | 9    | 148       | 7319                |
| Zheng, Gingham       | 9    | 416       | 4811                |
| Zhou, Ao             | 9    | 376       | 6951                |
| Khatua, Sunirmal     | 8    | 300       | 4135                |
| Lawey, Ahmed Q.      | 8    | 85        | 5768                |
| Masdar, Mohammad     | 8    | 190       | 1069                |
| Rahmani, Amir Masoud | 8    | 439       | 5954                |
| Subhajini, A.C.      | 8    | 232       | 6577                |
| Tata, Samir          | 8    | 10        | 1564                |
| Zekri, Ahmed         | 8    | 54        | 1719                |
| Elmirghani, Jaafar M. H. | 7 | 70 | 4219 |

each country’s research community.

Overall, this analysis provides a comprehensive overview of country performance in virtual machine placement research, offering valuable insights into the global distribution of expertise, collaboration networks, and research impact within the field. Figure 5 depicts the countries performance of studied papers between 2009 and 2024.

B. Thematic performance

Thematic performance analysis complements bibliometric analysis by delving deeper into the content and thematic evolution of scholarly literature within a specific research area. While bibliometric analysis quantifies the impact and dissemination of research, thematic analysis uncovers the underlying concepts, trends, and research clusters that shape the intellectual landscape.

The generated visualization map displays the keywords or terms in the data file based on the clustering techniques available in VOSViewer. The unified VOSviewer clustering technique can be seen as a kind of weighted variant of modularity-based clustering which has a resolution parameter to identify small clusters. Accordingly, the visualization map of term occurrence depicts the frequency of occurrence of certain key terms, hence called occurrence metric. The terms are represented as nodes of varying sizes, proportional to the terms’ recorded frequency. Additionally, the analysis indicates the frequency with which the terms appear in close proximity to one another. The co-occurrence of terms within a text network has a substantial effect on the construction of text clusters, also known as communities of terms.

A keyword co-occurrence analysis was conducted to delineate the knowledge production of VMP research. In this analysis, the keywords that authors list for their VMP articles are extracted and clustered according to their thematic similarities. The clusters of VMP articles revealed through a keyword co-occurrence analysis reflect the present of VMP research. Based on the analysis, a total of nine clusters of VMP research were revealed as depicted in Figure 6.

1) Efficient Resource Allocation and Service Provisioning for Virtual Machine Placement in Cloud Computing: 5G-Enabled Solutions: Cluster 1 encompasses a broad spectrum of topics related to 5G mobile communication, cloud computing services, and network infrastructure. It explores the convergence of communication technologies and computing resources, focusing on efficient application placement, bandwidth management, and quality of service requirements. The cluster delves into various aspects of mobile edge computing, fog computing, and distributed environments, emphasizing dynamic resource allocation, elasticity, and efficiency. Additionally, it addresses emerging paradigms such as edge clouds and federated clouds, highlighting the importance of feedback control, optimization strategies, and reinforcement learning approaches in optimizing resource utilization and service provisioning. The cluster also touches upon topics like network function virtualization (NFV), software-defined networking (SDN), and service function chaining, reflecting advancements in network architecture and virtualization technologies to support diverse applications and services in 5G environments. Furthermore, it explores the intersection of cloud computing and IoT, signal processing, and smart power grids, underscoring the interdisciplinary nature of research in this domain. Overall, Cluster 1 provides a comprehensive exploration of the interplay between communication networks, computing resources, and emerging technologies to enable efficient and reliable service delivery in 5G and cloud computing environments.

2) Virtual Machine Placement Strategies and Optimization in Cloud Computing Environments: Cluster 2 explores a diverse range of topics related to optimizing virtual machine placement and resource allocation within cloud computing infrastructures. It delves into various aspects of algorithms, optimization techniques, and infrastructure considerations crucial for enhancing performance and efficiency in cloud environments. Topics such as allocation algorithms, approximation algorithms, and combinatorial optimization problems underscore the importance of efficient resource utilization and cost minimization in virtual machine placement. Additionally, discussions on communication patterns, network optimization strategies, and bandwidth consumption highlight the need to optimize data transfers and reduce latency in virtual machine deployments. Furthermore, the cluster addresses emerging paradigms such as green networking and smart city applications, emphasizing sustainability and scalability in virtual machine placement strategies. Overall, Cluster 2 provides a comprehensive exploration of virtual machine placement strategies and optimization techniques essential for maximizing performance and resource utilization in cloud computing environments.
3) Optimizing Virtual Machine Placement and Energy Efficiency in Cloud Computing Environments: Cluster 3 delves into a comprehensive array of topics focused on virtual machine (VM) placement strategies and energy-efficient resource allocation within cloud computing infrastructures. It explores various strategies such as bin packing, genetic algorithms, and cuckoo search algorithms to address the challenges associated with optimal VM placement and resource utilization. Additionally, the cluster emphasizes energy-aware algorithms, green cloud computing, and reduction in energy consumption, highlighting the critical importance of energy efficiency in cloud data centers. Topics such as task scheduling, traffic-aware algorithms, and real-time applications underscore the need for responsive and scalable VM placement strategies to meet dynamic workload demands while minimizing energy consumption. Moreover, the cluster discusses the importance of predictive models and monitoring techniques in optimizing VM placement decisions and ensuring efficient resource allocation. Overall, Cluster 3 provides a comprehensive exploration of VM placement strategies and energy-efficient resource allocation techniques crucial for enhancing efficiency, scalability, and sustainability in cloud computing environments.

4) Enhancing Virtual Machine Placement Efficiency through Heuristic Algorithms and Advanced Analytics in Cloud Environments: Cluster 4 delves into the optimization of virtual machine (VM) placement efficiency within cloud computing environments using heuristic algorithms, adaptive techniques, and advanced analytics. It explores topics such as container placement, dynamic provisioning, and resource management problems, emphasizing the importance of efficiently allocating computing resources to meet workload demands while minimizing energy consumption. Additionally, the cluster addresses emerging paradigms such as hybrid clouds, software-defined clouds, and storage as a service, reflecting advancements in cloud infrastructure and service delivery models. Topics such as workload prediction, forecasting, and correlation methods underscore the need for predictive analytics and intelligent systems to anticipate resource requirements and optimize VM placement decisions. Moreover, the cluster discusses the significance of service level agreements and utility computing in ensuring quality of service and cost-effectiveness in cloud environments. Overall, Cluster 4 provides a comprehensive exploration of heuristic algorithms and advanced analytics techniques crucial for enhancing VM placement efficiency and resource utilization in cloud computing environments.

5) Optimizing Virtual Machine Placement and Energy Efficiency through Ant Colony Optimization in Cloud Environments: Cluster 5 focuses on leveraging ant colony optimization (ACO) algorithms and evolutionary techniques to enhance virtual machine (VM) placement efficiency and energy optimization within cloud computing infrastructures. It explores topics such as cloud federation, energy-efficient resource allocation, and heterogeneous servers, highlighting the importance of optimizing resource scheduling and reducing power consumption. Additionally, the cluster delves into multi-objective optimization and revenue maximization strategies, emphasizing the need to balance conflicting objectives and improve overall system performance. Topics such as network security and simulated annealing algorithms underscore the importance of ensuring robustness and reliability in VM placement decisions. Moreover, the cluster discusses the significance of objective functions and optimization solutions in achieving Pareto-optimal outcomes and improving cloud computing performance. Overall, Cluster 5 provides a comprehen-
Figure 6: Keyword Co-occurrence performance using VOSViewer. **Notes:** The size of nodes indicates the frequency of appearance of the keyword in VMP articles, wherein the bigger the node, the more frequent the keyword appears in VMP articles. The link between nodes indicates the similarity of keywords appearing in VMP articles, whereas the closeness of nodes indicates the degree of similarity of keywords appearing in VMP articles, wherein the closer the nodes, the greater the similarity of keywords appearing in VMP articles in an area of VMP research.

6) Strategies for Dynamic VM Consolidation in Cloud Environments: Cluster 6 focuses on dynamic virtual machine (VM) consolidation and related topics. This cluster likely encompasses discussions and research regarding the efficient management of VMs in cloud data centers to optimize resource utilization, energy consumption, and overall system performance. Key themes such as dynamic consolidation, energy efficiency, and quality of service constraints indicate a focus on developing strategies and algorithms for dynamically reallocating VMs to balance workloads, minimize energy consumption, and ensure adherence to service level agreements. The presence of keywords like machine selection, heuristics, and multi-agent systems suggests the exploration of intelligent and adaptive approaches to VM placement and consolidation. Through analyzing this cluster, researchers can gain insights into emerging techniques and best practices for orchestrating efficiency in cloud environments through dynamic VM consolidation and resource allocation strategies.

7) Leveraging Particle Swarm Intelligence for Performance Optimization in Cloud VM Placement: Cluster 7 underscores the pivotal role of particle swarm intelligence in driving performance optimization and resource allocation strategies. This cluster represents a rich tapestry of research and innovation aimed at optimizing performance and resource allocation through the application of advanced optimization techniques. Encompassing topics such as load balancing, dynamic migration, and scheduling tasks, the cluster highlights the paramount importance of leveraging particle swarm optimization (PSO) and swarm intelligence to tackle the complexities of VM placement in dynamic cloud environments. PSO, along with related concepts like self-adaptive algorithms and multi-resource scheduling, emerges as a cornerstone in the pursuit of efficient cloud resource management. With a focus on power conservation, system stability, and performance computing, the cluster underscores the critical role of PSO in driving intelligent VM
provisioning, workflow scheduling, and overall optimization of cloud computing environments. Analysis of this cluster promises valuable insights into cutting-edge approaches for balancing workload, optimizing performance objectives, and enhancing the overall efficiency of cloud computing environments through the application of particle swarm intelligence.

8) Safeguarding VM placement in Cloud Environments: Enhancing Security and Efficiency: Cluster 8 delves into the critical intersection of security and efficiency. This cluster encompasses a breadth of topics ranging from cloud security, cryptography, and denial of service attacks to resource optimization and power efficiency. Notably, the presence of bin packing algorithms and placement strategies underscores the importance of optimizing resource allocation while considering security requirements. The cluster sheds light on the significance of virtual machine placement strategies in mitigating risks associated with security breaches, co-location attacks, and side-channel attacks. Additionally, topics such as multi-tenancy and dynamic allocation highlight the complexities of balancing security concerns with the efficient utilization of cloud resources. Through an exploration of placement schemes, consolidation processes, and risk assessment techniques, the cluster offers valuable insights into approaches for enhancing security-aware VM placement and resource optimization in cloud environments. Analysis of this cluster promises to illuminate the evolving landscape of cloud security and its integration with efficient VM placement strategies to safeguard data privacy and infrastructure integrity.

9) Intelligent algorithms for Resource Management and VM placement: Cluster 9 offers a comprehensive exploration of strategies aimed at enhancing efficiency and intelligence in resource management. This cluster spans a wide array of topics, ranging from computer-supported cooperative work and distributed applications to performance evaluation and power modeling. Notably, the inclusion of keywords such as multi-objective optimization, resource provisioning, and network resource utilization reflects a keen focus on optimizing resource allocation while considering diverse objectives and constraints. Furthermore, the presence of terms like artificial intelligence, learning algorithms, and hybrid genetic algorithms underscores a shift towards intelligent approaches to VM placement and resource management. Through simulations, surveys, and evaluations, this cluster provides valuable insights into the future directions of research, offering potential advancements in VM migration, live migration, and overall resource-aware management systems. Analysis of this cluster promises to illuminate emerging trends and methodologies that leverage both efficiency-driven strategies and artificial intelligence techniques to optimize VM placement and resource management in cloud environments.

V. SYNTHESIS

The synthesized research landscape presents a dynamic ecosystem of studies aimed at optimizing virtual machine (VM) placement and resource management within cloud computing environments. Across the clusters, a rich tapestry of algorithms and techniques emerges, spanning from traditional bin packing algorithms to cutting-edge heuristic approaches like ant colony optimization (ACO) and particle swarm optimization (PSO). These techniques are harnessed to achieve diverse optimization objectives, including maximizing performance, minimizing energy consumption, maximizing resource utilization, minimizing cost, and ensuring quality of service (QoS). Moreover, the research emphasizes the importance of infrastructure considerations, energy efficiency, and security in shaping VM placement strategies.

Moreover, these clusters provide valuable insights into the future directions of research in cloud computing environments. One promising avenue lies in the integration of emerging technologies such as edge computing and Internet of Things (IoT) with cloud infrastructures, enabling more efficient and responsive service delivery. Additionally, there is a growing emphasis on incorporating machine learning and AI-driven approaches into VM placement and resource management, enabling systems to adapt dynamically to changing workloads and optimize resource allocation in real-time. Moreover, as sustainability becomes increasingly crucial, future research is likely to focus on developing novel energy-efficient algorithms and green computing strategies to minimize the environmental impact of cloud data centers. Furthermore, the intersection of security and efficiency will continue to be a key area of exploration, with a growing need for robust placement strategies that can mitigate evolving cybersecurity threats while optimizing resource utilization. By embracing these future insights and leveraging advancements in technology and methodology, researchers can drive innovation and propel the evolution of cloud computing towards more resilient, sustainable, and intelligent architectures.

VI. CONCLUSION

In conclusion, this systematic literature review has provided valuable insights into the landscape of VM placement within cloud computing environments. By utilizing the SPAR 4 SLR protocol and VOSViewer tool, we have identified nine distinct clusters representing the diverse array of algorithms and techniques employed to tackle VM placement challenges. These clusters not only highlight the breadth of research in this domain but also underscore the multifaceted nature of optimization objectives, ranging from performance enhancement to cost reduction, security enhancement, and quality of service assurance. The synthesis of these findings emphasizes the ongoing efforts within the research community to address the complexities of VM placement in cloud systems. Moving forward, this review serves as a foundation for future research endeavors aimed at developing more robust, efficient, and secure VM placement strategies, ultimately contributing to the continued evolution and advancement of cloud computing technologies.

REFERENCES

[1] A. C. Adamuthe, R. M. Pandharpatte, and G. T. Thampi. Multiobjective virtual machine placement in cloud environment. In 2013 International
Jin Han, Wanyu Zang, Songqing Chen, and Meng Yu. Reducing security

Abdullah Alouari, Aqsa Khalid, Muhammad Tahir, and Muhammad Sardaraz. Energy efficient virtual machines placement in cloud data-centers using genetic algorithm and adaptive thresholds. Plos one, 19(1):e0296399, 2024.

Wissal Attaoui and Essaid Sabir. Multi-criteria virtual machine placement in cloud computing environments: A literature review, 2018.

Sadoon Azizi, Maz‘har Zandsalimi, and Dawei Li. An energy-efficient algorithm for virtual machine placement optimization in cloud data centers. Cluster Computing, 23(3):3421–3434, 2020.

Anton Beloglazov, Jemal Abawajy, and Rajkumar Buyya. Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. Future generation computer systems, 28(5):755–768, 2012.

Anton Beloglazov and Rajkumar Buyya. Energy efficient resource management in virtualized cloud data centers. In 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, pages 826–831. IEEE, 2010.

Mouhebeddine Berrima, Aicha Khatijaa Nasr, and Narjes Ben Rajeb. Co-location resistant strategy with full resources optimization. In Proceedings of the 2016 ACM on Cloud Computing Security Workshop, CCSW ’16, page 3–10, New York, NY, USA, 2016. Association for Computing Machinery.

Nicolò Maria Calcavecchia, Ofer Biran, Erez Hadad, and Yosef Moatti. VM placement strategies for cloud scenarios. In Proceedings of the 2012 IEEE Fifth International Conference on Cloud Computing, CLOUD ’12, pages 852–859, Washington, DC, USA, 2012. IEEE Computer Society.

Zhibo Cao and Shoubin Dong. An energy-aware heuristic framework for virtual machine consolidation in cloud computing. The Journal of Supercomputing, 69(1):429–451, 2014.

S. Chaisiri, Bu-Sung Lee, and D. Niyato. Optimal virtual machine placement across multiple cloud providers. pages 103–110, Dec 2009.

Sivadon Chaisiri, Bu-Sung Lee, and Dusit Niyato. Optimization of resource provisioning cost in cloud computing. IEEE transactions on services Computing, 5(2):164–177, 2011.

R. Cohen, L. Lewin-Eytan, J. Seffi Naor, and D. Raz. Almost optimal virtual machine placement for traffic intense data centers. In 2013 Proceedings IEEE INFOCOM, pages 355–359, April 2013.

A. Dalvandi, M. Gurusamy, and K. C. Chua. Time-aware vm-placement and routing with bandwidth guarantees in green cloud data centers. In 2013 IEEE 5th International Conference on Cloud Computing Technology and Science, volume 1, pages 212–217, Dec 2013.

D. S. Dias and L. H. M. K. Costa. Online traffic-aware virtual machine placement in data center networks. In 2012 Global Information Infrastructure and Networking Symposium (GIIS), pages 1–8, Dec 2012.

Xiong Fu and Chen Zhou. Virtual machine selection and placement for dynamic consolidation in cloud computing environment. Frontiers of Computer Science, 9(2):322–330, 2015.

Yongjiang Gao, Haibing Guan, Zhengwei Qi, Yang Hou, and Liang Liu. A multi-objective ant colony system algorithm for virtual machine placement in cloud computing. Journal of Computer and System Sciences, 79(8):1230 – 1242, 2013.

Yongjiang Gao, Haibing Guan, Zhengwei Qi, Yang Hou, and Liang Liu. A multi-objective ant colony system algorithm for virtual machine placement in cloud computing. Journal of Computer and System Sciences, 79(8):1230–1242, 2013.

Lin Gu, Deze Zeng, Song Guo, Ahmed Barnawi, and Yong Xiang. Cost efficient resource management in fog computing supported medical cyber-physical system. IEEE Transactions on Emerging Topics in Computing, 5(1):108–119, 2017.

Jin Han, Wanyu Zang, Songqing Chen, and Meng Yu. Reducing security risks of clouds through virtual machine placement. In Giovanni Livraga and Shu-Chuen Zhu, editors, Data and Applications Security and Privacy XXXI, pages 275–292, Cham, 2017. Springer International Publishing.

N. T. Hieu, M. Di Francesco, and A. Y. Jääski. A virtual machine placement algorithm for balanced resource utilization in cloud data centers. In 2014 IEEE 7th International Conference on Cloud Computing, pages 474–481, June 2014.

Inkwon Hwang and Massoud Pedram. Hierarchical virtual machine consolidation in a cloud computing system. In Proceedings of the 2013 IEEE Sixth International Conference on Cloud Computing, CLOUD ’13, pages 196–203, Washington, DC, USA, 2013. IEEE Computer Society.

Chris Hyser, Brett McKee, Rob Gardner, and Brian J. Watson. Autonomic virtual machine placement in the data center. 2007.

Motlati i Isaac Thulo and J. H. P. Eloff. Towards optimized security-aware (o-sec) vm placement algorithms. In Proceedings of the 3rd International Conference on Information Systems Security and Privacy - Volume 1: ICISSP, pages 411–422. INSTICC, SciTePress, 2017.

B. Kantarci, L. Foschini, A. Corradi, and H. T. Mouftah. Inter-and-intra data center vm-placement for energy-efficient large-scale cloud systems. pages 708–713, Dec 2012.

H. Andrés Lagar-Cavilla, Joseph Andrew Whitney, Adin Scannell, Philip Patchin, Stephen M. Rumble, Eyal de Lara, Michael Brudno, and Mahadev Satyanarayanan. Snowﬂock: rapid virtual machine cloning for cloud computing. In European Conference on Computer Systems, 2009.

K. Li, J. Wu, and A. Blaise. Elasticity-aware virtual machine placement for cloud datacenters. In 2013 IEEE 2nd International Conference on Cloud Networking (CloudNet), pages 99–107, Nov 2013.

V. Natu and T. N. B. Duong. Secure virtual machine placement in infrastructure cloud services. In 2017 IEEE 10th Conference on Service-Oriented Computing and Applications (SOCA), pages 26–33, Nov 2017.

Justin Paul, Weng Marc Lim, Aron O’Cass, Andy Wei Hao, and Stefano Bresciani. Scientific procedures and rationales for systematic literature reviews (spar-4-slr). International Journal of Consumer Studies, 45(4):O1–O16, 2021.

J. T. Fiao and J. Yan. A network-aware virtual machine placement and migration approach in cloud computing. In 2010 Ninth International Conference on Grid and Cloud Computing, pages 87–92, Nov 2010.

Thomas Ristenpart, Eran Tromer, Hovav Shacham, and Stefan Savage. Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds. In Proceedings of the 16th ACM conference on Computer and communications security, pages 199–212, 2009.

Sachin Shetty, Xuexiao Yuchi, and Min Song. Security-Aware Virtual Machine Placement in Cloud Data Center, pages 13–24, Springer International Publishing, Cham, 2016.

Maolin Tang and Shenzhen Pan. A hybrid genetic algorithm for the energy-efficient virtual machine placement problem in data centers. Neural processing letters, 41:211–221, 2015.

Johan Tordsson, Rubén S Montero, Rafael Moreno-Vozmediano, and Ignacio M Llorente. Cloud brokering mechanisms for optimized placement of virtual machines across multiple providers. Future generation computer systems, 28(2):358–367, 2012.

J. Xu and J. A. B. Fortes. Multi-objective virtual machine placement in virtualized data center environments, pages 179–188, Dec 2010.

M. Zamanifar, N. Nasri, and M. H. Nadimir-Shahraki. Data-aware virtual machine placement and rate allocation in cloud environment. In 2012 Second International Conference on Advanced Computing Communication Technologies, pages 357–360, Jan 2012.

Xiaoping Zhang, Qiang Yue, and Zhongtang He. Dynamic Energy-Efficient Virtual Machine Placement Optimization for Virtualized Clouds, pages 439–448. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.