



FAKULTÄT II - INFORMATIK, WIRTSCHAFTS- UND RECHTSWISSENSCHAFTEN

DEPARTMENT FÜR INFORMATIK

## Identifikation von Freiheitsgraden und Wechselwirkungen in Rechenzentren unter Betrachtung elektrischer und thermischer Energie

Von der Fakultät für Informatik, Wirtschafts- und Rechtswissenschaften der Carl von Ossietzky Universität Oldenburg zur Erlangung des Grades und Titels eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

angenommene Dissertation

von Herrn Dipl.-Inform. Stefan Janacek

geboren am 23.09.1982 in Lübeck

*Gutachter*

Prof. Dr.-Ing. Wolfgang Nebel

Prof. Dr. Michael Sonnenschein

Tag der Disputation: 14.02.2017



# Inhaltsverzeichnis

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Einleitung</b>  | <b>1</b>  |
| 1.1      | Motivation . . . . .   | 1         |
| 1.2      | Beitrag dieser Arbeit . . . . .  | 4         |
| 1.3      | Aufbau der Arbeit . . . . .  | 8         |
| <b>2</b> | <b>Grundlagen</b>  | <b>11</b> |
| 2.1      | Das Rechenzentrum & seine Infrastruktur . . . . .                                  | 11        |
| 2.1.1    | Elektrische Leistungsaufnahme des Rechenzentrums . . . . .                         | 12        |
| 2.1.2    | IT-Hardware . . . . .  | 15        |
| 2.1.3    | Stromversorgungsarchitektur . . . . .  | 16        |
| 2.1.4    | Klimatisierung . . . . .   | 19        |
| 2.1.5    | Netzwerkarchitektur . . . . .  | 24        |
| 2.2      | Rechenzentrumsbetrieb . . . . .  | 25        |
| 2.3      | Virtualisierung . . . . .  | 27        |
| 2.4      | IT-Auslastung und Ressourcenbetrachtung . . . . .                                  | 29        |
| 2.4.1    | Ressourcen eines Servers . . . . .   | 29        |
| 2.4.2    | Ressourcenbedarf von virtuellen Maschinen . . . . .                                | 30        |
| 2.5      | Betriebszustände des Rechenzentrums . . . . .                                      | 32        |
| 2.6      | Optimierung durch Server-Konsolidierung . . . . .                                  | 34        |
| 2.7      | Zusammenfassung . . . . .  | 35        |
| <b>3</b> | <b>Verwandte Arbeiten</b>  | <b>37</b> |
| 3.1      | Lastmanagement- und Konsolidierungsverfahren . . . . .                             | 37        |
| 3.1.1    | Lokales Lastmanagement . . . . .   | 38        |
| 3.1.2    | Lastmanagement unter Berücksichtigung thermischer Aspekte . . . . .                | 39        |
| 3.1.3    | Lastmanagement als multi-kriterielles Problem . . . . .                            | 41        |
| 3.1.4    | Rechenzentrumsübergreifendes Lastmanagement . . . . .                              | 43        |
| 3.2      | Energetische Steuerung . . . . .   | 45        |
| 3.3      | Rechenzentrumsmodellierung . . . . .   | 47        |
| 3.4      | Zusammenfassung . . . . .  | 48        |
| <b>4</b> | <b>Energetische und betriebsstabile Optimierung als multi-kriterielles Problem</b> | <b>49</b> |
| 4.1      | Problembeschreibung . . . . .  | 49        |
| 4.1.1    | Energetische Einflussfaktoren im Rechenzentrum . . . . .                           | 51        |
| 4.1.2    | Identifikation der Kriterien . . . . .   | 58        |

|          |  |            |
|----------|--|------------|
| 4.2      | Kriterium: Optimieren der elektrischen Leistungsaufnahme . . . . . | 60         |
| 4.2.1    | Server . . . . .   | 61         |
| 4.2.2    | Kühlung . . . . .  | 63         |
| 4.2.3    | Lösungsansatz . . . . .  | 68         |
| 4.3      | Kriterium: Thermische Stabilität . . . . .                         | 69         |
| 4.3.1    | Server-Temperatur . . . . .  | 69         |
| 4.3.2    | Wärmeverteilung . . . . .  | 70         |
| 4.3.3    | Lösungsansatz . . . . .  | 73         |
| 4.4      | Kriterium: Sicherstellen der Flexibilität . . . . .                | 73         |
| 4.4.1    | Lokal: Zusätzliche freie Server-Ressourcen . . . . .               | 73         |
| 4.4.2    | Global: Zusätzliche Server bereithalten . . . . .                  | 74         |
| 4.4.3    | Begrenzen der Evakuierungszeit . . . . .                           | 77         |
| 4.4.4    | Lösungsansatz . . . . .  | 78         |
| 4.5      | Formale Darstellung . . . . .                                      | 79         |
| 4.6      | Kombinierter Lösungsansatz . . . . .                               | 81         |
| 4.7      | Zusammenfassung . . . . .  | 82         |
| <b>5</b> | <b>Rechenzentrumsmodellierung</b>                                  | <b>85</b>  |
| 5.1      | Modellanforderungen . . . . .                                      | 85         |
| 5.2      | Unterbrechungsfreie Stromversorgung . . . . .                      | 87         |
| 5.3      | Server . . . . .   | 88         |
| 5.3.1    | Modell der elektrischen Leistungsaufnahme . . . . .                | 88         |
| 5.3.2    | Temperaturmodell . . . . .   | 90         |
| 5.4      | Umluftkühlgerät . . . . .  | 91         |
| 5.4.1    | Modell der elektrischen Leistungsaufnahme . . . . .                | 91         |
| 5.4.2    | Kühlleistungsmodell . . . . .                                      | 92         |
| 5.5      | Kältemaschine . . . . .  | 97         |
| 5.6      | Auslastungs- und VM-Migrationsmodelle . . . . .                    | 98         |
| 5.6.1    | Auslastung und Vergleichbarkeit . . . . .                          | 99         |
| 5.6.2    | VM-Migrationsmodelle . . . . .                                     | 102        |
| 5.7      | Netzwerkmodellierung . . . . .                                     | 104        |
| 5.8      | Zusammenspiel der Modelle . . . . .                                | 106        |
| 5.9      | Zusammenfassung . . . . .  | 108        |
| <b>6</b> | <b>DaMucA: Data Center Multi-criteria Aware Allocation</b>         | <b>109</b> |
| 6.1      | Architektur . . . . .  | 109        |
| 6.2      | Anwendungsfälle . . . . .  | 110        |
| 6.2.1    | Kurzzeit- und Langzeitplanung . . . . .                            | 111        |
| 6.2.2    | Reaktion auf Notfälle: Evakuierung von Servern . . . . .           | 116        |
| 6.3      | Übersicht: DaMucA-Verfahren . . . . .                              | 118        |
| 6.4      | DaMucA-Metaheuristik: Kombination aktiver IT-Räume . . . . .       | 120        |
| 6.4.1    | Genetischer Algorithmus . . . . .                                  | 120        |
| 6.4.2    | Ablauf des Algorithmus . . . . .                                   | 122        |



|          |  |            |
|----------|--|------------|
| 6.5      | DaMucA-Server-Heuristik: Lastverteilung in aktiven IT-Räumen . . . . . | 125        |
| 6.5.1    | Algorithmus zur Lastverteilung auf Server in gewählten IT-Räumen       | 125        |
| 6.5.2    | IT-Raumwahl . . . . .  | 128        |
| 6.5.3    | Server-Wahl . . . . .  | 129        |
| 6.5.4    | Wahl der zu migrierenden virtuellen Maschinen . . . . .                | 133        |
| 6.5.5    | Sicherstellen der globalen Flexibilität . . . . .                      | 134        |
| 6.6      | Laufzeit und Komplexität . . . . .                                     | 136        |
| 6.6.1    | Laufzeit der DaMucA-Metaheuristik . . . . .                            | 137        |
| 6.6.2    | Laufzeit der DaMucA-Server-Heuristik . . . . .                         | 137        |
| 6.7      | Zusammenfassung . . . . .  | 138        |
| <b>7</b> | <b>Evaluation</b>  | <b>141</b> |
| 7.1      | Evaluationsszenario . . . . .  | 141        |
| 7.2      | DaMucA-Metaheuristik: IT-Raumwahl mittels genetischem Algorithmus . .  | 142        |
| 7.2.1    | Einfluss der Anzahl und Kombination aktiver IT-Räume . . . . .         | 143        |
| 7.2.2    | Lösungsfindung: Generationen bis zum Abbruchkriterium . . . . .        | 146        |
| 7.3      | DaMucA-Server-Heuristik: Optimierung in den aktiven IT-Räumen . . . .  | 149        |
| 7.3.1    | Elektrische Leistung auf Ebene der IT-Räume . . . . .                  | 149        |
| 7.3.2    | Thermische Stabilität und Wärmeverteilung in IT-Räumen . . . . .       | 157        |
| 7.3.3    | Evakuierungszeit . . . . .   | 161        |
| 7.3.4    | Energieeinsparung gegenüber Flexibilität . . . . .                     | 161        |
| 7.3.5    | Energieeinsparung bei der Kühlung . . . . .                            | 164        |
| 7.4      | Gesamtenergieeinsparung . . . . .                                      | 166        |
| 7.5      | Einfluss des Szenarios auf Leistungsfähigkeit des Verfahrens . . . . . | 166        |
| 7.6      | Zusammenfassung . . . . .  | 168        |
| <b>8</b> | <b>Zusammenfassung und Ausblick</b>                                    | <b>171</b> |
|          | <b>Literaturverzeichnis</b>  | <b>175</b> |
|          | <b>Abbildungsverzeichnis</b>   | <b>187</b> |
|          | <b>Tabellenverzeichnis</b>   | <b>189</b> |
|          | <b>Symbolverzeichnis</b>   | <b>191</b> |
|          | <b>Abkürzungsverzeichnis</b>   | <b>195</b> |



# Zusammenfassung

Aufgrund steigender Energiepreise und erhöhten Rechenbedarfs, der beispielsweise durch den Trend des Cloud-Computings zusätzlichen Steigerungen unterliegt, ist der Energiebedarf von Rechenzentren auch trotz immer effizienter werdender Hardware weiterhin ein ökonomisch und ökologisch diskutiertes Thema. Dabei werden und wurden mittlerweile vielfache Möglichkeiten der Energiebedarfsreduktion für Rechenzentren untersucht und angewandt, darunter auch Server-Konsolidierungsstrategien, die durch die Nutzung von Virtualisierungstechniken realisiert werden. Durch den Einsatz dieser Virtualisierungstechniken wurden nun auch erstmals umfangreiche Freiheitsgrade für Rechenzentren erschaffen, die eine Regelbarkeit des Rechenzentrums nach verschiedenen Kriterien erlauben. Freiheitsgrade entstehen dabei durch die Möglichkeit, eine spezifische, unveränderte IT-Last in unterschiedlichen Betriebszuständen abzuarbeiten, wobei jeder dieser Zustände andere Betriebsparameter aufweist. Diese Freiheitsgrade müssen jedoch zunächst untersucht und verstanden werden, damit sie optimal ausgenutzt werden können und eine einseitige Optimierung des Rechenzentrums vermieden wird. Eine solche kann sich negativ auf die Gesamtsituation im Rechenzentrum auswirken, zum Beispiel aufgrund von nicht beachteten Wechselwirkungen verschiedener Geräteklassen, beispielsweise im Bezug auf den Energiebedarf oder den thermisch sicheren Betrieb.

In dieser Dissertation wird gezeigt, dass eine einseitige Optimierung in Hinsicht auf die Minimierung des Energiebedarfs von IT-Hardware, insbesondere von Servern, nicht immer zielführend ist und vor allem zu signifikanten Ineffizienzen im Betrieb weiterer Geräteklassen führen kann. Um dies zu vermeiden, werden die Freiheitsgrade im Rechenzentrum, die ausschließlich durch die Verwendung von Virtualisierungstechniken und der damit verbundenen Möglichkeit der Lastverschiebung entstehen, untersucht und es werden die Auswirkungen auf mehrere Kriterien des Betriebszustands dargestellt. Die Arbeit bezieht sich dabei auf die Kriterien der Optimierung des Energiebedarfs des gesamten Rechenzentrums, der Wahrung der thermischen Stabilität und der Sicherstellung einer definierten Flexibilität des Betriebszustands des Rechenzentrums. Es werden Modelle aufgestellt, die für eine Bewertung jedes Betriebszustands hinsichtlich der einzelnen Kriterien genutzt werden, und es wird die Kombination dieser Bewertungen vorgenommen, wodurch eine multi-kriterielle Bewertung einzelner Lösungskandidaten vorgenommen werden kann.

Grundlage der eigentlichen Optimierung ist die Verteilung der Rechenlast im Rechenzentrum; dies geschieht mittels virtueller Maschinen und der Technik der Live-Migration, die ein dynamisches und annähernd unterbrechungsfreies Verschieben der IT-Last auf verschiedene physikalische Server und in verschiedene Server-Räume ermöglicht. Zur Energiebedarfsenkung können dann ungenutzte Server ausgeschaltet werden. Auch die Verteilung der Auslastung über verschiedene Server-Räume hinweg hat einen signifikanten Einfluss auf

den Energiebedarf des Rechenzentrums, sodass auch hier Optimierungen untersucht und vorgenommen werden.

Zur Lösungssuche wird ein spezielles neues Suchverfahren vorgestellt, das auf evolutionären Algorithmen basiert, und das die hierfür entwickelte *Data Center Multi-criteria aware Allocation (DaMucA)* verwendet, die eine schnelle Lösungskandidatenerzeugung erlaubt und dabei alle betrachteten Kriterien mit einbezieht. Durch dieses Verfahren kann eine Lösung für ein mittleres bis großes Rechenzentrum mit mehreren tausend Servern und zehntausenden virtuellen Maschinen in einer hinreichend kurzen Laufzeit gefunden werden. Das Verfahren wird schließlich simulativ evaluiert. Es wird deutlich, dass das Optimierungsverfahren signifikante Energieeinsparungen erreicht und gleichzeitig eine thermisch stabile und flexible Verteilung der IT-Auslastung vornimmt.

# Abstract

Due to rising energy prices and increased computational requirements that were fuelled for example by the trend of cloud computing, the energy efficiency of data centres is currently an economically and ecologically discussed topic. As a result a multitude of possibilities to reduce the energy demand of data centres have now been studied and applied, including server consolidation strategies. These are realized by the use of virtualization technologies. These virtualization techniques also create extensive degrees of freedom for data centres that allow the energetic adjustability of the data centre according to various criteria. Degrees of freedom arise here by the possibility to process a specific unchanged IT load in different operational states; with each of these states having different operating parameters. However, these degrees of freedom must first be examined and understood, so that they can be optimally utilized and a unilateral optimization of the data centre is avoided. A unilateral optimization may have a negative impact on the overall (energetic) situation in the data centre, for example due to interactions between different classes of devices. Especially the energy demand or the thermally safe operation may be impacted. In this thesis it is shown that a unilateral optimization in terms of minimizing the energy consumption of IT hardware, especially servers, is not always constructive and can lead to significant inefficiencies in the operation of other classes of devices. To avoid this, the degrees of freedom in the data centre, which are exploited solely by the use of virtualization technologies and the associated possibility of load shifting, are investigated. Especially the impact and effects on several criteria of the operating state are researched. The thesis hereby refers to the criteria of optimizing the power consumption of the entire data centre, maintaining the thermal stability and ensuring a defined flexibility of the operating state of the data centre. The thesis presents energetic models for the most important hardware devices in the data centre and for the utilization effects in virtual machines. These models are then used to evaluate different solutions according to the identified optimization criteria. Basis of the actual optimization is the distribution of computational load in the data centre. This is achieved by applying virtualization and by live-migrating virtual machines to different servers in the data centre. Live-migration allows for a dynamic and almost uninterrupted reallocation of the IT load on different physical servers in different server rooms in the data centre. Also, the load can be concentrated on few servers, while unused servers can be turned off to reduce the energy demand. Additionally, the distribution of the application load in different server rooms has a significant impact on the energy needs of the data centre. To find a solution that satisfies all the identified optimization criteria, the new optimization algorithm *Data Center Multi-criteria Allocation (DaMucA)* is presented. It is based on evolutionary algorithms and heuristics and allows for a quick generation of solutions. The algorithm is capable of solving the underlying problem for a data centre

with thousands of servers in a sufficiently short time. The algorithm is finally evaluated by simulation. Results show that the optimization method achieves significant energy savings while providing a thermally stable and flexible distribution of the computing load in the data centre.

# 1 Einleitung

Die Entwicklung von Rechenzentren innerhalb der letzten Jahre zeigt sowohl einen deutlich erhöhten Energiebedarf aller Rechenzentren weltweit aufgrund der zunehmenden Ausbreitung, als auch ein gesteigertes Preis- und Umweltbewusstsein der Betreiber. Trotzdem wird die Mehrheit der Rechenzentren auf der Welt immer noch ineffizient betrieben: Server sind nur sehr gering ausgelastet, die Stromversorgung und Klimatisierung sind häufig überdimensioniert. Zudem wird der Betrieb für eine theoretische Maximalbelastung ausgelegt, die in der Realität nicht vorkommt, was zu weiteren Ineffizienzen, vor allem bei den Servern und der Klimatisierung, führt.

Aktuelle Ansätze zur Verbesserung der Energieeffizienz von Rechenzentren beziehen sich auf den Einsatz moderner Hardware, die auch bei niedrigen Auslastungen einen hohen Wirkungsgrad aufweist, und eine Optimierung des Betriebs. Diese erfolgt häufig mittels Virtualisierung, Server-Konsolidierung und Lastmanagement. Einseitige Optimierungen, die die Wechselwirkungen zwischen verschiedenen Geräteklassen im Rechenzentrum ignorieren, können allerdings zu zusätzlichen Ineffizienzen oder instabilen Betriebszuständen führen.

Die vorliegende Dissertation untersucht zunächst die energetischen Einflüsse und Wechselwirkungen der Geräte verschiedener Gewerke eines Rechenzentrums. Durch die Möglichkeit der Lastverschiebung, die durch Virtualisierung erschaffen wurde, können äußerst unterschiedliche energetische Betriebszustände mit jeweiligen Vor- und Nachteilen erreicht werden. Auf dieser Basis wird dann eine Optimierungsstrategie entwickelt, die die energetischen Einflüsse und Wechselwirkungen gezielt ausnutzt, was sowohl zu energetischen Einsparungen als auch zu einer Verbesserung der Flexibilität und der Stabilität des Betriebs führen soll.

In den folgenden Abschnitten wird die Motivation und der Kontext der Arbeit sowie die Zielsetzung weiter konkretisiert.

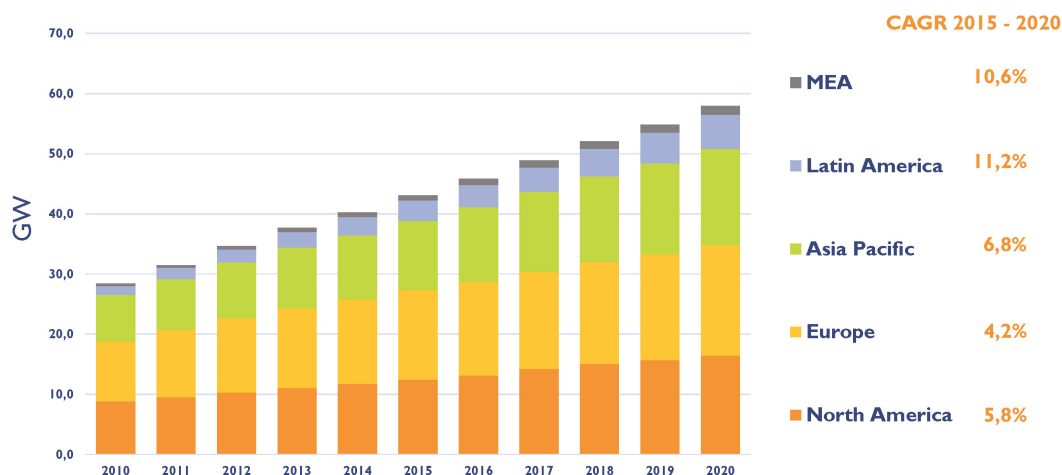
## 1.1 Motivation

Der weltweite Energiebedarf von Rechenzentren hat sich in den letzten Jahren deutlich gesteigert und wird nach Prognosen diesen Trend weiterhin fortsetzen [148], wie die Abbildung 1.1 zeigt. Insgesamt hat der Energiebedarf mittlerweile signifikante Ausmaße erreicht. Als ein Beispiel aus Deutschland können die lokalen Rechenzentren der Stadt Frankfurt am Main dienen. Es entfallen mittlerweile ca. 20% des gesamten elektrischen Energiebedarfs

der Stadt alleine auf Rechenzentren<sup>1</sup>.

### WORLDWIDE DATA CENTER FACILITIES – POWER NEEDS IN GW

(Source: New Technologies and Architectures for Efficient Data Center report, July 2015, Yole Développement)



**With no slowdown in new facility construction, data centers worldwide will have an increasing need for power.**



©2015 | www.yole.fr

Abbildung 1.1: *Weltweite Entwicklung des Energiebedarfs von Rechenzentren inklusive Prognose über die nächsten Jahre: In allen Regionen der Welt ist mit einem Anstieg zu rechnen. Nordamerika, Europa und Ozeanien/Ostasien sind absolut gesehen die dominierenden Regionen, während die größten relativen Steigerungen in den Ländern Lateinamerikas, des Nahen Ostens und Zentralasiens zu finden sind (Grafik unverändert übernommen aus [148]).*

Gleichzeitig steigt das Bewusstsein der Rechenzentrumsbetreiber für Energieeffizienz. So bewerteten nach einer aktuellen Studie [79] 50% der Betreiber die Effizienzsteigerung der Rechenzentren in der Zukunft als eine der wichtigsten Herausforderungen. Dies schlägt sich auch im Markt nieder. So ist von einem signifikantem Wachstum im Segment der „grünen Rechenzentren“ in den nächsten Jahren auszugehen [135] (dieser Trend ist in Abbildung 1.2 dargestellt). Dabei bezieht sich der Begriff des „grünen Rechenzentrums“ auf besonders effiziente Rechenzentren (insbesondere bezogen auf die Server, Kühlung und Stromversorgung), die durch steigende Anteile erneuerbarer Energie versorgt werden und eine Reduktion des Elektromülls durch vermehrtes Recycling anstreben. Neben den ökologischen

<sup>1</sup>Diese Zahl wurde im Rahmen eines Netzausbaus in einer Pressemitteilung des Netzbetreibers NRM Netzdienste Rhein-Main GmbH vom 08.06.2015 veröffentlicht: [https://www.mainova.de/static/de-mainova/downloads/PM\\_Mainova\\_Interxion.pdf](https://www.mainova.de/static/de-mainova/downloads/PM_Mainova_Interxion.pdf)



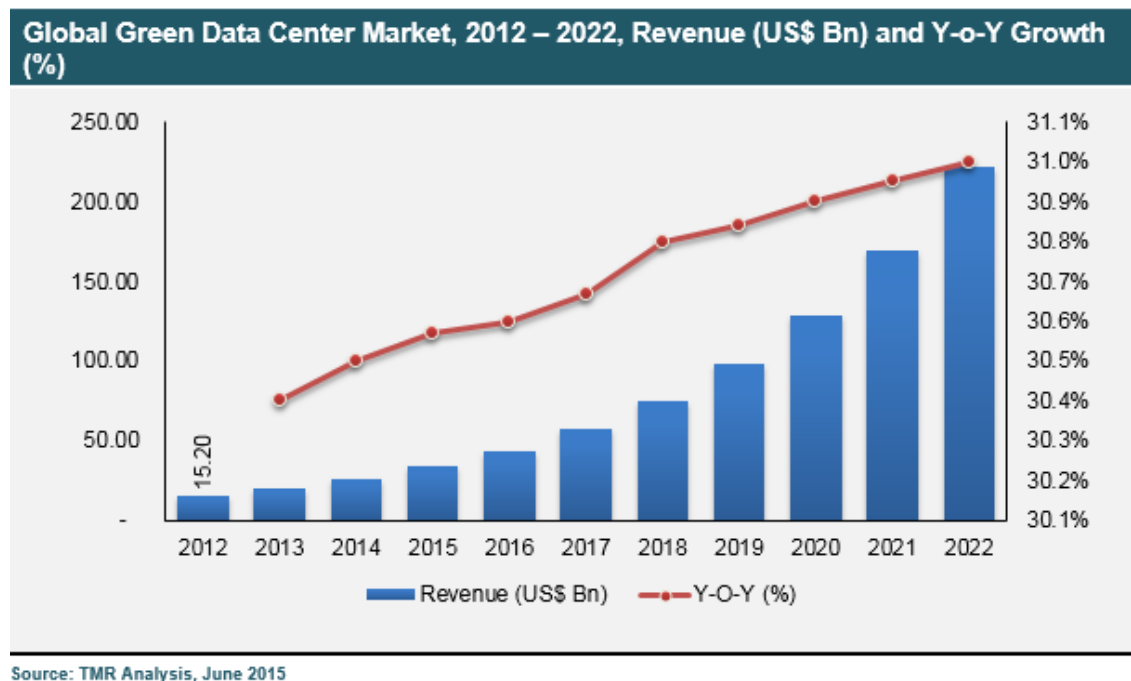


Abbildung 1.2: Der Markt für „grüne Rechenzentren“ hat deutlich zugelegt, was die Ist-Analyse bis zum Jahr 2015 zeigt. Für die Zukunft werden erhebliche Steigerungen prognostiziert (Grafik unverändert übernommen aus [135]).

Vorteilen dieser „grünen Rechenzentren“ ist auch von sinkenden laufenden Kosten auszugehen, da die Energiepreise mittlerweile einen signifikanten Anteil der Gesamtkosten darstellen [135].

Hier zeigen sich also zwei gegenseitige Entwicklungen: auf der einen Seite benötigen Rechenzentren immer mehr Energie, auf der anderen Seite steigt das Bewusstsein für Effizienz und Energieeinsparungen. Diese Kombination ist durch mehrere Trends zu erklären. Zum einen ist der Rechenzentrumsmarkt allgemein einem Wachstum unterlegen. Der Bedarf für Cloud-Anwendungen verstärkt diesen Trend zusätzlich [27], wodurch immer mehr Rechenzentren und immer mehr Server-Kapazitäten benötigt werden. Zum anderen wird die deutliche Mehrheit der Rechenzentren (95%) immer noch ineffizient betrieben [93], hier ist also ein enormes Verbesserungspotential vorhanden.

Diese Ineffizienz im Betrieb bezieht sich primär auf die Server, die die eigentliche Arbeitslast im Rechenzentrum tragen. So gehen verschiedene Studien [12, 55] davon aus, dass ein durchschnittlicher Server nur zwischen 10% und 50% ausgelastet ist. Einige Untersuchungen gehen sogar von deutlich geringeren 6-12% aus [61, 93]. Diese als „komatös“ bezeichneten Server [66] stellen einer aktuellen Studie aus dem Jahr 2015 nach einen Anteil von ca. 30% aller weltweit vorhandenen Server dar. Die sehr gering ausgelasteten Server benötigen trotzdem ca. 50% der maximalen Leistungsaufnahme [66], was insgesamt betrachtet zu einer niedrigen Effizienz im Betrieb führt.

Die geringe Auslastung der Server bietet demnach ein großes Potential zur Energieeinsparung und Optimierung, denn die Effizienz steigt bei aktuellen Server-Typen in der Regel mit zunehmender Auslastung. Diese Tatsache nutzen aktuelle Lastmanagementverfahren aus, indem sie darauf abzielen, möglichst wenige Server einzuschalten und diese dann möglichst vollständig auszulasten. Dadurch lassen sich erhebliche Einsparungen bei den Servern erreichen, die im Vergleich zum nicht optimierten Betrieb 50% oder mehr betragen können [53]. Allerdings beziehen sich diese Einsparungen nur auf die Server. Eine einseitige Optimierung dieser kann jedoch zu weiteren Ineffizienzen oder sogar instabilen Betriebszuständen führen.

Hier sind insbesondere die weiteren Geräte im Rechenzentrum zu nennen, die den Betrieb der Server ermöglichen, indem sie die notwendige Stromversorgung und Klimatisierung bereitstellen. Diese notwendigen infrastrukturellen Geräte zeigen ebenfalls einen signifikanten elektrischen Energiebedarf auf. Gerade Geräte, die zur Klimatisierung benötigt werden, sind ein enormer Energiekostentreiber und bereits seit einigen Jahren Gegenstand der Forschung und Entwicklung, damit diese Kosten gesenkt werden können. Dazu wurden sowohl vollständig neue Technologien zur Klimatisierung entwickelt, als auch die aktuell angewandten Systeme so verbessert, dass sie beispielsweise eine Regelung ihres Betriebs vornehmen können und so auch bei niedrigen Auslastungen einen erhöhten Wirkungsgrad zeigen. Diese Verbesserung ist deshalb besonders wichtig, da der Rechenzentrumsbetrieb einer hohen Dynamik ausgeliefert ist, die aus den unterschiedlichen Nutzungsprofilen und Zeiten resultiert. Greifen besonders viele Nutzer auf einen angebotenen Dienst zu, steigt die Server-Auslastung und damit dessen elektrische Leistungsaufnahme, was im Normalfall auch zu einer erhöhten Leistungsaufnahme der Klimatisierung und weiterer Geräte führt. Der bereits erwähnte Einsatz von Server-Konsolidierung setzt genau an dieser Stelle an und passt die genutzten Server-Kapazitäten an den jeweiligen Bedarf an. Diese Dissertation setzt an diesem Punkt an und erweitert die existierenden Konsolidierungsstrategien um eine gezielte Ausnutzung der energetischen Freiheitsgrade der unterstützenden Geräte im Rechenzentrum. Damit diese Betrachtung überhaupt erfolgen kann, ist eine Identifikation und Analyse der energetischen Einflüsse der einzelnen Geräte, der individuellen Regelbarkeit und möglicher energetischer Wechselwirkungen notwendig.

## 1.2 Beitrag dieser Arbeit

Die Entwicklung und Anwendung von Virtualisierungstechnologien bietet die Möglichkeit, die Anwendungen im Rechenzentrum zwischen verschiedenen physikalischen Servern zu verschieben. Die Auslastung, die durch diese Anwendungen auf den Servern erzeugt wird, bestimmt maßgeblich deren Energiebedarf. Durch unterschiedliche Kombinationen der Verteilung der virtuellen Maschinen auf die Server können so verschiedene energetische Zustände erreicht werden. Dazu zählt auch die räumliche Verortung der Auslastung im Rechenzentrum und auf den Servern. Dies gilt sowohl für die Verteilung der Auslastung in den einzelnen Server-Räumen, als auch über mehrere Räume hinweg. Der bereits angesprochene Trend der energetischen Regelbarkeit von weiteren Geräten, insbesondere der Stromversor-

gung und der Klimatisierung, führt dazu, dass die energetischen Zustände der Server auch diese weiteren Geräte beeinflussen und etwaige energetische Einsparungen auch hier zum Tragen kommen können. Jede Geräteklasse im Rechenzentrum, die energetisch regelbar ist, besitzt damit einen energetischen Freiheitsgrad, in dem es betrieben werden kann.

Damit eine energetische Optimierung des Rechenzentrums erfolgen kann, werden zunächst diese Freiheitsgrade der einzelnen Geräte untersucht. Insbesondere ist die Analyse der Einflüsse verschiedener Betriebsparameter einzelner Geräte auf den Energiebedarf eine zentrale Fragestellung. Durch die gezielte Ausnutzung der Freiheitsgrade der Geräte in Kombination mit der Möglichkeit, einzelne Parameter zur Veränderung des energetischen Zustands zu setzen, wird eine energetische Optimierung und Steuerung möglich. Allerdings müssen an dieser Stelle auch Wechselwirkungen zwischen den beteiligten Geräteklassen im Rechenzentrum untersucht werden. Denn wenn eine Optimierung eines einzelnen Gerätes zur Verschlechterung weiterer Teilnehmer führt, ist ein Abwägen der jeweiligen Einzeloptimierungen notwendig.

Die Erkenntnisse, die aus diesen Analysen hervorgehen, werden zur Entwicklung eines Optimierungsverfahrens genutzt, das sowohl den Energiebedarf eines Rechenzentrums senken soll, aber gleichzeitig auch weitere Kriterien berücksichtigen kann, die für den Betrieb des Rechenzentrums von Bedeutung sind. Denn beim Rechenzentrumsbetrieb spielen neben der Energieeffizienz weitere Aspekte eine zentrale Rolle. Rechenzentren werden in der Regel so entworfen, dass sie einen stabilen Betrieb der Server ermöglichen sollen. Dafür werden redundante Systeme für die Stromversorgung, Klimatisierung und Netzwerkanbindung verwendet. Eine Senkung des Energiebedarfs steht diesen Redundanz- und Stabilitätsanforderungen häufig direkt entgegen. Viele Betreiber weigern sich daher, Server-Konsolidierung und weitere energetische Optimierungen anzuwenden, da sie Ausfälle oder eine negative Beeinflussung des Betriebs befürchten. Hier gilt es, formale Kriterien zu identifizieren, mit denen eine Abwägung zwischen Energieeinsparung und zusätzlichen Reserveressourcen getroffen werden kann. Außerdem müssen die Kriterien und Betriebsparameter für einen betriebssicheren und stabilen Zustand des Rechenzentrums identifiziert und berücksichtigt werden. Dies gilt insbesondere für plötzlich und unerwartet auftretende Ereignisse, wie plötzliche Auslastungsschwankungen, thermische Anomalien oder Ausfälle von Geräten. Im Falle eines Fehlers im Rechenzentrum, muss dieser schnell behoben werden können und etwaige betroffene Systeme müssen durch Ersatzsysteme ersetzt werden können, beispielsweise müssen betroffene Anwendungen auf bereitstehende Reserverserver „evakuiert“ werden können.

Damit diese Anforderungen zusammen mit einer energetischen Optimierung berücksichtigt werden können, betrachtet diese Arbeit die folgenden drei grundsätzlichen Kriterien:

- *Optimierung der elektrischen Leistungsaufnahme:* Die Geräte im Rechenzentrum sollen so betrieben werden, dass der gesamte Energiebedarf des Rechenzentrums gesenkt werden kann.
- *Thermische Stabilität:* Die einzuhaltenden Parameter für einen thermisch sicheren Betrieb müssen identifiziert und deren Einhaltung gewährleistet werden.

- *Sicherstellen der Flexibilität:* Damit den genannten Befürchtungen der Betreiber entgegengewirkt werden kann, wird eine Flexibilität des Betriebszustands definiert und deren Einhaltung auch in einem konsolidierten Zustand sichergestellt. Diese Flexibilität führt dazu, dass eine lokale Überlastung von Servern verhindert werden soll und in einem plötzlich auftretenden Notfall genug Reserveserver vorhanden sind und eine rechtzeitige Evakuierung der betroffenen Systeme erfolgen kann.

Insbesondere zur Sicherstellung der genannten Flexibilität wird eine Metrik vorgestellt, die diese für einen gegebenen Betriebszustand des Rechenzentrums bestimmen und mittels des entwickelten Verfahrens gegebenenfalls korrigieren kann. Diese Flexibilität steht, wie bereits kurz erwähnt, der energetischen Optimierung direkt entgegen. Es muss also eine Abwägung zwischen einem flexiblen Betriebszustand und einem möglichst geringen Energiebedarf erfolgen. Die Flexibilitätsmetrik wurde dabei so entworfen, dass die gewünschte Flexibilität eines Betriebszustands erreicht wird, gleichzeitig aber ein dafür energetisch optimierter Betrieb des Rechenzentrums erfolgt. Damit kann die Flexibilität auch zur energetischen Regelung des Rechenzentrums verwendet werden. Das entwickelte Optimierungsverfahren erlaubt also gezielt Betriebszustände mit einem höheren Energiebedarf, wenn dadurch die Flexibilität des Betriebs signifikant steigt.

Dies ist gleichzeitig auch eine deutliche Abgrenzung zu bisherigen Optimierungsverfahren. Diese setzen meist auf Server-Konsolidierung oder einzelne thermische Verbesserungen, betrachten aber nicht die Gesamtheit der in dieser Arbeit behandelten Kriterien.

Ein weiterer wichtiger Aspekt ist die Betrachtung des zeitlichen Rahmens der energetischen Optimierung des Rechenzentrums. Die existierenden Verfahren des Lastmanagements in Rechenzentren konzentrieren sich auf kurze zeitliche Abstände, an denen eine Anpassung der Server-Kapazitäten vorgenommen werden kann. Ist beispielsweise die Auslastung der angebotenen Dienste im Rechenzentrum im Laufe eines Tages stark schwankend, so ist es möglich, dass ein Verfahren zu mehreren Migrationen und Aus- und Einschaltvorgängen von Servern an einem Tag führt. Dadurch kann auch auf kurze und eher schwach ausgeprägte Schwankungen der Auslastung reagiert werden und die Energieersparnis wird erhöht.

Die Betrachtungen und Analysen, die in dieser Arbeit vorgenommen werden, schließen allerdings träge Geräte, wie insbesondere die Geräte der Klimatisierung, mit ein. Diese können nicht in einem vergleichbar kurzen Zeitraum geregelt werden. Eine Anpassung verschiedener Parameter des Kühlsystems kann mehrere Tage in Anspruch nehmen, da einerseits das System eine Zeit benötigt, bis die Änderungen im gesamten System umgesetzt wurden. Andererseits tendieren Rechenzentrumsbetreiber dazu, aufgrund der Wahrung der Betriebsstabilität des Rechenzentrums insbesondere Änderungen an Parametern der Klimatisierung nur schrittweise durchzuführen. Dadurch muss für die in dieser Arbeit vorgeschlagenen Optimierungen ein längerer Zeitraum von mehreren Tagen bis Monaten betrachtet werden. Es ergibt sich die Fragestellung, inwieweit ein Optimierungsverfahren die nur langfristig mögliche Regelbarkeit von Geräten wie beispielsweise der Klimatisierung ausnutzen kann. Beispielsweise muss das Verfahren vorausschauende Entscheidungen treffen können, damit eine Optimierung über diese längeren Zeiträume möglich wird. Die

Auslastungen der Dienste und damit der Server müssen also langfristig betrachtet werden können. Die Optimierung nimmt dann eine Abwägung zwischen der Planungsdauer und dem entsprechenden energetischen Einsparpotential vor.

Eine weitere zentrale Abgrenzung des Optimierungsverfahrens, das in dieser Arbeit entwickelt wird, ist die Behandlung der vorhandenen Anwendungen im Rechenzentrum. Es gilt grundsätzlich die Voraussetzung, dass alle Anwendungen zu jeder Zeit ausgeführt werden können müssen und keine Pausierungen oder eine Verlagerung auf andere Rechenzentrumsstandorte vorgenommen wird.

Unter Berücksichtigung der beschriebenen Kriterien und der genannten Voraussetzungen wird das Optimierungsverfahren *DaMucA* (*Data Center Multi-criteria Aware Allocation*) entwickelt. Dieses nimmt sowohl die energetische Optimierung als auch die Wahrung der Stabilitätsanforderungen vor. Da der Lösungsraum des zugrundeliegenden Problems sehr groß ist, nutzt das Verfahren mehrere dazu entwickelte Heuristiken, die zu einer schnellen Berechnung der Verteilung der IT-Last auf die Server führen. Da eine Optimierung auf verschiedenen Ebenen stattfindet, werden diese Heuristiken mit einem genetischen Algorithmus verknüpft, der bereits eine Vorauswahl der zu nutzenden Ressourcen im Rechenzentrum vornimmt. Das Verfahren stützt seine Entscheidungen auf einer Rechenzentrumsimulation, deren Modelle sich zu Teilen aus existierenden Ansätzen aus der Forschung und eigenen Modellen, die durch Messungen erstellt wurden, zusammensetzen.

In der Evaluation wird dann gezeigt, dass das Optimierungsverfahren zu signifikanten Energieeinsparungen führt, die vor allem durch die Kombination der Server-Konsolidierung und der Optimierung der Kühlung erreicht werden. Dabei gelingt die Sicherstellung der weiteren Kriterien, die zu einem stabilen und flexiblen Rechenzentrumsbetrieb beitragen. Zusammenfassend werden in dieser Arbeit folgende Beiträge formuliert und beschrieben:

- *Identifikation und Analyse der energetischen Einflüsse*: Die Analyse des energetischen Zusammenspiels der Gewerke im Rechenzentrum bildet die Grundlage für energetische Optimierungen. Es werden somit Bereiche identifiziert, die das Potential für Optimierungen beinhalten, aber auch Bereiche, in denen keine Optimierungen möglich sind.
- *Identifikation von betriebsrelevanten und energetischen Kriterien*: Die Ergebnisse der energetischen Analyse und weitere Parameter, die den stabilen Betrieb des Rechenzentrums sicherstellen sollen, werden in Optimierungskriterien definiert.
- *Räumliche Optimierung*: Die Analyse der Einflüsse der räumlichen Verteilung der IT-Last (sowohl in den Räumen als auch über mehrere Räume hinweg) wird zunächst vorgenommen. Das Optimierungsverfahren soll dann eine gezielte Ausnutzung der räumlichen Verteilung der IT-Last vornehmen, was einerseits die Betriebsstabilität begünstigen und andererseits zu signifikanten Energieeinsparungen bei versorgenden Geräten führen soll.
- *Definition und Anwendung einer Flexibilitätsmetrik zur gezielten energetischen Steuerung*: Da energetische Optimierungen häufig einen negativen Einfluss auf die Flexi-

bilität des Rechenzentrumsbetriebes haben, wird eine Metrik zur Bestimmung dieser definiert und in den Optimierungsprozess mit einbezogen. Damit lässt sich eine Abwägung zwischen energetischer Optimierung und Flexibilität des Betriebszustands vornehmen.

- *Modellierung der wichtigsten beteiligten Geräte:* Bereits existierende und neue Modelle werden kombiniert, damit die identifizierten energetischen Einflüsse und Wechselwirkungen untersucht und simuliert werden können.
- *Entwicklung des DaMucA-Lösungsverfahrens zur Rechenzentrumsoptimierung:* Unter Berücksichtigung der identifizierten energetischen Einflüsse und der Definition der Optimierungskriterien wird das DaMucA-Lösungsverfahren entwickelt. Da ein großer Lösungsraum untersucht werden muss und das Verfahren zudem in großen Rechenzentren mit mehreren tausend Servern eingesetzt werden können soll, wird eine heuristische Herangehensweise angewandt. Dazu werden mehrere Heuristiken entwickelt und vorgestellt.

### 1.3 Aufbau der Arbeit

In Kapitel 2 erfolgt zunächst eine Beschreibung der Grundlagen, die für das weitere Verständnis dieser Arbeit notwendig sind. Es wird auf das Rechenzentrum und seine Infrastruktur eingegangen und es folgen einige grundlegende Definitionen, die in der späteren Arbeit aufgegriffen werden.

Kapitel 3 zeigt verwandte Arbeiten aus den Forschungsbereichen, die in dieser Arbeit untersucht oder als Vorarbeiten aufgegriffen werden. Hier werden insbesondere vorhandene Konzepte des Lastmanagements und der energetischen Steuerung von Rechenzentren behandelt. Auch werden existierende Ansätze zur Rechenzentrumsmodellierung genannt und in Relation zu den hier verwendeten Modellen gesetzt.

Die Untersuchung und Analyse der Freiheitsgrade, energetischen Einflussfaktoren und Wechselwirkungen erfolgt zunächst in Kapitel 4. Hier werden aufgrund dieser Analyse die einzelnen Optimierungs- und Einflusskriterien herausgearbeitet, die im entwickelten Optimierungsverfahren berücksichtigt werden sollen. Es erfolgt jeweils eine formale Definition der Kriterien und des jeweiligen Problems sowie eine kurze Beschreibung des Lösungsansatzes, wie er im Optimierungsverfahren Anwendung findet.

Damit die in Kapitel 4 identifizierten Optimierungen vorgenommen und evaluiert werden können, wird eine Modellierung der zentralen Komponenten eines Rechenzentrums benötigt. Diese Modelle werden in Kapitel 5 vorgestellt.

Das Optimierungsverfahren, das die Kriterien aus Kapitel 4 betrachtet und nach diesen einen optimierten Betrieb des Rechenzentrums vornehmen soll, wird in Kapitel 6 entwickelt. Hier werden zunächst kurz die Anwendungsfälle und Rahmenbedingungen genannt, unter denen das Verfahren zum Einsatz kommen kann. Danach folgt die Beschreibung des entwickelten DaMucA-Verfahrens.

Die Evaluation dieses Verfahrens sowie der energetischen Einflüsse auf die elektrische Leistungsaufnahme des Rechenzentrums, die in Kapitel 4 identifiziert wurden, erfolgt in Kapitel 7.

Abschließend wird in Kapitel 8 eine Zusammenfassung der Arbeit, sowie ein Ausblick über weiterführende Themen aus dem Kontext dieser Arbeit gegeben.





## 2 Grundlagen

In diesem Kapitel wird zunächst genauer auf das Rechenzentrum und seine Infrastruktur sowie die wichtigsten Geräteklassen, die in dieser Arbeit betrachtet werden, eingegangen. Anschließend werden das Prinzip der Virtualisierung und die damit verbundenen Techniken kurz erklärt. Das Kapitel legt damit die Grundlagen, die für das weitere Verständnis der Arbeit notwendig sind. Da nicht alle Themen in vollem Umfang beschrieben werden können, werden Literaturangaben zur weiteren Recherche angegeben.

### 2.1 Das Rechenzentrum & seine Infrastruktur

Da der Begriff des Rechenzentrums im wissenschaftlichen Sinne nicht eindeutig definiert ist und häufig unter unterschiedlichen Vorstellungen genutzt wird<sup>1</sup>, erfolgt in diesem Abschnitt zunächst eine für diese Arbeit zugrundeliegende Definition des Sachgegenstandes Rechenzentrum.

#### **Definition 2.1** *Rechenzentrum*

*Ein Rechenzentrum ist ein Gebäude, das hauptsächlich Geräte zur elektronischen Datenverarbeitung beherbergt und speziell für diesen Zweck entworfen wurde. Der Sinn und Zweck eines Rechenzentrums ist im Allgemeinen die Möglichkeit einer Abarbeitung einer bestimmten IT-Last in Form von Anwendungen oder Diensten zu schaffen. Diese Abarbeitung erfolgt durch die installierten Server, weitere Geräte unterstützen diese (z.B. Stromversorgung, Klimatisierung) oder stellen die Vernetzung zur Außenwelt bereit (z.B. Switches, Router). Wie groß die IT-Lastkapazität des Rechenzentrums ist, also wie viel Aufgaben innerhalb einer bestimmten Zeit erledigt werden können, hängt maßgeblich vom Typ und der Anzahl der installierten Server ab. Diese IT-Lastkapazität bestimmt nun auch die energetische Dimensionierung der unterstützenden Geräte.*

*Wenn in dieser Arbeit der Begriff „Rechenzentrum“ verwendet wird, ist damit das entsprechende Gebäude inklusive aller notwendigen Gewerke, die zur Versorgung der Server benötigt werden oder Einfluss auf diese haben, gemeint. Dies sind neben der IT-Hardware insbesondere die Klimatisierungstechnologien und die Stromversorgung.*

In Abbildung 2.1 wird die typische Aufteilung eines mittelgroßen bis großen Rechenzentrums dargestellt. Diese Kategorisierung beschreibt ein Rechenzentrum mit mehreren tausend Servern und einer elektrischen Anschlussleistung ab 240 kW bis in den Megawattbereich [52]. Die Server in einem solchen Rechenzentrum sind in separaten vollklimatisierten

---

<sup>1</sup>In [30, 52, 106] finden sich mehrere Definitionen und Klassifizierungen

IT-Räumen und dort in Rackreihen untergebracht. In diesen Räumen und ebenfalls in den Racks befindet sich die Netzwerkhardware und die Stromunterversorgung, die für die im Raum befindlichen Server benötigt wird. Die Racks im IT-Raum sind im Beispiel in Abbildung 2.1 nach dem Kühlkonzept einer Kaltgangeinhausung angeordnet. Dieses Prinzip wird in Abschnitt 2.1.4 detailliert erklärt.

Damit wird ein IT-Raum deutlich von anderen Räumen, wie beispielsweise Büros differenziert. In dieser Arbeit erfolgt eine detaillierte Betrachtung von IT-Räumen und deren Techniken, weshalb an dieser Stelle der Begriff IT-Raum ebenfalls klar definiert werden soll:

### **Definition 2.2 IT-Raum**

*Mit dem Begriff „IT-Raum“ sind in dieser Arbeit ausschließlich die Räume gemeint, die dediziert für die Unterbringung von Servern in Rack-Reihen zuständig sind. Diese Räume verfügen über jeweils eigene Umluftkühlgeräte (UKG), die die rauminterne Klimatisierung durchführen, sowie über entsprechende Stromunterverteilungen für die im Raum befindlichen Server. Außerdem befinden sich automatisierte Brandlöschanlagen in diesen Räumen. Der Zutritt durch Mitarbeiter ist meist deutlich reguliert und durch Anwesenheitssensoren, Vereinzelungsanlagen oder ähnliche Systeme überwacht.*

Mittlere und größere Rechenzentren beinhalten meist mehrere IT-Räume. In dieser Arbeit wird ein IT-Raum dann als aktiv bezeichnet, wenn mindestens ein Server in diesem eingeschaltet ist und damit auch die unterstützenden Geräte im IT-Raum eingeschaltet sein müssen. In einem inaktiven IT-Raum müssen alle Server ausgeschaltet sein, sodass auch alle unterstützenden Geräte in diesem IT-Raum ausgeschaltet werden können.

Je nach Konzeption und Aufbau des Rechenzentrums gibt es weitere dedizierte Räume für die unterbrechungsfreie Stromversorgung (USV) oder diese ist ebenfalls mit in den IT-Räumen platziert. Im gezeigten Beispiel gibt es einen Versorgungsraum, in dem sich die USV-Geräte und deren Batterien befinden. Auch diese Geräte müssen gekühlt werden, daher befindet sich auch in diesem Raum ein UKG. Des Weiteren befinden sich häufig auch kleinere Büros im Gebäude. Die Darstellung in Abbildung 2.1 ist aus Gründen der Übersichtlichkeit vereinfacht und kompakt gehalten. In der Realität versorgt beispielsweise ein Versorgungsraum mehrere IT-Räume und es existieren häufig weitere kleine Räume, in denen allgemeine Gebäudetechnik vorhanden ist (Hauptsicherung, Alarmsystem, etc.). Da in dieser Arbeit die IT-Räume und deren Zusammenhänge im Vordergrund stehen, wird auf die weitere Gebäudetechnik nicht detailliert eingegangen.

### **2.1.1 Elektrische Leistungsaufnahme des Rechenzentrums**

Wenn eine energetische Betrachtung und Optimierung eines Rechenzentrums erfolgen soll, dann muss zunächst einmal eine Ist-Analyse über den Anteil der elektrischen Leistungsaufnahme einzelner Geräte und Gewerke Auskunft geben. So können Bereiche mit großem Potential und im Gegensatz solche mit niedrigem Einsparpotential identifiziert werden. Dazu wird in Abbildung 2.2 die anteilige elektrische Leistungsaufnahme der größten Verbraucher im Rechenzentrum verglichen. Es ist deutlich zu erkennen, dass die Server und

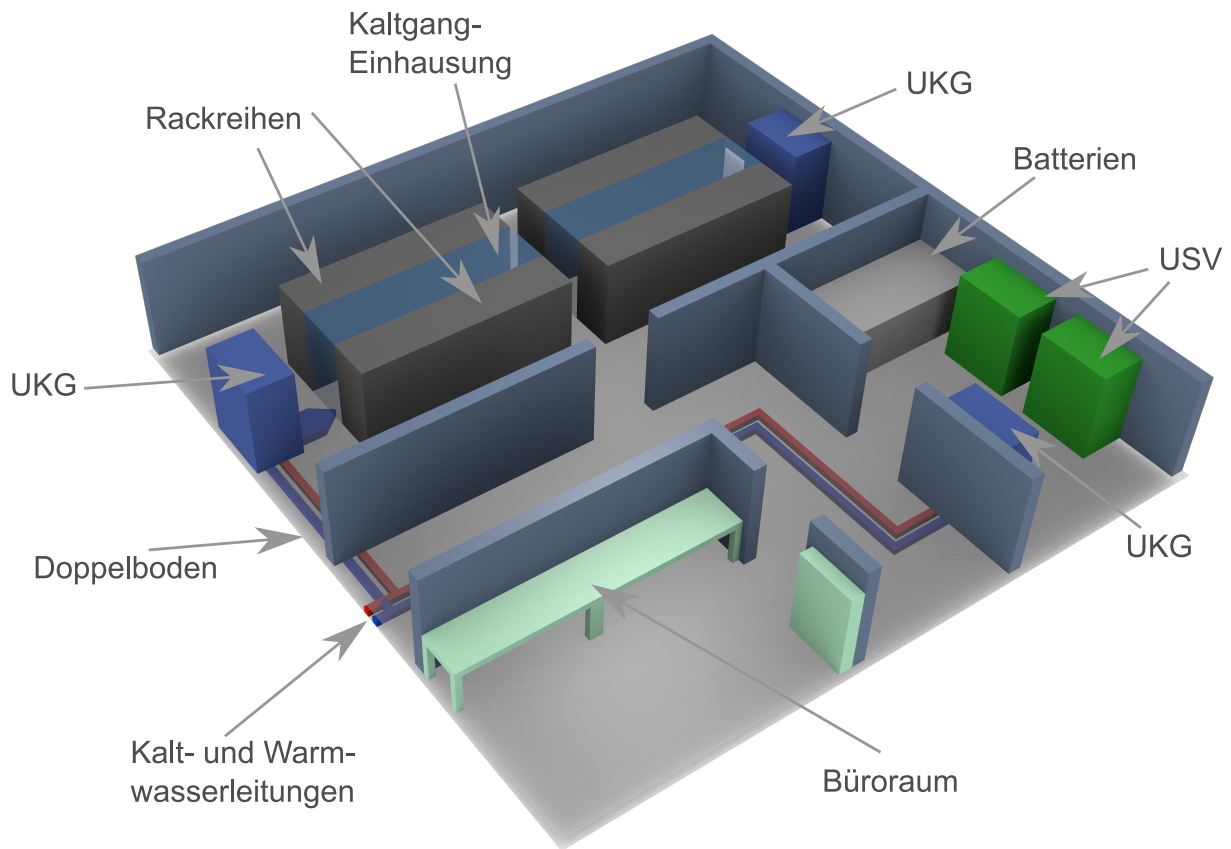


Abbildung 2.1: Vereinfachte Darstellung der typischen Aufteilung eines Rechenzentrums: Im IT-Raum (oben) befinden sich zwei Umluftkühlgeräte, die die kalte Luft durch den Doppelboden in die Kaltgangeinhausung blasen. Die Einhausung verbindet jeweils zwei Rackreihen, in denen sich die Server befinden. Diese saugen die kalte Luft aus dem Kaltgang ein und blasen erwärmte Luft in den Raum. Die beiden Umluftkühlgeräte saugen diese warme Luft wiederum ein und kühlen diese mit Hilfe des Kühlwassers ab. Die Versorgung mit Kühlwasser erfolgt über die im Boden liegenden Wasserleitungen. Im Versorgungsraum (rechts) befinden sich zwei unterbrechungsfreie Stromversorgungsgeräte sowie deren Batterien. Auch diese Geräte müssen mit einem Umluftkühlgerät gekühlt werden. Ein weiterer Raum dient als Bürofläche (unten).

die Klimatisierung die prozentual größten Verbraucher darstellen, gefolgt von der unterbrechungsfreien Stromversorgung. Der relativ geringe Anteil der Storage-Server erklärt sich dadurch, dass einige wenige Storage-Server ausreichend sind, um die Daten der Server, die die Anwendungen ausführen, zu speichern.

Zusätzlich zu den in Abbildung 2.1 erkennbaren Räumen und Geräten befinden sich noch weitere Einrichtungen im Rechenzentrum. Dazu zählen beispielsweise die Beleuchtung,

Alarmsysteme, Schließsysteme und weitere Gebäudetechnik. Diese Geräte werden in Abbildung 2.2 unter „sonstige Infrastruktur“ zusammengefasst. Da diese Geräte keinen direkten Einfluss auf den Betrieb der Server haben, keine lastsensitiven dynamisch regelbaren Systeme darstellen und mit einem Anteil von 4% eine eher geringe elektrische Leistungsaufnahme aufweisen, werden sie im weiteren Verlauf dieser Arbeit als statische Grundleistung betrachtet, ansonsten aber nicht weiter behandelt.

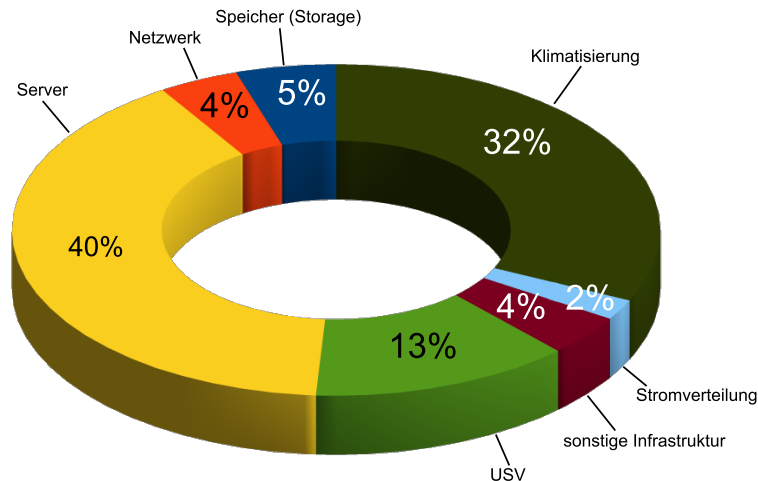


Abbildung 2.2: *Typische durchschnittliche elektrische Leistungsaufnahme nach Gewerken im Rechenzentrum (eigene Darstellung, Daten aus [51])*

Für eine Modellierung des Energiebedarfs eines Rechenzentrums ist wichtig, ob eine Komponente lastadaptiv ist, den eigenen Energiebedarf also bei unterschiedlichen Lastzuständen anpassen kann, oder ob der Energiebedarf annähernd statisch ist.

Die Adaptivität von Servern [56, 60, 113], der Klimatisierung und der Stromversorgung [134] wurde bereits untersucht. Storage-Server sind dagegen kaum regelbar [151]. In [2, 47, 123] wurde gezeigt, dass Netzwerkrouter und Switches zwar grundsätzlich eine leichte Änderung des Energiebedarfs bei unterschiedlicher Auslastung (hier Netzwerkverkehr) zeigen, diese allerdings sehr gering ist (1.55% Variabilität bei maximaler Laständerung). Die Verbesserung der Effizienz von Netzwerkgeräten ist zwar ein aktuelles Forschungsthema [89, 91], spielt aber bei der Modellierung von Rechenzentren nur eine untergeordnete Rolle. Daher werden diese Geräte in dieser Arbeit als statische Verbraucher eingeordnet, was im Kontext energetischer Rechenzentrumsmodellierung auch dem Stand weiterer Forschungsarbeiten entspricht [36, 145].

Damit können die einzelnen Geräte in folgende Klassen unterteilt werden:

- Lastsensitive Geräteklassen
  - Server
  - Klimatisierung
  - Unterbrechungsfreie Stromversorgung

- Nicht lastsensensitive Geräteklassen
  - Netzwerkgeräte
  - Storage-Server
  - Stromverteilung
  - Sonstige Infrastruktur

Die Modellierungen und Simulationen des Rechenzentrums in dieser Arbeit konzentrieren sich auf die drei lastsensitiven Geräteklassen, die nicht lastsensitiven Geräte werden als statische Grundlast des gesamten Rechenzentrums angesehen. Lediglich zur Verbesserung der Genauigkeit der Modellierung werden existierende Modelle zur Abschätzung der elektrischen Leistungsaufnahme von Netzwerkschaltern verwendet, allerdings wird auch diese Leistung statisch betrachtet.

Wiederum unter Berücksichtigung der Erhebung in [51] ergibt sich damit ein statischer Grundbedarf von ca. 15% der Anschlussleistung des Rechenzentrums, während der last-sensitive und damit dynamisch regelbare Anteil bei ca. 85% liegt. Im weiteren Verlauf der Arbeit wird zudem analysiert, inwieweit die lastsensitiven Geräte optimiert werden können und ob dies für alle Geräte aus dieser Klasse gilt.

Für die Berechnung der elektrischen Leistungsaufnahme des Rechenzentrums ergeben sich nun folgende Grundlagen: Die Grundleistung des Rechenzentrums wird, wie oben beschrieben, als statischer Wert angenommen und damit zeitunabhängig mit  $P_{base}$  angegeben. Die elektrische Leistungsaufnahme der Server wird mit  $P_{server}(t)$  angegeben, der der Klimatisierung mit  $P_{cooling}(t)$ . Dieser ergibt sich aus der jeweiligen konkreten Kühltechnologie und deren Dimensionierung, sowie der aufzubringenden Kühlleistung für die aktuelle Server-Auslastung und die damit entstehende Leistungsaufnahme. Ähnliches gilt für die Leistungsaufnahme der USV<sup>2</sup>  $P_{ups}(t)$ . Damit ergibt sich folgende Formel für die elektrische Leistungsaufnahme  $P_{dc}(t)$  des gesamten Rechenzentrums zum Zeitpunkt  $t$ :

$$P_{dc}(t) = P_{base} + P_{server}(t) + P_{ups}(t) + P_{cooling}(t) \quad (2.1)$$

Die energetischen Zusammenhänge dieser einzelnen Geräte sowie die Möglichkeiten der Regelung werden in Kapitel 4.1.1 analysiert und in Kapitel 5 werden die entsprechenden Modelle beschrieben. Alle weiteren Berechnungen beziehen sich dabei auf die zeitabhängigen, regelbaren Komponenten. Aus Gründen der Übersichtlichkeit wird in folgenden Gleichungen auf den Hinweis der Zeitabhängigkeit verzichtet.

## 2.1.2 IT-Hardware

Unter dem Begriff IT-Hardware werden die Geräte zusammengefasst, die ausschließlich für die Abarbeitung der Aufgaben zuständig sind. Dies sind insbesondere die Server, auf denen die Anwendungen und Dienste ausgeführt werden. Neben diesen Geräten existieren Storage-Server, die hauptsächlich aus Festplatten-Arrays bestehen und zur Datenaufnahme

---

<sup>2</sup>Die Variablenbezeichnung „UPS“ entstammt der englischen Bezeichnung: Uninterruptible Power Supply

oder zu Backups verwendet werden. In dieser Arbeit liegt der Fokus auf den „normalen“ Servern, also den Geräten, die die jeweiligen Anwendungen ausführen, da hier einerseits das größte Potential liegt und andererseits die elektrische Leistungsaufnahme lastsensitiv ist.

Auch der Begriff der Server soll an dieser Stelle für die weitere Verwendung in dieser Arbeit genau definiert werden:

### **Definition 2.3 Server**

*Mit dem Begriff „Server“ ist in dieser Arbeit immer der physikalische Server, also die Hardware gemeint. In vielen Texten zum Thema Rechenzentrum ist mit dem Begriff auch ein virtueller Server gemeint (siehe Unterkapitel 2.3), also beispielsweise eine Server-Anwendung (z.B. Webserver, Terminalserver, ...), die in einer virtuellen Maschine ausgeführt wird. Diese virtuellen Server werden in dieser Arbeit immer als „virtuelle Maschinen“ oder kurz „VM“ bezeichnet.*

Auf dem Markt existieren verschiedene Server-Typen, die sich in Kapazität, Größe und Energiebedarf teils deutlich unterscheiden. Neben den gängigen Einschubservern für Racks, deren Größe in Höheneinheiten gemessen wird, gibt es auch vollständige Server-Schränke (Mainframes) oder kompakte und austauschbare Blade-Server, die in eine dafür vorgesehene Chassis gesteckt werden. In dieser Arbeit werden primär Einschubserver behandelt, allerdings sind die analysierten Zusammenhänge und das entwickelte Verfahren auch auf andere Modelle übertragbar, wenn diese eine energetische und lastensitive Regelbarkeit aufweisen. Dies ist beispielsweise auch für Blade-Server gegeben [33, 133] und kann mit ähnlichen Modellen, wie sie auch in dieser Arbeit Verwendung finden, erreicht werden.

### **Netzwerkgeräte**

Zu der Kategorie der IT-Hardware werden ebenfalls die Netzwerkgeräte wie Switches und Router gezählt. Diese Geräte befinden sich meist ebenfalls mit in den Racks in den IT-Räumen. Wie bereits begründet, werden die Netzwerkgeräte als statische Verbraucher betrachtet, allerdings zählt auch die elektrische Leistungsaufnahme dieser Geräte mit zur Leistung der IT-Räume, was für die Auslegung der Kühlkapazität von Relevanz ist.

### **2.1.3 Stromversorgungsarchitektur**

Die Stromversorgungsarchitektur in Rechenzentren umfasst sowohl die Stromverteilung und Unterverteilung als auch die Infrastruktur zur Notstromversorgung und deren Redundanz. Eine typische Architektur ist in Abbildung 2.3 dargestellt. Es muss angemerkt werden, dass neben dieser dargestellten Architektur auch weitere, teils noch in Entwicklung befindliche Architekturen existieren, die beispielsweise USV-Geräte in die Racks [44] oder sogar in die Server [45] integrieren. Allerdings sind diese Konzepte noch nicht am breiten Markt angekommen und aktuell nur Gegenstand der Forschung.

Für Ausfälle des Netzstroms sind meist zwei Ebenen der Notstromversorgung vorhanden: Für kurze Ausfälle von maximal einigen Minuten werden Batterieaggregate vorgehalten,

die durch unterbrechungsfreie Stromversorgungsgeräte in einem solchen Fall hinzu geschaltet werden. In der Regel kann der Server-Betrieb dadurch ohne nennenswerte Beeinflussung gewährleistet werden, bis der Netzstrom wieder verfügbar ist. Da diese USV-Geräte noch weitere Aufgaben besitzen und zudem Einfluss auf die Energieeffizienz des Rechenzentrums haben, wird das Thema USV im folgenden Unterabschnitt *Unterbrechungsfreie Stromversorgung* detaillierter behandelt.

Bei längeren Ausfällen reicht die Kapazität der Batterien nicht aus, sodass hier auf Netzeratzanlagen zurückgegriffen werden muss. Diese sind bei Rechenzentren normalerweise große Dieselaggregate, die sich außerhalb des eigentlichen Rechenzentrumskomplexes befinden und in einem Zeitrahmen unterhalb einer Minute angefahren werden können [45]. Durch sie ist ein weiterer Rechenzentrumsbetrieb von teils mehreren Stunden bis wenigen Tagen möglich.

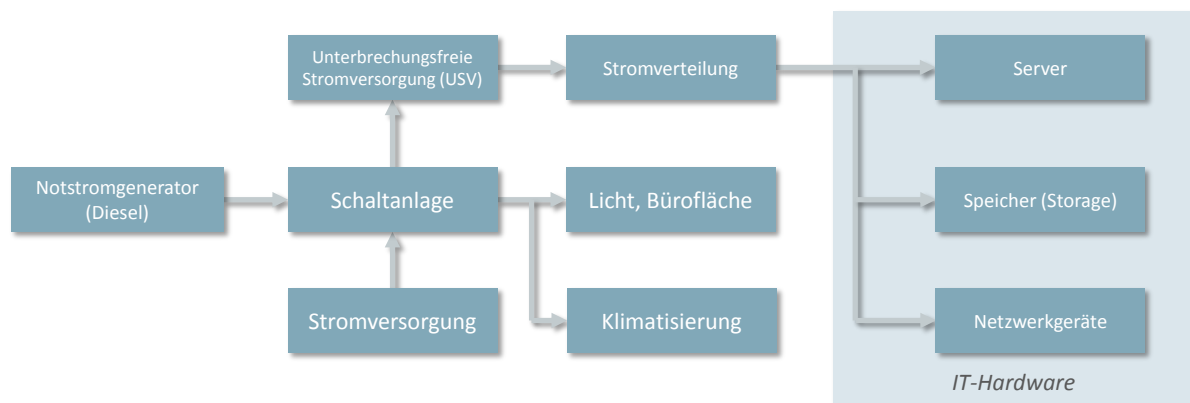


Abbildung 2.3: *Typische Stromversorgungsarchitektur im Rechenzentrum (eigene Darstellung, Daten aus [38])*

In der Abbildung 2.3 ist zu erkennen, dass die USV-Geräte die elektrische Versorgung der IT-Hardware übernehmen. Die energetische Dimensionierung der USV hängt also primär von den zu versorgenden Verbrauchern ab. Die Klimatisierung wird im Fall eines Stromausfalls durch den Notstromgenerator versorgt. Bis zum Start des Notstromgenerators erfolgt die Klimatisierung ausschließlich über die Server-Lüfter. Durch diese können mehrere Minuten eines Stromausfalls hinsichtlich der Server-Klimatisierung überbrückt werden.

### Unterbrechungsfreie Stromversorgung

Die USV-Geräte werden grundsätzlich zur Überbrückung von Stromausfällen genutzt, in der Regel können so mehrere Minuten eines Ausfalls kompensiert werden [3, 134]. In modernen Rechenzentren werden aber mittlerweile weitere Anforderungen an USV-Geräte gestellt: Schwankungen im Stromnetz, bei Spannung oder Frequenz, könnten der sensiblen Rechenzentrumshardware schaden und sollten daher durch die USV ausgeglichen werden. Auf dem Markt existieren verschiedene USV-Topologien, von denen nur einige diese Anfor-

derungen erfüllen. Alle genannten sind in der Lage, einen Stromausfall für eine bestimmte Zeit zu überbrücken.

Tabelle 2.1: USV-Topologien und deren jeweilige Funktionalität nach [134]

| Topologie            | Funktion   | Effizienz           | Spannungsschwankungen ausgleichen | Frequenzschwankungen ausgleichen     |
|----------------------|--|---------------------|-----------------------------------|--------------------------------------|
| Offline / Standby    | Einfache Schaltung zwischen Batterie und Netzstrom, dadurch keine Verlustleistung vorhanden  | -                   | Nein                              | Nein                                 |
| Inline / Line-active | Stromrichter und Batterie parallel zum Netzstrom, dadurch Spannungsausgleich möglich   | Auslastungsabhängig | Ja                                | Nicht oder nur eingeschränkt möglich |
| Double Conversion    | Doppelte Wandlung: Wechselstrom zu Gleichstrom, dann wieder zurück, wobei der resultierende Strom geglättet ist; vollständige elektrische Isolation der angeschlossenen Geräte vom Stromnetz möglich | Auslastungsabhängig | Ja                                | Ja                                   |
| Delta Conversion [8] | Patentiertes System, „Delta-Konverter“ steuert Sekundärspannung des Haupttransformators, dadurch direkte Regulierung von Spannung und Frequenz   | Auslastungsabhängig | Ja                                | Ja                                   |

In Tabelle 2.1 werden einige der am meisten verwendeten USV-Topologien und deren funktionale Möglichkeiten aufgelistet. Alle Anforderungen an moderne Rechenzentren erfüllen nur USV, die „Double-Conversion“ oder „Delta-Conversion“ nutzen. Nach Aussage von [44, 45, 134] ist die Variante „Double-Conversion“ in Rechenzentren am weitesten verbreitet, kleine Rechenzentren oder einzelne Workstations nutzen eher Offline-Varianten.

In dem in dieser Arbeit untersuchten Rechenzentrum kommen „Double-Conversion“-USV-Geräte zum Einsatz, diese werden dementsprechend in Kapitel 5.2 modelliert. Die Verlustleistungen, die bei dieser Variante auftreten, entstehen hauptsächlich durch die zweifache Wandlung (Wechselstrom  $\rightarrow$  Gleichstrom  $\rightarrow$  Wechselstrom) [44, 134]. Daher ist hier die



Betrachtung des Wirkungsgrades der USV von Bedeutung. Dieser Wirkungsgrad hängt von der anliegenden elektrischen Auslastung ab [32, 111, 134]. Für einen hohen Wirkungsgrad sollten USV-Geräte in der Regel eine Auslastung nahe an ihrer Kapazitätsgrenze aufweisen. Details zur Effizienz von USV-Geräten befinden sich in Kapitel 5.2.

Untersuchungen haben jedoch gezeigt, dass USV-Geräte meist nur geringfügig ausgelastet sind und somit in einem niedrigen Wirkungsgrad betrieben werden. Beispielsweise zeigten 80% der untersuchten USV eine Auslastung von unter 50% [134], der gemessene Auslastungsdurchschnitt lag bei 37,8%. Die in der Studie genannten Gründe dafür sind:

- *Redundanz*: Aufgrund von Redundanzanforderungen an das Rechenzentrum teilen sich mehrere USV-Anlagen die gesamte Auslastung. Damit kann eine einzelne Anlage nicht über eine relativ niedrige Auslastung hinauskommen.
- *Überdimensionierung*: Die USV-Geräte wurden für eine Auslastung angeschafft, die aufgrund der tatsächlich auftretenden Anwendungsauslastung im Rechenzentrum nicht erreicht werden kann.

Insgesamt kann also festgestellt werden, dass die USV zwar einen signifikanten Anteil des Energiebedarfs eines Rechenzentrums haben (13% nach Abbildung 2.2), der Einfluss auf die USV-Effizienz aber durch Entscheidungen bei der Planung maßgeblich bestimmt wird.

### 2.1.4 Klimatisierung

Die Klimatisierung bzw. Kühlung in Rechenzentren hat die Aufgabe, die Abwärme, die durch den Betrieb der Server und aller weiteren elektrischen Geräte im Rechenzentrum entsteht, abzuführen. Die Kühlung der Server in den IT-Räumen ist unter anderem aufgrund der steigenden Energiedichte in Servern und Racks [16, 100] eine immer größer werdende Herausforderung. So ist die Energiedichte von Servern seit dem Jahr 2000 von ca.  $1 \frac{kW}{m^2}$  auf ca.  $6 \frac{kW}{m^2}$  im Jahr 2014 gestiegen [120]. Da die elektrische Energie, die ein Server aufnimmt, in gleichem Maße auch in Form von Wärme wieder abtransportiert werden muss, ist bei steigender Leistungsaufnahme pro Server die Kühlung entsprechend anzupassen. Somit ist es nicht überraschend, dass nach der IT-Hardware die Klimatisierung den größten Energiebedarf besitzt. Gleichzeitig liegt hier auch ein großes Einsparpotential, was je nach eingesetzter Technologie und Bauart des Rechenzentrums auf 30-60% geschätzt wird [16].

#### Kühlkette im Rechenzentrum

In kleineren Rechenzentren oder einzelnen Server-Räumen, die sich inmitten von Büros befinden, werden zur Kühlung häufig Raum-basierte Klimageräte eingesetzt, die die vollständige in Abbildung 2.4 gezeigte Kühlkette in einem einzigen Gerät vereinen. In größeren Rechenzentren ist dies nicht möglich, da diese Geräte nicht genug Kühlkapazität bereitstellen oder diese Art nicht effizient wäre. Hier wird die Klimatisierung mittels einer zentralen Klimakette erreicht, wobei die Bereitstellung von kalter Luft in den IT-Räumen durch

Umluftkühlgeräte erreicht wird. Diese saugen die warme Luft, die durch die Abwärme der Server im IT-Raum entsteht, ein und kühlen diese mit Hilfe des Kühlwassers wieder ab. Das dazu benötigte Kühlwasser wird wiederum durch den zentralen Kaltwassersatz erzeugt, der hierfür meist Kältemaschinen oder - je nach Außentemperatur - Freikühlung verwendet. Auf dieses Prinzip wird im folgenden Abschnitt *Freikühlung* gesondert eingegangen. Zum Transport des Kühlwassers vom Kaltwassersatz zu den IT-Räumen werden Pumpen eingesetzt. Das Prinzip dieser Kühlkette ist in Abbildung 2.4 dargestellt.

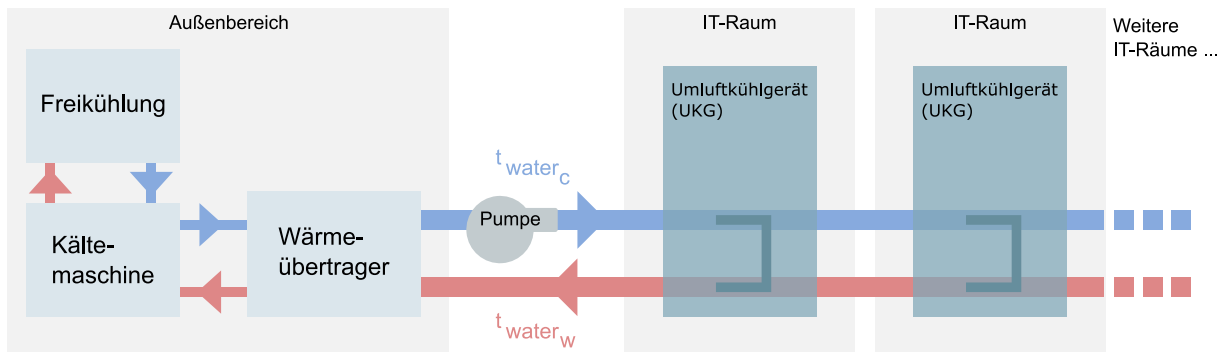


Abbildung 2.4: Kühlkette im Rechenzentrum, aufgeteilt in den Außenbereich mit der Kaltwasserversorgung und den beispielhaften Umluftkühlgeräten in den IT-Räumen

Das Kühlwasser, das meist aus einem Wasser-Glykol-Gemisch besteht, verlässt den Kaltwassersatz mit der Temperatur  $t_{water_c}$ . Es wird in dieser Arbeit angenommen, dass es auch mit dieser Temperatur in die Umluftkühlgeräte eintritt [32]. Dies ist in der Realität nicht ganz der Fall, allerdings ist der Erwärmungsgrad durch Isolationen der Rohrleitungen sehr gering. Das Kühlwasser wird in den Umluftkühlgeräten erwärmt, indem die Wärmeenergie der Warmluft auf das Kühlwasser übertragen wird und kommt dann mit der Temperatur  $t_{water_w}$  in den Rücklauf zum Kaltwassersatz. Die Energie, die benötigt wird, um das Kühlwasser von  $t_{water_w}$  auf  $t_{water_c}$  abzukühlen, muss durch die Kältemaschine oder durch die Freikühlungseinheit bereitgestellt werden. Bei Nutzung der Kältemaschinen wird hier der Großteil der Energie für die Kühlung benötigt [73].

Der Kühlwasserkreislauf wird für alle IT-Räume und Umluftkühlgeräte im Rechenzentrum verwendet. Damit sind die Parameter des Kreislaufs, wie beispielsweise die Kaltwassertemperatur und Flussrate, entscheidend für die Kühlleistung aller Umluftkühlgeräte im Rechenzentrum. Da die Flussrate des Kühlwassers nur geringen Einfluss auf den Energiebedarf der Geräte hat und außerdem nicht beliebig regelbar ist, ist sie in der Regel konstant [103]. Eine Regelung sollte über die Anpassung der Kühlwassertemperatur erfolgen.

In Rechenzentren werden meist Kompressionskältemaschinen verwendet [103], die mittels eines Kompressors ein Gas komprimieren und dann innerhalb eines Wärmeübertragers entspannen. Bei diesem Vorgang wird dem vorbeiströmenden (warmen) Kühlwasser die Wärme entzogen. Dieser Prozess benötigt den Großteil der Energie in der Kühlkette. Der Energiebedarf hängt dabei maßgeblich von der benötigten Temperatur  $t_{water_c}$  ab und kann

mittels des Coefficient of Performance (COP) [87] beschrieben werden. Je höher die Temperatur  $t_{water_c}$  ausfallen darf, umso weniger Energie wird für das Erreichen dieser Temperatur benötigt.

## Freikühlung

An Rechenzentrumsstandorten, bei denen die Außentemperatur im Jahr häufig relativ niedrig ist, kann Freikühlung verwendet werden. Hier gibt es grundsätzlich zwei verschiedene Möglichkeiten [16]:

- *Nutzung direkter Außenluft*: Die Außenluft wird direkt als Kaltluft im IT-Raum verwendet [152]. Diese Freikühlung kann dann angewandt werden, wenn die Außentemperatur unter der geforderten Kaltlufttemperatur im IT-Raum liegt. Es ist dann keine Kühlkette wie in Abbildung 2.4 gezeigt erforderlich. Dieses Verfahren ist allerdings problembehaftet. Die Luft muss aufwendig gefiltert werden, da Staub und anderer Schmutz entfernt werden muss [16]. Ansonsten würde dies die Server-Lüfter und Luftkanäle beeinträchtigen und zu Defekten führen. Auch muss ein relativ konstantes Level der Luftfeuchtigkeit gehalten werden, was bei direkter Außenluftkühlung nur schwer erreichbar ist. Daher findet diese Variante eher selten Anwendung, ist aber Untersuchungsgegenstand aktueller Forschung, da die Anwendung energetische Einsparungen ermöglichen kann [145, 152].
- *Nutzung kalter Außenluft zur Kaltwassererzeugung*: Diese gängige Methode behält das Raumluftklima bei, es wird also die gleiche Luft nach der Erwärmung durch die Geräte im IT-Raum wieder abgekühlt. Die in Abbildung 2.4 gezeigte Kühlkette bleibt erhalten, allerdings wird die Abkühlung des Kühlwassers anstatt durch die Kältemaschinen durch die Freikühlungseinheit vorgenommen. Diese besteht in aller Regel aus einem Kühlturm [103].

Die Nutzung von kalter Außenluft zur Kaltwassererzeugung ist die gängige und empfohlene Form der Freikühlung, vor allem für größere Rechenzentren [16]. Die möglichen Energieeinsparungen hängen hier direkt von der Kapazität des Kühlturms, der Außentemperatur und der benötigten Temperatur  $t_{water_c}$  ab [16, 103]. Es gibt allerdings Empfehlungen und Hochrechnungen, die die Möglichkeiten der Freikühlung in Abhängigkeit der Außentemperatur untersuchen. So nennt *the green grid* [50] in Zusammenhang mit den ASHRAE-Temperaturempfehlungen<sup>3</sup> [9] eine Außentemperatur von 35,0°C, bei der immer noch von Freikühlung profitiert werden kann. Dies gilt allerdings nur unter der Annahme, dass die höchsten empfohlenen Raumtemperaturen im Rechenzentrum angewandt werden.

Damit eine vollständige Freikühlung verwendet werden kann, nennt [105] eine ungefähre Außentemperatur, die 2-4 Kelvin unter der benötigten Temperatur  $t_{water_c}$  liegt<sup>4</sup>.

<sup>3</sup>Ursprünglich wurde die Organisation *American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE)* genannt, mittlerweile nur noch *ASHRAE*, siehe auch <https://www.ashrae.org/>

<sup>4</sup>Diese Abschätzung unterliegt mehreren Annahmen, vor allem der, dass das Außentemperaturniveau im Bereich von maximal ca. 20,0°C liegt. Auch sollte die relative Luftfeuchtigkeit umso niedriger sein, je wärmer die Außentemperatur ist

### IT-Raumklimatisierung

Die IT-Raumklimatisierung ist dafür verantwortlich, die durch die Geräte im IT-Raum entstehende Abwärme abzutransportieren. Hier wird davon ausgegangen, dass alle elektrische Leistung  $P_{el}$ , die in den IT-Raum hineingeht, in Wärmeleistung  $\dot{Q}_{therm}$  gewandelt wird und somit abtransportiert werden muss. Damit gilt dann:

$$P_{el} \approx \dot{Q}_{therm} \quad (2.2)$$

In weiteren Teilen der Arbeit wird diese Gesetzmäßigkeit als Voraussetzung angenommen. Die Klimatisierung muss nun dafür sorgen, dass ausreichend Kälteleistung bereitgestellt wird. An dieser Stelle wird von einer Klimakette wie in Abbildung 2.4 dargestellt ausgegangen. Demnach befinden sich Umluftkühlgeräte in den IT-Räumen, die diese Kälteleistung bereitstellen müssen, indem sie die Warmluft einsaugen, mittels des Kühlwassers abkühlen und diese kalte Luft wieder in den IT-Raum abgeben.

An dieser Stelle können unterschiedliche Konzepte zur Verteilung der Kaltluft im IT-Raum angewandt werden. Im einfachsten Fall wird die kalte Luft in den IT-Raum geblasen. Dies ist jedoch gleichzeitig am ineffizientesten, da dadurch eine Vermischung der kalten Luft mit der im Raum vorhandenen warmen Luft stattfindet. Die Temperatur der Luft, die dann tatsächlich bei den Servern ankommt, steigt damit signifikant an, insgesamt sinkt damit die Kühlleistung. Damit ein solcher Effekt vermieden wird, wurde das Konzept der Einhausung entwickelt. Dabei wird der IT-Raum in zwei unterschiedliche Bereiche getrennt: den Bereich der kalten Luft und den der warmen Abluft. Die Einhausung besteht aus einem Gang, der die einzelnen Racks umschließt. Das Prinzip ist am Beispiel einer Kaltgangeinhausung bildlich in Abbildung 2.5 dargestellt. Bei dieser Art der Einhausung strömt die kalte Luft durch Leitungen aus dem Doppelboden in den Gang zwischen jeweils zwei Rack-Reihen. Die Server können nun die kalte Luft einsaugen und geben die warme Luft in den offenen IT-Raum ab, wo diese sich im oberen Bereich sammelt und dann durch die UKG wieder eingesaugt wird. Dieses Konzept führt zu einer signifikanten Verbesserung der Kühlleistungseffizienz [98, 99] und ist in modernen und vor allem mittleren und größeren Rechenzentren mittlerweile weit verbreitet.

Neben der Kaltgangeinhausung existiert auch das gegenteilige Modell der Warmgangeinhausung. Hier wird dementsprechend die kalte Luft in den IT-Raum abgegeben und die warme Abluft der Server in einem Gang gesammelt und durch UKG abgeführt. Sowohl Kaltgang- als auch Warmgangeinhausung führen zu signifikanten Effizienzsteigerungen der Klimatisierung. Welches System eingesetzt werden sollte hängt von verschiedenen Parametern des Rechenzentrums und der einzusetzenden Server ab und kann nicht pauschal beantwortet werden[92].

In der Regel befinden sich mehrere UKG in einem IT-Raum, damit diese einerseits verschiedene Positionen des IT-Raumes kühlen und sich damit die Arbeit teilen können. Andererseits soll eine Redundanz erreicht werden, damit bei einem Ausfall eines UKG die Kühlung durch die verbleibenden Geräte garantiert werden kann. Im Rahmen dieser Arbeit wurde ein Rechenzentrum exemplarisch untersucht. Hier befinden sich beispielsweise immer zwei

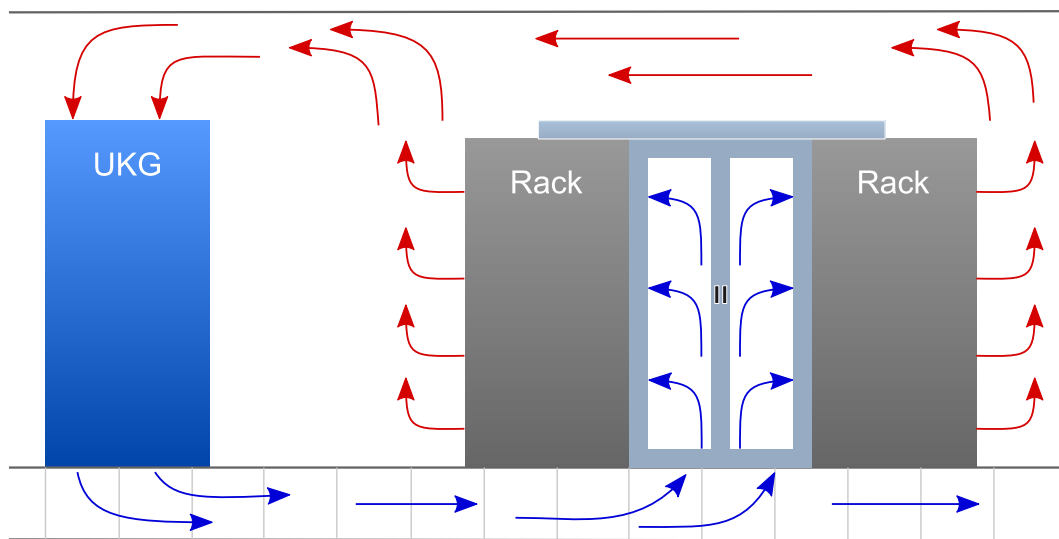


Abbildung 2.5: Schematische Darstellung einer Kaltgangeinhausung: Die kalte Luft verlässt das Umluftkühlgerät in den Doppelboden und tritt in den Kaltgang ein. Hier saugen die Server die Luft ein. Diese nimmt die Wärmeleistung der Server auf und erwärmt sich dadurch. Die warme Luft verlässt die Server und Racks auf der Rückseite und steigt in den oberen Bereich des IT-Raumes. Die Umluftkühlgeräte saugen nun diese Luft wieder ein und kühlen sie ab.

UKG in jedem IT-Raum. Aufgrund der beschriebenen Redundanz muss allerdings jedes einzelne Gerät in der Lage sein, die Kühlung für den gesamten IT-Raum zu übernehmen. Die Server in den IT-Räumen saugen die kalte Luft mit der Temperatur  $t_{air_c}$  ein und blasen sie mit der Temperatur  $t_{air_w}$  wieder heraus. Diese Temperaturen sowie deren Differenz sollen nach Empfehlungen in einem definierten Bereich liegen. So gibt die US-amerikanische Organisation ASHRAE folgende Werte an:

Die Ansaugtemperatur  $t_{air_c}$  soll nach der Empfehlung zwischen  $18^\circ\text{C}$  und  $27^\circ\text{C}$  betragen, erlaubt sind allerdings  $15^\circ\text{C}$  bis  $32^\circ\text{C}$  [9]. Für die Temperaturspreizung, also die Differenz  $t_{air_w} - t_{air_c}$ , wird ebenfalls eine Empfehlung ausgesprochen. Hier sollten im Fall von Einschubservern 15 Kelvin und im Fall von Blade-Servern 28 Kelvin eingehalten werden [9, 16]. Grundsätzlich gilt für diese Empfehlungen, dass eine möglichst hohe Temperatur  $t_{air_c}$  innerhalb des Rahmens der Empfehlung gewählt werden sollte, da dies die Leistungsaufnahme der Kältemaschinen senkt [9, 16].

### Mögliche Problemfälle der IT-Raumklimatisierung

Trotz mittlerweile ausgereifter Konzepte zur IT-Raumklimatisierung kann es zu Problemfällen kommen. Im Folgenden sind einige der am häufigsten auftretenden Problemfälle dargestellt und es wird kurz erläutert, welche Gegenmaßnahmen ergriffen werden können:

- *Ineffiziente Klimatisierung durch Verteilung*: IT-Räume sind meistens geometrisch gleichmäßig aufgebaut und die Umluftkühlgeräte und Rackreihen sollten sich ebenfalls abstandsgleich angeordnet im IT-Raum befinden. Wird nun jedoch die Auslastung auf die Server in nur einer Hälfte des IT-Raums konzentriert, kann dies zu ungleichmäßigen Luftströmungen und Wärmeverteilungen führen. Es kann der Fall auftreten, dass ein UKG im IT-Raum dadurch eine große Wärmemenge abführen muss, während ein anderes im selben IT-Raum nur geringfügig erwärmte Luft einsaugt. Dies führt zu einer ineffizienten Kühlung und sollte daher vermieden werden. Durch eine gleichmäßige Verteilung der Auslastung auf die Server im IT-Raum ist ein solcher Problemfall auch während des operativen Betriebs vermeidbar.
- *Hot-Spots*: Ein Hot-Spot ist die Extremform des oben beschriebenen Phänomens der ungleichmäßigen Wärmeverteilung. Hier konzentriert sich die Wärme auf einen kleinen Bereich, beispielsweise auf wenige Server oder benachbarte Racks. Meist liegt dies an einer starken bzw. stark angestiegenen Auslastung der betroffenen Server. Diese müssen in einem solchen Fall umgehend entlastet werden, damit es nicht zu automatischen Notabschaltungen oder Hardware-Defekten kommt.
- *Rezirkulation oder „Luftkurzschluss“*: Wenn eine Einhausung nicht korrekt abgedichtet wird, kann es dazu kommen, dass - im Beispiel der Kaltgangeinhausung - die kalte Luft direkt aus dem Kaltgang austritt, ohne durch die Server zu strömen und diese abzukühlen [19, 107]. Dadurch geht Kälteleistung in Form der kalten Luft „verloren“ und die Umluftkühlgeräte im IT-Raum saugen ein Gemisch aus der warmen Abluft der Server und der ausgeströmten Kaltluft ein. Es sinkt die Effizienz der Klimatisierung im IT-Raum, die „verlorene“ Kaltluft muss durch zusätzliche Kälteleistung kompensiert werden. Die Behebung dieses Problemfalls sollte durch bauliche Maßnahmen erfolgen (Abdichtungen, beispielsweise durch Blindleisten an den Positionen der Racks, an denen gerade keine Server eingebaut sind). Eine Behebung dieses Problemfalles nur durch die Verteilung der Auslastung ist schwierig, allerdings gibt es Forschungsarbeiten, die sich mit der Analyse und Reduzierung der Rezirkulation der kalten Luft durch Maßnahmen im Betrieb auseinandersetzen [95, 131].

Die ungünstige Verteilung der Wärmeleistung im IT-Raum und die auftretenden Hot-Spots sind Problemfälle, die in dieser Arbeit weiter aufgegriffen werden, da sie während des Betriebs durch Umverteilung der Auslastung auf andere Server oder IT-Räume gelöst werden können.

### 2.1.5 Netzwerkarchitektur

Die Netzwerkgeräte im Rechenzentrum (Switches, Router) sorgen für die Erreichbarkeit der Server und sind damit als kritische Geräte anzusehen. Daher werden diese meist redundant ausgelegt und auch die Topologie des Netzwerks beinhaltet mehrere redundante Routen. Aufgrund der Flexibilität und Skalierbarkeit kommt meist die VL2-Architektur [37, 46] zur Anwendung (siehe Abbildung 2.6). Diese besitzt unterschiedliche Ebenen mit

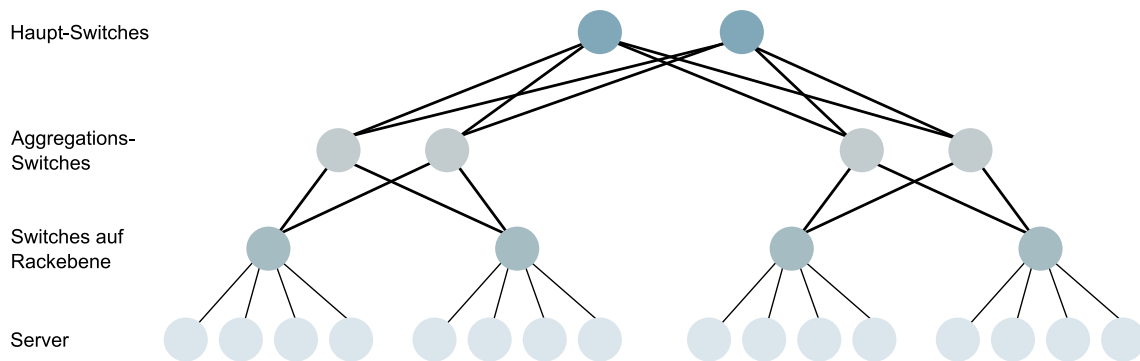


Abbildung 2.6: Die VL2-Netzwerktopologie (eigene Darstellung, Daten aus [37, 46])

jeweils unterschiedlichen Bandbreiten. Zusätzlich werden die Server in parallelen Netzen verknüpft. Typischerweise sind dies drei [5]:

1. *Datenverkehrsnetz:* Das Netzwerk, über das der reguläre Datenverkehr der Anwendungen abgewickelt wird.
2. *Backup-Netz:* Über dieses Netzwerk werden die regelmäßigen Backups durchgeführt.
3. *Management-Netz:* Netzwerk, über das Management und Wartung durchgeführt werden. Über dieses Netzwerk werden auch Migrationen von virtuellen Maschinen durchgeführt, sodass der Netzwerkverkehr dieser nicht den Anwendungsverkehr beeinflussen kann.

Aufgrund der physikalischen Trennung dieser Netzwerke treten keine gegenseitigen Störungen auf. Vor allem sorgt dies dafür, dass Migrationen und Live-Migrationen der virtuellen Maschinen durchgeführt werden können, ohne den Datenverkehr der Anwendungen zu verlangsamen. Dies ist eine wesentliche Voraussetzung zur einfachen Planbarkeit von Migrationen. Wären die Netze nicht getrennt, müsste der Einfluss einer VM-Migration auf den Datenverkehr vor jeder Migration analysiert und bei einer zu großen zu erwartenden Verzögerung die Migration ausgesetzt werden.

## 2.2 Rechenzentrumsbetrieb

Neben der technischen Betrachtung von Rechenzentren ist die Art der Betriebsform ein wichtiger Aspekt, da dieser häufig vertragliche Bestimmungen definiert, die im Betrieb eingehalten werden müssen. Diese Bedingungen können weiteren Einfluss auf mögliche energetische Optimierungen, den Einsatz von Virtualisierung und den gesamten Betriebszustand haben.

Nach einer Studie über die Nutzung von Rechenzentren in Deutschland [106] werden ca. die Hälfte aller Rechenzentren für mehrere Zwecke verwendet, während die andere Hälfte einem exklusiven Geschäftsmodell zugeordnet werden können. Diese speziellen Geschäftsmodelle sind:

- *Eigenbetrieb*: Der Betreiber unterhält das Gebäude und die Server für die eigenen, internen Anwendungen. Damit hat der Rechenzentrumsbetreiber auch die Möglichkeit der Realisierung von Virtualisierung und des Verfassens von vertraglichen Bedingungen und Betriebsparametern.
- *Hosting*: Der Rechenzentrumsbetreiber stellt Kunden fertige Anwendungen oder virtualisierte Systeme zur Verfügung, wobei sich die vertraglichen Bedingungen primär auf die Verfügbarkeit der Anwendung und Support beziehen. Der Rechenzentrumsbetreiber hat direkten Einfluss auf die eingesetzte Hardware und kann damit Optimierungen und Konsolidierungsstrategien umsetzen.
- *Colocation-Betrieb*: Der Rechenzentrumsbetreiber stellt lediglich das Gebäude und die Infrastruktur (Stromversorgung, Klimatisierung, Netzwerktechnik) zur Verfügung, während die Kunden ihre eigene Server-Hardware in den IT-Räumen platzieren. Abrechnungsmodelle reichen hier von der Anzahl der genutzten Höheneinheiten über Energie- und Übertragungskostenmodelle. Der Rechenzentrumsbetreiber hat keinen Einfluss auf die Server-Systeme und kann weder Virtualisierung noch weiterführende Optimierungen vornehmen. Der Betreiber verpflichtet sich vertraglich, Betriebsparameter wie beispielsweise IT-Raumtemperaturen einzuhalten.
- *High Performance Computing (HPC)*: Dieses Themenfeld umfasst aufwendige Berechnungen, wie beispielsweise wissenschaftliche Simulationen aus den Bereichen der Physik, Biologie oder Meteorologie. Die Berechnung der Ergebnisse benötigt in der Regel einige Zeit und zur Beschleunigung wird meist auf Parallelität gesetzt. Die dafür genutzte Hardware wird speziell für diese Anwendungen reserviert. Dabei kann diese Hardware von „normalen“ Servern bis hin zu speziell angefertigten Systemen reichen. Energetische Optimierungen und Konsolidierung können nur unter gesonderten Bedingungen durchgeführt werden [139].

Die Untersuchungen, die in dieser Arbeit durchgeführt werden, beziehen sich zwar allgemein auf Rechenzentren, das Optimierungsverfahren das daraus entsteht kann aber nur unter bestimmten Bedingungen angewandt werden. Dazu gehört grundsätzliche Virtualisierung sowie der Zugriff auf die Virtualisierungsschicht durch den Rechenzentrumsbetreiber. Damit ist das Optimierungsverfahren primär für die ersten beiden der genannten Betreibergruppen von Interesse, was nach [106] ein Potential von mehr als 70% der deutschen Rechenzentren bedeutet.

### **Service Level Agreement**

Die bereits im vorherigen Abschnitt angesprochenen vertraglichen Bedingungen, die ein Rechenzentrumsbetreiber seinen Kunden zusichert, werden als Service Level Agreement (SLA) bezeichnet. SLA beschreiben meist technische Details der Zusicherung des Betriebes eines Dienstes im Rechenzentrum und werden in der Regel genau definiert und überwacht. Beispielsweise können SLA exakte Grenzwerte der Server-Temperaturen, Antwortzeiten eines Dienstes oder maximale Ausfallzeiten beinhalten.



## 2.3 Virtualisierung

Die Technologie der Virtualisierung im Kontext von Rechenzentren beschreibt die Ausführung von Anwendungen in einem eigens dafür erstellten, virtuellen Kontext, der in der Regel von der zugrunde liegenden Hardware des Servers abstrahiert. Allerdings gibt es verschiedene Formen der Virtualisierung [37], bei denen die jeweilige Abstraktion unterschiedlich stark ausgeprägt ist. Beispielsweise können je nach Bedarf nur einzelne Anwendungen virtualisiert werden, die dann isoliert voneinander ausgeführt werden. Bei der Paravirtualisierung hingegen wird eine speziell angepasste Variante eines Betriebssystems im virtuellen Container installiert, die sehr ähnlich zum nativen Betriebssystem des Servers ist. Dadurch können Operationen entweder virtuell oder nativ ausgeführt werden, was zu einer höheren Performance führen soll [11, 37, 112].

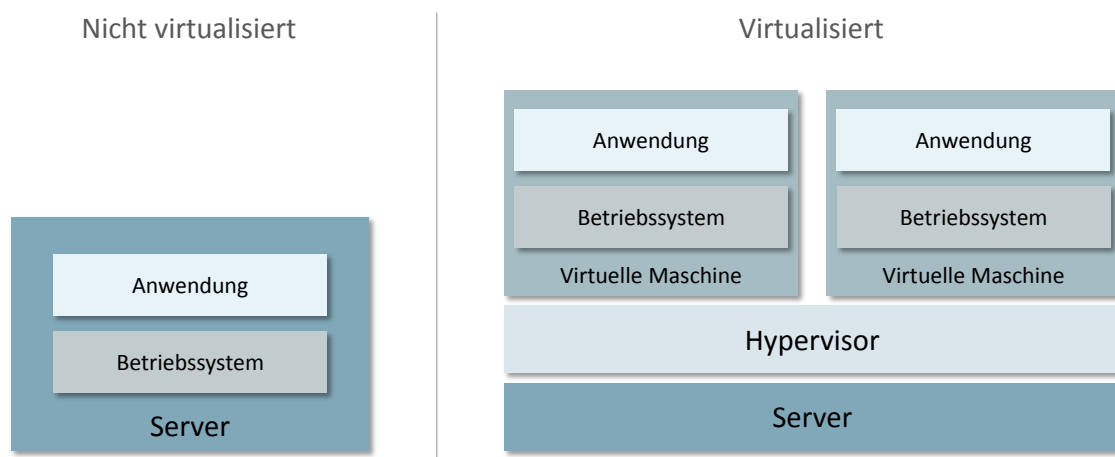


Abbildung 2.7: *In einem nicht virtualisierten System werden die Anwendungen direkt auf dem physikalischen Server installiert. Virtualisierung hingegen führt eine Managementebene ein, die Hypervisor genannt wird. Dieser kann mehrere virtuelle Maschinen verwalten, in denen die Anwendungen installiert sind. Dadurch wird es möglich, mehrere Anwendungen vollständig isoliert voneinander auf einem physikalischen Server zu betreiben.*

In Rechenzentren ist, wenn von Virtualisierung die Rede ist, häufig die Vollvirtualisierung gemeint, bei der ein virtueller Container einen eigenständigen Rechner nachbildet. In diesem Container muss zunächst ein Betriebssystem installiert werden, damit Anwendungen ausgeführt werden können. Auf dem Server selbst ist ein Hypervisor installiert, der die Verwaltung der verschiedenen virtuellen Maschinen übernimmt. Dieses Prinzip ist in Abbildung 2.7 dargestellt. Gegenüber der Paravirtualisierung liegen die Vorteile hier vor allem darin, dass eine vollständige Isolation der virtuellen Maschine erreicht wird und keine Modifikationen des Gast-Betriebssystems erforderlich sind.

Damit im weiteren Verlauf der Begriff „virtuelle Maschine“ eindeutig verwendet werden kann, erfolgt an dieser Stelle die Definition des Begriffs:

### Definition 2.4 Virtuelle Maschine

Eine virtuelle Maschine oder kurz „VM“ ist eine durch einen Hypervisor und auf einem physikalischen Server ausgeführte, vollständig gekapselte Betriebssystemumgebung, auf der beliebige Anwendungen ausgeführt werden. Im Allgemeinen wissen diese Anwendungen selbst nicht, dass sie in einer virtuellen Maschine laufen, normalerweise ist es für sie auch unerheblich.

### Live-Migration

Ein bedeutender Vorteil der Nutzung von Virtualisierung ist die Möglichkeit, virtuelle Maschinen auf verschiedenen physikalischen Servern ausführen zu können. Dies wird durch die Abstraktion der Hardware durch den Hypervisor ermöglicht. Eine besondere Art dieses Verschiebens von virtuellen Maschinen ist die Live-Migration [28, 127]. Diese Technik erlaubt eine direkte Migration der VM während des Betriebs ohne eine Pausierung oder ein Ausschalten. Bei einer Live-Migration muss nur der jeweilige Inhalt des Arbeitsspeichers der VM übertragen werden. Die Daten der VM befinden sich auf einem zentralen Storage-Server. Dadurch wird eine Migrationszeit von wenigen Sekunden erreicht, wobei die Zeit der Nicht-Erreichbarkeit der VM nur einige hundert Millisekunden beträgt und durch den Schaltvorgang des Ausführungs- und Netzwerkkontextes begründet ist [55].

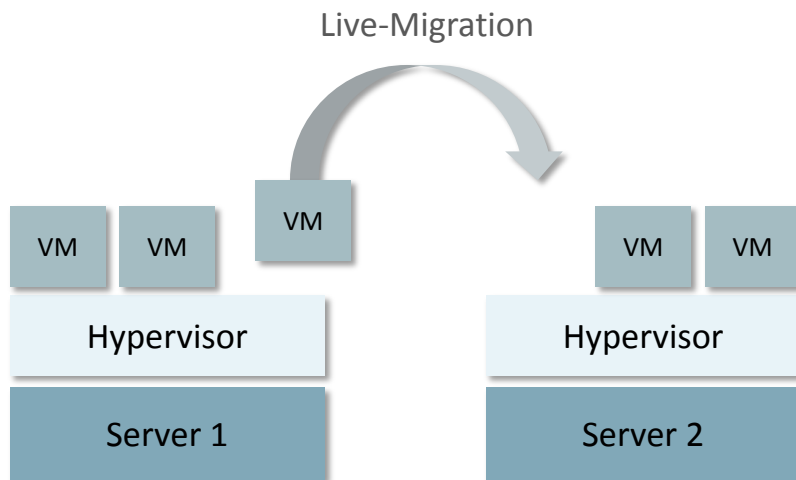


Abbildung 2.8: Durch Migrationen können einzelne virtuelle Maschinen von einem Server auf einen weiteren verschoben werden. Die Technik der Live-Migration erlaubt dies ohne eine Unterbrechung der Ausführung der virtuellen Maschine vorzunehmen, es entsteht lediglich eine kurze Nichterreichbarkeit von wenigen Millisekunden.

Die Dauer einer Live-Migration bezieht sich in dieser Arbeit auf den gesamten Kopiervorgang und die Umschaltung des Kontextes und beträgt damit meist einige Sekunden.

Weitere Untersuchungen und die Modellierung zu dieser Dauer folgt im Modellierungskapitel 5.6.2.

Die Technik der Live-Migration ist die Basis für dynamische Server-Konsolidierung und viele weitere Optimierungsstrategien in diesem Kontext. Daher sind Voraussetzungen für die Anwendung des in dieser Arbeit entwickelten Optimierungsverfahrens sowohl der Einsatz von Virtualisierung generell, als auch die Möglichkeit, Live-Migrationen durchführen zu können.

## 2.4 IT-Auslastung und Ressourcenbetrachtung

Die Server in Rechenzentren werden dazu verwendet, Aufgaben abzuarbeiten oder Dienste bereitzustellen. Beides erfolgt durch die installierten Anwendungen, die sich im hier betrachteten Fall in den virtuellen Maschinen befinden. Durch die Nutzung dieser Anwendungen entsteht eine Auslastung auf dem Server, die auf verschiedene Ressourcen bezogen werden kann. Meist werden hier der Prozessor, der Arbeitsspeicher, die Festplatte und der Netzwerkanschluss genannt.

Anwendungen, die einzelne Aufgaben abarbeiten, können unter Umständen pausiert werden, wenn Ressourcen knapp sind, Problemfälle auftreten oder die Energiepreise steigen. Anwendungen, die dagegen Dienste anbieten, müssen immer auf ihre benötigten Ressourcen zugreifen können. Kann dies nicht erfolgen, liegt in der Regel ein Ressourcenengpass oder ein Fehler vor, was vermieden werden soll.

In dieser Arbeit wird die IT-Auslastung als nicht pausierbare und nicht zeitlich verschiebbare Ressourcennutzung betrachtet. Den Anwendungen müssen also immer die Ressourcen zur Verfügung stehen, die sie für die jeweilige Aufgabenlast benötigen. Dadurch sind die Untersuchungen und das entwickelte Verfahren auf viele verschiedene Anwendungsarten übertragbar, und nicht auf pausierbare Tasks beschränkt.

In den weiteren Abschnitten folgen formale Definitionen der Anwendungslast und der Bereitstellung der Ressourcen durch die Server. Diese Definitionen werden im Laufe der Arbeit immer wieder aufgegriffen oder referenziert.

### 2.4.1 Ressourcen eines Servers

Die gesamte Auslastung des Servers  $s$  wird mit einem Vektor  $\vec{l}_s$  mit einem Auslastungswert für jeden betrachteten Ressourcentyp des Servers angegeben. In dieser Arbeit sind dies die beiden Indikatoren CPU-Auslastung  $u_s$  und Arbeitsspeicherbelegung (Random Access Memory, RAM)  $o_s$ :

$$\vec{l}_s = (u_s, o_s) \quad (2.3)$$

Die Beschränkung auf diese beiden Ressourcen erfolgt deshalb, weil bereits mehrfach durch Messungen und Modellierungen gezeigt wurde, dass eine energetische Betrachtung eines Servers nur anhand der CPU-Auslastung möglich und sinnvoll ist [60, 113], da hier eine signifikante Korrelation zur Leistungsaufnahme vorliegt. Eigene Messungen und Experimente, die zur Modellbildung durchgeführt wurden (siehe Kapitel 5.3.1), bestätigen dies.

Eine Erweiterung der Ressourcenbetrachtung auf die Lese- und Schreibzyklen der Festplatte sowie des Arbeitsspeichers und die Netzwerkübertragung resultiert nur in sehr speziellen Fällen in einer signifikanten Verbesserung der energetischen Server-Modellierung [113]. Die Arbeitsspeicherbelegung ist ein notwendiger Parameter der betrachtet werden muss, da diese eine feste Schranke für die Anzahl der virtuellen Maschinen auf einem Server bildet, die weder überschritten werden kann, noch darf. Damit wird Arbeitsspeicherteilung, wie es einige Hypervisoren anbieten, nicht betrachtet.

Jeder Server  $s$  besitzt eine maximale Ressourcenkapazität  $\vec{l}_{scap} = (u_{scap}, o_{scap})$ , wobei  $u_{scap}$  die Summe der CPU-Ressourcen aller physikalischen Prozessoren und Kerne beschreibt.

Das in dieser Arbeit entwickelte Optimierungsverfahren soll heterogene Server-Umgebungen unterstützen. Dies bedeutet, dass verschiedene Server-Typen mit einer unterschiedlichen Anzahl von Prozessoren und Kernen und unterschiedlicher Leistungsfähigkeit genutzt werden. Diese müssen auch untereinander verglichen werden können. Daher kann die CPU-Auslastung nicht prozentual betrachtet, sondern muss absolut angegeben werden. Damit wird der CPU-Ressourcenbedarf  $u$  grundsätzlich in Megahertz (MHz) angegeben. Damit eine Vergleichbarkeit von verschiedenen Servern korrekt durchgeführt werden kann, werden Untersuchungen vorgenommen, die die absolute und prozentuale Auslastung von virtuellen Maschinen auf verschiedenen Server-Typen vergleichen. Diese Untersuchungen werden in Kapitel 5.6.1 vorgestellt.

Der verfügbare Arbeitsspeicher eines Servers wird absolut in Megabyte (MB) angegeben. Hierfür werden keine Vergleichsmodelle benötigt, da der durch eine virtuelle Maschine auf einem Server belegte Speicher festgelegt wird und auf allen Systemen als absoluter Wert direkt vergleichbar ist.

Eine Ressourcenbelegung des Servers  $s$  ist dann konfliktfrei bzw. legal, wenn gilt:

$$\vec{l}_{scap} \geq \vec{l}_s + \vec{\varepsilon}_s \quad (2.4)$$

Dabei ist  $\vec{\varepsilon}_s = (u_\varepsilon, o_\varepsilon) | u_\varepsilon, o_\varepsilon \geq 0$  ein Pufferwert, der verhindert, dass der Server mit dem theoretischen Maximum ausgelastet wird, was in der Praxis zu Performanzproblemen führen kann, vor allem dann, wenn noch Migrationen durchgeführt werden sollen. Die Zusammensetzung der Ressourcenbelegung eines Servers  $s$  wird in Abbildung 2.9 beispielhaft dargestellt.

## 2.4.2 Ressourcenbedarf von virtuellen Maschinen

In dieser Arbeit wird davon ausgegangen, dass die Auslastung eines Servers ausschließlich durch die darauf ausgeführten virtuellen Maschinen erzeugt wird. Tatsächlich erzeugt auch der Hypervisor eine bestimmte Auslastung, die auch als Virtualisierungsoverhead bezeichnet wird. Bei den aktuellen Hypervisoren ist dieser im Allgemeinen aber gering [143], sodass dieser vernachlässigt werden kann oder er tritt signifikant nur in speziellen Anwendungen auf, wenn beispielsweise direkt konkrete Hardwarekomponenten angesprochen werden [84, 143] (siehe auch Kapitel 2.3 und 5.6.1). Der Ressourcenbedarf einer virtuellen

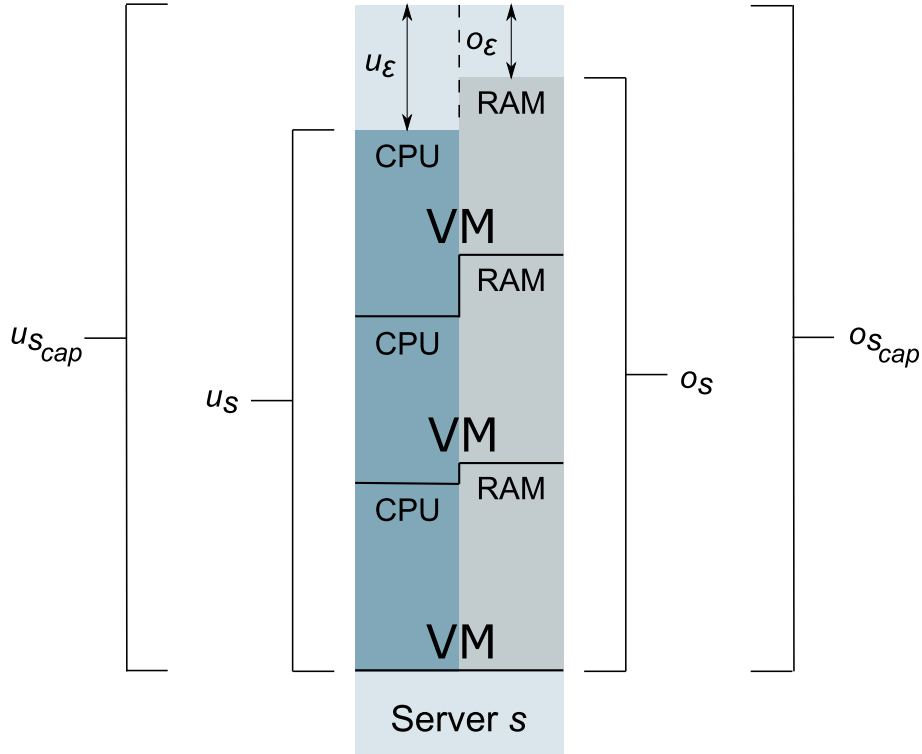


Abbildung 2.9: Ressourcenbelegung eines Servers durch VMs: Die Ressourcen CPU-Auslastung und Belegung des Arbeitsspeichers (RAM) werden bei der Ausführung einer VM auf einem Server betrachtet. Beide Ressourcen sind durch die Kapazität des Servers beschränkt. Im gezeigten Beispiel befinden sich drei VMs auf dem Server.

Maschine  $v$  wird mit

$$\vec{l}_v = \begin{pmatrix} u_v \\ o_v \end{pmatrix} \quad (2.5)$$

angegeben. Auch hier ist die Notwendigkeit der Vergleichbarkeit der tatsächlich erreichten Auslastung einer virtuellen Maschine auf verschiedenen Server-Typen gegeben. Die Untersuchungen dazu befinden sich ebenfalls in Kapitel 5.6.1.

Die insgesamt durch virtuelle Maschinen in Anspruch genommenen Ressourcen des Servers  $s$  ergeben sich dabei wie folgt, wobei  $n_v$  die Anzahl der virtuellen Maschinen auf Server  $s$  ist:

$$\vec{l}_s = \sum_{i=1}^{n_v} \vec{l}_{v_{i_s}} \quad (2.6)$$

Weiterhin wird in dieser Arbeit davon ausgegangen, dass die elektrische Leistungsaufnahme  $P_s$  eines Servers  $s$  von der jeweiligen CPU-Auslastung  $u_s$  aus  $\vec{l}_s$  abhängt:

$$P_s = \text{spm}_s(u_s) \quad (2.7)$$

Hierbei ist  $spm_s(u_s)$  die Modellfunktion, die für das spezifische Server-Modell charakterisiert werden muss. Dies erfolgt in Kapitel 5.3.1. Dabei ist, wie bereits erwähnt, die CPU-Auslastung  $u_s$  die Teilkomponente von  $\vec{l}_s$ , die das primäre Merkmal zur Modellierung der elektrischen Leistungsaufnahme darstellt.

## 2.5 Betriebszustände des Rechenzentrums

Energetisch gesehen sind Rechenzentren dynamische Akteure und können nicht statisch betrachtet werden. Aufgrund einer dynamischen IT-Auslastung der Server verändern sich ebenfalls die energetischen Zustände weiterer Geräte. Ein Betriebszustand des Rechenzentrums beschreibt den jeweiligen Zustand aller Geräte, der sich aus den aktuellen Betriebsparametern zusammensetzt. Zu diesen Parametern zählen interne und externe Einflüsse, wie beispielsweise die momentane IT-Auslastung der Server oder die Außentemperatur. Einige dieser Einflüsse können sich schnell ändern und damit Reaktionen der Geräte im Rechenzentrum hervorrufen. Dabei können sowohl Auswirkungen auf die Stabilität des Betriebs des Rechenzentrums, als auch auf den Energiebedarf auftreten. Beispielsweise kann bei einer ansteigenden IT-Auslastung eine stärkere Klimatisierung notwendig werden. Auch die Verteilung der IT-Auslastung auf die Server bestimmt den Betriebszustand des Rechenzentrums und kann zu teils sehr unterschiedlichen energetischen Situationen führen. Es ist auch Gegenstand dieser Arbeit, diese Einflüsse darzustellen - dies erfolgt in Kapitel 4.1.1.

### Allokation und Betriebszustand

Im weiteren Verlauf dieser Arbeit ist die Differenzierung der Verteilung der IT-Auslastung und dem daraus und aus weiteren Parametern resultierenden Betriebszustand wichtig.

Wie bereits beschrieben, kann aufgrund der Technologie der Virtualisierung die durch die Anwendungen anfallende IT-Auslastung annähernd beliebig auf verschiedene Server verteilt werden, solange genug Ressourcen zur Verfügung stehen. Die Verteilung dieser IT-Auslastung auf die Server wird hier als Allokation bezeichnet. Da in dieser Arbeit von einer konkreten IT-Auslastung, die beispielsweise durch eine Datenbankanwendung oder einen Webserver verursacht wird, abstrahiert wird, erfolgt eine Beschreibung einer Allokation ausschließlich auf Ebene der Virtualisierung. Damit beschreibt eine Allokation die Zuordnung von virtuellen Maschinen zu physikalischen Servern, die Auslastung der virtuellen Maschine entspricht der, die durch die darin ausgeführte Anwendung entsteht.

Im weiteren Verlauf dieser Arbeit beschreibt  $S$  die Menge aller Server im Rechenzentrum und  $V$  die Menge aller virtuellen Maschinen. Ausgehend hiervon wird eine Allokation wie folgt definiert:

#### Definition 2.5 Allokation

Die Allokation  $\vec{a}$  ist ein Vektor der Dimension  $n$  mit

$$\vec{a} = (\vec{a}_{v_1}, \vec{a}_{v_2}, \vec{a}_{v_3}, \dots, \vec{a}_{v_n}) \Big| n = |S| \quad (2.8)$$

wobei jedes Element  $\vec{a}_{v_i}$  ebenfalls ein Vektor ist und dem Server  $s_i$  zugeordnet wird.  $\vec{a}_{v_i}$  enthält die virtuellen Maschinen, die unter der Allokation  $\vec{a}$  auf dem Server  $s_i$  ausgeführt werden:

$$\vec{a}_{v_i} = (v_1, v_2, \dots, v_k) \mid v_1, v_2, \dots, v_k \in V \quad (2.9)$$

Wenn  $\vec{a}_{v_i} = \emptyset$ , dann gilt der Server  $s_i$  als ausgeschaltet, er wird dann als inaktiv bezeichnet.

Grundsätzlich ist eine Allokation nur dann legal, wenn sich jede virtuelle Maschine aus  $V$  auf genau einem Server befindet. Ein Server ist aktiv, wenn auf ihm mindestens eine virtuelle Maschine ausgeführt wird. Es wird dann implizit vorausgesetzt, dass dieser Server auch eingeschaltet ist. Wenn ein Server keine virtuelle Maschine ausführt, gilt er als inaktiv. Dann wird implizit angenommen, dass der Server ausgeschaltet wird.

Ein Betriebszustand eines Rechenzentrums beschreibt nicht nur die Allokation der virtuellen Maschinen, sondern zusätzlich den Zustand weiterer Geräte im Rechenzentrum:

### Definition 2.6 Betriebszustand

*Ein Betriebszustand des Rechenzentrums resultiert aus einer eingenommenen Allokation, beschreibt über die Allokation hinaus aber weitere Betriebsparameter des gesamten Rechenzentrums. So beinhaltet ein Betriebszustand auch die Zustände weiterer Geräteklassen wie der USV oder der Kühlung, die sich aus der benötigten Versorgung der Server ergeben, deren Zustände wiederum durch eine Allokation beschrieben wurden.*

Nicht jede Allokation, die alle virtuellen Maschinen auf die Server verteilt, ist legal. Auch Allokationen, die auf den ersten Blick korrekt erscheinen, können zu illegalen Betriebszuständen führen. Ein Betriebszustand ist dann illegal, wenn er zur Verletzung von SLA führt oder die Betriebsstabilität des Rechenzentrums gefährdet. Damit eine solche Potentielle Gefährdung erkannt und bewertet werden kann, werden in dieser Arbeit Kriterien definiert, die in einem Betriebszustand eingehalten werden müssen. Dies erfolgt in Kapitel 4.

Für eine Allokation  $\vec{a}$  muss also, bevor diese aktiv eingenommen wird, feststellbar sein, wie der Betriebszustand der wichtigsten Geräte des Rechenzentrums voraussichtlich aussehen wird. Dazu wird eine Rechenzentrumssimulation verwendet, deren Modelle in Kapitel 5 vorgestellt werden.

### Zustandsänderungen

Die Veränderung des Betriebszustands des Rechenzentrums erfolgt durch eine Variation der Allokation. Dazu können nur folgende Aktionen durchgeführt werden:

- *Ausschalten eines Servers:* Befinden sich keine virtuellen Maschinen auf einem Server, kann dieser ausgeschaltet werden. In dieser Arbeit werden grundsätzlich alle „leeren“ Server als ausgeschaltet betrachtet.
- *Einschalten eines Servers:* Werden zusätzliche Server-Kapazitäten benötigt, muss ein weiterer Server eingeschaltet werden.

- *Live-Migrationen von virtuellen Maschinen:* Eine virtuelle Maschine kann auf einen eingeschalteten Server verschoben werden, dies führt in der Regel dazu, dass die elektrische Leistungsaufnahme des Quellserver sinkt und die des Zielservers steigt.

Das Ein- und Ausschalten eines Servers benötigt einige Zeit, teilweise bis zu mehreren Minuten. Werden also zusätzliche Server-Kapazitäten benötigt, stehen diese nicht sofort zur Verfügung, sondern es muss erst die Boot-Zeit des Servers abgewartet werden. Erst dann kann eine Migration einer virtuellen Maschine auf den Server erfolgen. Damit Engpässe bei schwankenden Lasten und der Notwendigkeit von zusätzlichen Server-Kapazitäten vermieden werden, wurden Lastvorhersagemethoden entwickelt, die ein rechtzeitiges Einschalten der Server ermöglichen [22, 53]. Eine solche Methode wird auch in dieser Arbeit angewandt und in Kapitel 3 genauer vorgestellt.

In dieser Arbeit werden energetische Zusammenhänge analysiert, und das entwickelte Optimierungsverfahren betrachtet Zeiträume von mehreren Tagen bis zu einigen Monaten. Die Dauer des Ein- und Ausschaltens von Servern sind in diesen Analysen eher nebensächlich. In der Rechenzentrumssimulation und Evaluation der in dieser Arbeit entwickelten Verfahren werden diese Zeiten berücksichtigt, allerdings werden sie im weiteren Verlauf dieser Arbeit nicht mehr erwähnt.

## 2.6 Optimierung durch Server-Konsolidierung

Die Möglichkeiten der energetischen Optimierung von Rechenzentren, die durch die Technik der Virtualisierung geschaffen wurden, werden bereits in Form von Konsolidierungsstrategien ausgenutzt. Dabei geht es prinzipiell darum, die existierende IT-Last auf möglichst wenigen aktiven Servern zu bündeln und möglichst viele, ungenutzte Server auszuschalten. Dies erlaubt signifikante Energieeinsparungen. Ein Beispiel dieses Prinzips ist in Abbildung 2.10 dargestellt. Sinkt zu einem Zeitpunkt die Auslastung einer virtuellen Maschine, sodass eine weitere auf den Server passt, kann eine Migration vorgenommen werden. Im nächsten Schritt kann ein Server, der dann keine virtuellen Maschinen mehr ausführt, abgeschaltet werden. Im Beispiel ist dies Server 5. Das kombinatorische Problem, eine Allokation zu finden, die möglichst wenige aktive Server benötigt, entspricht dem Bin-Packing-Problem und ist NP-vollständig [17, 53, 54, 63, 83].

Das Thema Konsolidierung ist komplex und kann unter vielen verschiedenen Aspekten und mit unterschiedlichen Schwerpunkten betrachtet werden. Eine detaillierte Aufführung dieser Facetten ist an dieser Stelle aufgrund des Umfangs nicht möglich, allerdings werden in Kapitel 3.1 mehrere Verfahren aus diesem Kontext vorgestellt und es werden Literaturhinweise zur weiteren Vertiefung genannt.

Auch das in dieser Arbeit entwickelte Optimierungsverfahren nutzt die Server-Konsolidierung, um Energie zu sparen. Das Verfahren stützt sich allerdings auf die Analyse der energetischen Einflüsse im Rechenzentrum und verbindet die Erkenntnisse daraus mit der Optimierungsstrategie. Damit wird das Verfahren im Vergleich zu existierenden Konsolidierungsstrategien deutlich erweitert, was neben einer erhöhten Flexibilität und Dynamik



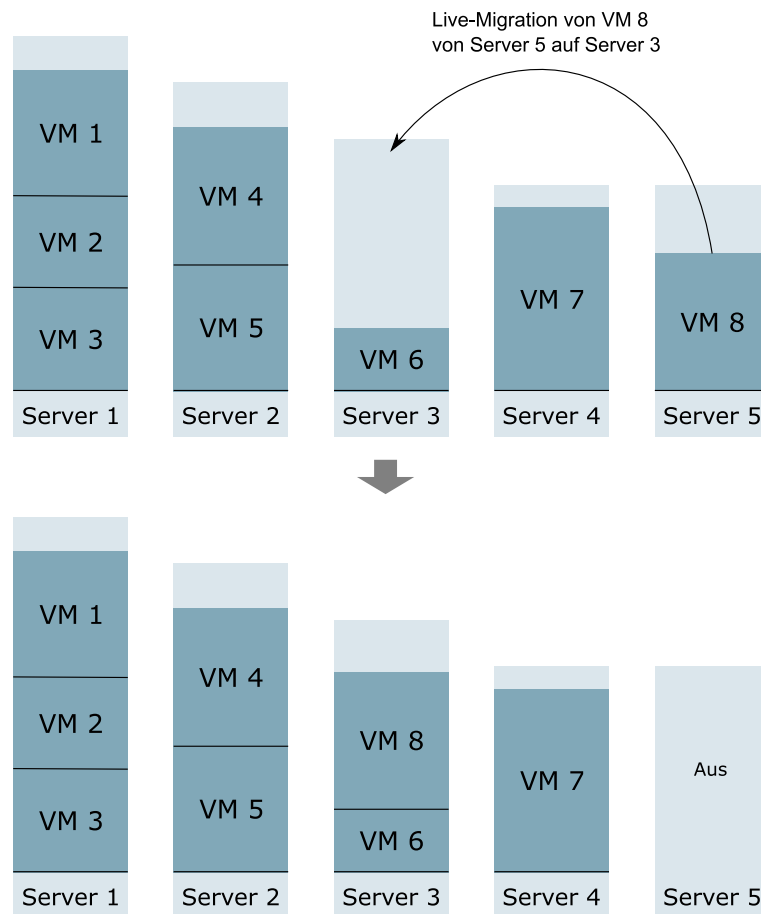


Abbildung 2.10: *Beispiel einer Server-Konsolidierung: Die VM 8 wird von Server 5 auf Server 3 verschoben, da Server 3 genug freie Ressourcen zur Verfügung hat. Danach kann Server 5 ausgeschaltet werden. Die Höhe des Rechtecks des jeweiligen dargestellten Servers zeigt exemplarisch die verfügbaren Ressourcen, die Höhe der Rechtecke der virtuellen Maschinen soll die Ressourcennutzung repräsentieren. Aufgrund einer besseren Übersichtlichkeit wurde in der Grafik keine Differenzierung einzelner Ressourcen vorgenommen.*

des Energiebedarfs eines Rechenzentrums auch zusätzliche Einsparungen ermöglicht. Das Verfahren wird in Kapitel 6 vorgestellt.

## 2.7 Zusammenfassung

In diesem Kapitel wurden die Grundlagen für das Verständnis der weiteren Arbeit gelegt. Wichtige und zentrale Begriffe wie das Rechenzentrum, die IT-Räume und Server sowie Allokationen von virtuellen Maschinen und Betriebszustände wurden eingeführt und de-

finiert. Der Ressourcenbedarf von virtuellen Maschinen, die auf einem Server ausgeführt werden, wurde formal definiert. Diese Definition wird vor allem in Kapitel 4 und 5 wieder aufgenommen und erweitert oder konkretisiert.

Die beschriebenen Zusammenhänge der Rechenzentrumsinfrastruktur und insbesondere der Kühlung sind ebenfalls wichtige Grundlagen für das Verständnis der energetischen Analysen in Kapitel 4.

Das folgende Kapitel 3 behandelt verwandte Arbeiten aus dem Kontext dieser Dissertation.

## 3 Verwandte Arbeiten

In dieser Arbeit werden energetische Freiheitsgrade, Einflüsse und Abhängigkeiten der Geräte im Rechenzentrum untersucht. Diese Erkenntnisse haben zentralen Einfluss auf die Möglichkeiten der energetischen Steuerung eines Rechenzentrums und werden später im Optimierungsverfahren genutzt. Die energetische Steuerung von Rechenzentren ist ein aktuelles Forschungsthema, das unter mehreren Aspekten untersucht wird. Diese werden in Unterkapitel 3.2 vorgestellt und im Kontext dieser Arbeit verglichen.

Eine spezielle Form der energetischen Steuerung von Rechenzentren ist die gezielte Konsolidierung mittels Lastmanagementverfahren. Hierzu existieren ebenfalls bereits mehrere Verfahren und Ansätze, die mit dem in dieser Arbeit verfolgten Konzept im folgenden Abschnitt 3.1 verglichen werden.

Zur Untersuchung der Auswirkung einzelner Optimierungsschritte im Rechenzentrum und damit eine Optimierung, beispielsweise durch ein Lastmanagementverfahren, im Rechenzentrum vorgenommen werden kann, werden Modelle der beteiligten Geräte benötigt. Auch in diesem Bereich gibt es einige Vorarbeiten, auf denen teilweise aufgebaut wird. Diese Modelle und Simulationen von Rechenzentren oder einzelnen Subsystemen werden in Abschnitt 3.3 vorgestellt.

### 3.1 Lastmanagement- und Konsolidierungsverfahren

Das Thema der energetischen Optimierung von Rechenzentren durch Lastmanagement und Konsolidierung ist und war Untersuchungsgegenstand der Forschung. Es existieren dazu verschiedenen Verfahren mit unterschiedlichen Schwerpunkten, die in diesem Abschnitt miteinander verglichen und im Kontext dieser Arbeit betrachtet werden. Dabei betrachten viele der Ansätze virtualisierte Rechenzentren, wobei aufgrund der Auslagerung der Anwendungen in virtuelle Maschinen erst die Möglichkeit der Verlagerung mehrerer Anwendungen auf einen Server entsteht. Dieses Prinzip wird ausgenutzt, um die vorhandenen Anwendungen auf möglichst wenige aktive Server zu verschieben und die ungenutzten auszuschalten. Dadurch kann Energie eingespart werden. Da IT-Last häufig Schwankungen unterliegt, ist eine Herausforderung, auf diese zu reagieren, ohne dass es zu Ressourcen-Engpässen kommt.

Dieses Konzept kann nicht nur lokal angewandt werden, sondern kann auch auf mehrere Rechenzentren an verschiedenen Standorten, die in einem Verbund organisiert sind, ausgeweitet werden. In den folgenden Abschnitten wird zunächst das lokale Lastmanagement und dann das rechenzentrumsübergreifende vorgestellt.

### 3.1.1 Lokales Lastmanagement

Das lokale Lastmanagement bezieht sich primär auf die Verteilung einer Menge virtueller Maschinen auf die vorhandenen physikalischen Server. Dabei soll zur Energieeinsparung eine möglichst kleine Anzahl an Servern verwendet werden. Das zugrundeliegende Problem wird dabei als Set-Packing oder Bin-Packing beschrieben, bei dem die virtuellen Maschinen die Elemente darstellen, die auf die Container (Server) verteilt werden müssen [54, 63, 83]. Die Verteilung von virtuellen Maschinen auf physikalische Server wurde auch in [17, 53] als Bin-Packing-Problem beschrieben. Die Autoren legen den Fokus aber auf eine Vorhersagemethodik, mit der periodisch schwankende IT-Auslastungen erkannt werden können. Damit lässt sich ein dynamisches Lastmanagement realisieren, bei dem ungenutzte Server ausgeschaltet, aber vor ansteigenden Auslastungen früh genug Server wieder eingeschaltet werden. Die beschriebene Vorhersagemethodik aus [53] wird in dieser Arbeit ebenfalls angewandt und wird in Kapitel 6.2.1 wieder aufgegriffen.

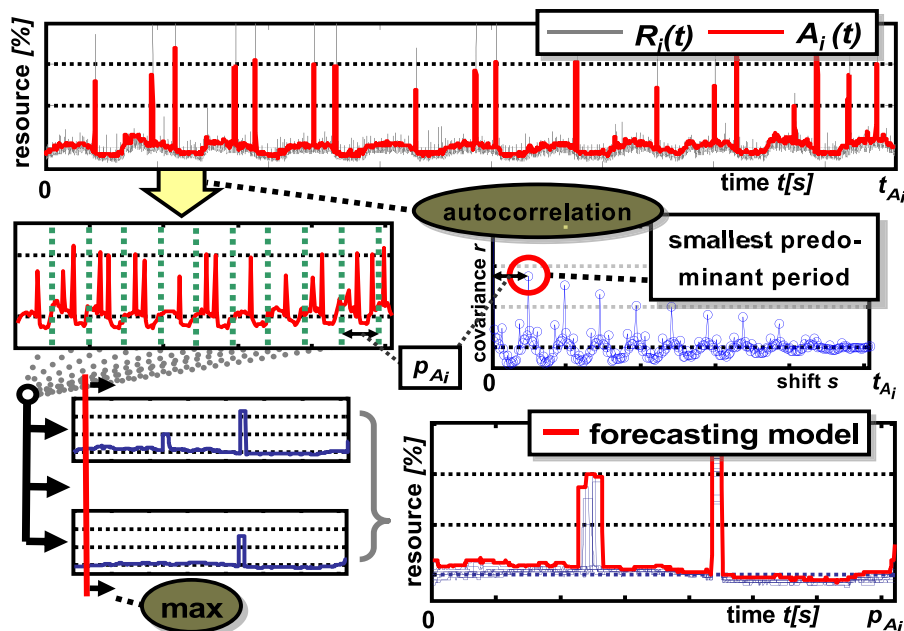


Abbildung 3.1: Die Vorhersagemethodik betrachtet die prozentuale IT-Auslastung über einen definierten Zeitraum und extrahiert mittels Autokorrelationsanalyse periodische Schwankungen. Diese können nun für zukünftige Zeiträume als weitergeführte Vorhersage verwendet werden (Grafik unverändert übernommen aus [53]).

Mehrere Ansätze greifen diese klassische Konsolidierung auf, setzen aber unterschiedliche Schwerpunkte. So löst [132] die dynamische Allokation mit einem verteilten Verfahren anstatt mit einem zentralen Managementsystem. [68] beschränkt sich auf eintreffende Aufgaben (Jobs).

Die Autoren von [31, 37, 77, 146] verbinden die Server-Konsolidierung mit der Untersu-

chung von Verteilungsstrategien für virtuelle Maschinen, die aufgrund eines optimierten Netzwerkroutings den Energiebedarf des gesamten Rechenzentrums reduzieren sollen.

In [20] wird die Ausführung von sporadisch auftretenden Aufgaben (Tasks und Jobs) auf verschiedenen Servern energetisch optimiert. Es wird davon ausgegangen, dass diese Aufgaben aufgeteilt und parallel abgearbeitet werden können. Die Aufgaben werden nun auf unterschiedlich performante Server verteilt, sodass insgesamt eine Minimierung des Energiebedarfs erreicht werden soll. Das Problem wird auch hier als Bin-Packing formuliert.

#### **Kritik und Abgrenzung**

Die genannten Verfahren lösen teils grundlegende dynamische Allokationsprobleme, betrachten aber keine weiteren energetischen Kriterien und ignorieren auch die räumliche Verortung. So können die Lösungen zu Hot-Spots oder Ineffizienzen bei der Kühlung führen. Außerdem erfolgt keine Betrachtung einer Flexibilität der Allokation. Es werden meist statische Puffer auf den Servern freigehalten, die Auslastungsschwankungen abfangen sollen. Sind diese Puffer zu klein gewählt oder trifft eine Erwartung, wie beispielsweise die Vorhersage der Auslastung nicht zu, so müssen die Verfahren im Fehlerfall mit dem Einschalten von Servern reagieren. Damit diese Probleme vermieden werden, bleibt nur eine gesamte Erhöhung der Puffergröße.

Zudem werden teils sehr aufwendige Optimierungen in Bereichen vorgenommen, die nur geringe Auswirkungen auf den gesamten Energiebedarf haben (beispielsweise die Netzwerkoptimierung). Keines der Verfahren versucht die Geräte im Rechenzentrum mit dem größten Potential gemeinsam zu optimieren.

Aufgrund der kombinatorischen Komplexität wird für das Bin-Packing-Problem der Allokation der virtuellen Maschinen auf möglichst wenige Server von den genannten Verfahren eine Lösungsheuristik eingesetzt. Diese ist meist aus der Kategorie der Best-Fit und First-Fit-Heuristiken (FF) und häufig die First-Fit-Decreasing-Heuristik (FFD) [63] oder eine Variation dieser. Die Heuristik sortiert die virtuellen Maschinen nach Ressourcenbedarf und verteilt diese dann nacheinander auf möglichst wenige Server. Auch wenn die Lösungen, die durch diese Heuristiken erzeugt werden, häufig unter der Betrachtung weiterer (energetischer) Kriterien zu keiner validen Lösung führen, wird in der Evaluation in Kapitel 7 des Optimierungsverfahrens, das in dieser Arbeit entwickelt wird, ein Vergleich mit den Ergebnissen von FFD vorgenommen. Dies erfolgt aufgrund der großen Popularität dieser Heuristik.

#### **3.1.2 Lastmanagement unter Berücksichtigung thermischer Aspekte**

Die bisher vorgestellten Arbeiten zum lokalen Lastmanagement haben die Kühlung und Thermik im IT-Raum ignoriert. Mehrere Arbeiten haben speziell die Wärmeverteilung im IT-Raum untersucht und daraus ein Lastmanagement abgeleitet. So schlagen [87, 122, 129, 130] ein Task-Scheduling vor, das die Last nach thermischen Aspekten auf die Server verteilt. Lokale Überhitzungen sollen so vermieden werden.

Die genannten Arbeiten stützen sich dabei auf die Modelle aus [87], die auch die Luftströmungen innerhalb von IT-Räumen beachten und modellieren. Dazu wurden Messungen und aufwendige Computational-Fluid-Dynamics-Simulationen (CFD) durchgeführt. Die Arbeiten behandeln sowohl die Vermeidung von Hot-Spots, als auch die thermische Optimierung, wodurch Kühlkosten gespart werden können. Allerdings bezieht sich diese auf die Kältemaschine und die generelle Reduktion der zu kühlenden Wärmemenge, also der Leistungsaufnahme der Server. Die Autoren haben dazu als eine der ersten den Coefficient of Performance (COP) der Kältemaschine untersucht, vermessen und charakterisiert. Dieses Modell wird auch in dieser Arbeit verwendet und in Kapitel 5.5 aufgegriffen.

Die Autoren von [87, 129] kommen zu dem Schluss, dass eine ungleiche Temperaturverteilung im IT-Raum zu steigenden Kühlkosten führt. Der in dieser Arbeit in Kapitel 4.3 vorgestellte Ansatz zur Gleichverteilung der Wärmeleistung im IT-Raum folgt demnach der Empfehlung dieser Publikationen.

Pakbaznia et al. [94, 95] verbinden klassische Server-Konsolidierung mit der Reduktion der Kühlkosten, indem sie das in [87] vorgestellte Modell der Kältemaschinen verwenden und die Einsparungen der Server mit denen der Kältemaschinen verbinden. Es findet jedoch keine explizite thermische Optimierung statt, sondern lediglich eine Anwendung der Modelle zur Berechnung der Einsparungen, die durch die Konsolidierung erreicht werden. Ein ähnlicher Ansatz wird auch in [96] verfolgt.

#### **Kritik und Abgrenzung**

Die Modelle und Erkenntnisse aus [87, 130, 131] bilden die Grundlagen für weitere Forschungen der thermischen Optimierung in Rechenzentren [4, 74, 97, 103]. Die darauf aufbauenden Arbeiten haben die Modelle verifiziert und teils eigene Charakterisierungen veröffentlicht. Einige dieser Modelle werden auch in dieser Arbeit verwendet und in Kapitel 5 vorgestellt.

Kernpunkt der Arbeiten ist die Untersuchung der thermischen Situation im IT-Raum mittels Messungen und CFD-Simulationen. Die Erkenntnisse aus diesen Untersuchungen werden in dieser Arbeit aufgegriffen und beispielsweise in Kapitel 4.3 zur Formulierung der thermischen Kriterien erweitert.

Die bisherigen Optimierungen der Forschung beziehen sich primär auf Task-Scheduling, bei dem eingehende Tasks den jeweils thermisch sinnvollsten Regionen in einem IT-Raum zugewiesen werden. Die Verteilung der IT-Last erfolgt dann nach dem Prinzip, den kühlest Server zuerst zu wählen. Einsparungen in der Kühlung erfolgen dann aufgrund der homogenen Temperaturverteilung, weitere energetische Optimierungen werden aber nicht vorgenommen.

Sowohl die Modelle als auch die Optimierungen der Forschung beziehen sich auf die Luftbewegungen und Wärmeleistungen im IT-Raum. Es findet zwar eine Modellierung der Kältemaschinen statt, aber eine genauere Betrachtung des Kaltwasserkreislaufs, eine gesamtheitliche Analyse und raumübergreifende Optimierungen fehlen ebenfalls.

### 3.1.3 Lastmanagement als multi-kriterielles Problem

In [146] wird Server-Konsolidierung und energetische Optimierung im Rechenzentrum als multi-kriterielles Problem betrachtet. Die Formulierung des Problems bezieht sich dabei explizit auf die Platzierung der virtuellen Maschinen auf die Server im Rechenzentrum. Die Arbeit der Autoren ist auf mehrere Publikationen aufgeteilt, in [146] wird die Platzierung von neu hinzukommenden virtuellen Maschinen behandelt. Dabei werden folgende Kriterien betrachtet:

- Durch Server-Konsolidierung Energie einsparen (virtuelle Maschinen auf möglichst wenigen Servern ausführen)
- Effiziente Ausnutzung der multi-dimensionalen Ressourcen der Server
- Vermeidung von Hot-Spots in IT-Räumen

Die IT-Auslastung wird hier in Form von Anwendungen betrachtet, die in virtuellen Maschinen gekapselt sind. Damit entspricht eine Anfrage für eine neue VM einer neuen Anwendung (siehe auch Abbildung 3.2 oben). Das Problem der Verteilung wird als Kombination einer multi-kriteriellen Optimierung und einer kombinatorischen Optimierung betrachtet. Das kombinatorische Problem wird dabei wie auch in einigen der bereits genannten Arbeiten als Bin-Packing formuliert.

Die Abbildung 3.2, die unverändert aus [146] übernommen wurde, zeigt die Architektur des Verfahrens. Neue VM-Anfragen werden auf den jeweils günstigsten und kühlest pas-

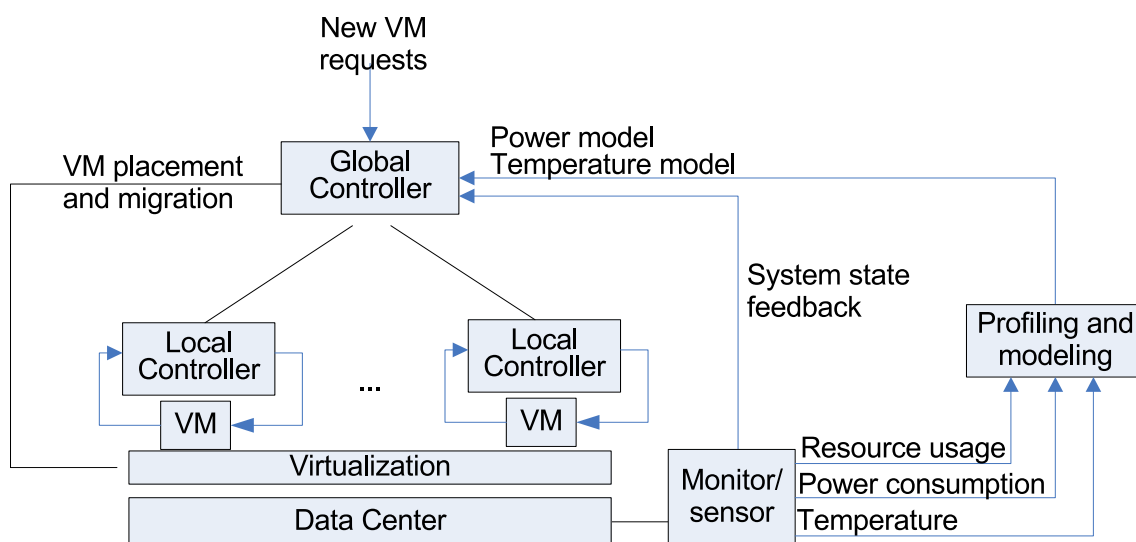


Abbildung 3.2: Architektur des autonomen Ressourcenmanagements für virtualisierte Rechenzentren (Grafik unverändert übernommen aus [146])

senden Server verteilt. Ein lokaler Controller, der in den virtuellen Maschinen ausgeführt

wird, überwacht die Ressourcennutzung der Anwendung. Je nach Bedarf werden ihr die geforderten Ressourcen bereitgestellt.

Der Lösungsansatz verwendet den gruppierenden genetischen Algorithmus (GGA) aus [35], der die direkte Zuordnung der virtuellen Maschinen auf die Server löst. Die Länge der Chromosomen steigt dabei mit zunehmender Anzahl von virtuellen Maschinen.

In [147] beschreiben die selben Autoren von [146] das dynamische Lastmanagement, das unter Betrachtung derselben Kriterien realisiert wurde. Bei Betrachtung des Kontrollflusses des Verfahrens in Abbildung 3.3 wird deutlich, dass das genannte Vorgehen einer klassischen Konsolidierung mit Berücksichtigung der Problematik von Hot-Spots entspricht. Bei

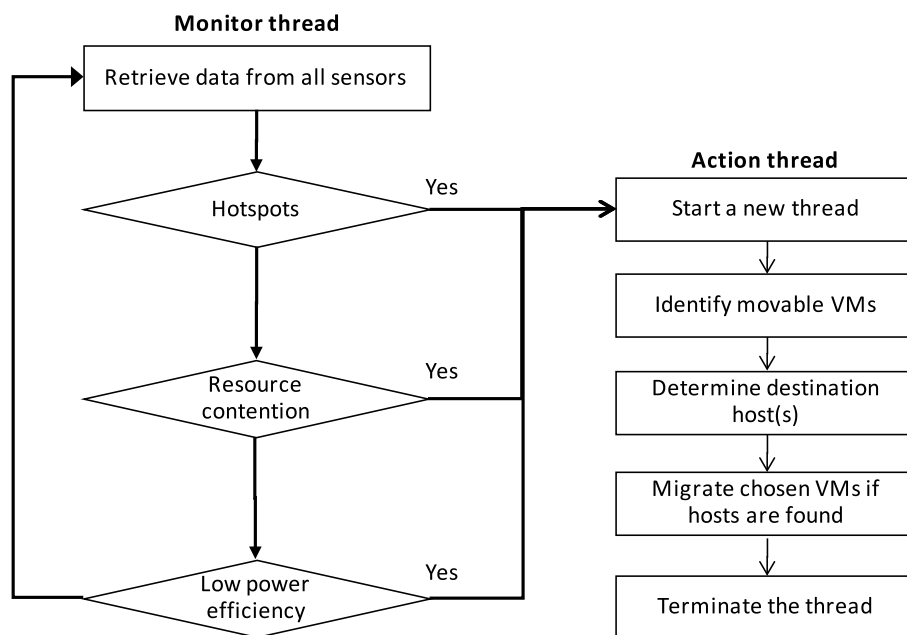


Abbildung 3.3: Kontrollfluss der multi-kriteriellen Optimierung (Grafik unverändert übernommen aus [147])

Lastschwankungen oder Überhitzungen erfolgt eine Reaktion, die eine Umverteilung der virtuellen Maschinen nach sich zieht.

### Kritik und Abgrenzung

Das Verfahren zur Platzierung von virtuellen Maschinen nutzt zur schnellen Abschätzung der thermischen Belastung eines Servers ähnliche Modelle, wie sie auch in dieser Arbeit (siehe Kapitel 5.3.2) zur Anwendung kommen. Es erfolgt dann aber ausschließlich eine Priorisierung kühlerer Server aufgrund dieses Modells, ohne eine thermische Betrachtung der Gesamtsituation vorzunehmen. Eine energetische Optimierung hinsichtlich der Kühlung findet nicht statt.

Der Ansatz benötigt einen lokalen „Agenten“ in jeder virtuellen Maschine, der die Ressourcenbelegung und den Bedarf überwacht. Damit ist der Ansatz invasiv, was in dieser Arbeit



vermieden werden soll. Zudem erlaubt das Verfahren eine Beeinflussung der Anwendungen. Erst wenn der lokale Controller die Verfügbarkeit zu weniger Ressourcen feststellt, erfolgt eine Korrektur.

Das dynamische Verfahren agiert rein reaktiv und behebt beispielsweise Hot-Spots oder Ressourcenengpässe erst nach dem Auftreten bzw. beim Übertreten eines Grenzwertes. Eine Garantie für eine schnelle Evakuierung bzw. Auflösung des Engpasses wird jedoch nicht behandelt.

Aufgrund der Kodierung der Chromosomen des GGA steigt deren Länge mit zunehmender Anzahl virtueller Maschinen. Die Autoren evaluieren den Algorithmus mit maximal 2000 virtuellen Maschinen. Eine Skalierung des Verfahrens auf die Bedürfnisse eines großen Rechenzentrums mit mehreren zehntausend virtuellen Maschinen erscheint allerdings fragwürdig.

Aufgrund der genannten Punkte entspricht das Verfahren einem Lastmanagement mit Server-Konsolidierung, wobei reaktiv thermische Grenzwerte eingehalten werden. Die Abwägung mehrerer Kriterien bezieht sich also primär auf dieses Einhalten der Grenzwerte, ohne weitere Optimierungen in energetischer Sicht durchzuführen.

#### 3.1.4 Rechenzentrumsübergreifendes Lastmanagement

Neben der Möglichkeit, lokales Lastmanagement in einem Rechenzentrum durchzuführen, kann die Lastverteilung auch auf mehrere Rechenzentren, die in einem Verbund organisiert sind, ausgeweitet werden. Dieses Konzept wird unter anderem für die Integration von Rechenzentren in Smart Grids verwendet, was auch in Abschnitt 3.2 thematisiert wird.

Grundsätzlich kann ein rechenzentrumsübergreifendes Lastmanagement von unterschiedlichen Standortparametern profitieren. Hier sind insbesondere die Strompreise [109] und klimatische Bedingungen [145] von Bedeutung. Letztere haben signifikanten Einfluss auf die Kosten der Kühlung. Die Optimierung erfolgt hier, indem die IT-Last zum jeweils günstigsten Standort verschoben wird, wobei eine Abwägung gegenüber den Migrationskosten (Energiekosten, die durch die Netzwerkübertragung entstehen, sowie direkte finanzielle Kosten aufgrund der benötigten Bandbreite) [80] erfolgt.

Damit ein einzelner Standort möglichst optimal ausgelastet werden kann, untersucht [117] das Erreichen eines vorgegebenen Auslastungsprofils durch Kombination mehrerer virtueller Maschinen, die einem globalen VM-Pool entnommen werden. Dieses Prinzip wird in [118] mit dem Last- und Powermanagement (LPM) aus [53] kombiniert und für ein rechenzentrumsübergreifendes Lastmanagement erweitert.

#### Kritik und Abgrenzung

Das rechenzentrumsübergreifende Lastmanagement benötigt zwingend mehrere Rechenzentrumsstandorte, damit überhaupt eine Optimierung erreicht werden kann. Sind mehr Standorte mit einer größeren Distanz und größerer Verschiedenheit im Bezug auf Stromkosten und Klima vorhanden, profitieren die Verfahren. Allerdings sind die angenommenen

Anwendungsfälle teilweise unrealistisch. So wird in [119] davon ausgegangen, dass die Datenquellen der zu migrierenden Anwendungen ebenfalls mit migriert werden können, was in der Realität nur selten der Fall sein dürfte. Hier müssen Abhängigkeiten weiterer Anwendungen zu der gleichen Datenquelle betrachtet werden. Auch spielen Datenschutzgesetze und SLA eine Rolle, wenn es um automatisiertes Verschieben von Anwendungen in andere Regionen oder Länder geht.

Aus diesen Gründen betrachtet das in dieser Arbeit entwickelte DaMucA-Verfahren nur ein einzelnes Rechenzentrum, untersucht die energetischen Einflüsse und Freiheitsgrade und nimmt hier eine lokale Optimierung vor. Dadurch vergrößert sich die Menge der potentiellen Anwender und selbst bei verschiedenen Rechenzentrumsstandorten desselben Anbieters ist eine individuelle Optimierung unabhängig voneinander möglich.

#### Vergleich der Lastmanagementverfahren

In Tabelle 3.1 ist ein zusammenfassender Vergleich der unterschiedlichen Verfahren dargestellt. Die einzelnen Verfahren sind dabei entsprechend der Unterkapitel aus diesem Kapitel klassifiziert. Die lokale Anwendung bezieht sich auf die Möglichkeit, ein einzelnes Rechen-

Tabelle 3.1: Vergleich der Lastmanagementverfahren

| Verfahren                                   | Lokale Anwendung | Optimierung Energiebedarf | Thermische Optimierung | Flexibilitätsbetrachtung |
|---|------------------|---------------------------|------------------------|--------------------------|
| Lokales Lastmanagement                      | Ja               | Ja                        | Nein                   | Nein                     |
| Lastmanagement + thermische Aspekte         | Ja               | Teilweise                 | In IT-Räumen           | Nein                     |
| Multi-kriterielles Lastmanagement           | Ja               | Ja                        | In IT-Räumen           | Nein                     |
| Rechenzentrumsübergreifendes Lastmanagement | Nein             | Ja                        | Nein                   | Nein                     |
| <b>DaMucA</b>                               | <b>Ja</b>        | <b>Ja</b>                 | <b>Ja</b>              | <b>Ja</b>                |

zentrum zu behandeln, anstatt mehrere Rechenzentren im Verbund zu optimieren. Das Lastmanagement mit der Berücksichtigung thermischer Aspekte nimmt zwar auch energetische Optimierungen vor, allerdings liegt der Fokus auf der thermischen Lastverteilung. Zudem werden teilweise implizite Annahmen einer möglichen Pausierung oder Unterbrechung der Anwendungen getroffen, wodurch die genannten Einsparungen erreicht werden. Die thermische Optimierung beinhaltet sowohl die thermische Stabilität im IT-Raum, als auch die gesamte Optimierung des Kühlkreislaufs. Die genannten Arbeiten beziehen sich

jedoch nur auf die Untersuchung und Optimierung innerhalb der IT-Räume. Keines der Verfahren nimmt eine Flexibilitätsbetrachtung vor.

Insgesamt ist zu erkennen, dass nur das DaMucA-Verfahren, das in dieser Arbeit entwickelt wird, alle geforderten Aspekte unterstützt.

## 3.2 Energetische Steuerung

In dieser Arbeit werden energetische Einflüsse in Rechenzentren analysiert. Die aus diesen Analysen resultierenden Ergebnisse und Strategien können zur gezielten energetischen Steuerung von Rechenzentren verwendet werden. Das hier entwickelte Verfahren nutzt diese Ergebnisse, fokussiert aber auf einer energetischen Optimierung, also auf einem möglichst niedrigen Energiebedarf des Rechenzentrums. Allerdings wird unter bestimmten Umständen auch ein höherer Energiebedarf erlaubt, wenn dadurch andere Vorteile entstehen. Dadurch wird auch eine energetische Steuerung des Rechenzentrums möglich. Diese und die Regelbarkeit von Rechenzentren sind auch Gegenstände aktueller Forschung. Die in der aktuellen Forschung formulierte energetische Steuerung von Rechenzentren beschreibt primär die Möglichkeit, die elektrische Leistungsaufnahme innerhalb festgelegter Schranken aufgrund äußerer Eingaben einem Zielwert anzupassen. Dies ist insbesondere bei der Integration von Rechenzentren in Smart Grids oder Smart Cities notwendig. Hier geht es darum, ein Rechenzentrum als einen Teilnehmer von Demand Response-Verfahren zu etablieren [3, 39, 42, 43], um einerseits von unterschiedlichen Marktpreisen profitieren zu können [15, 40] und andererseits eine Stabilisierung des Stromnetzes zu unterstützen. Dies wird beispielsweise durch zeitliches Verschieben der Aufgaben [74] oder durch Regulierung der CPU-Leistung der Server [23, 24, 71] erreicht. Auch gibt es Ansätze, diese Regelung über mehrere verteilte Rechenzentrumsstandorte zu realisieren [117, 140, 141].

In [41] wird die Integration eines Rechenzentrums in ein allgemeines Stromnetz mit erneuerbaren Energiequellen vorgeschlagen, die Reaktion auf Stromschwankungen erfolgt über eine Verteilung der anfallenden IT-Last auf verschiedene Standorte je nach Verfügbarkeit von erneuerbarer Energie. Ähnlich geht auch [150] vor. In [85] schlagen die Autoren vor, dass ein Rechenzentrum das Stromnetz in Bezug auf Stabilität und Lastverteilung unterstützen soll, dazu werden *Service Request Routings* für Rechenzentren äquivalent zu *Power Flow Routings* in Smart Grids eingesetzt. Damit soll das gezielte Auslasten von Rechenzentren erreicht werden, um so den Energiebedarf zu steuern. Allerdings basieren diese Verfahren auf der Verfügbarkeit von mehreren Rechenzentrumsstandorten, wobei mindestens ein Standort für einen gegebenen Zeitpunkt nutzbar sein muss.

Mehrere Arbeiten beschäftigen sich mit der Möglichkeit, durch Scheduling auf Stromschwankungen zu reagieren. In [88] wird ein Aufgaben-Scheduling für virtualisierte Rechenzentren vorgeschlagen. Das Konzept schlägt ein zeitliches Verschieben und Pausieren der Aufgaben je nach Verfügbarkeit von energieeffizienter Hardware und thermischer Situation im Rechenzentrum vor. Dadurch werden Energieeinsparungen im Kühlsystem erzielt. Ähnliche Ansätze finden sich in [1] und [67], wobei hier der Ressource-Scheduler *Torque* [126] für das Taskmanagement verwendet wird.

#### **Abgrenzung**

Das Verfahren, das in dieser Arbeit entwickelt wird, unterscheidet sich deutlich von den bereits existierenden aus dem Kontext der energetischen Steuerung von Rechenzentren. Es nutzt langfristige Planungen, während die energetische Steuerung, beispielsweise zur Unterstützung von Smart Grids, auf sehr kurzfristige Anpassungen setzt. Außerdem bezieht sich das Verfahren in dieser Arbeit auf einen einzelnen Rechenzentrumsstandort und zeitliche Verschiebungen der IT-Auslastung können nicht vorgenommen werden, da diese durch SLA verhindert werden und die untersuchten Anwendungen zudem keine einmaligen, pausierbaren Aufgaben darstellen, sondern permanent verfügbare Dienste beschreiben. Allerdings können Teile des Verfahrens und die in dieser Arbeit vorgenommene Analyse der energetischen Einflüsse die bisher in der Forschung existierenden Steuerungskonzepte unterstützen.

#### **Power Budgets und Power Capping**

Im Gegensatz zur zeitlichen oder örtlichen IT-Lastverschiebung zur energetischen Steuerung von gesamten Rechenzentren können auch technische Mittel zur energetischen Regelung der Server eingesetzt werden. Eine solche technische Möglichkeit ist das Power Capping [49, 90]. Dabei wird die Energieaufnahme eines Servers gezielt reduziert, indem die maximale Leistung der Prozessoren künstlich beschränkt wird. Dies kann beispielsweise durch das Setzen von P-States (CPU-Power-States) oder durch Dynamic Voltage and Frequency Scaling (DVFS) [70, 104] geschehen.

Power Capping wird neben der Regulierung des Energiebedarfs des Rechenzentrums auch dazu genutzt, die maximale Energieaufnahme mehrerer Server zu beschränken und somit Überlastungen von unterstützenden Geräten (USV, Stromverteilung) zu verhindern. Gleichzeitig kann eine energetische Optimierung erreicht werden, indem allen Geräten Power Budgets zur Verfügung gestellt werden, die zu einem effizienten Betrieb führen und nicht überschritten werden dürfen [149]. Diese Regelung kann auch für das gesamte Rechenzentrum durchgeführt werden, wodurch eine hierarchische Stromsteuerung erreicht werden kann [133, 142].

Allerdings führt der Einsatz von Power Capping im Normalfall zu einer Verlangsamung des Servers und damit auch der Anwendungen. Insofern ist eine vollständige energetische Regelung des Rechenzentrums mit Power Capping nur in bestimmten Fällen möglich, wenn Auswirkungen auf die Anwendungen nicht durch SLA untersagt sind. Im Kontext dieser Arbeit wird der Einsatz von Power Capping für eine energetische Optimierung aus diesen Gründen ausgeschlossen. Allerdings kann Power Capping dazu eingesetzt werden, energetische Lastspitzen in einem optimierten Betriebszustand abzufangen, um damit Überlastungen beispielsweise bei der Kühlung zu verhindern. In Kapitel 6.2.1 wird dieser Ansatz im Kontext des Optimierungsverfahrens aus dieser Arbeit noch einmal kurz aufgegriffen.

## 3.3 Rechenzentrumsmodellierung

Die in den Abschnitten 3.1 und 3.2 vorgestellten Forschungsarbeiten fokussieren häufig auf Optimierungsverfahren für Rechenzentren, nutzen aber grundsätzlich Modelle zu einzelnen Geräten in Rechenzentren. Dies sind meist einzelne Modelle zur elektrischen Leistungsaufnahme von Servern oder Kühlgeräten oder die speziellen Modelle zur Abschätzung der Thermik im IT-Raum.

Einige Arbeiten haben sich auf die Modellierung der Zusammenhänge aller Geräte im Rechenzentrum spezialisiert, ohne dabei die jeweiligen Modelle zur Anwendung einer Optimierung zu bringen. So nimmt [103] eine allgemeine Modellierung der wichtigsten Geräte im Rechenzentrum vor. Die Autoren kombinieren hier existierende Modelle und erstellen so eine ganzheitliche Rechenzentrumsmodellierung. Es werden auch Wärmeleistung und Wärmeströme der Kältemaschinen und einer möglichen Freikühlung behandelt. Die Betrachtung von Umluftkühlgeräten erfolgt aber nur sehr rudimentär und die verwendeten Modelle werden lediglich zur Abschätzung der elektrischen Leistungsaufnahme der Ventilatoren verwendet. Kältemaschinen werden mit Hilfe der Modelle von [86] betrachtet. Ähnlich geht auch [32] vor.

Modelle der elektrischen Leistungsaufnahme für multi-core Server in virtualisierten Rechenzentren finden sich in [102]. Diese Modelle beziehen sich auf detaillierte Untersuchungen der Prozessoren unter einer hohen zeitlichen Auflösung. In [48, 69] erfolgt die Modellierung der elektrischen Leistungsaufnahme eines Servers unter Berücksichtigung der Nutzung von Festplatte, Speicher und Prozessor. Diese Arbeiten untersuchen unter anderem die energetischen Auswirkungen von Speicherzugriffen und Prozessorzuständen auf Chip-Ebene. Eine derartige Genauigkeit wird in dieser Arbeit nicht benötigt und kann aufgrund der dafür benötigten Daten und Messungen auch nicht realisiert werden. Die in Kapitel 5.3.1 vorgestellten Modelle orientieren sich daher an denen aus [32, 103], wobei die Charakterisierungen auf eigenen Messungen beruhen.

### Netzwerkmodelle und VM-Migrationskosten

Das DaMucA-Verfahren, das in dieser Arbeit zur Optimierung des Rechenzentrumsbetriebs entwickelt wird, berücksichtigt auch die Dauer von VM-Migrationen. Dazu gibt es bereits einige Vorarbeiten. So behandelt [77, 125] eine Abschätzung der zu übertragenden Datenmengen bei Live-Migrationen aufgrund der VM-Größe und der auftretenden Page-Dirty-Rate des VM-Speichers. Ähnlich geht auch [72] vor, wobei die Autoren eine energetische Kostenabschätzung von Live-Migrationen hinzufügen. Die veröffentlichten Modelle benötigen ein detailliertes Auslesen der Page-Dirty-Rate, was nach aktuellem Stand nur mit Eingriff in die VM and damit mit invasiven Methoden möglich ist. Daher können die Modelle aus den Arbeiten für das Verfahren in dieser Arbeit nicht verwendet werden.

In [57, 115, 127] wurde eine Vermessung der Dauer und der energetischen Auswirkungen von Live-Migrationen vorgenommen, wobei [55] auch die Auswirkungen unterschiedlicher Hypervisoren untersucht. Allerdings wurden die Untersuchungen unter Anwendung veralteter Netzwerktechnik vorgenommen. Dies führt zu der Angabe von hohen Migrationszeiten.

## 3.4 Zusammenfassung

In diesem Kapitel wurde eine Betrachtung der verwandten Arbeiten vorgenommen. Dabei wurden sowohl direkt vergleichbare Lastmanagementverfahren vorgestellt, als auch Techniken und Konzepte aus dem Bereich der energetischen Steuerung von Rechenzentren, die nicht direkt mit dem Vorgehen aus dieser Arbeit vergleichbar sind, aber trotzdem im selben Kontext zu verorten sind. Es wurde herausgestellt, dass die existierenden Lastmanagementverfahren zwar einzelne Optimierungen vornehmen und ebenfalls thermische Untersuchungen von Rechenzentren durchgeführt wurden, aber die kombinierte und ganzheitliche Betrachtung fehlt.

Schließlich wurden Arbeiten aus dem Kontext der Rechenzentrumsmodellierung vorgestellt, von denen einige in Kapitel 5 wieder aufgegriffen werden.

In Kapitel 4 folgt nun die Analyse der energetischen Einflüsse im Rechenzentrum und die Beschreibung der Optimierungskriterien.

# 4 Energetische und betriebsstabile Optimierung als multi-kriterielles Problem

In diesem Kapitel wird zunächst die Problemstellung dargelegt, die in dieser Arbeit behandelt wird und es wird kurz auf das Optimierungsverfahren eingegangen, das dieses Problem lösen soll. Da das Verfahren grundsätzlich einen Fokus auf die energetisch optimierte Nutzung des Rechenzentrums legen soll, wird zunächst analysiert, welche Einflussfaktoren auf die elektrische Leistungsaufnahme des Rechenzentrums existieren. Hierbei wird ebenfalls untersucht, welche Freiheitsgrade und Wechselwirkungen aus diesen Einflüssen entstehen, damit diese im Optimierungsverfahren gezielt ausgenutzt werden können.

Zusätzlich werden weitere Bedingungen für den korrekten und stabilen Rechenzentrumsbetrieb formuliert und schließlich in den Kriterien für das multi-kriterielle Optimierungsproblem zusammengefasst. Diese Kriterien werden dann in jeweils eigenen Unterkapiteln genau formuliert und formal dargelegt.

## 4.1 Problembeschreibung

Die Nutzung von Virtualisierung erlaubt die Lastverteilung im Rechenzentrum in unterschiedlichen Allokationen. Die Möglichkeit, verschiedene Allokationen mit der bestehenden Anwendungslast vorzunehmen ohne die Anwendungen einzuschränken oder abzuändern, wird hier als ein Freiheitsgrad des Rechenzentrums bezeichnet. Die Anwendungslast, die in einzelnen virtuellen Maschinen gekapselt ist, kann also auf beliebige physikalische Server im ganzen Rechenzentrum verteilt werden, solange jede einzelne virtuelle Maschine ihre beanspruchten Ressourcen auch tatsächlich bekommen kann. Eine Überbelastung von Servern soll also ausgeschlossen werden; sollte diese aufgrund plötzlicher Lastschwankungen auftreten, so muss der Server umgehend entlastet werden, was eine Migration von mindestens einer der virtuellen Maschinen des Servers auf einen anderen bedeutet.

Diese relativ freie Verteilung der Last im Rechenzentrum kann nun dazu verwendet werden, Server-Konsolidierung zu betreiben, also die virtuellen Maschinen auf so wenig Servern wie möglich unterzubringen und gleichzeitig die nicht genutzten Server auszuschalten (vgl. Abbildung 4.1 a) und b)). Durch diese Technik lassen sich Energieeinsparungen bei den Servern erreichen. Die bisherige Forschung hat sich beim Thema Konsolidierung auf die minimal benötigte Anzahl der aktiven Server konzentriert, ohne dabei mögliche Seiteneffekte zu betrachten, die entweder zu Problemen oder sogar erhöhtem Energiebedarf des gesamt-

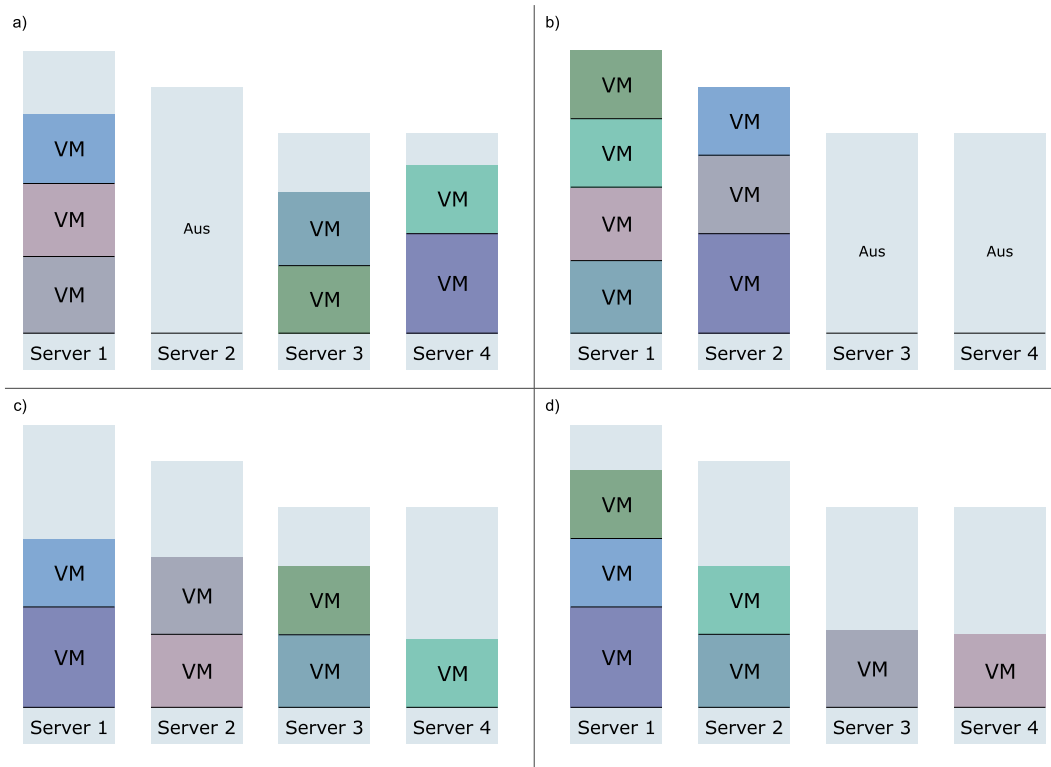


Abbildung 4.1: *Unterschiedliche Allokationen der gleichen virtuellen Maschinen: In den Allokationen a) und b) wurden Server-Konsolidierungen vorgenommen, wobei Allokation b) eine maximale Konzentration der virtuellen Maschinen auf möglichst wenige Server zeigt. Eine Flexibilität im Hinblick auf Lastveränderungen ist hier nicht gegeben. Hier muss also bereits eine Abwägung stattfinden, ob die zusätzlichen Energieeinsparungen in Allokation b) gegenüber a) tatsächlich zu vertreten sind. Gleichzeitig können auch flexiblere Allokationen, wie in c) und d) gezeigt, unterschiedliche Auswirkungen haben. Allerdings kann ein Optimierungsverfahren nicht alle beliebigen VM-Kombinationen überprüfen, sodass hier eine Vereinfachung der Bewertung gefunden werden muss.*

ten Rechenzentrums führen können. Genau diese Effekte sollen in dieser Arbeit untersucht werden, wobei die möglichen Betriebszustände dabei im gesamten Raum der energetischen Freiheitsgrade der Geräte im Rechenzentrum betrachtet werden.

Als mögliche energetische Regelung des Rechenzentrums steht ausschließlich die unterschiedliche Verteilung von virtuellen Maschinen auf physikalische Server und die Abschaltung ungenutzter Geräte zur Verfügung. In Abbildung 4.1 werden dazu einige mögliche Fälle beispielhaft betrachtet. So kann die gleiche Menge virtueller Maschinen sehr dicht auf möglichst wenige Server konzentriert werden (Abbildung 4.1 a) und b) ) oder es werden mehr Server als eigentlich benötigt eingeschaltet und die virtuellen Maschinen dann auf



diesen verteilt (Abbildung 4.1 c) und d)). Im ersten Fall führt dies zu Energieeinsparungen bei den Servern, während im zweiten aufgrund der zusätzlich verfügbaren Server-Ressourcen plötzlich auftretende Lastspitzen besser abgefangen werden können. Auch nur sich gering unterscheidende Allokationen, wie in Abbildung 4.1 c) und d) dargestellt, können aufgrund weiterer Einflussfaktoren zu teils signifikant unterschiedlichen Betriebszuständen des Rechenzentrums führen. Dazu zählt unter anderem die räumliche Verteilung der Server, auf denen sich die virtuellen Maschinen befinden oder die Kombination der tatsächlich genutzten Server und deren Effizienz. Diese Abhängigkeiten und Wechselwirkungen werden in Kapitel 4.1.1 dediziert analysiert und in weiteren Kapiteln immer wieder aufgegriffen. Es wird also deutlich, dass prinzipiell eine Vielzahl von unterschiedlichen Allokationen möglich ist, eine Bewertung dieser jedoch je nach Sichtweise und Priorität unterschiedlich ausfallen kann. Damit ein Optimierungsverfahren deterministisch vorgehen kann, um eine gewünschte Allokation zu finden, werden genau diese verschiedenen Sichtweisen mit Hilfe von formal definierten Kriterien abgebildet. Dabei beinhalten diese Kriterien sowohl mögliche Vorgaben, die ein Rechenzentrumsbetreiber oder Kunde formuliert hat, als auch harte Bedingungen, die für den betriebssicheren Zustand des Rechenzentrums immer gelten müssen, beispielsweise um Überhitzung oder Überlastung zu vermeiden. Die Identifikation dieser Kriterien erfolgt nach der Analyse der Einflussfaktoren in Unterkapitel 4.1.2.

Auf Basis der Analysen aus diesem Kapitel wird in Kapitel 6 ein Optimierungsverfahren entwickelt, das die identifizierten Kriterien anwendet und dementsprechend die virtuellen Maschinen auf die Server im Rechenzentrum verteilt, wodurch eine nach diesen Kriterien optimierte Allokation eingenommen wird, die dann in einem optimierten Betriebszustand des Rechenzentrums resultiert. Damit das Optimierungsverfahren angewendet werden kann, gilt die Voraussetzung, dass der Rechenzentrumsbetreiber auch tatsächlich vollen Zugriff auf alle Server und virtuelle Maschinen im Rechenzentrum hat, damit er beliebige Migrationen vornehmen kann. Ein Zugriff auf den Inhalt der virtuellen Maschinen, also auf die Anwendungen, ist dafür aber nicht notwendig und wird sogar gezielt ausgeschlossen. Das Verfahren besitzt also keine Informationen über die ausgeführten Anwendungen, sondern nur auf Ebene der Virtualisierung.

Zusammenfassend ergibt sich aus dieser Problembeschreibung eine zentrale Fragestellung, die in dieser Arbeit behandelt wird und deren Details in diesem Kapitel weiter ausgearbeitet werden:

*Welche Kriterien gibt es bei der Ausnutzung des Freiheitsgrades des Rechenzentrums bezüglich der Verteilung der Anwendungslast und wie kann ein Verfahren diese Kriterien in der Allokationsplanung vereinen, sodass ein optimierter Betriebszustand des Rechenzentrums hinsichtlich dieser Kriterien erreicht werden kann?*

### **4.1.1 Energetische Einflussfaktoren im Rechenzentrum**

In diesem Abschnitt werden die energetischen Einflussfaktoren im Rechenzentrum untersucht, die zum einen maßgeblich die elektrische Leistungsaufnahme des gesamten Rechenzentrums bestimmen und zum anderen für den stabilen Betrieb untersucht und beachtet werden müssen. Diese energetischen Analysen gelten sowohl für elektrische Energie als auch

für die thermische Energie, die im Normalfall in Form von Abwärme anfällt und aus den IT-Räumen und dem gesamten Gebäude abtransportiert werden muss.

In Abbildung 4.2 wird der Energiefluss der energetisch gesehen wichtigsten Geräteklassen im Rechenzentrum dargestellt. Außerdem sind in der Abbildung die Abhängigkeiten untereinander dargestellt. Anhand der Gruppen in der Darstellung werden nun die Einfluss-

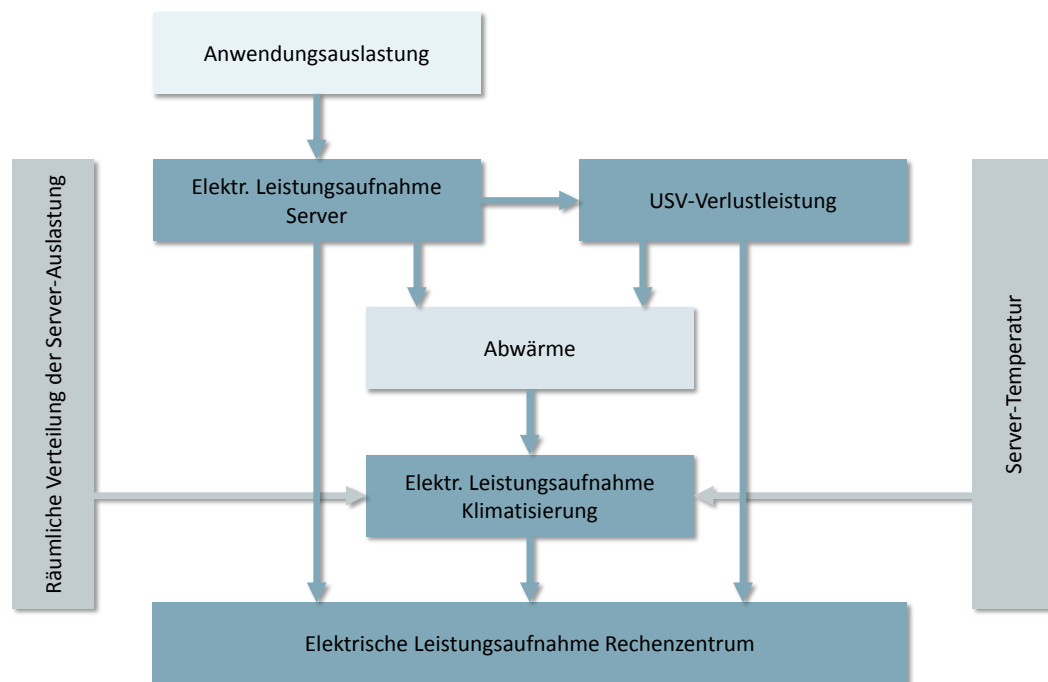


Abbildung 4.2: *Energetische Einflussfaktoren im Rechenzentrum: Die Anwendungsauslastung ist die primäre Größe, die die elektrische Leistungsaufnahme eines Servers bestimmt. Letztere wirkt sich auf die Verlustleistung der USV aus und bestimmt mit diesem direkt die Menge an Abwärme, die durch die Klimatisierung abgeführt werden muss. Wie viel Energie die Klimatisierung dafür benötigt, hängt jedoch von weiteren Faktoren ab, wie den Wärmemengen in den einzelnen IT-Räumen und den maximalen Temperaturen bzw. Grenzwerten im IT-Raum und damit von den Servern.*

faktoren auf den Energiehaushalt des Rechenzentrums untersucht und auftretende Wechselwirkungen identifiziert.

### Anwendungsauslastung & Server

Die tatsächlich auftretende Anwendungsauslastung der virtuellen Maschinen auf einem Server bestimmt primär die elektrische Leistungsaufnahme des Servers, wobei hier die CPU-Auslastung die entscheidende Komponente ist. Dies wird in den Modellen in Kapitel 5.3.1 quantifiziert. Bei genutzter Konsolidierung gibt es zudem einen indirekten Einfluss:

je höher die gesamte Anwendungsauslastung im Rechenzentrum ist, um so mehr Server müssen eingeschaltet werden, wodurch der Energiebedarf steigt. Die Server selbst sind nach Erhebungen (siehe Kapitel 2.1) die Einzelgruppe mit dem größten Energiebedarf des Rechenzentrums. Damit ist die tatsächlich anfallende Anwendungslast ein erster Einflussfaktor auf die elektrische Leistungsaufnahme des Rechenzentrums.

Ein weiterer Faktor ist das Verhältnis der tatsächlich anfallenden Anwendungsauslastung zu der Auslastung, für die das Rechenzentrum ausgelegt und geplant wurde. Bei der Planung eines Rechenzentrums wird häufig die maximale Auslastung aller Server als Basis genommen und dementsprechend die Dimensionierung der unterstützenden Geräte wie der Klimatisierung und der USV vorgenommen. Wenn diese Dimensionierung nun signifikant über der tatsächlich anfallenden Auslastung liegt, entstehen energetische Ineffizienzen aufgrund des Betriebs dieser Geräte in einem ineffizienten Arbeitspunkt. Ein weiterer Einflussfaktor auf die elektrische Leistungsaufnahme des gesamten Rechenzentrums ist damit also das Verhältnis der geplanten Auslastung der Server zu der tatsächlichen Auslastung und der daraus resultierende Wirkungsgrad aller beteiligten Geräte. Ein direkt damit zusammenhängender Einflussfaktor ist die individuelle Regelbarkeit insbesondere von USV- und Klimageräten. Sind diese Geräte in einem größeren Bereich regelbar und können unter Umständen sogar einzelne Module deaktiviert werden, so kann selbst bei deutlicher Unterauslastung ein effizienter Wirkungsgrad erreicht werden.

Allerdings hat ein Optimierungsverfahren, wie es in dieser Arbeit entwickelt wird, auf die tatsächlich anfallende Anwendungsauslastung keinen Einfluss, es dürfen weder Anwendungen pausiert noch entfernt werden. Das Optimierungsverfahren kann aber über die Verteilung eine Anpassung der Wirkungsgrade der Server und einzelner weiterer Geräteklassen vornehmen, um damit auch in einem ungünstigen Fall der tatsächlich vorhandenen Anwendungsauslastung trotzdem noch energetische Effizienzverbesserungen vornehmen zu können. Außerdem steht die Server-Konsolidierung zur Verfügung, sodass bei geringerer Anwendungsauslastung weniger Server eingeschaltet werden können, wodurch hier zunächst die elektrische Leistungsaufnahme der Server sinkt. Aufgrund der Abhängigkeiten zu den unterstützenden Geräten wie den USV-Geräten und der Klimatisierung können auch hier Energieersparnisse erreicht werden. Dies hängt jedoch deutlich von deren Regelbarkeit und weiteren Faktoren ab, die in den entsprechenden folgenden Abschnitten beschrieben werden.

### **Stromversorgung**

Energetisch gesehen leiten die Geräte der unterbrechungsfreien Stromversorgung nicht nur den elektrischen Strom weiter, sondern sie sorgen auch für eine Art „Aufbereitung“, durch die wiederum ein eigener Energiebedarf entsteht (siehe dazu auch die Beschreibungen in Kapitel 2.1.3). Dieser ist ebenfalls signifikant hinsichtlich des Gesamtanteils des Rechenzentrumsbedarfs (siehe auch hier die Aufteilung in Kapitel 2.1). Die Effizienz der USV-Geräte beschreibt das Verhältnis der Nutzlast zu diesem Eigenbedarf und hängt primär von deren prozentualer Auslastung ab, je geringer diese ist, desto höher ist im Allgemeinen die Verlustleistung. Eine Quantifizierung dieses Effekts findet sich in Kapitel 5.2.

Die Effizienz und damit der Energiebedarf der USV hängen also wiederum primär vom Energiebedarf der Server, die von der USV versorgt werden, ab. Allerdings besitzen moderne und praktisch alle größeren Rechenzentren zentrale USV-Anlagen mit großen Kapazitäten, die mehrere IT-Räume versorgen. Gleichzeitig sind diese USV aus Zertifizierungsgründen redundant aufgebaut. Ein Abschalten einzelner Geräte ist damit nicht möglich. Somit ist auch eine direkte Regelbarkeit der Energieaufnahme und der Verlustleistung der USV-Geräte nicht möglich, der Energiebedarf stellt sich vielmehr implizit durch den Energiebedarf der angeschlossenen Geräte ein.

Vor allem da solche USV-Anlagen häufig überdimensioniert angeschafft werden, um beispielsweise zukünftigen Anforderungen zu genügen, kann eine hohe USV-Auslastung nur selten erreicht werden. Allerdings sind solche Anlagen intern meist modular aufgebaut, sodass eine automatische Regelung je nach Leistungsbereich erfolgen kann. Auch muss beachtet werden, dass aufgrund einer geringeren elektrischen Leistungsaufnahme der Server zwar unter Umständen die Effizienz der USV sinken kann, trotzdem werden aber absolute Energieeinsparungen auch auf Seiten der USV erreicht. Diese verhalten sich eben nur nicht proportional zur gesunkenen Nutzlast.

Im Optimierungsverfahren wird aus diesen genannten Gründen keine direkte Optimierung hinsichtlich der USV vorgenommen. Allerdings wird der energetische Einfluss der USV hinsichtlich der Energieaufnahme und Verlustleistung modelliert und im Bezug auf den Energiebedarf der Klimatisierung und den gesamten Energiebedarf des Rechenzentrums beachtet.

### **Klimatisierung**

Die gesamte Klimatisierung stellt im Rechenzentrum nach den Servern den zweitgrößten Energiekonsumenten dar (siehe Kapitel 2.1). Der Einfluss auf den gesamten Energiebedarf des Rechenzentrums ist aber komplex und muss differenziert betrachtet werden: Zum einen muss der zentral vorhandene Kältekreislauf inklusive Kälteerzeugung betrachtet werden, zum anderen die Kälteverteilung in den IT-Räumen, die hauptsächlich durch die Umluftkühlgeräte gewährleistet wird. Die Abbildung 4.3 zeigt diesen Sachverhalt als eine erweiterte Darstellung der Abbildung 2.4 aus Kapitel 2.1.4. Die Kälteerzeugung, die als wesentliche Stromverbraucher die Kältemaschinen besitzt, wird dazu verwendet, das Kaltwasser bereitzustellen. Über dieses wird die benötigte Kühlleistung in die IT-Räume transportiert und dort in den Umluftkühlgeräten mittels Wärmeübertragern als Kaltluft in die Kaltgänge verteilt. Dies bedeutet aber auch, dass die Menge der Kälteleistung durch die anfallende Wärmemenge in den IT-Räumen bestimmt wird. Diese Wärmemenge hängt nach Gleichung 2.2 direkt von der elektrischen Leistungsaufnahme der IT-Räume ab.

Die elektrische Leistungsaufnahme  $P_r$  des IT-Raums  $r$  setzt sich aus der Summe der Leistungsaufnahmen aller Geräte zusammen, die sich in diesem IT-Raum befinden. Thermodynamisch gesehen wird der Raum also als geschlossenes System betrachtet, bei dem alle Energie, die primär<sup>1</sup> in Form von Elektrizität dem Raum zugeführt, in Wärme gewandelt wird und damit auch als solche wieder abgeführt werden muss. Damit und mit den

---

<sup>1</sup>Praktisch gesehen gibt es weitere Einflüsse, wie beispielsweise menschliche Mitarbeiter, die einen Raum

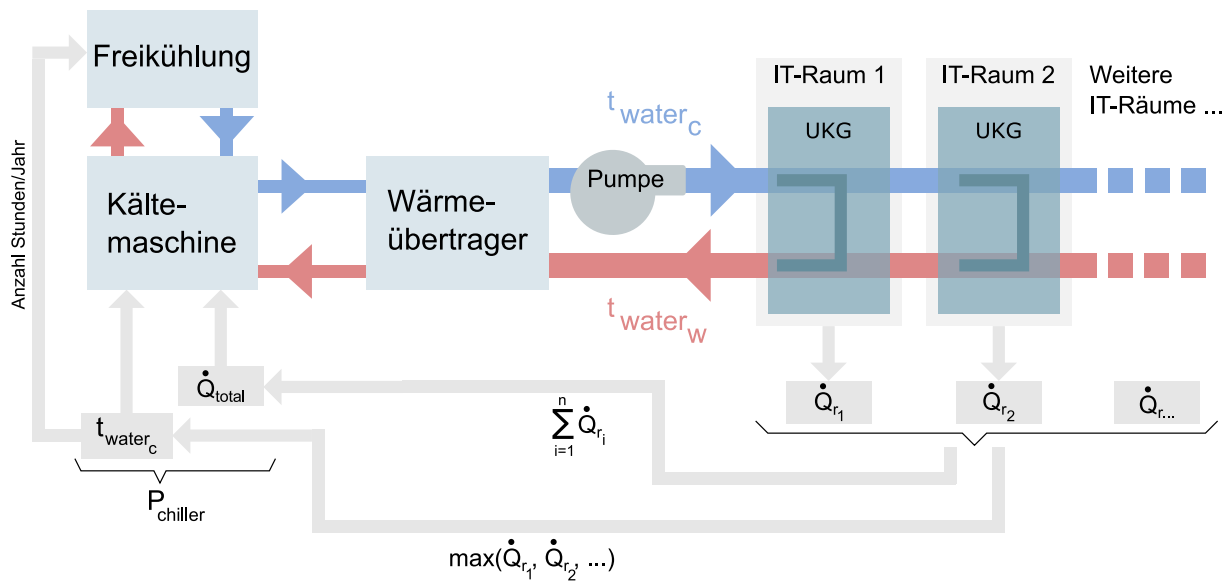


Abbildung 4.3: Die Abbildung zeigt eine vereinfachte Darstellung der Klimakette im Rechenzentrum, allerdings mit den jeweiligen Kenngrößen, die den Wärmefluss und die elektrische Leistungsaufnahme bestimmen. Die IT-Räume werden mittels Umluftkühlgeräten mit Kälte versorgt, diese Geräte hängen allerdings alle am selben Kühlkreislauf, sodass hier eine Abstimmung untereinander erfolgen muss. Damit ergibt sich ein Einfluss auf die elektrische Leistungsaufnahme der Kältemaschinen aus der Summe der Bedarfe der IT-Räume und dem jeweiligen Maximum, da dieses die Kaltwassertemperatur bestimmt.

Feststellungen aus Gleichung 2.2 gilt:

$$P_r \approx \dot{Q}_r \quad (4.1)$$

Somit befindet sich in jedem IT-Raum  $r$  eine anfallende Wärmemenge  $\dot{Q}_r$ , die durch die Umluftkühlgeräte abtransportiert werden muss. Anders ausgedrückt muss ein Umluftkühlgerät in der Lage sein, eine entsprechende Kälteleistung bereitzustellen. Diese Kälteleistung ergibt sich maßgeblich aus der Temperatur des Kühlwassers  $t_{water_c}$ . Auch die Flussrate des Kühlwassers besitzt einen Einfluss, allerdings ist diese für Umluftkühlgeräte bauartbedingt nur in einem sehr engen Bereich regelbar und in der Praxis meist konstant. Zudem muss die Kälteleistung über das Kühlwasser durch die Kältemaschinen bereitgestellt werden und diese zeigen deutliche Energieeinsparungen bei erhöhten Kaltwassertemperaturen. Eine Optimierung durch die Anpassung der Kaltwassertemperatur ist also aus zwei Gründen sinnvoll.

betreten und damit Wärmeenergie zuführen. Diese Einflüsse sind allerdings kaum messbar und werden in dieser Arbeit nicht weiter betrachtet.

Die Kältemaschinen benötigen den größten Anteil der Energie in der gesamten Klimakette. Hier ergibt sich also auch das größte Einsparpotential. Die elektrische Leistungsaufnahme  $P_{chiller}$  der Kältemaschinen wird maßgeblich durch zwei Faktoren bestimmt:

1. *Wärmemenge*: Die gesamte Wärmemenge  $\dot{Q}_{total}$ , die abtransportiert, also gekühlt werden muss, entspricht gleichzeitig der benötigten Kältemenge. Dieser Wert setzt sich aus der Summe der Wärmemengen aller IT-Räume (und grundsätzlich aller zu kühlenden Räume) zusammen:

$$\dot{Q}_{total} = \sum_{i=1}^n \dot{Q}_{r_i} \quad (4.2)$$

mit  $n$  als der Anzahl der zu kühlenden Räume.

2. *Kaltwassertemperatur*: Die Kaltwassertemperatur  $t_{water_c}$  hat einen signifikanten Einfluss auf die elektrische Leistungsaufnahme der Kältemaschinen [87, 103], was auch in der Modellierung in Kapitel 5.5 gezeigt wird. Je höher diese Temperatur ist, desto weniger Energie wird für das Aufbringen der Kältemenge benötigt [18]. Hier ergeben sich also direkte Energieeinsparungen bei einer Erhöhung der Kaltwassertemperatur. Gleichzeitig ist die Kaltwassertemperatur auch für die Anzahl der Stunden im Jahr maßgeblich, an denen Freikühlung eingesetzt werden kann und somit die Kältemaschinen ausgeschaltet bzw. reduziert werden können, was die Klimatisierungskosten für diesen Zeitraum signifikant senkt [103, 145]. Je höher die Kaltwassertemperatur, desto häufiger kann Freikühlung eingesetzt werden, da eine maximale Außentemperatur nicht überschritten werden darf, um diese Kaltwassertemperatur durch Freikühlung zu erreichen.

Ein offensichtliches Optimierungsziel im Kontext der Klimakette hinsichtlich der Energieeffizienz ist also das Anheben der Kaltwassertemperatur auf einen Wert, der einerseits die benötigte Kältemenge für die Umluftkühlgeräte sicherstellt und andererseits so hoch wie möglich liegt, um Energiekosten zu senken. Daher wird nun analysiert, welche Einflüsse die Kaltwassertemperatur bestimmen und welche Faktoren berücksichtigt werden müssen, damit diese angehoben werden kann.

In Abbildung 4.3 ist zu erkennen, dass alle IT-Räume im Rechenzentrum am gleichen, zentralen Kaltwassersystem angeschlossen sind, aus dem die Umluftkühlgeräte ihre Kälteleistung beziehen. Eine unabhängige Regelung für einzelne Räume ist damit nicht möglich, die Kaltwassertemperatur im Kreislauf gilt für alle IT-Räume und die sich darin befindlichen Umluftkühlgeräte. Gleichzeitig wird die von den Umluftkühlgeräten benötigte Kaltwassertemperatur durch die Wärmemenge im jeweiligen IT-Raum bestimmt. Damit wird deutlich, dass sich die Kaltwassertemperatur des gesamten Kühlkreislaufs an dem IT-Raum orientieren muss, der die höchste Wärmeleistung besitzt, da auch für diesen Raum die adäquate Kühlung gewährleistet werden muss. Es ergibt sich also für  $t_{water_c}$  die Abhängigkeit

$$t_{water_c} = f(\max(\dot{Q}_{r_1}, \dot{Q}_{r_2}, \dots, \dot{Q}_{r_n})) \quad (4.3)$$

mit  $n$  als der Anzahl der IT-Räume im Rechenzentrum und der fiktiven Funktion  $f$ , die mittels Modellgleichungen die genaue Abhängigkeit quantifiziert. In Kapitel 5.4.2 wird die Modellierung der Geräte der Klimatisierung vorgenommen. Dort wird dann auch die Zusammensetzung der Funktion  $f$  analysiert.

Aus dieser Analyse heraus ergibt sich ein maßgeblicher Einflussfaktor von  $t_{water_c}$  auf die elektrische Leistungsaufnahme der Kühlung und damit auf einen signifikanten Anteil der Leistungsaufnahme des gesamten Rechenzentrums. Dieser Einfluss kann über die Verteilung der Wärmeleistung und deren Maxima und damit nach Gleichung 4.1 auch die Verteilung der elektrischen Leistung in den IT-Räumen im Rechenzentrum beeinflusst werden. Das Optimierungsziel ist hier also, eine möglichst gleichmäßige elektrische Leistungsverteilung mit gleichzeitigem Absenken des Maximums zu erreichen, damit die Kaltwassertemperatur  $t_{water_c}$  nicht durch IT-Räume mit Leistungsspitzen nach oben gesetzt werden muss. Gleichzeitig muss garantiert werden können, dass diese Leistungen eingehalten werden, da eine dynamische Regelung der Kaltwassertemperatur nur sehr langsam möglich ist.

### Zusammenfassender Vergleich der Einflussfaktoren

In der Tabelle 4.1 werden die beschriebenen Faktoren noch einmal zusammengefasst und gleichzeitig wird der Grad der Auswirkungen und die jeweilige Regelbarkeit kurz beschrieben. Insbesondere zeigt sich, dass einige Einflussfaktoren zwar eine signifikante Auswirkung auf den Energiebedarf des gesamten Rechenzentrums besitzen, jedoch keine oder nur eine sehr eingeschränkte Möglichkeit der Regelung besteht. Beispielsweise kann die Anwendungslast nicht verändert werden. Dies ist eine deutliche Abgrenzung zu existierenden Verfahren, die Energieeinsparungen erreichen, indem sie Anwendungen pausieren [88] oder diese auf andere - teils energetisch günstigere - Rechenzentrumsstandorte verteilen [41, 119].

Die USV-Verlustleistung, die aufgrund der meist niedrigen Auslastung der einzelnen USV-Geräte auftritt, besitzt zwar einen signifikanten Anteil an der Leistungsaufnahme des Rechenzentrums, ist aber aufgrund der genannten Redundanz und Dimensionierung der Geräte nicht explizit steuerbar. Hier muss auf die zunehmende Effizienz der neu auf den Markt kommenden Geräte auch in geringeren Auslastungsbereichen verwiesen werden. Zusätzlich gibt es Ansätze, USV-Geräte direkt in die Racks oder die Server zu integrieren. Dadurch wird eine feingranulare Regelung möglich. Allerdings befinden sich diese Ansätze aktuell noch in der Forschung oder sie werden in nur wenigen Rechenzentren testweise eingesetzt.

Die regelbaren Einflussfaktoren sind damit die elektrische Leistungsaufnahme der Server und die Verteilung der Wärmeleistung in den IT-Räumen, durch die eine Anpassung des Energiebedarfs der Klimatisierung erreicht werden kann. Damit die Freiheitsgrade dieser Geräte jedoch ausgenutzt werden können, muss eine Optimierung nach definierten Kriterien stattfinden, die zwar eine energetische Optimierung priorisieren, aber gleichzeitig auch einen stabilen Betriebszustand ermöglichen. Diese Kriterien werden im folgenden Abschnitt identifiziert.

Tabelle 4.1: Vergleich der Einflussfaktoren

| <b>Einflussfaktor</b>  | <b>Auswirkung</b>  | <b>Regelbarkeit</b>  |
|--|--|--|
| Tatsächliche Anwendungsauslastung  | Sehr hoch  | Keine, eine Änderung der Anwendungsausführung ist nicht vorgesehen                                     |
| Verhältnis tatsächlicher Anwendungsauslastung zu geplanter   | mittel - hoch, Auswirkung vor allem auf Wirkungsgrad unterstützender Geräte                  | Keine im laufenden Betrieb, sondern nur bei Planung des Rechenzentrums oder Neuanschaffung von Geräten |
| <b>Leistungsaufnahme der Server, Einfluss durch individuelle Auslastung, Effizienz und Anzahl aktiver Server</b> | <b>sehr hoch, sowohl auf Leistungsaufnahme der Server als auch auf unterstützende Geräte</b> | <b>Regelbarkeit durch Server-Konsolidierung, Lastmanagement und Priorisierung effizienter Server</b>   |
| Effizienz der USV bzw. USV-Verlustleistung   | mittel - hoch  | Keine explizite Regelung möglich   |
| <b>Wärmemenge in IT-Räumen und Leistungsmaxima</b>   | <b>sehr hoch</b>   | <b>Anpassung durch Lastverteilung über IT-Räume hinweg</b>   |

### 4.1.2 Identifikation der Kriterien

In diesem Abschnitt werden die einzelnen Kriterien, die sich aus der Analyse und eventuellen weiteren Anforderungen ergeben und die im Optimierungsprozess abgebildet werden sollen, identifiziert und kurz dargestellt. Es wird jeweils begründet, warum das jeweilige Kriterium gewählt wurde und welche Bereiche im Rechenzentrum bzw. im Optimierungsprozess durch das Kriterium abgedeckt werden.

Bei der Formulierung der Kriterien werden unterschiedliche und teilweise gegensätzliche Betriebsparameter des Rechenzentrums berücksichtigt. Grundsätzlich gilt für das zu entwickelnde Optimierungsverfahren jedoch immer, dass ein Betriebszustand des Rechenzentrums energetische Einsparungen im Vergleich zum nicht optimierten Betrieb erreichen soll. Es wird also im Optimierungsverfahren eine Priorisierung der Optimierung der elektrischen Leistungsaufnahme vorgenommen.

Insgesamt wurden folgende Kriterien identifiziert:



### **Optimierung der elektrischen Leistungsaufnahme**

Dieses Kriterium ergibt sich direkt aus der Motivation dieser Arbeit, da die Energieeffizienz und gezielte Energieeinsparung hier im Mittelpunkt steht. Damit soll grundsätzlich eine Allokation und ein daraus resultierender Betriebszustand mit möglichst geringem elektrischen Energiebedarf gefunden werden. Dies führt zu direkten Kosteneinsparungen und je nach Stromquelle zu deutlichen ökologischen Vorteilen wie  $CO_2$ -Einsparungen. Die zu erreichenden Energieeinsparungen gelten für alle Geräteklassen und beinhalten sowohl das Abschalten von Servern als auch eine Betrachtung der Effizienz der unterstützenden Geräte. Dabei erfolgt eine Konzentration auf die in der Analyse herausgearbeiteten Einflussfaktoren, die sowohl einen signifikanten Einfluss auf den gesamten Energiebedarf des Rechenzentrums haben und gleichzeitig auch einen Freiheitsgrad besitzen und damit regelbar sind. Damit liegt der Schwerpunkt bei der Optimierung der elektrischen Leistungsaufnahme der Server und der gezielten Effizienzsteigerung der Klimatisierung, die durch konkrete Lastverteilungen in den IT-Räumen im Rechenzentrum erreicht werden soll. Die Details des Kriteriums werden in Kapitel 4.2 beschrieben.

### **Thermische Stabilität**

Wenn aufgrund von energetischen Einsparungen die Leistungsaufnahme der Klimatisierung gesenkt werden soll, muss klar definiert werden, welche thermischen Grenzen einzuhalten sind, damit ein stabiler Betriebszustand des Rechenzentrums auch bei verminderter Leistung garantiert werden kann. Genau diese Bedingungen werden in diesem Kriterium formuliert. Diese Bedingungen und Grenzwerte resultierten zuteilen direkt aus den Anforderungen, die von Rechenzentrumsbetreibern und Kunden vertraglich festgehalten wurden. Diese regeln meist beispielsweise eine maximale Server-Temperatur und redundante Kühlung.

Unabhängig von der vertraglichen Situation wird auch bei historischer Betrachtung der Entwicklung von Rechenzentren schnell deutlich, dass durch eine steigende Energiedichte [64] eine stabile Kühlung immer komplexer wird. In [34] werden die Verfügbarkeit sowie die thermische und elektrisch-energetische Dichte im Rechenzentrum als einige der größten Herausforderungen von Rechenzentrumsbetreibern angegeben. Daraus kann abgeleitet werden, dass eine thermische Stabilität im Rechenzentrum einer der wichtigsten Aspekte für den Betrieb geworden ist, weshalb diese Anforderungen in diesem eigenen Kriterium behandelt werden.

Eine Allokation und der daraus resultierende Betriebszustand sind nur dann legal, wenn die Kühlgeräte die anfallende Wärmemenge auch abtransportieren können und zudem die Temperaturgrenzwerte für die einzelnen Server eingehalten werden. Dies und eine gleichmäßige Verteilung der Wärmemenge im IT-Raum tragen maßgeblich dazu bei, die Wahrscheinlichkeit des Auftretens von Hot-Spots signifikant zu senken. In diesem Kriterium spielt die energetische Effizienz der Kühlung keine Rolle, diese wird im vorher beschriebenen Kriterium behandelt. Allerdings muss das Kriterium der thermischen Stabilität auch eine energetische Betrachtung vornehmen, dies gilt aber im Bezug zur Wärmeenergie bzw. der

anfallenden Wärmemenge und den entsprechenden Kapazitäten der Kühlgeräte. In diesem Kriterium werden mehrere feste Bedingungen formuliert, die nicht verletzt werden dürfen und damit die Freiheitsgrade der anderen Kriterien unter Umständen einschränken. Eine detaillierte Beschreibung des Kriteriums findet in Kapitel 4.3 statt.

### **Flexibilität eines Betriebszustands**

Gerade wenn es um den Einsatz von Server-Konsolidierungsstrategien geht, die im Zuge der energetischen Optimierung vorgenommen werden, ist eine von Rechenzentrumsbetreibern häufig geäußerte Befürchtung die fehlende Flexibilität des konsolidierten Betriebszustands. Durch das Ausschalten von ungenutzten Servern und die Konzentration der virtuellen Maschinen auf wenige aktive Server steigt die Reaktionszeit auf unvorhergesehene Ereignisse. Treten beispielsweise Notfallsituationen wie Überhitzungen, Defekte oder plötzliche Lastspitzen auf, müssen in einer unflexiblen Allokation erst Server hochgefahren werden, bevor betroffene virtuelle Maschinen mittels Live-Migration von den betroffenen Servern evakuiert werden können. Wenn dagegen eine Konsolidierungsstrategie die Flexibilität einer Allokation mit einplant, kann diese Reaktionszeit von vorne herein gesenkt und für jeden einzelnen Server genau bestimmt werden.

Eine hohe Flexibilität einer Allokation bedeutet dabei, dass diese Verteilung noch genügend freie Ressourcen zur Verfügung hat, um auf plötzliche Lastschwankungen vorbereitet zu sein, ohne dass beispielsweise weitere Server eingeschaltet werden müssen. Damit sinkt also gleichermaßen auch die Energieersparnis durch die Konsolidierung. Hier ist also ein Abwägen je nach Bedürfnissen notwendig.

Ein weiterer Aspekt der Flexibilität bezieht sich direkt auf den Evakuierungsfall von virtuellen Maschinen von einem Server in einer Notsituation. Die Zeit, die benötigt wird, um einen betroffenen Server vollständig zu evakuieren, ist als kritisch zu betrachten und muss für jeden Server im Vorfeld angegeben und zudem unter einem maximalen Sollwert gehalten werden können. Auch diese Betrachtung wird unter dem Kriterium der Flexibilität vorgenommen.

Eine formale Definition der Flexibilität einer Allokation und die detaillierte Beschreibung des Kriteriums findet in Kapitel 4.4 statt.

## **4.2 Kriterium: Optimieren der elektrischen Leistungsaufnahme**

Prinzipiell haben alle elektrischen Geräte im Rechenzentrum Einfluss auf dessen Leistungsaufnahme. Allerdings gibt es Geräteklassen mit besonders hohem Bedarf und gleichzeitiger Möglichkeit der Regelbarkeit, während andere Geräte einen relativ geringen und eher statischen Bedarf aufweisen. Diese Unterscheidung wurde in Kapitel 2.1 genauer dargestellt und in 4.1.1 im Bezug auf die Einflussfaktoren analysiert. Für das Kriterium der Optimierung der elektrischen Leistungsaufnahme werden daher die dort identifizierten Geräteklassen der Server und der Klimatisierung betrachtet.

Die Optimierung und damit Reduktion der elektrischen Leistungsaufnahme bezieht sich auf das gesamte Rechenzentrum, Ziel ist also,  $P_{dc}$  aus Gleichung 2.1 zu reduzieren. Dies kann auch bedeuten, dass einzelne Bereiche des Rechenzentrums eine höhere elektrische Leistungsaufnahme haben, andere Bereiche aber derart signifikante Einsparungen verbuchen, sodass die gesamte Leistungsaufnahme damit gesenkt wird.

### 4.2.1 Server

Die Server im Rechenzentrum sind mit einem durchschnittlichen anteiligen Energiebedarf von 40% (siehe Kapitel 2.1) die größte Verbrauchergruppe und gleichzeitig ein signifikanter Faktor, da sich Einsparungen hier auf weitere Geräteklassen, insbesondere die Kühlung, fortsetzen.

Energetische Einsparungen im Bereich der Server lassen sich hauptsächlich durch zwei Strategien erreichen: Reduktion der Anzahl aktiver Server und Steigerung der Effizienz der dann noch aktiven Server.

Im Normalfall wird die größte Ersparnis dadurch erreicht, dass Server ausgeschaltet werden und die Last auf möglichst wenige Server konzentriert wird. Dies entspricht der klassischen Konsolidierungsstrategie [53]. Wird nur dieser Fall alleine betrachtet, entspricht dies dem klassischen Bin-Packing-Problem [10, 29], wobei die Server hierbei die Container (Bins) darstellen und deren Größe der verfügbaren Ressourcen entspricht. Die VMs stellen dann die Objekte (Items) dar, deren Größe wiederum dem Ressourcenbedarf entspricht. Ziel ist dann, alle Items durch dichtes Packen auf so wenige Bins wie möglich zu verteilen. Bezogen auf die Konsolidierungsstrategie bedeutet dies dann, dass alle ungenutzten Server ausgeschaltet werden können.

Gleichzeitig liegen signifikante energetische Potentiale in der Effizienzsteigerung von aktiven Servern. Diese Betrachtung der Effizienz muss dabei zwei Aspekte berücksichtigen. Zum einen sind verschiedene Server unterschiedlich effizient. Vor allem neuere Server können eine höhere Auslastung bei vergleichsweise geringerem Energiebedarf abarbeiten. Zum anderen hängt der Energiebedarf eines Servers von dessen Auslastung ab, die elektrische Leistungsaufnahme steigt mit der Last. Diese Steigerung ist aber nicht immer linear (siehe Kapitel 5.3.1), hier gilt es also, den relativen Mehrbedarf bei zusätzlicher Auslastung abzuwägen, sodass ein Server möglichst nah an seinem optimalen Wirkungsgrad betrieben wird. Diese Effekte müssen auch bei der Konsolidierung berücksichtigt werden, denn wenn Server deaktiviert und deren Last auf andere verteilt werden muss, steigt auch deren elektrische Leistungsaufnahme. Im Allgemeinen ist dieser Anstieg aber geringer, als die Einsparung, die durch die Deaktivierung der Server entsteht.

Bezogen auf das formale Problem des Bin-Packing ergibt sich für diese Situation nun ein zusätzlicher Aspekt, der dadurch abgebildet werden kann, indem die Server, die mit VMs gefüllt werden sollen, zu Beginn nach ihrer Effizienz sortiert werden. Formal wird diese Konsolidierungsstrategie wie folgt dargestellt:

Gegeben sei die Menge der virtuellen Maschinen  $V = \{v_1, v_2, \dots, v_m\}$  und deren entsprechende Ressourcenbedarfe  $L = \{\vec{l}_{v_1}, \vec{l}_{v_2}, \dots, \vec{l}_{v_m}\}$ . Der Vektor der verfügbaren Server ist mit

$\vec{S} = \{s_1, s_2, \dots, s_k\}$  gegeben, wobei die Server nach Energieeffizienz sortiert werden. Diese wird im nachfolgenden Abschnitt definiert. Die Ressourcenkapazität aller Server wird mit der Menge  $X = \{\vec{l}_{scap_1}, \vec{l}_{scap_2}, \dots, \vec{l}_{scap_k}\}$  angegeben. Die Funktion  $f : V \rightarrow \vec{S}, v \rightarrow s$  weist jedem  $v$  aus  $V$  genau einen Server  $s \in \vec{S}$  zu, wobei gilt:

$$\forall s \in \vec{S} : \sum_{i=1}^m \vec{l}_{v_i} + \vec{\varepsilon} < \vec{l}_{scap} | f(v_i) = s \quad (4.4)$$

Dabei ist  $\vec{\varepsilon}$  der in Gleichung 2.4 definierte Puffer. Ziel ist nun eine Allokation  $\vec{a}$  zu finden, unter der alle  $v \in V$  genau einem Server  $s \in \vec{S}$  zugeordnet sind und gleichzeitig die Anzahl der aktiven Server minimiert wird. Die Server werden zudem so gewählt, dass diejenigen mit der höchsten Energieeffizienz zuerst verwendet werden.

### Energieeffizienz von Servern - eine Metrik

Grundsätzlich gilt in dieser Arbeit für die Effizienz eines Servers, dass diese für die Server-Hardware bestimmt wird und unabhängig von einer spezifischen Art der Anwendung ermittelt wird. Beispielsweise ist es in der Praxis durchaus möglich, dass ein bestimmter Server sehr gut für Datenbankanwendungen geeignet ist, nicht aber für numerische Berechnungen. Hier soll stattdessen eine möglichst allgemeine Metrik der Effizienz verwendet werden, damit diese auf alle virtuellen Maschinen in einem Rechenzentrum angewendet werden kann. Dazu kann für einen Server  $s$  als einfachste Metrik die maximale CPU-Auslastung  $u_{scap}$  und die dabei auftretende elektrische Leistungsaufnahme  $P_{smax}$  verwendet werden:

$$e_s = \frac{u_{scap}}{P_{smax}} \quad (4.5)$$

In einem solchen Fall wird allerdings der Wirkungsgrad des Servers in niedrigeren Auslastungsbereichen ignoriert. Also sollte zuerst die Auslastung  $u_s \in [0, u_{scap}]$  gefunden werden, bei der

$$e_s = \frac{u_s}{spm_s(u_s)} \quad (4.6)$$

maximiert wird. Dabei ist  $spm_s(u_s)$  die Modellfunktion, die die elektrische Leistungsaufnahme unter der CPU-Auslastung  $u_s$  angibt und in Kapitel 5.3.1 beschrieben wird. Die Abbildung 4.4 verdeutlicht diese Effizienzmetrik anhand von vier Server-Beispielen. Die Daten für die Darstellung stammen aus dem frei verfügbaren *SPECpower\_ssj 2008 benchmark* [124]. Der Wert  $e_s$  gibt die Energieeffizienz des Servers in der Form Auslastung in Megahertz pro Watt elektrische Leistungsaufnahme an. In der Abbildung ist zu erkennen, dass die einzelnen Effizienzprofile durchaus unterschiedlich sind. So besitzt der Server von Huawei Technologies seinen effizientesten Arbeitspunkt bei ca. 70% seines Maximums. Wird dieser Server also darüber hinaus ausgelastet, steigen die Energiekosten überproportional an. Dagegen sollte der gezeigte Server der Hewlett-Packard Company immer bis zum Maximum ausgelastet werden, da seine Effizienz mit der Auslastung stetig steigt.

Bei der Betrachtung des Kriteriums der energetischen Optimierung der Server alleine kann also ein klassisches Bin-Packing-Problem in Kombination mit der Effizienzbetrachtung der

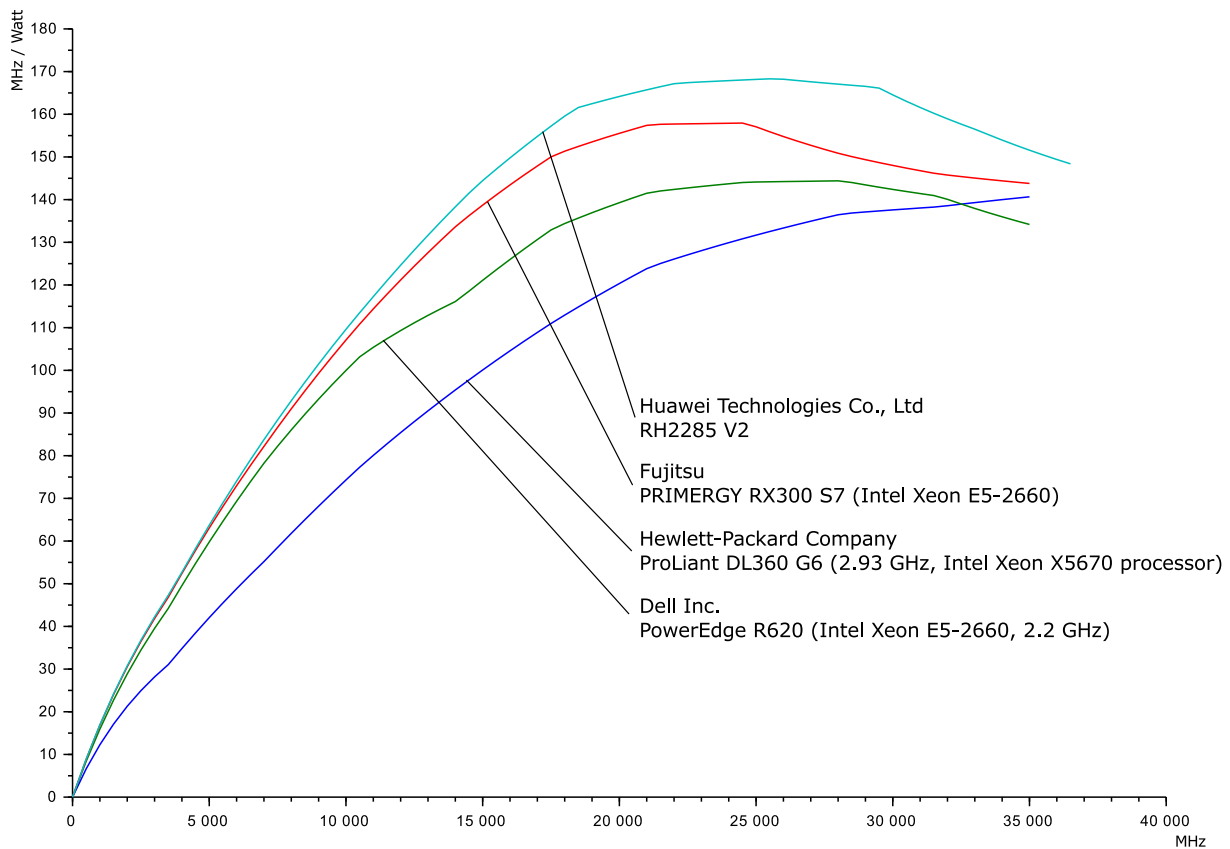


Abbildung 4.4: *Server-Energieeffizienz nach CPU-Auslastung. Die Gesamtkapazität der CPU der Server ergibt sich durch Addition aller Kerne der physikalischen Prozessoren. Dadurch entsteht die gezeigte Kapazität bis zu knapp 40 GHz. Diese Betrachtung ist immer dann legitim, wenn Virtualisierung oder vollständig parallelisierbare Anwendungen verwendet werden.*

Server angenommen werden. Allerdings betrachtet die Gleichung 4.4 nicht die weiteren Kriterien und Einschränkungen, die sich in den weiteren Kriterien ergeben, sodass eine alleinige Betrachtung des klassischen Bin-Packing-Problems für das Optimierungsverfahren nicht zielführend ist. Um eine Erweiterung dieses Problems vornehmen zu können, werden zunächst die weiteren Kriterien betrachtet und dann in einem kombinierten Lösungsansatz formuliert.

### 4.2.2 Kühlung

Die Geräte zur Kühlung im Rechenzentrum besitzen durchschnittlich einen Energiebedarf von 32% (siehe Kapitel 2.1) und stellen damit die energetisch zweitgrößte Gruppe nach den Servern dar. Energetische Optimierungen in diesem Bereich besitzen damit ein großes Einsparpotential.

Allerdings ist die Kühlkette im Rechenzentrum ein komplexes Konstrukt aus unterschied-

lichen Geräteklassen, von denen nur einige effektiv regelbar sind und damit für eine energetische Optimierung in Frage kommen. Im Kapitel 2.1.4 wurde der grundsätzliche Aufbau einer typischen Kühlkette dargestellt und in der Analyse in Kapitel 4.1.1 wurden einige Regelungsmöglichkeiten herausgestellt. Anhand dieser wird nun betrachtet, bei welchen Geräten und durch welche Aktionen energetische Optimierungen vorgenommen werden können.

Grundsätzlich muss bei der Kühlkette zwischen der Produktion und Verteilung des Kaltwassers und der Raumluftkühlung unterschieden werden. In den Räumen stehen Umluftkühlgeräte, die in Wärmeübertragern die Raumluft mit Hilfe des Kaltwassers auf ein definiertes und einheitliches Temperaturniveau abkühlen. Zur Erzeugung der erforderlichen Luftbewegung werden Ventilatoren eingesetzt. Eine Regelbarkeit der Umluftkühlgeräte ist nur eingeschränkt möglich und bezieht sich im Allgemeinen auf den Zufluss des Kaltwassers und die Drehzahl der Ventilatoren. Diese bestimmt letztendlich auch die elektrische Leistungsaufnahme des Umluftkühlgerätes. In der Praxis hat sich gezeigt, dass die Luftbewegung und damit die Drehzahl konstant bleibt, während nur über Ventile der Kaltwasserzufluss geringfügig angepasst wird. Dies hat auf die elektrische Leistungsaufnahme des Gerätes aber keinen Einfluss, sodass Umluftkühlgeräte in der Praxis als konstante Verbraucher angesehen werden können. Zusätzlich ist die Größenordnung des Energiebedarfs der Umluftkühlgeräte vor allem im Vergleich zu den Kältemaschinen eher gering.

Neben diesen Geräten existieren noch die Pumpen, die für die Verteilung des Kaltwassers benötigt werden. Auch hier hat die Praxis gezeigt, dass eine Anpassung der Fließgeschwindigkeit praktisch nicht vorgenommen wird [103]. Dies liegt vor allem daran, dass die beteiligten Geräte auf einen bestimmten Durchfluss hin optimiert sind und es nicht sinnvoll ist, von diesem Wert abzuweichen. Auch sind die zu erwartenden energetischen Ersparnisse bei den Pumpen bei langsamerer Flussrate nur gering. Daher erfolgt nun zunächst eine Betrachtung der Umluftkühlgeräte und im Anschluss eine Betrachtung der Kältemaschinen zur Kälteerzeugung.

### Umluftkühlgeräte

Die Umluftkühlgeräte können wegen der genannten Gründe als statische Verbraucher angesehen werden. Damit ergibt sich für die elektrische Leistung der Umluftkühlgeräte  $P_{crah_r}$  im IT-Raum  $r$  folgendes:

$$P_{crah_r} = \begin{cases} 0 & \text{falls } P_s = 0 \forall s \in S_r \\ c_{crah_r} & \text{sonst} \end{cases} \quad (4.7)$$

Es wird also davon ausgegangen, dass Umluftkühlgeräte im IT-Raum  $r$  ausgeschaltet werden können, wenn alle Server im IT-Raum  $r$  ebenfalls ausgeschaltet wurden. Ansonsten entspricht die elektrische Leistung der Konstante  $c_{crah_r}$ .  $P_{crah_r}$  bezieht sich dabei auf alle im IT-Raum  $r$  befindlichen Umluftkühlgeräte. Eine Einzelbetrachtung der Geräte ist nicht sinnvoll, da aufgrund der Redundanzbedingungen im Regelfall alle Geräte eingeschaltet sind, wenn der IT-Raum genutzt wird.

Die Umluftkühlgeräte müssen in der Lage sein, die anfallende Wärmeleistung  $\dot{Q}_r$  abzutransportieren, die nach Gleichung 4.1 der eingehenden elektrischen Leistung  $P_r$  im IT-Raum  $r$  entspricht.  $P_r$  setzt sich aus folgenden Teilen zusammen:

$$P_r = \underbrace{\sum_{i=1}^n P_{s_i}}_{\text{Server im Raum}} + \underbrace{\sum_{j=1}^m P_{crah_r_j}}_{\text{Alle UKG im Raum}} + \underbrace{P_{c_r}}_{\text{Konstant (PDU, Netzwerk, ..)}} \quad |P_{s_i} \in S_r \quad (4.8)$$

Neben den Servern und den weiteren Geräten im IT-Raum muss also auch der Eigenbedarf der Umluftkühlgeräte mit betrachtet werden. Jedes einzelne UKG im IT-Raum  $r$  muss für einen korrekten Betrieb genug Kälteleistung zur Verfügung stellen, damit die anfallende Wärmeleistung  $\dot{Q}_r$  abtransportiert werden kann. Dazu muss folgende Bedingung erfüllt sein:

$$\dot{Q}_{crah_r} \geq \dot{Q}_r \quad (4.9)$$

Dies gilt, da aufgrund der Redundanz zwar alle UKG gleichzeitig aktiv sein müssen, im Fehlerfall aber jedes einzelne die geforderte Kälteleistung bereitstellen muss. Diese Anforderung ergibt sich aus der oftmals gewünschten bzw. benötigten Zertifizierung für Rechenzentren. Beispielsweise befinden sich im während dieser Arbeit betrachteten Rechenzentrum immer zwei UKG in jedem IT-Raum, welche sich im Normalbetrieb die Raumkühlung teilen. Für den Fehlerfall muss aber ohne Änderung der Betriebsparameter jedes einzelne der beiden die gesamte Raumkühlung übernehmen können.

Daher muss für das hier entwickelte Optimierungsverfahren bestimmt werden, welche Kälteleistung  $\dot{Q}_{crah}$  (dies entspricht der anfallenden Wärmeleistung) ein einzelnes Umluftkühlgerät in Abhängigkeit von der Kaltwassertemperatur  $t_{water_c}$  liefern kann:

$$\dot{Q}_{crah} = ukm(t_{water_c}) \quad (4.10)$$

Die Modellfunktion  $ukm(t_{water_c})$  muss dazu für das Umluftkühlgerät charakterisiert werden und wird in Kapitel 5.4 beschrieben. Die Funktion betrachtet nur die Kaltwassertemperatur, die Ventilatorendrehzahl wird wie bereits beschrieben als statischer Wert angenommen.

Optimierungsziel ist nun,  $t_{water_c}$  zu maximieren aber die Bedingungen aus Gleichung 4.9 einzuhalten, da aufgrund dieser Anpassung Energieeinsparungen bei der Kältemaschine zu erwarten sind. Durch eine Anhebung dieser Temperatur sind außerdem zusätzliche Einsparungen durch Freikühlung zu erwarten.

Damit die Kaltwassertemperatur  $t_{water_c}$  angehoben werden kann, ist das sich daraus ergebende Optimierungsziel, die Leistungsaufnahme der IT-Räume, die gleichzeitig auch der anfallenden Wärmemenge entspricht (siehe Gleichung 4.1), anzugleichen. Da sich  $t_{water_c}$  nach dem am höchsten ausgelasteten Raum richten muss (siehe Gleichung 4.3), gilt es ebenfalls, das Maximum zu senken. Es sollen also IT-Räume mit stark ausfallenden Leistungsaufnahmen vermieden werden. Dazu ist in Abbildung 4.5 ein Beispiel dargestellt, das

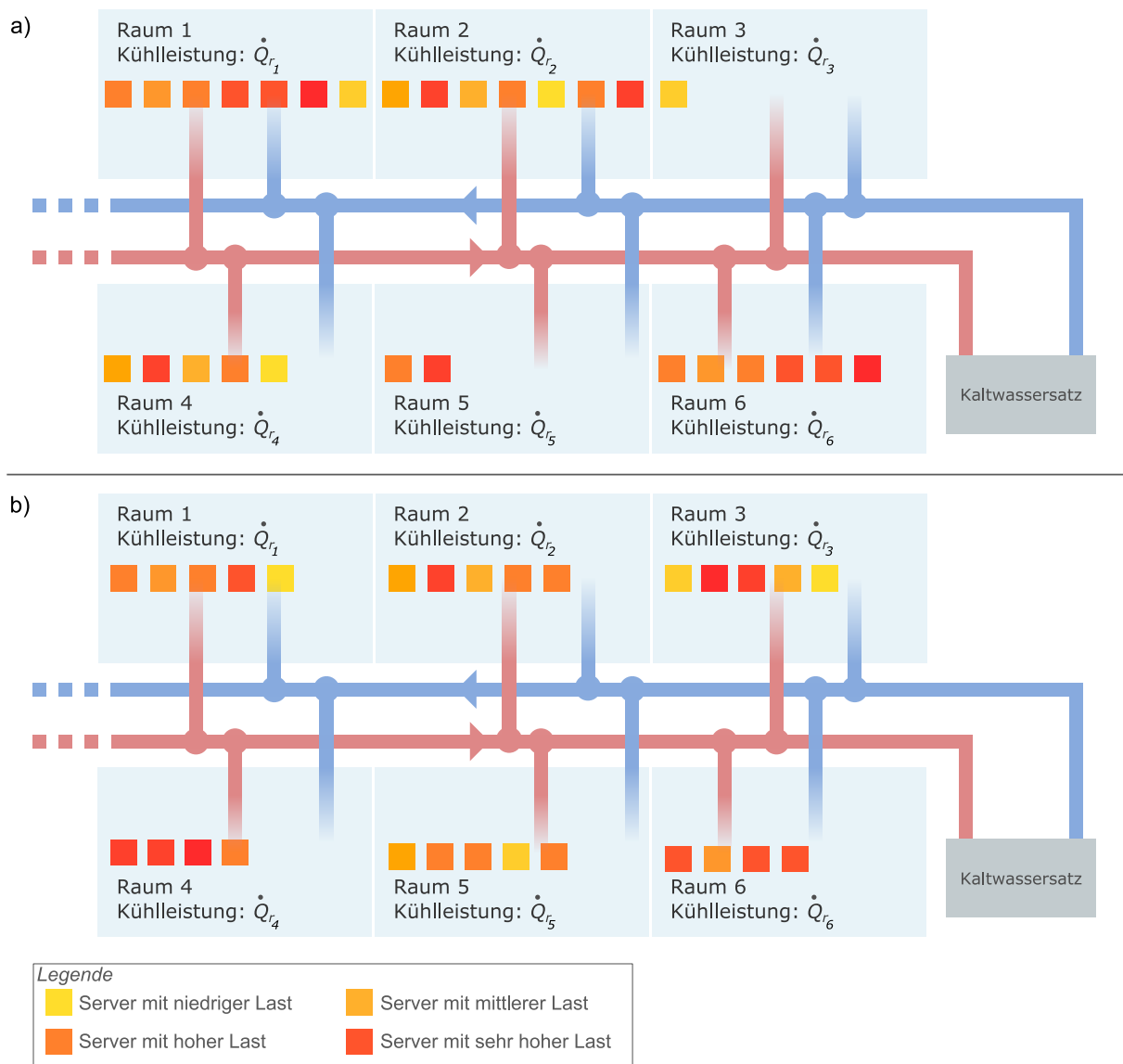


Abbildung 4.5: Verteilung der Wärmeleistung: Im Beispiel hängen 6 IT-Räume an der gleichen Kaltwasserversorgung. In Situation a) (oben) muss sich die Kaltwassertemperatur an IT-Raum 1 orientieren, da dieser die höchste Wärmemenge beinhaltet. Die großen Differenzen der Wärmemengen der IT-Räume sorgen für eine überdimensionierte Kühlung in den weiteren IT-Räumen, insbesondere in IT-Raum 3. Situation b) (unten) zeigt eine gleichmäßig verteilte Wärmemenge, die Kühlung kann einem niedrigeren Maximum angepasst werden und es tritt keine Überdimensionierung für einzelne IT-Räume auf.

zwei verschiedenen Situationen enthält. In Beispiel a) ist der nicht optimierte Zustand dar-



gestellt, bei dem IT-Raum 1 die höchste Wärmeleistung aufweist und im Gegensatz dazu IT-Raum 3 nur sehr gering ausgelastet ist. Beispiel b) hingegen zeigt eine gleichmäßige Verteilung der gleichen Gesamtauslastung, wodurch das Maximum gesenkt werden konnte. Damit  $t_{water_c}$  angehoben werden kann, muss also die Minimierung des Maximums der einzelnen Wärmeleistungen in den IT-Räumen erreicht werden. Dies geschieht durch eine Gleichverteilung der Lasten. Mit statistischen Mitteln kann diese bewertet werden. Dazu wird zunächst der Vektor  $\vec{Q}_R$  betrachtet, dessen Komponenten die jeweilige Wärmeleistung aller aktiven IT-Räume im Rechenzentrum darstellen. Mit der Menge der aktiven IT-Räume  $R_{active}$  gilt folgendes:

$$\vec{Q}_R = (\dot{Q}_{r_1}, \dot{Q}_{r_2}, \dots, \dot{Q}_{r_n}) \mid n = |R_{active}|, \dot{Q}_r \in R_{active} \quad (4.11)$$

Die Gleichverteilung der Wärmemengen in  $\vec{Q}_R$  kann dann zunächst mit der Standardabweichung  $\sigma_Q$  bestimmt werden, die aus der Varianz von  $\vec{Q}_R$  gebildet wird:

$$\sigma_Q = \sqrt{var(\vec{Q}_R)} \quad (4.12)$$

Aus dieser und dem Mittelwert  $avg(\vec{Q}_R)$  wird dann der Variationskoeffizient, der die Standardabweichung normalisiert, berechnet:

$$\rho_Q = \frac{1}{|\vec{Q}_R| - 1} \cdot \frac{\sigma_Q}{avg(\vec{Q}_R)} \mid 0 \leq \rho_Q < 1 \quad (4.13)$$

Diese gibt ein normalisiertes Maß der Ungleichheit an, wobei sich der Wert mit zunehmender Gleichverteilung der Zahl 0 nähert. Ziel ist damit,  $\rho_Q$  zu minimieren. Der Variationskoeffizient ermöglicht einen Vergleich der Bewertung der Gleichverteilung, was im entwickelten Optimierungsverfahren genutzt wird.

### Kältemaschinen & Kaltwassersatz

Im Gegensatz zu den Umluftkühlgeräten benötigen die Kaltwasserproduktion und insbesondere die Kältemaschinen den größten Teil der Energie in der Klimakette. Bei den Kältemaschinen, die in Rechenzentren meist mit Kompressionselementen arbeiten, ist eine Abhängigkeit des eigenen Energiebedarfs von der zu kühlenden Energiemenge  $\dot{Q}_{total}$  aus Gleichung 4.2 gegeben. Gleichzeitig sind diese Geräte einerseits über die Anzahl der gleichzeitig aktiven Kompressoren und andererseits über die Menge des Kompressionsgases regelbar, sodass eine Adaptivität zum tatsächlich benötigten Kühlbedarf vorliegt. Bisherige Forschungen [87, 129] haben zudem gezeigt, dass eine signifikante Abhängigkeit der elektrischen Leistungsaufnahme zur Kaltwassertemperatur besteht. Daher fließen beide Parameter in die Modellfunktion  $kpm(\dots)$  ein, die die elektrische Leistungsaufnahme der Kältemaschine schätzt und in Kapitel 5.5 beschrieben wird:

$$P_{chiller} = kpm(t_{water_c}, \dot{Q}_{total}) \quad (4.14)$$

Es ist also anzumerken, dass die Optimierungsschritte, die im vorangegangenen Abschnitt der Umluftkühlgeräte beschrieben wurden und die Anhebung von  $t_{water_c}$  anvisieren, zwar durch die UKG bestimmt werden, allerdings die Energieersparnisse dann bei den Kältemaschinen anfallen.

Ein weiterer wichtiger Aspekt, der mit der Anhebung von  $t_{water_c}$  einhergeht und eine signifikante Energieersparnis in der Kaltwasserproduktion ermöglicht, ist die Freikühlung. Die Häufigkeit im Jahr, mit der Freikühlung verwendet werden kann, hängt direkt von der benötigten Kaltwassertemperatur und der Außentemperatur ab. Je höher jedoch die Kaltwassertemperatur ist, desto häufiger kann auch Freikühlung eingesetzt werden, da dann höhere Außentemperaturen noch zur Freikühlung genutzt werden können. An dieser Stelle entsteht aber kein weiteres Optimierungsziel, da diese Ersparnisse mit den bereits beschriebenen Aktionen ebenfalls erreicht werden.

Die Annahmen und Beschreibungen, die in diesem Abschnitt formuliert wurden, um eine effiziente und energetisch optimierte Kühlung zu erreichen, gelten nur unter der Bedingung, dass die in Kapitel 4.3 beschriebenen Maßnahmen zur sicheren und stabilen Wärmeverteilung im IT-Raum auch tatsächlich eingehalten werden. Daher ist das genannte Kriterium auch ein hartes, nicht zu verletzendes Constraint. Des Weiteren gibt es zusätzliche Anforderungen, die erfüllt werden müssen, damit die in diesem Abschnitt genannten Einsparungen durch das Anheben von  $t_{water_c}$  umgesetzt werden können. Es wurde bereits kurz erwähnt, dass eine Regelung der Kaltwassertemperatur einige Zeit in Anspruch nimmt und daher nicht innerhalb kürzester Zeit dynamisch erfolgen kann. Beispielsweise können Server-Konsolidierungen mehrmals täglich durchgeführt werden, eine Regelung der Kaltwassertemperatur ist aber unter anderem aufgrund der Trägheit im System nur für deutlich längere Zeiträume sinnvoll. Daher müssen hier Möglichkeiten im Optimierungsverfahren gefunden werden, die eine langfristige Planung erlauben. Diese Untersuchung und die dazu verwendete Lösung wird in Kapitel 6.2.1 beschrieben.

### 4.2.3 Lösungsansatz

Damit die identifizierten Optimierungen erreicht werden können, bietet sich eine Aufteilung der Lösungssuche an. Energieeinsparungen bei den Servern sollen durch Konsolidierungsstrategien und Effizienzverbesserungen erreicht werden, dies ist aber unabhängig von der Verteilung der Auslastung über die IT-Räume hinweg. Durch diese soll wiederum eine signifikante Ersparnis bei der Kühlung erreicht werden. Daher ist der Ansatz hier, zwei heuristische Verfahren einzusetzen, die die jeweiligen Bereiche abdecken. Diese einzelnen Verfahren können dann je nach Notwendigkeit mit den weiteren Kriterien verknüpft werden.

#### **Energieeinsparung: Server**

Analog zu einer Heuristik wie First Fit Decreasing (FFD) für Bin-Packing, werden Server nach Effizienz sortiert und dann mit virtuellen Maschinen „befüllt“. Allerdings muss diese Heuristik auch die weiteren Kriterien berücksichtigen können. Daher kann ein Be-

wertungssystem genutzt werden, das eine energetische Priorisierung von effizienten Servern vornimmt, ansonsten aber die Bewertungen der weiteren Kriterien berücksichtigt. Damit Energieeinsparungen erreicht werden, wird also auch der Aspekt der klassischen Server-Konsolidierung aufgegriffen, bei der möglichst wenige aktive Server verwendet werden sollen. Allerdings kann das Kriterium der Energieeffizienz alleine nicht bestimmen, welche Server zu welchem Grad der Auslastung verwendet werden. An dieser Stelle greifen die weiteren Kriterien ein.

### Energieeinsparung: Klimatisierung

Für die Anwendung der Klimatisierungsoptimierungen muss eine Kombination aktiver IT-Räume gesucht werden, bei der einerseits durch Abschalten ungenutzter IT-Räume Energie eingespart werden kann und andererseits ein Anheben der Kaltwassertemperatur ermöglicht wird. Dazu muss die Gleichverteilung der Auslastung über die aktiven IT-Räume stattfinden. Diese Gleichverteilung für eine gegebene Kombination aktiver IT-Räume könnte mittels Greedy-Verfahren erfolgen. Wenn jedoch eine optimierte Kombination in einem großen Suchraum gefunden werden soll, dann sollte die Suche mittels heuristischer Suchverfahren erfolgen.

## 4.3 Kriterium: Thermische Stabilität

Die thermische Stabilität einer Allokation und damit eines Betriebszustandes definiert harte Bedingungen, die aufgrund von Kundenverträgen oder grundlegenden physikalischen Eigenschaften eingehalten werden müssen. Zusätzlich gibt es Aspekte, die dazu beitragen, dass die Wahrscheinlichkeit des Auftretens von Hot-Spots oder ähnlichen thermischen Problemen gesenkt wird. Diese werden ebenfalls in diesem Kriterium definiert und in den folgenden Abschnitten beschrieben.

### 4.3.1 Server-Temperatur

Für jeden Server wird im Normalfall bereits durch den Hersteller ein Temperaturfenster angegeben, für das der Betrieb garantiert wird. Wird beispielsweise die maximale Temperatur überschritten, kann es zu Garantieverlust oder Strafzahlungen an Kunden kommen. Als Grundlage für die Temperaturangaben werden häufig die ASHRAE-Empfehlungen [9] verwendet. Diese geben zum Beispiel eine maximale empfohlene Temperatur  $t_{s_{cold}}$  der eingehenden (kalten) Luft von 18 °C bis 27 °C an. Damit gilt für  $t_{s_{cold}}$  die Bedingung:

$$t_{s_{cold}} < t_{s_{cmax}} \quad (4.15)$$

wobei  $t_{s_{cmax}}$  hier das durch den Betreiber festgelegte Maximum darstellt und im Allgemeinen im oben genannten Temperaturbereich liegen sollte.

Innerhalb des Servers wird die durchgehende Luft erwärmt und tritt dann mit der Temperatur  $t_{s_{warm}}$  wieder aus. Auch diese Temperatur ist durch Garantiebedingungen geregelt

und bildet daher eine harte Grenze. Der Grad der Erwärmung der Luft innerhalb des Servers hängt direkt von der elektrischen Leistungsaufnahme  $P_s$  des Servers ab, da über die Luft die zu Wärme gewandelte Energie abtransportiert wird.

Moderne Server besitzen mehrere Temperatursensoren, wodurch beide Temperaturen direkt feststellbar sind. Mittels der Messung einiger Kenngrößen und grundlegender thermodynamischer Zusammenhänge kann ein Modell für diese Temperaturen ermittelt werden. Dieses Modell wird für den Server  $s$  durch die Funktion  $stm_s(\dots)$  bereitgestellt und in Kapitel 5.3.2 definiert. Damit ergibt sich folgende Abhängigkeit:

$$t_{s_{warm}} = stm_s(t_{s_{cold}}, P_s) \quad (4.16)$$

Das entwickelte Optimierungsverfahren nutzt dieses Modell in Kombination mit dem elektrischen Leistungsmodell des Servers (siehe dazu Gleichung 2.7 und die Definition in Kapitel 5.3.1), um eine Abschätzung vorzunehmen, ob ein Server eine konkrete VM-Auslastung nach diesen thermischen Kriterien verkraften kann. Dabei gilt, dass die Temperatur  $t_{s_{warm}}$  immer unter dem vertraglich bzw. durch den Hersteller festgelegten Maximum liegt:

$$t_{s_{warm}} < t_{s_{max}} \quad (4.17)$$

Die Überprüfung dieser Grenze erfolgt dabei je nach Anwendungsfall: das Optimierungsverfahren nutzt das Modell für eine Planung, aber im laufenden Betrieb wird die Temperaturgrenze permanent mittels der Sensoren im Server überwacht. Tritt hier ein plötzlicher Temperaturanstieg auf, muss der Server umgehend entlastet und eventuell sogar vollständig evakuiert werden. Dieser Fall wurde in [26] simuliert, wobei hier in einem Beispielszenario als Ergebnis ermittelt wurde, dass bei einer plötzlichen signifikanten Temperatursteigerung der betrachtete Server innerhalb von 200-220 Sekunden evakuiert werden musste. Aus diesen Beobachtungen ergibt sich beispielsweise ein weiterer Aspekt, der im Kriterium der Flexibilität (Kapitel 4.4) betrachtet wird: die maximale Evakuierungszeit, die für alle Server garantiert werden muss. Für eine proaktive Erkennung möglicher Überhitzungen bietet [26] ebenfalls Vorschläge, allerdings ist diese Betrachtung nicht mehr Teil dieser Arbeit. Hier ist die Einhaltung der Grenzen und im Fehlerfall eine sofortige Evakuierung durch einen flexiblen Zustand der Fokus.

### 4.3.2 Wärmeverteilung

Bisherige Forschungen [59, 87, 129, 130, 131] haben gezeigt, dass eine möglichst gleichmäßige Verteilung der Wärmeleistung im IT-Raum dazu beiträgt, die Wahrscheinlichkeit des Auftretens von Hot-Spots oder anderen Wärme-Problemen signifikant zu senken. Dazu wurden detaillierte Messungen und Computational-Fluid-Dynamics-Simulationen (CFD) durchgeführt [99] um damit einerseits die Wärmeverteilung im IT-Raum zu untersuchen und andererseits das Auftreten von Hot-Spots zu simulieren und geeignete Gegenmaßnahmen vorzuschlagen. Mit der Annahme, dass alle IT-Räume geometrisch rechteckig aufgebaut sind und auch die Positionen der Racks in den Räumen symmetrisch sind (dies ist im untersuchten Rechenzentrum der Fall), kann aus den Ergebnissen der Forschungen eine

gleichmäßige Verteilung der Wärmeleistung über die Racks im IT-Raum als Ziel geschlossen werden [98, 99].

Sollte ein Rechenzentrum deutlich unterschiedliche Raumgeometrien oder verwinkelte IT-Räume besitzen, so kann diese Annahme nicht ohne weiteres getroffen werden. Stattdessen müssen zunächst Gewichtungen für einzelne Raumpositionen bestimmt werden, die die unterschiedlichen Luftströmungen berücksichtigen. Dies kann durch Messungen oder CFD-Simulationen, wie in den genannten Forschungsarbeiten beschrieben, geschehen, wird aber an dieser Stelle nicht weiter vertieft.

Das Optimierungsziel der gleichmäßigen Wärmeverteilung ist beispielhaft in Abbildung 4.6 dargestellt. Die nicht optimierte Situation in 4.6 a) weist starke Wärmeschwerpunkte auf, sodass hier eine erhöhte Überhitzungsgefahr entsteht. Zudem liegt ein weiteres Problem vor: tritt hier beispielsweise eine Überhitzung in der unteren linken Rackregion auf, so müssen sofort mehrere Server entlastet werden. Im Gegenzug dazu zeigt 4.6 b) eine optimierte Situation, in der aufgrund der gleichmäßigen Verteilung die Gefahr lokaler Überhitzungen niedriger ist. Tritt hier in einem Rack trotzdem eine Überhitzung auf, so müssen nur wenige Server entlastet werden.

Zur Umsetzung des Optimierungsziels wird auf Rack-Ebene eine annähernd gleiche Wärmeleistung und nach Gleichung 2.2 damit elektrische Leistungsaufnahme angestrebt. Formal kann dies wie folgt ausgedrückt werden: Mit der Menge aller Racks  $K_r$  im IT-Raum  $r$  kann die elektrische Leistungsaufnahme  $P_k$  eines einzelnen Racks  $k \in K_r$  aus dem summierten Bedarf der Menge der Server  $S_k$  im Rack  $k$  angegeben werden:

$$P_k = \sum_{i=1}^n P_{s_i} \mid s_i \in S_k, n = |S_k| \quad (4.18)$$

Die Komponenten des Vektors  $\vec{P}_{K_r}$  enthalten nun die elektrische Leistungsaufnahme der Racks im IT-Raum  $r$ :

$$\vec{P}_{K_r} = (P_{k_0}, P_{k_1}, \dots, P_{k_i}) \quad (4.19)$$

Analog zur Gleichverteilung der Wärmemenge in Abschnitt 4.2.2 wird auch hier die Gleichverteilung der elektrischen Leistungsaufnahmen in  $\vec{P}_{K_r}$  mit der Standardabweichung  $\sigma_K$  bestimmt:

$$\sigma_K = \sqrt{\text{var}(\vec{P}_{K_r})} \quad (4.20)$$

Aus dieser und dem Mittelwert  $\text{avg}(\vec{P}_{K_r})$  wird dann ebenfalls der Variationskoeffizient berechnet:

$$\rho_K = \frac{1}{|\vec{P}_{K_r}| - 1} \cdot \frac{\sigma_K}{\text{avg}(\vec{P}_{K_r})} \mid 0 \leq \rho_K < 1 \quad (4.21)$$

Diese gibt auch hier ein normalisiertes Maß der Ungleichheit an, wobei sich der Wert mit zunehmender Gleichverteilung der Zahl 0 nähert. Ziel ist also,  $\rho_K$  zu minimieren. Auch hier kann der Variationskoeffizient zum Vergleich mehrerer Lösungskandidaten genutzt werden.

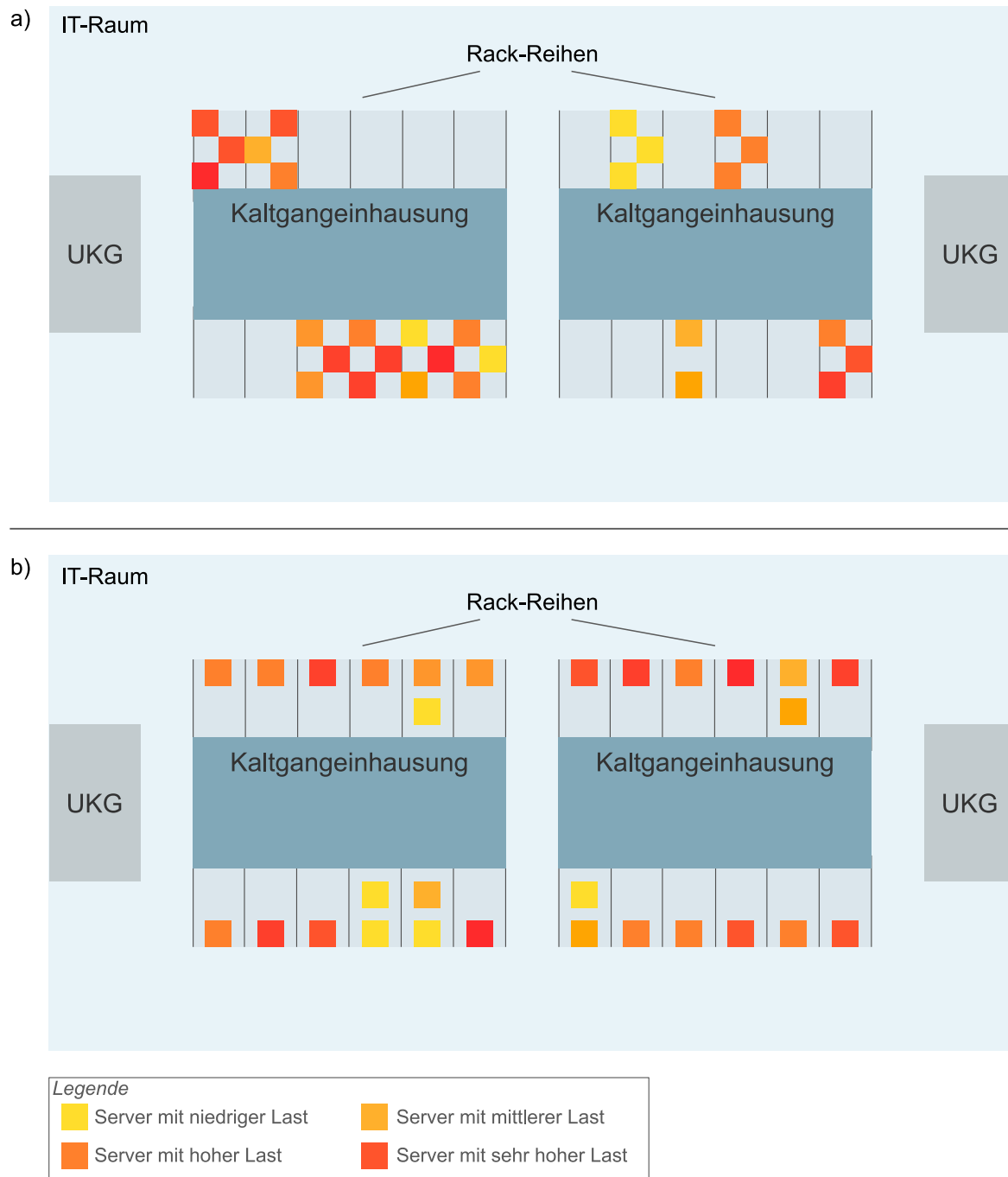


Abbildung 4.6: Verteilung der Wärmeleistung: Situation a) (oben) zeigt starke Wärmeschwerpunkte, die zu Hot-Spots führen können. Situation b) (unten) zeigt eine gleichmäßige Wärmeleistungsverteilung.

### 4.3.3 Lösungsansatz

Die harten Grenzen der Server-Temperaturen müssen immer eingehalten werden und werden daher als Beschränkungen im Verfahren behandelt. Dazu wird vor einem möglichen Optimierungsschritt durch Modelle die zu erwartende Änderung der Temperatur prognostiziert und nur bei Einhaltung der Grenzen wird der Schritt tatsächlich ausgeführt. Die zeitlich begrenzte Reaktion auf Überhitzungen bzw. steigende Temperaturen wird im folgenden Kriterium der Flexibilität und Evakuierungszeit behandelt.

Die gleichmäßige Verteilung der Wärmeleistung innerhalb jedes einzelnen IT-Raums wird durch Priorisierung der kühleren Bereiche vorgenommen. Dazu wird eine normalisierte Bewertung der einzelnen IT-Raumpositionen relativ untereinander verwendet und in den Entscheidungsprozess des Optimierungsverfahrens mit eingebunden.

## 4.4 Kriterium: Sicherstellen der Flexibilität

Die Flexibilität einer Allokation bezieht sich auf die folgenden drei Aspekte:

- *Lokale Puffer*: Jeder Server soll einen Puffer freier Ressourcen besitzen, der Lastschwankungen von den auf dem Server befindlichen virtuellen Maschinen abfangen kann.
- *Reserveserver bereithalten*: Im vorherigen Kriterium der thermischen Stabilität wurden mehrere Situationen geschildert, bei denen es zu Evakuierungen von einzelnen Servern kommen kann. In einem solchen Fall müssen im gesamten Rechenzentrum freie Server-Ressourcen verfügbar sein. Die virtuellen Maschinen des zu evakuierenden Servers werden dann auf diese freien Ressourcen verteilt, sodass eine direkte Evakuierung möglich ist, ohne erst zusätzliche Server einschalten zu müssen.
- *Evakuierungszeit begrenzen*: Speziell im Hinblick auf eine solche Evakuierung muss die dafür benötigte Zeit abschätzbar und für eine Allokation für jeden Server auf ein Maximum beschränkbar sein.

Das Kriterium der Sicherstellung der Flexibilität einer Allokation bildet diese Bereiche ab. Eine Beschreibung und Definition erfolgt in den folgenden Unterkapiteln.

### 4.4.1 Lokal: Zusätzliche freie Server-Ressourcen

Bei Anwendung von Server-Konsolidierungsstrategien werden mehrere virtuelle Maschinen auf einen einzelnen Server konzentriert. Die Grenze der Anzahl der virtuellen Maschinen wird durch die Auslastungen der virtuellen Maschinen und die verfügbaren Ressourcen des Servers angegeben. Die Auslastungen der virtuellen Maschinen verhalten sich allerdings in der Regel dynamisch und auch bei Anwendung von Vorhersageverfahren oder ähnlichen

Methoden können unvorhergesehene Lastschwankungen nicht ausgeschlossen werden. Damit dies nicht zu Ressourcenengpässen führt, soll jeder Server einen Puffer freier Ressourcen besitzen, der diese Schwankungen auffangen kann. Genau dieser Puffer wird durch die lokale Flexibilität beschrieben.

Der aktuelle Stand der Forschung betrachtet diesen lokalen Puffer als festen Wert  $\bar{\epsilon}$ , wie in den Grundlagen in Gleichung 2.4 dargestellt. Dadurch wird allerdings die Art des Auslastungsprofils der einzelnen virtuellen Maschinen nicht berücksichtigt. In der Auslastung stark schwankende und konstant niedrige virtuelle Maschinen werden gleich behandelt. In dieser Arbeit wird die lokale Flexibilität  $f_{l_v}$  für jede einzelne virtuelle Maschine  $v$  betrachtet. Jede virtuelle Maschine  $v$  auf einem Server soll die eigene Auslastung um einen Faktor  $f_{l_v} | f_{l_v} \in \mathbb{R}, f_{l_v} \geq 0$  steigern können, ohne dass beim Server Ressourcenengpässe auftreten. Diese Steigerung bezieht sich auf die CPU-Auslastung  $u_v$  der virtuellen Maschine  $v$ , sodass für jeden Server  $s$  gelten muss:

$$\sum_{i=1}^n (u_{v_i} \cdot (1 + f_{l_{v_i}})) \leq u_{s_{cap}} \quad (4.22)$$

Dabei ist  $n$  die Anzahl der virtuellen Maschinen auf dem Server  $s$ . Die Bedingung in Gleichung 2.4, die die Ressourcenkapazität des Arbeitsspeichers  $o_v$  definiert, gilt weiterhin, da der Speicher als statische Größe betrachtet und keine Überprovisionierung angewandt wird.

Dieser Ansatz der Definition der lokalen Flexibilität sorgt in Kombination mit der Betrachtung der VM-Auslastungsprofile, wie in Kapitel 6.2.1 beschrieben, dafür, dass für virtuelle Maschinen mit starken Lastschwankungen oder grundsätzlich höherem Ressourcenbedarf mehr Ressourcen bereit gehalten werden als für virtuelle Maschinen mit geringerem Ressourcenbedarf.

#### 4.4.2 Global: Zusätzliche Server bereithalten

Gerade wenn es um die Reaktion auf Notfälle geht und ein oder mehrere Server plötzlich ausgeschaltet werden müssen, ist die zeitliche Dauer der Evakuierung dieser Geräte von höchster Bedeutung. Die Zeit, die für die Migrationen der virtuellen Maschinen benötigt wird, ist unvermeidbar. Wenn aber zunächst neue Server für die virtuellen Maschinen des zu evakuierenden Systems eingeschaltet werden müssen, so steigt die Evakuierungszeit signifikant an. Bei den untersuchten Servern fallen hier beispielsweise mindestens zehn Minuten für den Boot-Vorgang der einzuschaltenden Server zusätzlich zu der Migrationszeit an. Diese zusätzliche Zeit soll durch eine flexible Allokation vermieden werden. Dazu sollen zusätzliche Server bereitgehalten werden, sodass in einem Notfall der Einschaltprozess entfällt oder zumindest zeitlich hinausgezögert werden kann. Damit wird dem Argument vieler Rechenzentrumsbetreiber begegnet, das besagt, dass ein konsolidierter Betrieb in Rechenzentren zu unsicher sei, da in Fehlerfällen nicht rechtzeitig reagiert werden könne. Die globale Flexibilität gibt dazu die Anzahl der Server an, die evakuiert und ausgeschaltet werden können, ohne dass zusätzliche Server eingeschaltet werden müssen. In Abbildung



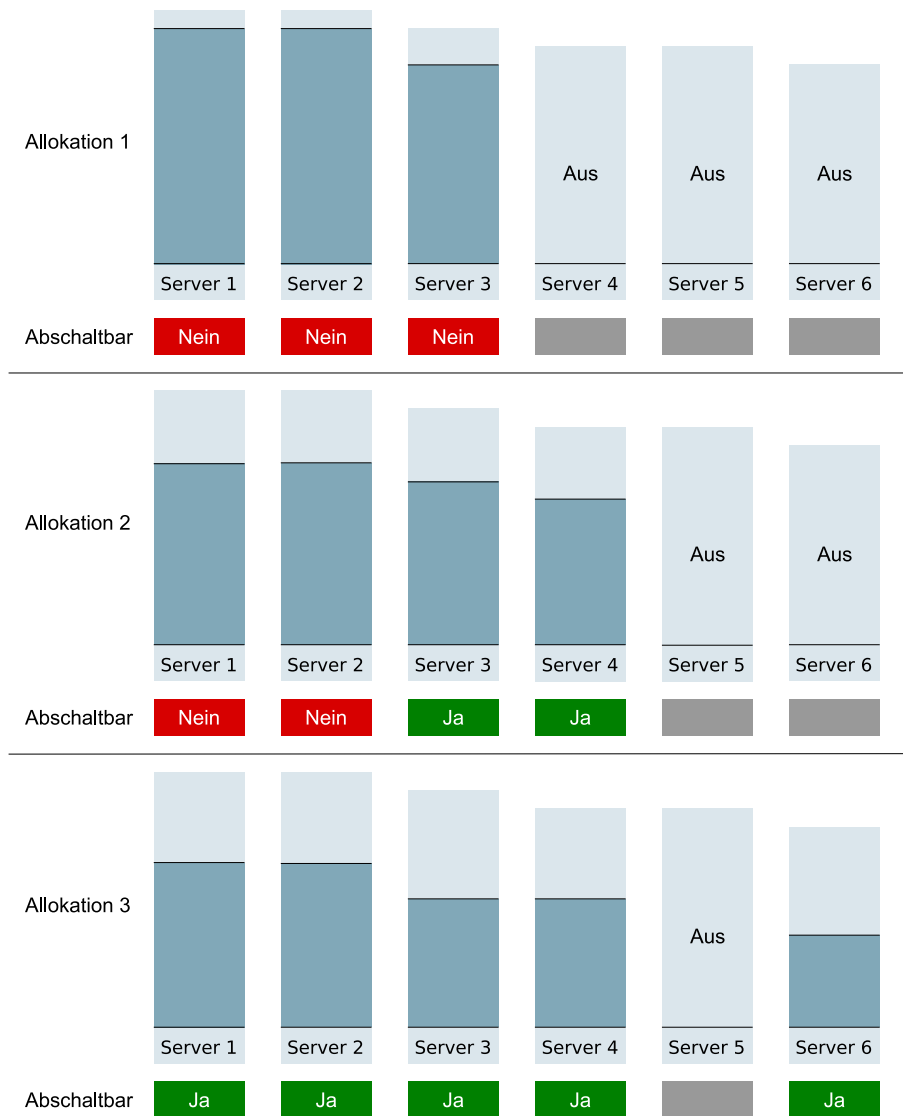


Abbildung 4.7: Die globale Flexibilität gibt die Anzahl der Server an, die ausgeschaltet werden können, ohne dass neue hinzu geschaltet werden müssen. Im gezeigten Beispiel in Allokation 1 kann keiner der Server ausgeschaltet werden, die Allokation ist minimal flexibel. In Allokation 2 wurde ein Server hinzu geschaltet und die Last verteilt, nun können zwei Server ausgeschaltet werden. Die globale Flexibilität beträgt immer noch null, da nicht mindestens ein beliebiger Server ausgeschaltet werden kann. Dies ist erst in Allokation 3 erreicht.

4.7 wird dieses Prinzip an einem Beispiel mit insgesamt drei verschiedenen Allokationen dargestellt. Die Allokation 1 in der Abbildung stellt einen maximal konsolidierten Zustand dar, dieser ist gleichzeitig minimal flexibel. Keiner der Server könnte in einem Notfall

ausgeschaltet und dessen virtuelle Maschinen auf die anderen verteilt werden, ohne dass ein weiterer Server hochgefahren werden müsste. In der *Allokation 2* wurde nun ein zusätzlicher Server eingeschaltet und die Last gleichmäßig auf allen aktiven Servern verteilt. In diesem Zustand können nun zwei der Server (Server 3 und Server 4) ausgeschaltet werden, für die anderen reichen die zusätzlichen freien Ressourcen nicht aus. Der Flexibilitätsgewinn dieser Allokation ist also minimal. Erst nachdem in *Allokation 3* ein weiterer Server aktiviert wurde, kann nun ein einzelner, beliebiger Server in einem Evakuierungsfall ausgeschaltet werden, ohne dass ein weiterer Server eingeschaltet werden muss.

Zum gezeigten Beispiel sei angemerkt, dass im dargestellten Fall in Allokation 3 fünf von sechs Servern eingeschaltet und diese nur mittelmäßig ausgelastet sind. Es wird also deutlich, dass das Erreichen der globalen Flexibilität in direktem Widerspruch zum energieoptimierten Zustand steht. Allerdings ist das Beispiel in dieser Hinsicht extrem und kann nur exemplarisch betrachtet werden, da im Beispiel nur eine sehr geringe Anzahl von sechs Servern vorhanden ist. In einem Rechenzentrum mit mehreren hunderten bis tausenden Servern ist die Anzahl der zusätzlich einzuschaltenden Server deutlich geringer und steht in einem deutlich kleineren Verhältnis zur Gesamtzahl der aktiven Server. Damit ist auch der zusätzliche Energiebedarf bei steigender globaler Flexibilität geringer, als bei Betrachtung des Beispiels aus Abbildung 4.7 vermutet werden könnte. Eine Auswertung dieses Sachverhalts findet in der Evaluation in Kapitel 7.3.4 statt.

Die globale Flexibilität  $f_g \in \mathbb{N}$  einer Allokation  $\vec{a}$  gibt also die Anzahl der Server an, die gleichzeitig in einer beliebigen Kombination ausgeschaltet werden können, wobei die darauf befindlichen virtuellen Maschinen auf weitere bereits aktive Server migriert werden.

Im gezeigten Beispiel in Abbildung 4.7 beträgt die globale Flexibilität der Allokation 1 und 2  $f_g = 0$ , da in beiden Allokationen nicht ein oder mehr beliebige Server ausgeschaltet werden können. Erst in Allokation 3 wird  $f_g = 1$  erreicht.

### Komplexität der Metrik und deren Berechnung

Es gibt mehrere Aspekte, die zur erhöhten Komplexität der Metrik der globalen Flexibilität beitragen und dazu führen, dass in dieser Arbeit im Lösungsverfahren ein heuristischer Ansatz zur Bestimmung der Flexibilität und zur Wahl der jeweils zuzuschaltenden Server zum Einsatz kommt. Diese sind:

- *Globale Flexibilität nur mit bekannter Allokation erreichbar:* Die globale Flexibilität  $f_g$  ist erst dann berechenbar, wenn eine Allokation  $\vec{a}$  gefunden wurde, die alle anderen Kriterien bereits erfüllt. Erst dann ist die Anzahl und konkrete Menge der aktiven Server und die jeweilige Verteilung der virtuellen Maschinen bekannt. Ausgehend davon muss dann der Ist-Wert von  $f_g$  bestimmt und im Falle eines zu geringen Wertes müssen zusätzliche Server eingeschaltet werden, sodass der Zielwert erreicht werden kann.
- *Kombinatorik der im Notfall auszuschtenden Server:* Der Flexibilitätswert muss durch den Rechenzentrumsbetreiber festgelegt werden. Dies muss in Form der absoluten Zahl  $f_g$  erfolgen. Hier ergibt sich also formalisiert eine Beschreibung, dass  $f_g$

von  $n$  Servern ausgeschaltet werden können müssen. Für dieses Problem gilt das kombinatorische Auswahlproblem ohne Wiederholung, es gibt also die folgende Anzahl an möglichen Kombinationen:

$$\frac{n!}{(n - f_g)! \cdot f_g!} = \binom{n}{f_g} \quad (4.23)$$

Allerdings ist die Gesamtzahl  $n$  der Server, die tatsächlich eingeschaltet sein werden, vor der Findung einer Allokation nicht bekannt. Daher kann nur die Anzahl der Server angegeben werden, die gleichzeitig ausgeschaltet werden können müssen. Hier sollte als Richtwert beispielsweise die maximale Anzahl der Server pro Rack verwendet werden, da je nach Notfall beispielsweise ein vollständiges Rack betroffen sein könnte.

- *Verteilung der Last auf die zusätzlich einzuschaltenden Server:* Diese Verteilung muss so erfolgen, dass die anderen Kriterien ebenfalls berücksichtigt werden, insbesondere im Hinblick auf den Energiebedarf. Daher sollte die Verteilung so erfolgen, dass jeder Server in seinem optimalen Wirkungsbereich ausgelastet wird. Kann dies nicht für alle erfolgen, sollte eine Näherung an diese Betriebszustände erfolgen. Dies alleine ist wiederum ein komplexes kombinatorisches Problem.

Wird also die globale Flexibilitätsmetrik in einem größeren Rechenzentrum mit mehreren tausend Servern angewandt, ist eine Berechnung und Sicherstellung der globalen Flexibilität nur durch Heuristiken in realistischer Zeit realisierbar.

Die Sicherstellung der globalen Flexibilität unterliegt darüber hinaus einigen Einschränkungen. So sollten die zusätzlich verfügbaren Server-Ressourcen in einem Notfall nur während der Evakuierungszeit genutzt werden. Danach muss das DaMucA-Verfahren vollständig angewandt werden, um eine neue Lösung zu finden. Dies liegt daran, dass der Betriebszustand, der im Falle eines Notfalls unter Ausnutzung der globalen Flexibilität eingenommen wird, kein optimierter Zustand hinsichtlich aller Kriterien mehr ist. Dieser Zustand gilt also nur als Überbrückung, um schnell den Notfall auflösen zu können und dann mit Hilfe des DaMucA-Verfahrens wieder in einen optimierten Betriebszustand zu gelangen. Die konkrete Reaktion und weitere Planung in einem solchen Notfall hängt stark vom Typ und der Ursache des Notfalls ab. In Kapitel 6.2.2 wird genauer auf die möglichen Reaktionen und deren Planungen eingegangen.

### 4.4.3 Begrenzen der Evakuierungszeit

Falls eine Notsituation eintreten sollte, müssen alle beteiligten Server so schnell wie möglich entlastet und unter Umständen ausgeschaltet werden. Dazu müssen die sich darauf befindenden virtuellen Maschinen auf andere Server verschoben (evakuiert) werden, was eine bestimmte Zeit in Anspruch nimmt. Diese Zeit soll für jeden Server gering gehalten werden, da grundsätzlich die Annahme gilt, dass die Wahrscheinlichkeit eines Notfalls in jedem Bereich des Rechenzentrums gleich groß ist, also praktisch jeder Server betroffen sein kann.

Ein flexibler Zustand kann also auf einen solchen möglichen Notfall, egal wo dieser Auftritt, gleichermaßen reagieren. Damit sollte auch die Zeit, die die Entleerung eines Servers benötigt, bei maximaler Flexibilität für alle Server sowohl ähnlich als auch möglichst niedrig sein.

Wenn in diesem Kriterium die Evakuierungszeit eines Servers und der sich darauf befindlichen virtuellen Maschinen bestimmt werden soll, ist noch nicht bekannt, auf welche Zielservers die virtuellen Maschine migriert werden sollen. Daher kann hier nur die mindestens benötigte Zeit, die eine Migration auf Grund der Modelldaten der virtuellen Maschine auf dem Quellserver benötigt, bestimmt werden. Hier spielt beispielsweise die Auslastung der zu migrierenden virtuellen Maschine eine entscheidende Rolle.

Das Kriterium der Begrenzung der Evakuierungszeit kann nur in Kombination mit der globalen Flexibilität betrachtet werden. Nur wenn Letztere gewährleistet ist, gelten die Gleichungen in diesem Abschnitt. Es muss also ein weiterer Server verfügbar sein, der die virtuellen Maschinen des zu evakuierenden Servers aufnimmt. In diesem Fall bestimmt das Migrationszeitmodell die Evakuierungszeit. Die Netzwerkarchitektur stellt, wie in Kapitel 2.1.5 dargestellt, die benötigte Bandbreite im dedizierten Netz zur Verfügung und wirkt damit nicht beschränkend.

Die Zeit, die für eine Migration von VM  $v$  benötigt wird, wird mit  $\psi_{mig_v}$  angegeben. Diese wird mittels eines Modells abgeschätzt, das in Kapitel 5.6.2 und insbesondere in Gleichung 5.27 beschrieben wird.

Die Zeit, die benötigt wird, um den Server  $s$  komplett zu leeren und damit zu evakuieren, also alle VMs von diesem zu entfernen, entspricht der benötigten Migrationszeit  $\psi_{eva_s}$  aller virtueller Maschinen auf diesem:

$$\psi_{eva_s} = \sum_{i=1}^{n_{VM_s}} \psi_{mig_{v_i}} \quad (4.24)$$

Es muss nun für alle Server  $s \in S$  gelten:

$$\psi_{eva_s} \leq \psi_{eva_{max}} \quad (4.25)$$

Dabei gibt  $\psi_{eva_{max}}$  die in den SLA bzw. durch den Betreiber festgelegte maximale Zeitdauer einer Evakuierung der Server an. Der Wert muss durch den Betreiber festgelegt werden und sollte möglichst auf Erfahrungen oder bestenfalls thermischen Simulationen basieren, da hierdurch konkret festgestellt werden kann, wie lange in einem thermischen Notfall eine (Teil-)Evakuierung dauern darf. Allerdings sollte die Grenze auch so hoch wie möglich liegen, da jeder Server nur eine Migration gleichzeitig stabil durchführen kann (siehe dazu Migrationsmodelle in Kapitel 5.6.2). Bei kleiner werdendem  $\psi_{eva_{max}}$  sinkt damit auch automatisch die maximale Anzahl der virtuellen Maschinen pro Server.

#### 4.4.4 Lösungsansatz

In diesem Kapitel 4.4 wurden bereits während der Beschreibung der einzelnen Aspekte und Metriken Lösungsansätze genannt bzw. darauf hingewiesen, dass nur eine heuristische

Lösungsstrategie in realistischer Zeit möglich ist. In diesem Abschnitt werden noch einmal die Lösungsansätze der drei behandelten Bereiche zusammengefasst:

- *Lokale Flexibilität*: Die lokale Flexibilität bildet eine harte Grenze der Ressourcenbelastung und damit der möglichen VM-Kapazität eines Servers. Das Kriterium kann deterministisch bestimmt werden, es wird dafür vor der Allokation einer virtuellen Maschine auf einen Server überprüft, ob durch diese Allokation die lokale Flexibilität verletzt wird. Nur wenn dies nicht der Fall ist, kann diese Allokation erfolgen.
- *Globale Flexibilität*: Der zu erreichende Wert der globalen Flexibilität  $f_g$  einer Allokation ist ebenfalls eine fest definierte Zahl, die nicht unterschritten werden darf. Allerdings ist das Erreichen dieses Werts nur heuristisch möglich, da eine direkte Berechnung für eine realistische Anzahl von Servern in einem Rechenzentrum in realistischer Zeit nicht möglich ist. Daher muss eine Bewertungsfunktion gefunden werden, die Server bevorzugt, die zu einer größeren globalen Flexibilität beitragen und gleichzeitig den anderen Kriterien entsprechen.
- *Evakuierungszeit*: Dieses Kriterium gilt nur in Verbindung mit einer erfüllten globalen Flexibilität und ist dann als feste, zu erreichende Grenze anzusehen. Mittels der Modelle zur Abschätzung der Migrationszeit kann für jeden Server unter einer Allokation die Evakuierungszeit bestimmt werden. Ein Absenken dieser Zeit ist nur durch Entfernen von virtuellen Maschinen eines Servers möglich. Die Modelle, die für diese Bestimmung der Migrationszeit verwendet werden, werden in Kapitel 5.6.2 beschrieben.

## 4.5 Formale Darstellung

Die in diesem Kapitel bisher vorgestellten Analysen und Problembeschreibungen werden hier zusammengefasst und als multi-kriterielles Optimierungsproblem formal beschrieben. Diese Optimierung hat zum Ziel den Vektor  $\vec{F}(\vec{a})$ , der die einzelnen Zielfunktionen enthält, zu minimieren, wobei die Allokation  $\vec{a}$  als Eingabe dient:

$$\text{Minimiere } \vec{F}(\vec{a}) = \left[ f_1(\vec{a}), f_2(\vec{a}), f_3(\vec{a}) \right] \quad (4.26)$$

Im Folgenden werden die einzelnen Zielfunktionen definiert und es werden die jeweiligen Nebenbedingungen dargelegt, die während der Analyse identifiziert wurden.

Die Funktion  $f_1(\vec{a})$  beschreibt die Optimierung der Leistungsaufnahme der Server nach den Gleichungen 2.7, 2.8 und 2.9.:

$$f_1(\vec{a}) = \sum_{i=1}^n P(s_i) \Big|_{s_i \in S_{active}}, P(s_i) = spm_{s_i} \left( \sum_{j=1}^m u_{v_j} \right), v_j \in \vec{a}_{v_s} \quad (4.27)$$

Für die Minimierung der Funktion  $f_1$  gelten folgende Nebenbedingungen:

Jede virtuelle Maschine muss einem Server zugeordnet werden:

$$\forall v \in V : \exists \vec{a}_v \in \vec{a} : v \in \vec{a}_v \quad (4.28)$$

Die Kapazitätsgrenzen der Server inklusive lokaler Flexibilität müssen eingehalten werden:

$$\forall s \in S_{active} : \sum_{i=1}^n (u_{v_i} \cdot (1 + f_{t_{v_i}})) \leq u_{scap} \quad (4.29)$$

Die Temperaturgrenzen der Server müssen eingehalten werden:

$$\forall s \in S_{active} : t_{swarm} = stm_s(t_{scold}, P_s) < t_{swmax} \quad (4.30)$$

Die maximal erlaubte Evakuierungszeit jedes Servers muss eingehalten werden:

$$\forall s \in S_{active} : \psi_{eva_s} \leq \psi_{eva_{max}} \quad (4.31)$$

Die Funktion  $f_2(\vec{a})$  beschreibt die Optimierung der Leistungsaufnahme der Kühlung:

$$f_2(\vec{a}) = \sum_{i=1}^n P_{crah_i} + P_{chiller} \quad (4.32)$$

mit  $P_{crah_i}$  als der jeweiligen Leistungsaufnahme der aktiven Umluftkühlgeräte und der Leistungsaufnahme der Kältemaschinen  $P_{chiller}$  nach Gleichung 4.14:

$$P_{chiller} = kpm \left( t_{water_c}, f_1(\vec{a}) + P_{ups_{loss}}(\vec{a}) + \sum_{i=1}^n P_{crah_i} + \sum_{j=1}^m P_{r_j} \right) \Big|_{r_j \in R_{active}} \quad (4.33)$$

$P_{ups_{loss}}(\vec{a})$  beschreibt die zusätzliche (Eigen-)Leistungsaufnahme der USV, die sich implizit durch die eingenommene Allokation einstellt und nach der Analyse in diesem Kapitel keiner dedizierten Optimierung unterliegt. Für jeden aktiven IT-Raum fällt eine grundlegende Leistungsaufnahme  $P_r$  an, die durch die benötigten unterstützenden Geräte in diesem IT-Raum zustande kommt. Die Maximierung von  $t_{water_c}$  ist zudem ein Ziel, wie in diesem Kapitel analysiert, wobei hier die Abhängigkeit wiederum von der Verteilung der elektrischen Leistungsaufnahmen in den IT-Räumen nach Gleichung 4.3 gilt. Voraussetzung für die Minimierung von  $f_2$  ist damit auch die Anhebung von  $t_{water_c}$  und die in Gleichung 4.3 formulierte Angleichung der Wärmeleistung der einzelnen IT-Räume.

Für die Optimierung der Leistungsaufnahme der Kühlung gilt, dass eine Senkung der Kühlleistung die Nebenbedingung aus Gleichung 4.9 einhalten muss, also jedes Umluftkühlgerät eine Kühlkapazität bereitstellen kann, die mindestens der Wärmemenge im jeweiligen IT-Raum entspricht.

Die Funktion  $f_3(\vec{a})$  beschreibt die Minimierung des Variationskoeffizienten der Gleichverteilung der Wärmeleistung in den IT-Räumen nach Gleichung 4.21:

$$f_3(\vec{a}) = max \left( \rho_{K_1}, \rho_{K_2}, \dots, \rho_{K_n} \right) \Big|_{n = |R_{active}|} \quad (4.34)$$

### Problemklassen

Die Verteilung der virtuellen Maschinen auf die Server zur Minimierung der Funktion  $f_1(\vec{a})$  (Gleichung 4.27) entspricht einem Bin-Packing-Problem. Dies wurde bereits in Abschnitt 4.2 in Gleichung 4.4 dargestellt. Da die ausschließliche Minimierung der Anzahl der aktiven Server (dies entspricht dem klassischen Bin-Packing) nicht dem Optimierungsziel der Analyse aus diesem Kapitel entspricht, kann stattdessen eine Kostenfunktion verwendet werden, die minimiert werden soll. Dies entspräche dann der Funktion  $f_c(\vec{a}) = f_1(\vec{a}) + f_3(\vec{a})$  und würde die Verteilung der virtuellen Maschinen auf die Server abdecken. Dabei wird sowohl die elektrische Leistungsaufnahme minimiert, als auch eine thermische Gleichverteilung innerhalb der IT-Räume angestrebt.

Die Optimierung der Leistungsaufnahme der Kühlung hängt maßgeblich von dem jeweiligen Wert von  $f_1(\vec{a})$  und der Kaltwassertemperatur  $t_{water_c}$  ab. Die Verteilung der elektrischen Leistungsaufnahme auf die einzelnen IT-Räume zur Erhöhung von  $t_{water_c}$  kann mit der oben angegebenen Kostenfunktion  $f_c(\vec{a})$  aber nicht mehr effizient kombiniert werden.

Daher wird als Lösungsansatz eine getrennte Betrachtung der einzelnen Optimierungen vorgeschlagen. Für alle Lösungsverfahren der genannten Optimierungen gilt jedoch, dass eine heuristische Herangehensweise gewählt wird, da die Suche nach dem Optimum bei den Dimensionen von heutigen Rechenzentren nicht in einer hinreichend kurzen Zeitspanne realisiert werden könnte.

## 4.6 Kombiniertes Lösungsansatz

Bisher wurden in diesem Kapitel die Analysen und identifizierten Einzelkriterien, die das Optimierungsverfahren berücksichtigen soll, sowie die formale Problemstellung beschrieben. An dieser Stelle erfolgt nun die Betrachtung der Kombination der genannten Kriterien. In Abbildung 4.8 findet sich dazu eine Übersicht über die Kriterien und deren einzelner Parameter. Einige dieser Parameter bilden feste Grenzwerte, die entweder nicht überschritten werden dürfen oder erreicht werden sollen. Hierzu zählen die Server-Temperaturgrenzen, die Evakuierungszeit und die lokale Flexibilität. Diese Parameter werden mit Modellen abgeschätzt und während der Zusammenstellung einer Allokation überprüft. Wird einer der Werte überschritten, kann die gefundene Allokation nicht als Lösung verwendet werden. Entweder muss dann eine Veränderung der Allokation erfolgen, oder eine Lösung ist mit den vorgegebenen Parametern nicht möglich. Auch der effizienteste Auslastungsbereich eines Servers ist ein fester Wert, der allerdings möglichst genau erreicht werden sollte. Ein Überschreiten führt hier aber nicht sofort zu einer illegalen Allokation.

Weitere Parameter der einzelnen Kriterien stellen jedoch Bewertungen dar, zwischen denen abgewogen werden muss. Dazu zählen die Effizienz eines jeden Servertyps, also die Effizienz, die ein Server grundsätzlich besitzt, unabhängig von dessen Auslastungszustand. Zusätzlich gehört die Wahl der aktiven Server nach der Temperaturverteilung im IT-Raum und der globalen Flexibilität zu den Parametern, die einer Bewertung unterliegen. Diese drei Parameter müssen jeweils einzeln bewertet, normalisiert und dann miteinander kom-

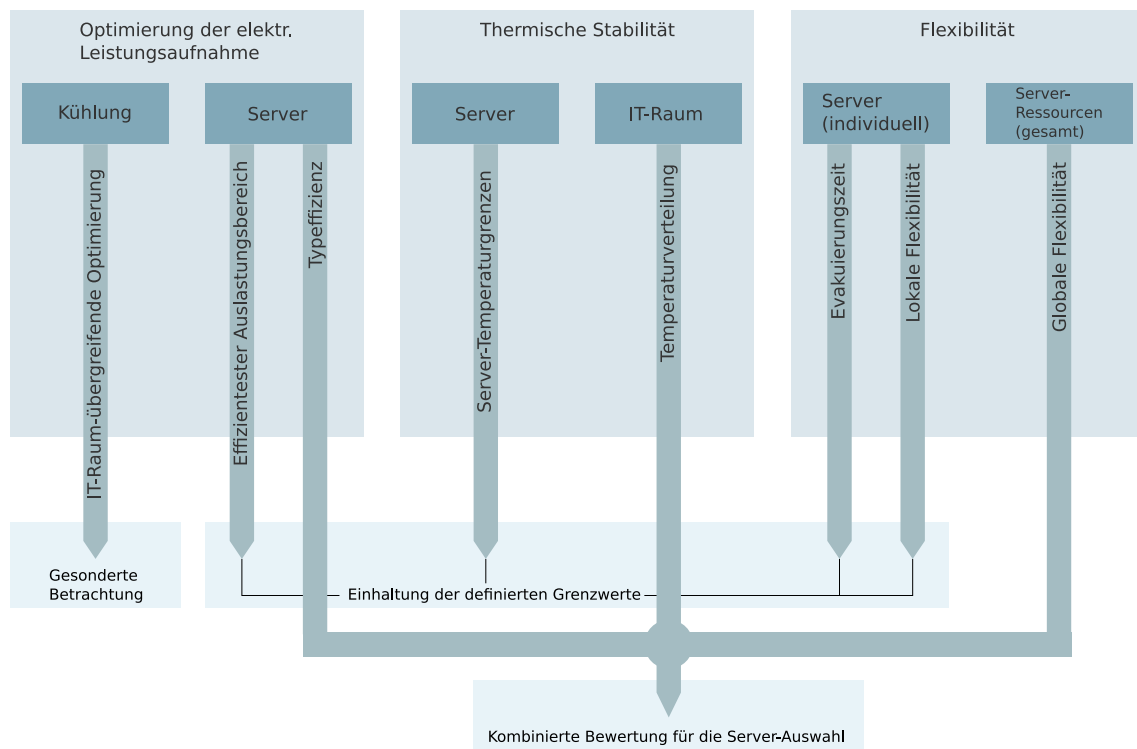


Abbildung 4.8: Kombination der verschiedenen Kriterien und Einflüsse zur Lösungssuche

biniert werden. Dadurch ergibt sich eine ganzheitliche Server-Bewertung, die die Wahl der aktiven Server maßgeblich bestimmt.

Die IT-Raum-übergreifende Optimierung bezieht sich auf die Regelung der Kaltwassertemperatur und wird aufgrund des signifikanten Optimierungspotentials gesondert betrachtet. Das Optimierungsverfahren nimmt bei der Suche nach einer optimierten Allokation diese einzelnen Überprüfungen und Bewertungen der Parameter vor. Daher erfolgt die detaillierte und formalisierte Beschreibung der einzelnen Parameter in Kapitel 6.

## 4.7 Zusammenfassung

In diesem Kapitel wurden die energetischen Einflussfaktoren auf den Energiebedarf des gesamten Rechenzentrums analysiert. Als zentrale regelbare Faktoren wurden dabei die Server und die Kühlung identifiziert. Die USV-Geräte tragen zwar auch maßgeblich zum Energiebedarf bei, sind aber aufgrund von Redundanz und zentraler Architektur nicht regelbar. Somit hat sich das Zusammenspiel der aktiven Server und die Verteilung in den IT-Räumen als signifikanter Faktor für die Kühleffizienz und für den Energiebedarf des Rechenzentrums herausgestellt.

Damit ein stabiler Betriebszustand im Rechenzentrum eingenommen wird, wurden neben der energetischen Optimierung weitere Kriterien definiert, die insbesondere die thermische



Sicherheit fokussieren. Einem aktuell häufig genannten Argument von Rechenzentrumsbetreibern, dass Konsolidierung zu unflexibel sei, wird mit der Definition von Flexibilitätskriterien entgegnet.

Die Modellierung der beteiligten Geräte im Rechenzentrum sowie des Verhaltens von virtuellen Maschinen folgt nun in Kapitel 5.



# 5 Rechenzentrumsmodellierung

In Kapitel 4 wurden mehrere Kriterien definiert, die das zu entwickelnde Optimierungsverfahren berücksichtigen soll, um einen optimierten Betriebszustand im Rechenzentrum zu erreichen. Für die Bewertung einer Allokation hinsichtlich der einzelnen Kriterien werden Modelle benötigt, die den Energiebedarf und die thermische Leistung von verschiedenen Geräteklassen abschätzen, sowie das Verhalten von virtuellen Maschinen beschreiben können. Auch werden energetische Modelle von Geräten benötigt, die nicht dem Optimierungsprozess unterliegen, aber deren Energiebedarf trotzdem einen signifikanten Einfluss auf den Gesamtbedarf des Rechenzentrums besitzen und damit in die Gesamtdarstellung mit einfließen sollen. Diese Modelle werden in diesem Kapitel vorgestellt und definiert.

Es werden zunächst die Modelle aus den einzelnen Gerätebereichen unabhängig voneinander vorgestellt. Am Ende des Kapitels wird dann dargelegt, wie die Modelle zusammenhängen und in einer Rechenzentrumssimulation verwendet werden. Das Kapitel beginnt mit der Beschreibung der Anforderungen an die Modelle, der Modellierung der USV und danach der Server. Es folgen dann die elektrischen und thermischen Modellierungen der Umluftkühlgeräte und der Kältemaschinen. Danach werden die Auswirkungen der Auslastungen virtueller Maschinen auf verschiedene Server und auf Live-Migrationen untersucht und es erfolgt eine kurze Darstellung der elektrischen Leistungsaufnahme der Netzwerkgeräte.

## 5.1 Modellanforderungen

In diesem Abschnitt werden die Anforderungen an die Modelle aufgelistet, die sich aus der Analyse aus Kapitel 4 und dem darin beschriebenen Optimierungsproblem ergeben.

### Modellierung der USV

In Kapitel 4 wurde die Funktion und Bedeutung der USV-Geräte analysiert und beschrieben. Für die Modellierung ergibt sich daraus die Anforderung, dass die Leistungsaufnahme der USV-Geräte abgeschätzt werden soll, damit eine möglichst vollständige Sicht auf die Leistungsaufnahme des gesamten Rechenzentrums möglich wird. Gleichzeitig unterliegen die USV-Geräte aber keiner dedizierten Optimierung, sodass hier lediglich die Verlustleistung unter einer sich einstellenden Auslastung und die Kapazitätsgrenzen dieser Geräte betrachtet werden sollen. Weitergehende Untersuchungen bezüglich Netzfrequenz, Phasenschieflage oder ähnlichen Effekten, die beim Betrieb von USV-Geräten auftreten können, werden hier nicht weiter beachtet.

### Modelle für Server

Die Verteilung der virtuellen Maschinen auf die Server ist einer der Kernpunkte der energetischen Optimierung und zudem ein zentraler Bestandteil des Aspekts der Betriebssicherheit im Rechenzentrum. Damit eine Optimierung hinsichtlich der Energieeffizienz der Server erfolgen kann, müssen Modelle entwickelt und verwendet werden, die eine Untersuchung der Auswirkungen verschiedener Auslastungen ermöglichen. Hier ist insbesondere das Modell der Leistungsaufnahme der Server zu nennen, das in bisherigen Gleichungen bereits mit der Funktion  $spm_s(\dots)$  beschrieben wurde. Dieses Modell muss für einzelne Server erstellt und charakterisiert werden können, damit eine energetische Betrachtung der Server und der IT-Räume insgesamt erfolgen kann. Die Abschätzungen dieses Modells fließen weiterhin in die Temperaturbetrachtung der Server und in die Bestimmung der benötigten Kühlenergie mit ein. Dadurch kommt diesem Modell eine hohe Bedeutung innerhalb der gesamten Modellierung zu.

Die Nebenbedingungen, die im Optimierungsproblem in Kapitel 4.5 für den Betrieb der Server formuliert wurden, müssen ebenfalls mittels Modellen überprüft werden können. Hier gilt für alle Nebenbedingungen ein pessimistischer Ansatz; die Modelle dürfen also das Ergebnis nicht unterschätzen. Ansonsten ist eine Modellierung mit geringerer Genauigkeit ausreichend.

### Modellierung der Klimatisierung

Die elektrische Leistung im IT-Raum bestimmt die darin anfallende Wärmeleistung. Diese muss für den Betrieb der Umluftkühlgeräte und der dementsprechend benötigten Dimensionierung der Kälteleistung ermittelt werden können. Daher werden auch beispielsweise Netzwerkgeräte im IT-Raum betrachtet, die ansonsten aber keiner Optimierung unterliegen.

Damit die Umluftkühlgeräte in den IT-Räumen die entsprechende anfallende Wärmeleistung abtransportieren können, müssen sie eine äquivalente Kälteleistung bereitstellen. Diese hängt maßgeblich von der Kaltwassertemperatur des Kühlmediums ab. Die Modellierung der Umluftkühlgeräte muss also diesen Zusammenhang fokussieren, wobei hier insbesondere der direkte Einfluss der Kaltwassertemperatur auf die damit verbundene Kälteleistung modelliert werden muss.

Die Modellierung der Kältemaschinen beschränkt sich auf deren elektrische Leistungsaufnahme, da weitere Effekte des Betriebs dieser Geräte für die Analysen in dieser Arbeit nicht relevant sind.

### Vergleichbarkeit verschiedener Server-Typen

Da das Optimierungsverfahren in dieser Arbeit heterogene Server-Umgebungen betrachten soll, muss zunächst die Vergleichbarkeit verschiedener Server-Typen untersucht und mittels entsprechender Modelle abgebildet werden. Dies gilt insbesondere im Kontext der Virtualisierung. Hier entsteht die Fragestellung, wie sich äquivalente virtuelle Maschinen mit einer vergleichbaren Anwendung auf verschiedenen Server-Typen hinsichtlich der CPU-

Auslastung verhalten. Dazu muss zunächst eine vergleichbare Anwendungsauslastung gefunden werden, durch die eine einheitliche Bewertung des Nutzens ermöglicht wird. Die entsprechenden Modelle müssen dabei den Zusammenhang dieses Nutzens der Anwendung und der dabei entstehenden CPU-Auslastung auf dem jeweiligen Server herstellen, da die CPU-Auslastung in dieser Arbeit als maßgeblicher Wert für die energetische Modellierung von Servern verwendet wird.

## 5.2 Unterbrechungsfreie Stromversorgung

Die unterbrechungsfreie Stromversorgung sorgt neben der Überbrückung von kurzzeitigen (in der Regel bis zu mehreren Minuten) Stromausfällen auch für eine Glättung der Spannungskurve, sie gleicht also Spannungsspitzen und Frequenzschwankungen aus. Die dazu notwendigen Spannungswandlungen führen allerdings zu einem Verlust, ebenfalls besitzt die USV unter anderem aufgrund der Steuerungselektronik einen Eigenenergiebedarf. In

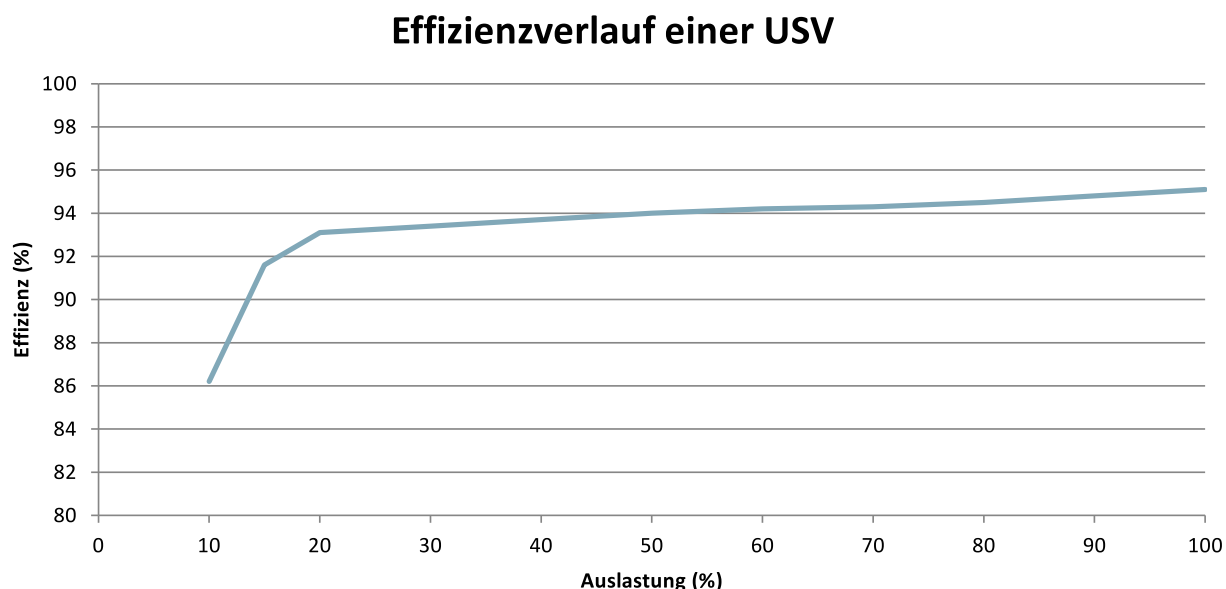


Abbildung 5.1: *Effizienzverlauf einer typischen USV (eigene Darstellung, Daten aus [134])*

Abbildung 5.1 ist der Effizienzverlauf in Abhängigkeit von der prozentualen Auslastung einer typischen USV angegeben [134]. Als Auslastung wird hier die gesamte angeschlossene elektrische Last betrachtet, also alle Geräte, die durch die USV versorgt werden. Diese sind maßgeblich die Server und die Netzwerkgeräte (siehe dazu auch die Abbildung 2.3 der Stromversorgungsarchitektur in Kapitel 2.1.3). Die prozentuale Auslastung bezieht sich auf die maximale Anschlusskapazität  $P_{ups_{cap}}$  der USV. Es ist zu erkennen, dass die Effizienz mit der Auslastung steigt, vor allem im Bereich einer sehr niedrigen Auslastung ( $< 30\%$ ) ist die Steigerung signifikant, darüber hinaus nur noch marginal. Die Effizienz der USV beschreibt den Anteil der elektrischen Leistung, die die USV als Nutzleistung an die ange-

schlossenen Geräte abgibt. Ist die Effizienz beispielsweise 90%, so gibt die USV 90% der Leistung, die sie selbst aufnimmt, an die Geräte weiter.

Die gesamte elektrische Leistungsaufnahme  $P_{ups}$  einer USV mit der Gesamtkapazität  $P_{ups\,cap}$  setzt sich wie folgt zusammen:

$$P_{ups} = \frac{P_{ups\,load}}{e_{ups}(P_{ups\,load})} \mid P_{ups\,load} \leq P_{ups\,cap} \quad (5.1)$$

Dabei gibt die Funktion  $e_{ups}$  die Effizienz unter der anliegenden Auslastung  $P_{ups\,load}$  an. Die Effizienzfunktion entspricht dabei der in Abbildung 5.1 dargestellten Funktion. Alternativ formuliert gibt die Differenz von  $P_{ups}$  und  $P_{ups\,load}$  also die Verlustleistung der USV an.

Wie bereits im vorherigen Kapitel erwähnt, erfolgt keine spezielle Optimierung hinsichtlich der USV-Geräte. Das USV-Leistungsmodell wird aber zur Berechnung der gesamten elektrischen Leistungsaufnahme des Rechenzentrums verwendet.

### Stromverteilung

Zur Stromverteilung innerhalb der Racks werden Power Distribution Units (PDU) verwendet. Diese Geräte können über eine eigene Netzwerkanbindung gesteuert werden und beispielsweise einzelne Ports ein- und ausschalten. Auch bieten mehrere Hersteller integrierte Messgeräte für die elektrische Leistung und Temperatur an, wodurch eine durchgängige Protokollierung der elektrischen Leistungsaufnahmen der angeschlossenen Server und der jeweiligen Lufttemperatur ermöglicht wird. Diese PDUs benötigen für diese Funktionalität ebenfalls Energie, allerdings ist die benötigte Leistung sehr gering (moderne PDUs benötigen aufgrund Energie-optimierter Relais unter 10 Watt [110]) und zudem annähernd konstant. Aus diesem Grund wird die Leistungsaufnahme von PDUs nicht modelliert und in der Gesamtbetrachtung nur als konstanter Wert berücksichtigt.

## 5.3 Server

Die Server im Rechenzentrum stellen eine wesentliche Gruppe energetischer Verbraucher dar und sind zudem energetisch regelbar, indem ihre Auslastung verändert wird. Die Modelle müssen genau diese Veränderungen abbilden können. In diesem Unterkapitel wird zunächst die elektrische Modellierung und anschließend die thermische Modellierung von Servern vorgenommen.

### 5.3.1 Modell der elektrischen Leistungsaufnahme

Das Modell der elektrischen Leistungsaufnahme der Server wird dazu verwendet, auf Basis einer zu erwartenden Ressourcenauslastung die dabei entstehende elektrische Leistungsaufnahme eines Servers abzuschätzen. Als maßgeblicher Indikator, durch den die vollständige Auslastung eines Servers definiert wird, wurde dazu die Prozessorauslastung identifiziert. Es wurde bereits mehrfach gezeigt, dass diese stellvertretend für das gesamte System verwendet werden kann [82, 103, 113]. Auf Basis dieser Untersuchungen kann die elektrische

Leistungsaufnahme mit ausreichender Genauigkeit durch eine quadratische Gleichung beschrieben werden:

$$P_s = \text{spm}_s(u_s) = a + b \cdot u_s + c \cdot u_s^2 \quad (5.2)$$

Das Modell geht von einer statischen Grundleistungsaufnahme eines Servers (entspricht der Leistungsaufnahme im Idle-Zustand) und einem dynamischen Anteil aus. Letzterer ist demnach von der Auslastung  $u_s$  des Servers abhängig. Die quadratische Komponente resultiert aus dem Einsatz von Dynamic Voltage and Frequency Scaling (DVFS) [81, 104, 121].

## Charakterisierung

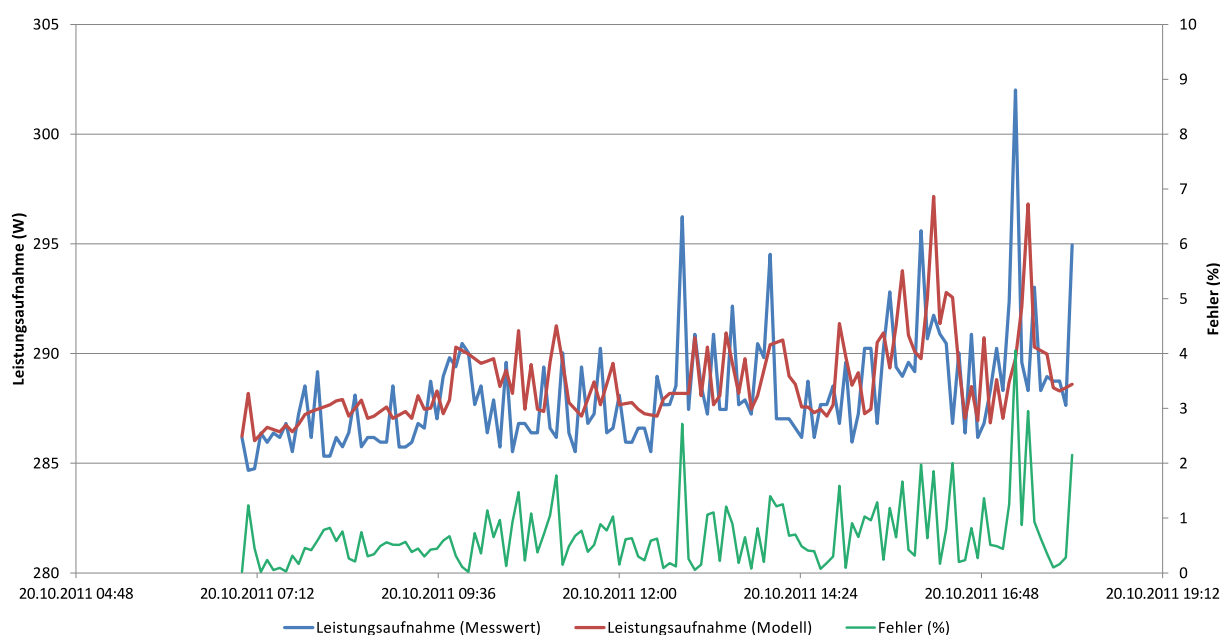


Abbildung 5.2: Die Modellgenauigkeit des Server-Leistungsmodells: die blaue Kurve zeigt die real gemessene Leistungsaufnahme des untersuchten Servers, die rote den Wert des Modells. Die grüne Kurve zeigt den prozentualen Modellfehler (Quelle: eigene Messungen).

Zur Charakterisierung dieses Modells wurden Messreihen in dem während dieser Arbeit untersuchten realen Rechenzentrum durchgeführt und in [56] veröffentlicht. Die Anwendungen, die im Messzeitraum ausgeführt wurden, waren reale Geschäftsanwendungen, wie beispielsweise Webserver, Applikationsserver und Terminalserver, wodurch die gemessenen Daten realen Ansprüchen entsprechen. Die mittels Regression aus diesen Messdaten bestimmte Charakterisierung beträgt:

$$a = 274,679; b = 0,00307297; c = 8,14862 \cdot 10^{-8} \quad (5.3)$$

Damit wurde eine Modellgenauigkeit von über 97% mit maximalen Fehlern (Peaks) von 9% erreicht. In Abbildung 5.2 ist sowohl ein Ausschnitt der Messdaten (blaue Kurve) als auch der Wert des Modells (rote Kurve) dargestellt. Die grüne Kurve zeigt den prozentualen Modellfehler.

## SPECpower-Modelle

Da das in dieser Arbeit entwickelte Optimierungsverfahren heterogene Server-Umgebungen unterstützen und gezielt ausnutzen soll, ist es sinnvoll, eine möglichst breite Palette unterschiedlicher Server-Typen und deren Modelle zu nutzen. Eine frei verfügbare Datenbank, die eine Auswertung der elektrischen Leistungsaufnahmen in Abhängigkeit von der Auslastung von verschiedenen Servern enthält, ist die *SPECpower-ssj 2008* [124]. In dieser befinden sich mehr als zweihundert Charakterisierungen für verschiedene Server verschiedener Hersteller. Die Daten liegen in Form einer Tabelle vor, wobei der prozentualen Auslastung im Intervall  $[0, 100]$  im Abstand von 10%-Schritten jeweils ein Energiebedarfswert zugeordnet wird. Für Zwischenwerte der Auslastung wird eine lineare Interpolation vorgenommen.

Diese Modelle werden primär für die Bewertung verschiedener Heuristiken zur Lösungssuche in Kapitel 6 eingesetzt, wenn eine Vielzahl von Servern als Vergleich benötigt wird.

### 5.3.2 Temperaturmodell

Das Server-Temperaturmodell muss die Temperatur der ausgehenden warmen Luft eines Servers abschätzen. Das Optimierungsverfahren nutzt dieses Modell, um zu entscheiden, ob ein Server mit einer bestimmten Auslastung betrieben werden darf, oder ob damit die in Kapitel 4.3.1 definierten Grenzwerte überschritten werden. Das Modell wird auf Basis der thermodynamischen Grundlagen erstellt, wonach eine Temperaturerhöhung eines Mediums von der Flussrate und den Eigenschaften des Mediums sowie der zugeführten Energiemenge abhängt. Die zugeführte Wärmemenge  $\dot{Q}_s$  entspricht hier wiederum nach Gleichung 2.2 der vom Server aufgenommenen elektrischen Leistung  $P_s$ . Demnach kann die Temperatur der ausgehenden warmen Luft  $t_{s_{warm}}$  des Servers  $s$  wie folgt berechnet werden:

$$t_{s_{warm}} = \frac{\dot{Q}_s}{\dot{m}_{air_s} \cdot c_{p_{air}}} + t_{s_{cold}} \quad (5.4)$$

Dabei ist  $t_{s_{cold}}$  die Temperatur der eingehenden kalten Luft und  $\dot{m}_{air_s}$  der Massenstrom der Luft. Dieser Wert wurde im untersuchten Rechenzentrum beispielsweise mit  $\dot{m}_{air_s} = 0,02 \frac{kg}{s}$  ermittelt. Die spezifische Wärmekapazität der Luft wird als Konstante mit  $c_{p_{air}} = 1,002 \frac{kJ}{kg \cdot K}$  angenommen. Die entsprechende Modellfunktion  $stm_s(t_{s_{cold}}, P_s)$ , die in Kapitel 4.3.1 bereits erwähnt wurde, kann damit durch Gleichung 5.4 abgebildet werden. Die Wärmeleistung  $\dot{Q}_s$  entspricht hier wiederum  $P_s$ .

Dieses Modell wird vom Optimierungsverfahren nur für zukünftige Abschätzungen während der Planung verwendet. Damit im laufenden Betrieb Hot-Spots frühzeitig erkannt werden,



muss die Temperatur der Server mittels der in den Servern integrierten Temperatursensoren permanent überwacht werden. Anomalien im Temperaturverlauf können dann gezielt ermittelt werden. Hier sollten existierende Verfahren der Datenstromanalyse zur Anwendung kommen, beispielsweise Hauptkomponentenanalyse (Principal Component Analysis) [58, 78]. Als Erweiterung kann auch ein Temperaturmodell auf Basis dieser Analysen erstellt werden, das dann dynamisch während des Betriebs lernt. Damit können weitere Einflüsse, wie beispielsweise überhitzte Nachbarserver unter Umständen besser erkannt werden. In dieser Arbeit werden diese Erweiterungen allerdings nicht weiter betrachtet.

## 5.4 Umluftkühlgerät

Die Umluftkühlgeräte befinden sich in den IT-Räumen und versorgen die Server direkt mit Kaltluft, die in die Kaltgänge geblasen wird. Die Geräte selbst bestehen allerdings nur aus einem Wärmeübertrager (hier meist Kreuzwärmeübertrager) und Ventilatoren, die für die Luftbewegung zuständig sind. Die Kälteenergie wird über das Kühlwasser geliefert.

Die im untersuchten Rechenzentrum genutzten Geräte sind alle vom Typ *Emerson Network Power L15UC*, wobei in jedem IT-Raum jeweils zwei Geräte platziert sind.

In den folgenden Abschnitten wird eine Modellierung der Umluftkühlgeräte nach elektrischer Leistungsaufnahme und Kühlleistung am Beispiel des genannten Geräts vorgenommen. Das Verfahren ist auf andere Geräte vergleichbarer Bauart direkt übertragbar und damit allgemeingültig. Da die Modellierung und Charakterisierung des Kühlleistungsmodells aber komplex ist und aus mehreren Schritten besteht, werden aus Gründen der Übersichtlichkeit und Verständlichkeit direkt gemessene Werte und Charakterisierungen bei der Beschreibung des Modells mit angegeben.

### 5.4.1 Modell der elektrischen Leistungsaufnahme

Die elektrische Leistungsaufnahme von Umluftkühlgeräten besteht fast ausschließlich aus der Leistungsaufnahme der Ventilatoren. Die Geräte besitzen zwar eine Steuerelektronik, deren Energiebedarf ist aber so gering, dass er vernachlässigt werden kann. Prinzipiell muss bei den Ventilatoren zwischen automatisch drehzahlgeregelten und konstanten Motoren unterschieden werden. Die elektrische Leistungsaufnahme von drehzahlgeregelten Motoren kann mittels der Ventilatorproportionalitätsgesetze [32] berechnet werden. Die hier verwendeten Geräte werden jedoch konstant betrieben, wobei die Drehzahl der Motoren manuell auf ca. 50% des Maximums reguliert wurde. Dadurch liegt eine nahezu konstante Leistungsaufnahme vor, die durch Messungen festgestellt wurde. Die elektrische Leistungsaufnahme  $P_{crah}$  eines einzelnen Umluftkühlgeräts wurde so als ein konstanter Wert von

$$P_{crah} = 3,8kW \quad (5.5)$$

ermittelt. Dieser Wert wird durch das vom Hersteller erhaltene Datenblatt des Geräts bestätigt. Hier ist eine maximale Leistungsaufnahme der drei Ventilatoren von 6,84kW angegeben, diese gilt allerdings nur bei maximaler Drehzahl.

Grundsätzlich kann es sinnvoll sein, die Leistungsaufnahme der Ventilatoren nach dem genannten Gesetz zu bestimmen, da hierdurch die Genauigkeit des Modells erhöht wird. Allerdings sind hierfür auch durchgehend vorliegende Messwerte der Drehzahl notwendig, während der Einfluss dieser Modellverbesserung auf die gesamte Leistungsaufnahme des IT-Raums und des Rechenzentrums nur gering ist. Auch aus diesem Grund wird in dieser Arbeit und für das hier entwickelte Optimierungsverfahren diese vereinfachte Betrachtung vorgenommen.

### 5.4.2 Kühlleistungsmodell

Das Kühlleistungsmodell für ein Umluftkühlgerät soll angeben, welche reale Kühlleistung mit dem spezifischen Gerät, den gegebenen Medien, Temperaturen und Volumenströmen erreicht werden kann. Umgekehrt kann dann bei bekannter benötigter Kühlleistung ermittelt werden, welche Temperaturen eingestellt werden müssen. Vor allem durch die Erhöhung der Temperatur des Kühlmediums, im hier betrachteten Fall dem Wasser-Glykolegemisch (34% Glykol), können signifikante Energieeinsparungen bei den Kältemaschinen erreicht werden (vgl. Kapitel 5.5). Die im Folgenden beschriebenen Verfahren und Gleichungen entstammen der Fachliteratur für Thermodynamik und Wärmetechnik [14, 21, 138] und können dort zur Vertiefung nachgeschlagen werden.

#### Ziel des Kühlleistungsmodells:

Es gilt, mittels des Modells die wärmste noch verwendbare Kaltwassertemperatur  $t_{water,c}$  auf Basis der vom IT-Raum  $r$  benötigten Kühlleistung  $\dot{Q}_r$  zu bestimmen.

Nach Gleichung 2.2 wird davon ausgegangen, dass die Wärmemenge im IT-Raum und damit die benötigte Kühlleistung der elektrischen Leistungsaufnahme entspricht. Dabei wird die elektrische Leistungsaufnahme nach Gleichung 4.8 durch die Geräte im IT-Raum bestimmt. Weitere Einflussfaktoren wie beispielsweise die Beleuchtung oder weitere Wärmequellen werden in dieser Betrachtung nicht berücksichtigt, da diese keine signifikanten Größenordnungen betragen.

Das zur Bestimmung der Kühlleistung verwendete Modell bezieht sich direkt auf die thermodynamischen Eigenschaften des Wärmeübertragers im Umluftkühlgerät. In der Regel und auch in diesem Fall werden in diesen Geräten Kreuzwärmeübertrager eingesetzt, sodass die entsprechenden physikalischen Gesetzmäßigkeiten zur Anwendung kommen.

Bei der Charakterisierung und Beschreibung des Modells werden mehrere Werte benötigt, die teils konstant sind und teils ermittelt wurde. Zur Übersicht werden die verwendeten Werte in Tabelle 5.1 dargestellt und es wird die jeweilige Datenquelle genannt.

So kann die insgesamt erreichte Wärmeübertragungsleistung  $\dot{Q}_{crah}$ , die auch der Kühlleistung entspricht, durch folgende Gleichung ermittelt werden:

$$\dot{Q}_{crah} = f \cdot U \cdot A \cdot \overline{\Delta T} \quad (5.6)$$

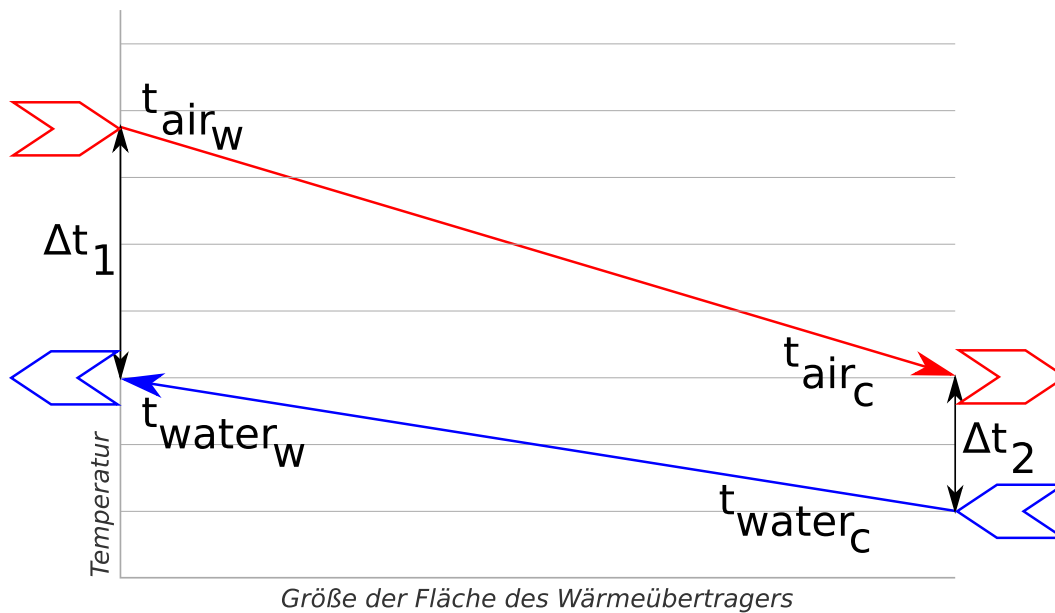


Abbildung 5.3: Erklärung des logarithmischen Mittels der Temperaturdifferenz: Im Wärmeübertrager findet ein Energieaustausch statt, vom warmen (rot) zum kalten (blau) Medium. Die warme Luft tritt in den Wärmeübertrager ein und wird am Kühlwasser entlang geführt. Dabei wird die Wärmeenergie auf das Kühlwasser übertragen, die Luft kühlt sich ab. Diese kalte Luft wiederum tritt aus dem Übertrager aus. Umgekehrt dazu gilt dies für das Kühlwasser. Messbar sind nur die beiden Differenzen  $\Delta t_1$  und  $\Delta t_2$ . Das logarithmische Mittel der Temperaturdifferenzen gibt nun die mittlere Temperaturdifferenz der beiden Medien „dazwischen“ an, also über die Fläche des Wärmeübertragers hinweg.

Hierbei ist  $U \cdot A$  die spezifische Charakterisierung der Wärmeübertragerflächen, bestehend aus dem Wärmeübergangskoeffizienten  $U$  und der Größe der Fläche  $A$ . Da das verwendete Umluftkühlgerät als Kreuzwärmeübertrager modelliert wird, muss ein Korrekturfaktor  $f$  angewendet werden [21].  $\overline{\Delta T}$  ist das logarithmische Mittel der Temperaturdifferenz (im Englischen häufig mit *LMTD* abgekürzt, von *Logarithmic Mean Temperature Difference*):

$$\overline{\Delta T} = \frac{\Delta t_1 - \Delta t_2}{\ln\left(\frac{\Delta t_1}{\Delta t_2}\right)} \quad (5.7)$$

Die einzelnen Temperaturdifferenzen entsprechen:

$$\Delta t_1 = t_{air_w} - t_{water_w}, \Delta t_2 = t_{air_c} - t_{water_c} \quad (5.8)$$

Das logarithmische Mittel stellt die mittlere Temperaturdifferenz der beiden Medien dar, die sich im Wärmeübertrager befinden. Der zugrundeliegende Prozess wird in Abbildung 5.3 dargestellt.

Tabelle 5.1: Benötigte Messwerte für die Charakterisierung des Umluftkühlgerätes

| Beschreibung                                    | Formelzeichen     | Wert  | Datenquelle             |
|---|-------------------|---|-------------------------|
| Charakterisierung der Wärmeübertragerfläche     | $U \cdot A$       | -   | Wird errechnet          |
| Kaltwassertemperatur                            | $t_{water_c}$     | 6,0°C   | Datenblatt + Messung    |
| Warmwassertemperatur                            | $t_{water_w}$     | 12,0°C  | Datenblatt + Messung    |
| Kaltlufttemperatur                              | $t_{air_c}$       | 10,8°C  | Datenblatt + Messung    |
| Warmlufttemperatur                              | $t_{air_w}$       | 25,0°C  | Datenblatt + Messung    |
| Dichte Luft                                     | $\rho_{air}$      | 1,205 $\frac{kg}{m^3}$  | Konstanter Referenzwert |
| Massestrom Luft                                 | $\dot{m}_{air}$   | 25400 $\frac{m^3}{h} = 7,05 \frac{m^3}{s} = 8,5 \frac{kg}{s}$ | Datenblatt + Messung    |
| Massestrom Kühlwasser                           | $\dot{m}_{water}$ | 6,3 $\frac{l}{s} \approx 6,3 \frac{kg}{s}$                    | Messung                 |
| Spezifische Wärmekapazität Luft                 | $c_{p_{air}}$     | 1,005 $\frac{kJ}{kg \cdot K}$                                 | Konstanter Referenzwert |
| Spezifische Wärmekapazität Wasser-Glykolgemisch | $c_{p_{water}}$   | 3,7 $\frac{kJ}{kg \cdot K}$                                   | Konstanter Referenzwert |

### Ermittlung von $U \cdot A$

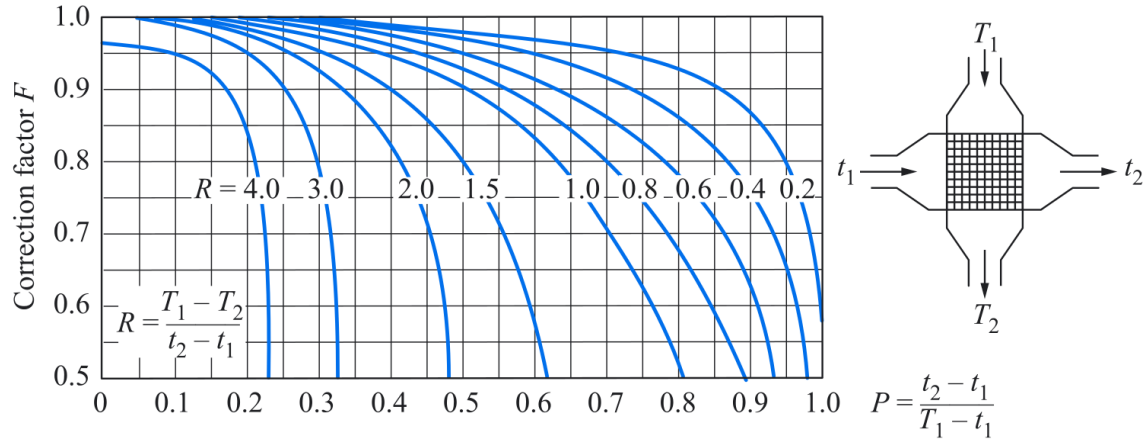
Die spezifische Charakterisierung  $U \cdot A$  ist für das untersuchte Gerät nicht bekannt, sodass diese erst durch Messungen und Berechnungen bestimmt werden muss. Hierzu werden die bekannten Messdaten des Geräts aus der Ist-Situation verwendet und damit die anderen Variablen der Gleichung 5.6 berechnet, woraus dann schließlich  $U \cdot A$  bestimmt werden kann.

Die theoretisch bereitgestellte Kälteleistung durch ein UKG im Idealfall kann durch folgende Gleichung errechnet werden:

$$\begin{aligned}
 \dot{Q}_{craH} &= \dot{m}_{air} \cdot \rho_{air} \cdot (h_1 - h_2) \\
 &= 7,05 \frac{m^3}{s} \cdot 1,205 \frac{kg}{m^3} \cdot (48,3 \frac{kJ}{kg} - 28,1 \frac{kJ}{kg}) \\
 &= 171,74 kW
 \end{aligned} \tag{5.9}$$

Die Dichte und der Massestrom der Luft entsprechen den in Tabelle 5.1 aufgeführten Werten. Zusätzlich wurden die Enthalpien der Luft bei Eintritt und Austritt,  $h_1$  und  $h_2$ , aus

einem Mollier-h-x-Diagramm<sup>1</sup> abgelesen. Die entsprechenden relativen Luftfeuchtigkeiten betragen  $\phi_1 = 45,0\%$  und  $\phi_2 = 96,9\%$ .



(c) Single-pass cross-flow with both fluids *unmixed*

Abbildung 5.4: Diagramm zum Ablesen des Korrekturfaktors  $f$  für einfache Kreuzwärmeübertrager ohne Vermischung der beiden Medien, Grafik aus [21]. Die in der Grafik dargestellten Temperaturen entsprechen:  $t_1 = t_{water_c}$ ,  $t_2 = t_{water_w}$ ,  $T_1 = t_{air_w}$ ,  $T_2 = t_{air_c}$

Der in Gleichung 5.6 benötigte Korrekturfaktor  $f$  kann mittels des Diagramms in Abbildung 5.4 bestimmt werden. Dazu müssen folgende Werte errechnet werden:

$$P = \frac{t_{water_w} - t_{water_c}}{t_{air_w} - t_{water_c}} = \frac{12,0^\circ\text{C} - 6,0^\circ\text{C}}{25,0^\circ\text{C} - 6,0^\circ\text{C}} = 0,316 \quad (5.10)$$

$$R = \frac{t_{air_w} - t_{air_c}}{t_{water_w} - t_{water_c}} = \frac{25,0^\circ\text{C} - 10,8^\circ\text{C}}{12,0^\circ\text{C} - 6,0^\circ\text{C}} = 2,37 \quad (5.11)$$

Durch Ablesen der Werte aus dem Diagramm 5.4 ergibt sich ein Korrekturfaktor von

$$f = 0,88 \quad (5.12)$$

Die mittlere logarithmische Temperaturdifferenz kann nach Gleichung 5.7 bestimmt werden:

$$\overline{\Delta T} = \frac{(25,0^\circ\text{C} - 12,0^\circ\text{C}) - (10,8^\circ\text{C} - 6,0^\circ\text{C})}{\ln\left(\frac{13,0^\circ\text{C}}{4,8^\circ\text{C}}\right)} = 8,23\text{K} \quad (5.13)$$

Durch Umstellen der Gleichung 5.6 kann nun die Charakterisierung des UKG ermittelt werden:

$$UA = \frac{\dot{Q}_{crah}}{\overline{\Delta T} \cdot f} = \frac{171,74\text{kW}}{8,23\text{K} \cdot 0,88} = 23,7 \frac{\text{kW}}{\text{K}} \quad (5.14)$$

<sup>1</sup>Mollier-h-x-Diagramm für atmosphärischen Luftdruck, zu finden in physikalischen Standardwerken wie [137, 138]

Da die Ermittlung von  $\dot{Q}_{crah}$  durch die beschriebene Gleichung 5.6 aber noch keine Verluste betrachtet, muss eine Berechnung der tatsächlichen Effektivität des Wärmeübertragers vorgenommen werden. Diese kann dann dazu verwendet werden, die reale Kühlleistung zu berechnen. Die Effektivität eines Kreuzwärmeübertragers kann mit Hilfe der NTU-Methode (aus dem Englischen: **N**umber of **T**ransfer **U**nits) erfolgen.

### Effektivität und Energiefluss mit der NTU-Methode

Die NTU-Methode bietet die Möglichkeit, die Effektivität und den Energiefluss über die beiden Medien in einem Wärmeübertrager zu bestimmen und ist dazu geeignet, die Effektivität eines Umluftkühlgeräts zu bestimmen [18, 32, 103]. Zunächst müssen die Wärmekapazitätsströme der beiden Medien Luft ( $C_{air}$ ) und Wasser-Glykolgemisch ( $C_{water}$ ) berechnet werden:

$$C_{air} = \dot{m}_{air} \cdot c_{p_{air}} = 8,5 \frac{kg}{s} \cdot 1,005 \frac{kJ}{kg \cdot K} = 8,5425 \frac{kW}{K} \quad (5.15)$$

$$C_{water} = \dot{m}_{water} \cdot c_{p_{water}} = 6,3 \frac{kg}{s} \cdot 3,7 \frac{kJ}{kg \cdot K} = 23,31 \frac{kW}{K} \quad (5.16)$$

Das Wärmestromverhältnis ist dann:

$$C_r = \frac{C_{air}}{C_{water}} = \frac{8,5425 \frac{kW}{K}}{23,31 \frac{kW}{K}} = 0,37 \quad (5.17)$$

Die einheitenlose NTU-Kennzahl gibt nun die Anzahl der Übertragungseinheiten an. Mit Hilfe dieser Kennzahl kann die Menge der Wärmeübertragung, also der Energiefluss, in Wärmeübertragern berechnet werden. Dazu wird diese wie folgt berechnet:

$$NTU = \frac{U \cdot A}{C_{air}} = \frac{23,7 \frac{kW}{K}}{8,5425 \frac{kW}{K}} = 2,77 \quad (5.18)$$

Die Effektivität  $\varepsilon_{NTU}$  des Umluftkühlgeräts wird mittels der NTU-Kennzahl und des Wärmestromverhältnisses durch folgende Gleichung (spezifisch für Kreuzwärmeübertrager) berechnet:

$$\varepsilon_{NTU} = 1 - \exp \left[ \frac{\exp(-NTU^{0,78} \cdot C_r) - 1}{NTU^{-0,22} \cdot C_r} \right] = 1 - \exp \left[ \frac{\exp(-2,77^{0,78} \cdot 0,37) - 1}{2,77^{-0,22} \cdot 0,37} \right] = 0,85 \quad (5.19)$$

Als Kontrolle können die Werte der idealen Berechnungen aus Gleichung 5.9 mit denen des Datenblatts und der Messungen verglichen werden. Wird die theoretisch mögliche Kälteleistung mit der ermittelten Effektivität  $\varepsilon_{NTU}$  kombiniert, so ergibt sich eine berechnete reelle Leistung von 145 kW. Im Datenblatt sind hier 142,3 kW angegeben, was lediglich einer Abweichung von ca. 2% entspricht.

Die Effektivität  $\varepsilon_{NTU}$  wird nun dazu verwendet, den tatsächlichen Wärmestrom zu bestimmen. Dazu wird folgende Gleichung verwendet, die eine Berechnung anhand des Temperaturunterschiedes zwischen der warmen Luft und dem kalten Wasser vornimmt (in diesem

Fall muss  $C_{air}$  verwendet werden, weil dies der kleinere Wert ist und somit den Wärmestrom beschränkt):

$$\dot{Q}_{crah} = \varepsilon_{NTU} \cdot C_{air} \cdot (t_{air_w} - t_{water_c}) \quad (5.20)$$

Die Modellgleichung für das Umluftkühlgerät, mit der nun die benötigte Kaltwassertemperatur bei gegebener effektiver erforderlicher Kühlleistung und Warmlufttemperatur berechnet werden kann, lautet:

$$t_{water_c} = \frac{\varepsilon_{NTU} \cdot C_{air} \cdot t_{air_w} - \dot{Q}_{crah}}{\varepsilon_{NTU} \cdot C_{air}} \quad (5.21)$$

Analog dazu kann nach thermodynamischen Grundsätzen auch die Warmwassertemperatur bestimmt werden:

$$t_{water_w} = \frac{\dot{Q}_{crah} + C_{water} \cdot t_{water_c}}{C_{water}} \quad (5.22)$$

Ebenfalls kann die Kaltlufttemperatur bestimmt werden, mit der die Luft das Umluftkühlgerät verlässt:

$$t_{air_c} = \frac{C_{air} \cdot t_{air_w} - \dot{Q}_{crah}}{C_{air}} \quad (5.23)$$

Ein beispielhafter Verlauf dieses Modells ist in Abbildung 5.5 dargestellt. Die unveränderte Temperatur der warmen Luft ist nicht eingezeichnet, diese beträgt weiterhin  $25,0^\circ\text{C}$ . Das Diagramm zeigt die notwendigen Temperaturen der eingesetzten Medien Luft und Wasser in Abhängigkeit der benötigten Kälteleistung  $\dot{Q}_{crah}$ . Die Kaltwassertemperatur  $t_{water_c}$  (blaue Linie) zeigt den durch das Modell ermittelten Wert, der direkt von der benötigten Kälteleistung abhängt. Wird also beispielsweise eine Kälteleistung von  $\dot{Q}_{crah} = 110\text{kW}$  benötigt, so kann das untersuchte Umluftkühlgerät diese Leistung mit einer Kaltwassertemperatur von ca.  $10^\circ\text{C}$  und den genannten weiteren Parametern aus Tabelle 5.1 bereitstellen. Es ergibt sich dann eine Kaltlufttemperatur von  $t_{air_c} \approx 12^\circ\text{C}$ . Das erwärmte Kühlwasser tritt dann aus dem UKG mit einer Temperatur von  $t_{water_w} \approx 14^\circ\text{C}$  wieder aus.

## 5.5 Kältemaschine

Die Kältemaschinen sorgen für die Bereitstellung des Kaltwassers. Wie bereits in der Analyse in Kapitel 4.1.1 beschrieben wurde, hängt der Energiebedarf dieser Geräte primär von der Kaltwassertemperatur ab. Eine Modellierung dieses Zusammenhangs wird über den Coefficient of Performance (COP) [130, 103] abgebildet. Dieser Wert gibt das Verhältnis der benötigten Kälteleistung  $\dot{Q}_{total}$  zu der dazu benötigten elektrischen Leistung  $P_{chiller}$  an.

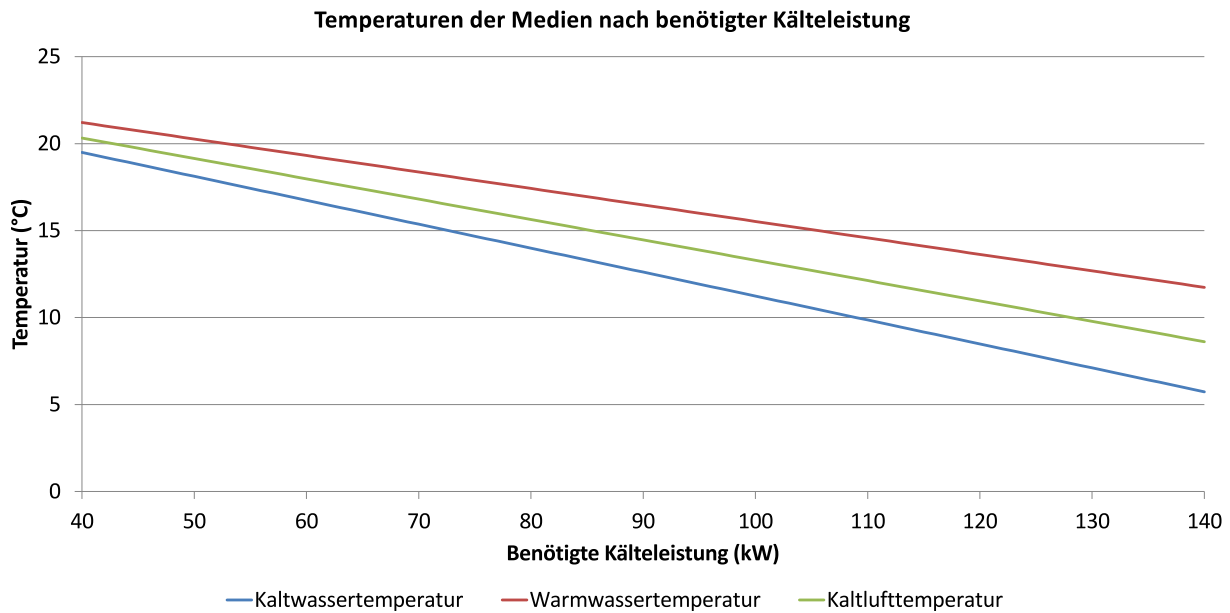


Abbildung 5.5: *Temperaturen des Kühlwassers und der Luft bei entsprechender benötigter Kälteleistung*

Anders ausgedrückt kann mit Hilfe des COP die elektrische Leistungsaufnahme abgeschätzt werden, die für die Kühlleistung  $\dot{Q}_{total}$  anfällt:

$$P_{chiller} = \frac{\dot{Q}_{total}}{COP_{chiller}(t_{water_c})} \quad (5.24)$$

Dabei hängt die Charakterisierung des COP von der Kaltwassertemperatur  $t_{water_c}$  ab. Für die Charakterisierung des COP muss auf Werte aus der Literatur zurückgegriffen werden. In [87] wurde die Vermessung und Charakterisierung einer Kältemaschine vorgenommen:

$$COP_{chiller}(t_{water_c}) = 0.0068 \cdot t_{water_c}^2 + 0.0008 \cdot t_{water_c} + 0.458 \quad (5.25)$$

Diese Charakterisierung wurde in mehreren wissenschaftlichen Arbeiten aufgegriffen [94, 95, 128, 130]. in Abbildung 5.6 ist die genannte Charakterisierung des COP in Abhängigkeit der Kaltwassertemperatur  $t_{water_c}$  dargestellt. Je höher der COP ausfällt, desto weniger Energie wird für die Kältemaschinen benötigt. Bei einer Steigerung der Kaltwassertemperatur von  $t_{water_c} = 9^\circ\text{C}$  auf  $t_{water_c} = 15^\circ\text{C}$  wird zum Beispiel der Energiebedarf der Kältemaschinen halbiert.

## 5.6 Auslastungs- und VM-Migrationsmodelle

In diesem Unterkapitel wird zunächst die Vergleichbarkeit von CPU-Auslastungsangaben auf verschiedenen physikalischen Servern untersucht. Da in dieser Arbeit Rechenzentren



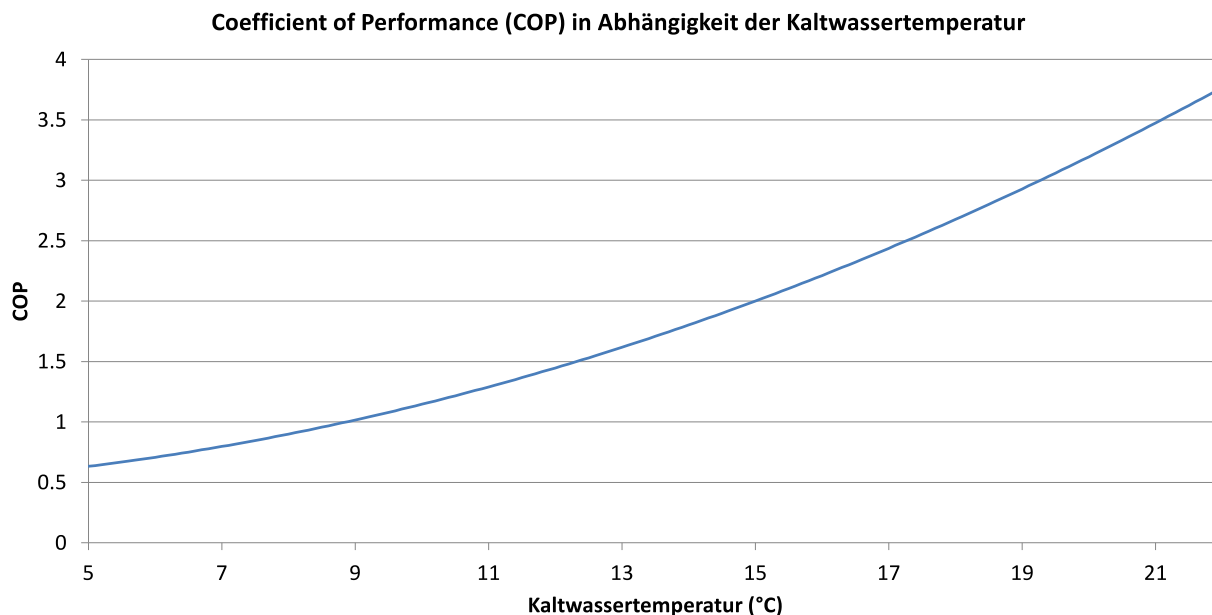


Abbildung 5.6: *Coefficient of Performance einer Kältemaschine nach der Charakterisierung aus [87]*

mit heterogenen Servern betrachtet werden, muss eine Methode für diese Vergleichbarkeit gefunden werden, da ansonsten keine zuverlässige Allokationsplanung der virtuellen Maschinen durchgeführt werden kann.

Anschließend wird die Dauer von Migrationen in Abhängigkeit des jeweiligen Ressourcenbedarfs von virtuellen Maschinen untersucht. Dies ist vor allem für die Berechnung der Evakuierungszeit, die in Kapitel 4.4.3 beschrieben wurde, notwendig.

### 5.6.1 Auslastung und Vergleichbarkeit

Die CPU-Auslastung einer virtuellen Maschine wird in dieser Arbeit absolut in Megahertz angegeben, wie bereits in Kapitel 2.4 definiert, wodurch eine Betrachtung heterogener Server gleicher Architektur vorgenommen werden kann. Bei einer prozentualen Betrachtung der CPU-Auslastung ist dies nicht möglich. Beispielsweise ist es wahrscheinlich, dass eine virtuelle Maschine mit einer CPU-Auslastung von 5% auf einem Server mit mehreren Prozessoren und großem Ressourcenangebot, auf einem deutlich kleineren Server mehr als diese 5% Auslastung erzeugen wird. Damit aber eine Allokationsplanung vorgenommen werden kann, muss das Optimierungsverfahren eine Methode zur Verfügung haben, mit der die erwartete Auslastung einer virtuellen Maschine auf einem Zielsystem prognostiziert werden kann. Die hier beschriebene Methode kann allerdings nur bei Servern mit der gleichen Architektur angewandt werden.

Zur Untersuchung der Vergleichbarkeit der CPU-Auslastung auf verschiedenen Servern wurden Experimente durchgeführt, bei denen die gleiche Anwendungslast auf eben diesen Servern ausgeführt wurde. Gleichzeitig wurde die auftretende CPU-Auslastung gemessen

und protokolliert. Damit dieses Experiment durchgeführt werden konnte, musste zunächst eine möglichst repräsentative Anwendungslast definiert werden, die zudem in ihrer Quantität steuerbar ist. Dazu wurde der Webserver *Apache HTTP* [6] gewählt, auf dem eine definierte Seite aufgerufen wird. Die Seitenaufrufe wurden dabei mittels des Lastgenerators *Apache jmeter* [7] erzeugt, wodurch die Anzahl und der zeitliche Abstand zwischen den Aufrufen geregelt werden konnte. Mittels dieses Szenarios konnte ein definierter Workload, der aus der Anzahl der simulierten Benutzeraufrufe pro definierter Zeit bestand, auf jedem Server ausgeführt werden. Die drei untersuchten Server besitzen folgende Hardware:

- Server AMD Phenom II X6 1055T, 6 Kerne,  $u_{scap} = 16800$  MHz
- Server Intel Core 2 Duo E8400, 2 Kerne,  $u_{scap} = 6000$  MHz
- Server 2 x Intel Xeon L5520, 8 Kerne (beinhaltet virtuelle Kerne),  $u_{scap} = 36100$  MHz

Die maximale Kapazität der CPU-Ressource  $u_{scap}$  in MHz jedes Servers wird dabei wie folgt berechnet:

$$u_{scap} = n_{core} \cdot u_{core} \cdot n_{cpu} \quad (5.26)$$

Hier ist  $n_{core}$  die Anzahl der nutzbaren Kerne eines Prozessors,  $u_{core}$  die maximale Taktfrequenz eines einzelnen Kerns in MHz und  $n_{cpu}$  die Anzahl der physikalisch vorhandenen Prozessoren. Diese Berechnung der CPU-Ressourcen eines Servers wird auch in [49, 132] vorgenommen, allerdings fehlt bei den Autoren das Vorgehen zur Vergleichbarkeit der Auslastung auf verschiedenen Servern.

Die erste Auswertung der Untersuchung zeigt in Abbildung 5.7 die prozentuale CPU-Auslastung in Abhängigkeit zu der Anzahl simulierter Benutzeranfragen. Prozentual fällt die Auslastung erwartungsgemäß für den jeweiligen Workload auf den verschiedenen Servern unterschiedlich aus. Beispielsweise erzeugt ein Workload mit Zugriffen von 600 Benutzern auf dem 6-Kern-Server ca. 19% CPU-Auslastung, während der kleinste Server bereits mit ca. 50% ausgelastet ist. Wird allerdings aus der prozentualen Auslastung eine absolute Auslastung in Megahertz errechnet, wie sie als Angabe in dieser Arbeit verwendet wird, so kann in Abbildung 5.8 erkannt werden, dass hier alle Server sehr eng beieinander liegen, also der gleiche Workload auf den Servern direkt in absoluten Zahlen vergleichbar ist.

Durch diese Messungen wird belegt, dass eine Umrechnung der prozentualen Auslastung in eine absolute zu korrekten Ergebnissen führt, wenn die Server die gleiche Architektur besitzen, und damit in den hier genutzten Modellen als Angabe verwendet werden kann.

Diese Betrachtung unterliegt aber auch spezifischen Einschränkungen. Grundsätzlich ist die gezeigte Berechnung genau und kann für typische Workloads angenommen werden. Signifikante Abweichungen können aber dann auftreten, wenn Anwendungen nicht parallelisiert sind und nur einen Kern beanspruchen können. Allerdings wird durch die Virtualisierung diese Einschränkung teilweise wieder aufgehoben, da hier jede VM entweder einem physikalischen Kern zugeordnet wird oder einen vollständig virtuellen Prozessor erhält, der auch im Hinblick auf die maximale CPU-Auslastung konfiguriert werden kann. Ein weiterer Fall, der Abweichungen erzeugen kann, ist eine stark optimierte Anwendung auf spezifische Systeme, die auf einem „normalen“ Server diese Optimierungen nicht ausnutzen kann.

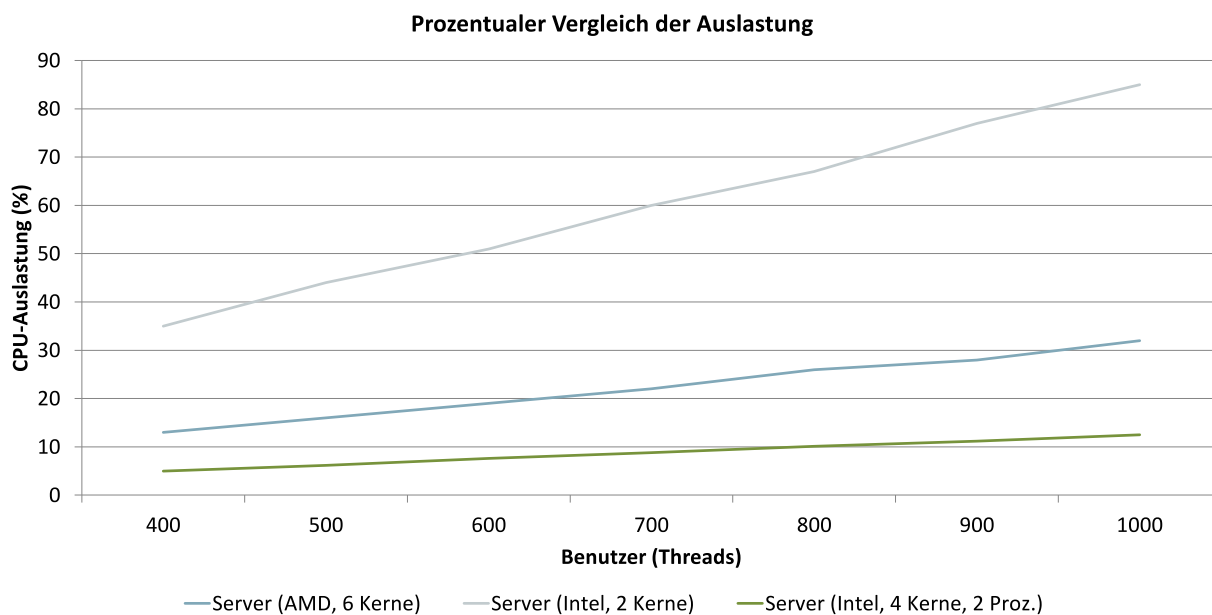


Abbildung 5.7: *Prozentualer Vergleich der Auslastung des gleichen Workloads auf verschiedenen Servern*

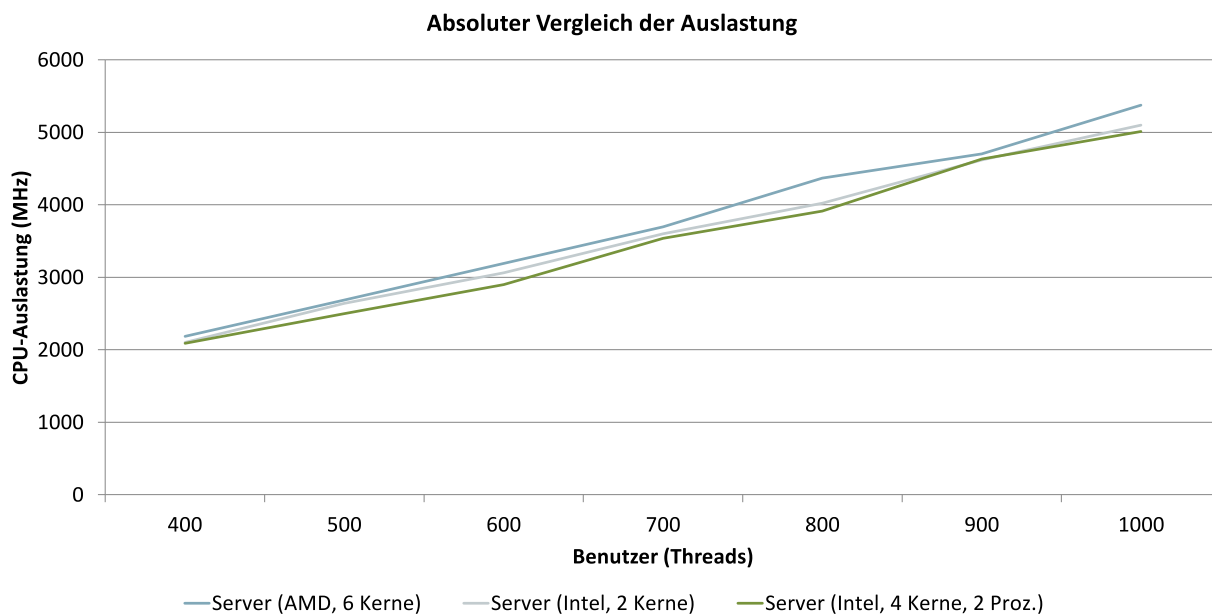


Abbildung 5.8: *Absoluter Vergleich der Auslastung des gleichen Workloads auf verschiedenen Servern*

Derartige Spezialanwendungen müssen gesondert betrachtet werden, allerdings ist es fraglich, ob diese Anwendungen überhaupt virtualisiert werden können. Daher erfolgt keine Betrachtung dieser Spezialfälle in dieser Arbeit.

## Overhead des Hypervisors

Das beschriebene Auslastungsmodell betrachtet nicht den Overhead [144]<sup>2</sup>, der durch den Einsatz des Hypervisors entsteht, also den Grad der Verringerung der Anwendungsperformance bei Virtualisierung. Dieser hängt signifikant von der Anwendung und deren Ressourcenprofil sowie vom eingesetzten Hypervisor [112, 136] ab. Untersuchungen haben allerdings gezeigt, dass moderne Hypervisoren bei typischen Geschäftsanwendungen keinen signifikanten Overhead mehr erzeugen [112]. Daher wird dieser in der Modellbetrachtung ignoriert.

Neben dem Overhead können auch Performanzeinbußen bei einzelnen virtuellen Maschinen entstehen, wenn diese mit weiteren virtuellen Maschinen mit spezifischen, konkurrierenden Ressourcenprofilen auf einem physikalischen Server ausgeführt werden [65]. Diese Untersuchungen sind ein eigenes, komplexes Themenfeld, für das außerdem invasive Messmethoden in den virtuellen Maschinen benötigt werden, was in dieser Arbeit vermieden werden soll. Außerdem treten diese Einbußen nur in speziellen Fällen auf, weshalb auch diese Effekte in dieser Arbeit nicht betrachtet werden.

### 5.6.2 VM-Migrationsmodelle

Für das in dieser Arbeit entwickelte Verfahren ist die Dauer von Live-Migrationen von VMs von Bedeutung, da diese die Evakuierungszeit eines Servers bestimmt. Bei Live-Migrationen wird grundsätzlich davon ausgegangen, dass sich die Daten einer VM auf einem Storage-Server befinden und damit nicht übertragen werden müssen. Nur der aktuelle Inhalt des Arbeitsspeichers muss auf den Zielservers übertragen werden, bevor die VM auf diesem ausgeführt werden kann.

Es existieren mehrere Forschungsarbeiten, die die Dauer sowie die zusätzliche Ressourcenbelastung und damit auch die Energiekosten von Live-Migrationen untersucht haben [55, 72, 77, 127]. Die Forschung ist sich grundsätzlich darüber einig, dass die Dauer einer Live-Migration primär von der Arbeitsspeichergröße und Auslastung der VM und des physikalischen Servers sowie der entsprechenden Netzwerkanbindung abhängt. Insbesondere die Page-Dirty-Rate einer VM ist hier von Bedeutung, da sich mit einer steigenden Rate aufgrund der häufigen Änderungen die Größe der zu übertragenden Daten aus dem RAM-Speicher vergrößert, was wiederum zu einer längeren Migrationsdauer führt. Mehrere Modelle zur Abschätzung der Migrationsdauer nutzen daher die Page-Dirty-Rate einer VM und die Netzwerkbandbreite der beteiligten Server [25, 125, 127]. Allerdings ist die Messung der Page-Dirty-Rate direkt vom eingesetzten Hypervisor abhängig und häufig nicht ohne zusätzliche Werkzeuge möglich [77], was einer invasiven Messung entspricht, die in dieser Arbeit vermieden werden soll.

Ein weiteres Problem dieser Modelle ist die Annahme, dass grundsätzlich der gesamte Arbeitsspeicher einer VM übertragen wird. Wurde eine VM mit einer festen Hauptspeichergröße erzeugt, so wird aber nicht immer diese Datenmenge übertragen. In eigenen Experi-

---

<sup>2</sup>Auch wenn [144] primär den Overhead bei Cloud-Lösungen (also rechenzentrenübergreifend) untersucht, findet sich hier auch eine Übersicht über technische Gründe von Overheads bei lokaler Virtualisierung

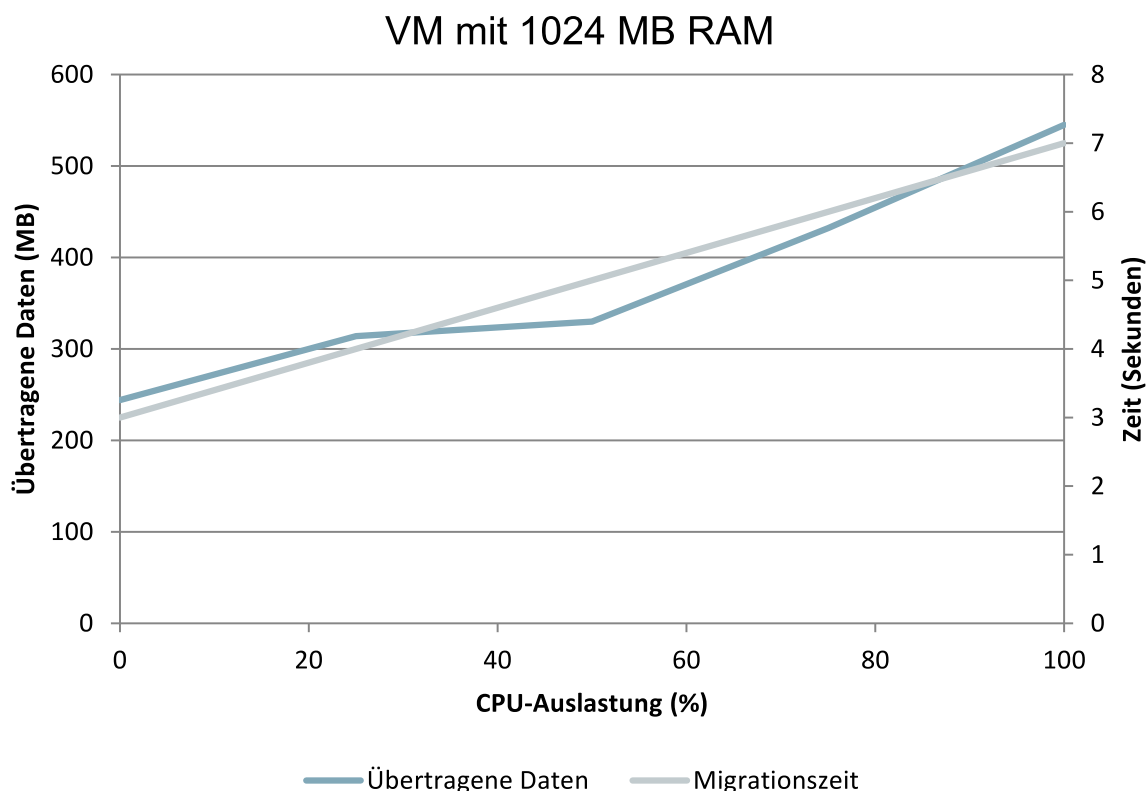


Abbildung 5.9: VM-Migrationszeiten nach CPU-Auslastung

menten wurde unter Nutzung des Hypervisors KVM gezeigt, dass nur der tatsächlich von der Anwendung verwendete und teils erheblich kleinere Teil des Speichers übertragen wird, wie in Abbildung 5.9 zu erkennen ist. Dazu wurde die in Kapitel 5.6.1 beschriebene Anwendungslast eines Webservers verwendet, um die zu migrierende VM in unterschiedliche Auslastungszustände zu versetzen. Diese Auslastung betrifft dabei sowohl die CPU, als auch den Arbeitsspeicher und die Festplatte. Die Messungen zeigen eine annähernd lineare Abhängigkeit der zu übertragenden Datenmenge von der CPU-Auslastung. Dementsprechend steigt auch die benötigte Zeit der Migration. Zu beachten ist, dass die angegebenen Zeiten der vollständigen Migrationszeit entsprechen. Der Zeitraum, in dem die VM nicht ansprechbar ist, ist deutlich kürzer und liegt nach Literaturangaben bei maximal einigen hundert Millisekunden [72, 101]. Im Experiment war dieser Zeitraum so kurz, dass er durch Messungen nicht festgestellt werden konnte.

Die in Abbildung 5.9 gezeigten Beobachtungen gelten allerdings nur für die hier verwendete Anwendungslast, Anwendungen mit anderen Ressourcenprofilen können sich von diesen Beobachtungen unterscheiden. Um eine genauere Modellierung unabhängig von der Anwendung erreichen zu können, müsste ein VM-internes Messverfahren verwendet werden, das mindestens die aktuelle RAM-Belegung und Page-Dirty-Rate protokolliert. Ein solches Verfahren ist in dieser Arbeit nicht vorgesehen, da die VMs als geschlossene Einheiten betrachtet werden sollen, auf deren Inhalt kein Zugriff vorhanden ist. Lediglich Messungen

durch den Hypervisor sind vorgesehen. Daher wird ein Modell auf Basis der vorgestellten Messungen verwendet, das die benötigte Migrationszeit  $\psi_{mig_v}$  der virtuellen Maschine  $v$  aufgrund der CPU-Auslastung von  $v$  und der maximalen RAM-Größe der virtuellen Maschine  $v$  errechnet. Dafür wird folgende Gleichung verwendet:

$$\psi_{mig_v} = \frac{o_{v_{alloc}} \cdot (1 + p_v)}{b_{mbit} \cdot \frac{1}{8}} \quad (5.27)$$

In der Gleichung gibt  $p_v | 0 \leq p_v \leq 1$  die relative CPU-Auslastung der virtuellen Maschine  $v$  an.  $o_{v_{alloc}}$  gibt den reservierten Arbeitsspeicher der VM  $v$  in Megabyte an.  $b_{mbit}$  bezeichnet die Bandbreite der Netzwerkanbindung in Millionen Bits pro Sekunde. Das Modell berücksichtigt damit auch eine Datenübertragungsmenge, die größer als  $o_{v_{alloc}}$  ist. Dieser Fall tritt bei einer hohen Auslastung der virtuellen Maschine auf.

Das beschriebene Modell gibt eine heuristische, pessimistische Schätzung ab, die eine Obergrenze der Migrationszeit zeigt. Für den Anwendungsfall in dieser Arbeit ist dieses Modell aber genau einsetzbar, da die Schätzung für die im Fall einer Evakuierung benötigte Migrationsdauer verwendet wird. Diese Schätzung darf nicht optimistisch ausfallen, da in einem solchen Fall bei Evakuierungen Ausfälle aufgrund einer unterschätzten Migrationszeit auftreten könnten.

Wenn mehrere Migrationen an einem Server gleichzeitig durchgeführt werden, steigt die Migrationszeit mitunter signifikant an [25], was durch Experimente bestätigt werden konnte. Dabei wurde auch eine erhöhte Abbruchrate der Migrationen beobachtet, was die Stabilität eines Zustandswechsels gefährdet. Daher wird hier auf mehrere parallele Migrationen zu oder von dem gleichen Server verzichtet, pro Server darf immer nur eine Migration gleichzeitig durchgeführt werden.

### **Zusätzliche elektrische Leistungsaufnahme während Migrationen**

Es wurde auch die zusätzliche elektrische Leistungsaufnahme der Server während Live-Migrationen untersucht. In Messungen sowohl auf dem sendenden als auch auf dem empfangenden Server konnte nur eine minimale Änderung der Leistungsaufnahme festgestellt werden, diese lag unter 4%. Beispielsweise konnte durch die Messungen festgestellt werden, dass unter der pessimistischen Annahme einer durchschnittlichen Migrationszeit von 10 Sekunden eine Energiebedarfserhöhung um eine Kilowattstunde durch ungefähr 60.000 Migrationen erreicht werden würde. Zum Vergleich weist das in der Simulation untersuchte Rechenzentrum einen täglichen Bedarf von mehreren tausend Kilowattstunden auf. Der Einfluss dieser Größe auf das Gesamtergebnis ist dementsprechend gering, sodass dieser Einfluss im Optimierungsprozess und in der Modellierung nicht berücksichtigt wird.

## **5.7 Netzwerkmodellierung**

Die energetische Modellierung des Netzwerks in einem Rechenzentrum erfolgt hier über die elektrische Leistungsaufnahme der Switches. Diese Modellierung wird im folgenden

Unterkapitel vorgenommen. Die in Kapitel 2.1.5 beschriebene Netzwerkarchitektur definiert dazu die Anzahl und den Typ der Switches je nach Netzwerkebene.

Da Migrationen in heutigen Rechenzentren in einem eigenen Wartungsnetzwerk durchgeführt werden, das vollständig vom Datenverkehrsnetz getrennt ist [5], muss bei der Betrachtung von Migrationen nicht die bereits anstehende Netzwerklast der Anwendung betrachtet werden. Daher wird in dieser Arbeit die Berücksichtigung des Netzwerks auf die elektrische Leistungsaufnahme der Switches beschränkt.

### Modellierung von Switches

In dieser Arbeit werden Switches als nicht lastensitive Geräte betrachtet, da die Leistungsaufnahme bei unterschiedlicher Netzwerkauslastung keinen signifikanten Schwankungen unterliegt [2, 75, 76]. Derartige Schwankungen können nur erreicht werden, wenn Switches in Schlafzustände versetzt werden [47, 91], dies wird in dieser Arbeit aber nicht angewandt, da hierzu ein komplexes Netzwerkmanagement durchgeführt werden muss und die Möglichkeiten dieser Schlafzustände deutlich von der eingesetzten Hardware abhängen. Allerdings kann die Leistungsaufnahme von Switches anhand der Anzahl aktiver Ports variieren. Dazu wurde eine Vermessung der Leistungsaufnahme von Switches sowie eine Modellierung in [75] vorgenommen. Dieses Modell hat sich in weiteren Untersuchungen als praktikabel und repräsentativ herausgestellt [31, 89] und wird auch hier angewandt. Dabei wird zwischen Switches auf Rack-Ebene und den Switches auf höheren Ebenen (Aggregations- und Hauptebene, siehe hierzu auch die Netzwerktopologie in Kapitel 2.1.5) unterschieden. Switches in Racks besitzen im Allgemeinen eine feste Anzahl Ports, während Switches auf höheren Ebenen flexibel aufgebaut sind und durch sogenannte Linecards erweitert werden können, auf denen dann die entsprechenden Ports verbaut sind. Für beide Typen kann grundsätzlich das gleiche Modell verwendet werden, bei Switches ohne Linecards entfällt dann der entsprechende Teil der Modellformel. Die elektrische Leistungsaufnahme eines Switches kann demnach wie folgt ermittelt werden:

$$P_{switch} = P_{chassis} + n_{linecards} \cdot P_{linecard} + n_{ports} \cdot P_{port} \quad (5.28)$$

Dabei steht  $P_{chassis}$  für die grundsätzlich auftretende Leistungsaufnahme,  $n_{linecards}$  bezeichnet die Anzahl der verbauten Linecards und  $P_{linecard}$  die zusätzliche Leistungsaufnahme einer einzelnen Linecard. Die Anzahl der Ports  $n_{ports}$  wird für den gesamten Switch angegeben und  $P_{port}$  gibt die Leistungsaufnahme eines einzelnen Ports an.

### Charakterisierung

An dieser Stelle werden auch die Charakterisierungen aus [75] sowie der Vorarbeit des selben Autors [76] verwendet, da diese für typische Switches erstellt wurden und im Rahmen der Messungen für diese Arbeit keine detaillierteren Daten erstellt werden konnten. Die gewählte Charakterisierung erfolgte jeweils für Switches mit einer Bandbreite von 1Gbps pro Port.

Charakterisierung für Switches mit Linecard (ab Aggregationsebene):

$$P_{switch} = 54W + n_{linecards} \cdot 39W + n_{ports} \cdot 0.9W \quad (5.29)$$

Charakterisierung für Switches ohne Linecard (Rack-Ebene):

$$P_{switch} = 146W + n_{ports} \cdot 0.87W \quad (5.30)$$

Die Leistungsaufnahme der Switches fließt dabei als konstanter Teil  $P_{c_r}$  in die Leistungsaufnahme des IT-Raums  $r$  nach Gleichung 4.8 ein. Damit ergeben sich nur dann energetische Ersparnisse, wenn ein vollständiger IT-Raum deaktiviert werden kann und damit auch die Switches ausgeschaltet werden können.

## 5.8 Zusammenspiel der Modelle

Die einzelnen vorgestellten Modelle werden für Einzelbewertungen verwendet, beispielsweise wenn die Auswirkungen einer möglichen Allokation auf einen einzelnen Server festgestellt werden sollen. Die Modelle werden aber auch für eine Gesamtsimulation des Rechenzentrums unter einer gefundenen Allokation und einer Abschätzung des damit verbundenen Energiebedarfs verwendet. In Abbildung 5.10 wird dazu die Zusammenwirkung der einzelnen energetischen Modelle dargestellt. Die Abbildung enthält nur die Modelle, die für eine energetische Abschätzung des Rechenzentrums verwendet werden. Einige der in diesem Kapitel beschriebenen Modelle kommen nur in speziellen Fällen zum Einsatz, beispielsweise zur Kontrolle der Einhaltung von Grenzwerten oder im Falle von Migrationen. So wird beispielsweise das Server-Temperaturmodell während der Suche nach einer Allokation eingesetzt, um die geforderte Temperaturhöchstgrenze einzuhalten.

Für die energetische Abschätzung der Leistungsaufnahme des gesamten Rechenzentrums und einzelner Teilbereiche unter einer gegebenen Allokation werden die gezeigten Modelle verwendet, wobei der Fokus hier auf den dynamisch agierenden Geräten liegt. In jedem IT-Raum befinden sich Server, die durch die jeweilige Anzahl und Auslastung der virtuellen Maschinen einen schwankenden Energiebedarf besitzen. Zusätzliche Geräte wie Switches werden statisch betrachtet. Die Umluftkühlgeräte im IT-Raum benötigen Energie für die Ventilatoren. Dies zählt zusammen zum Energiebedarf des IT-Raumes, wobei nur die IT-Geräte an den USV-Geräten hängen und somit auch deren Energiebedarf und die entsprechende Verlustleistung beeinflussen. Auch die Geräte im USV-Raum müssen gekühlt werden und beeinflussen damit die benötigte Kälteleistung der Umluftkühlgeräte. Hier wird der in Kapitel 4.1.1 in Gleichung 4.3 beschriebene Zusammenhang der Bestimmung der Kaltwassertemperatur aufgrund des Leistungsmaximums der einzelnen Räume angewandt. Zusammen mit der insgesamt benötigten Kälteleistung wird damit die elektrische Leistungsaufnahme der Kältemaschinen bestimmt.

In der Abbildung sind beispielhaft drei IT-Räume und ein USV-Raum gezeigt. In der späteren Evaluation wird ein Rechenzentrum mit mehr IT-Räumen untersucht, die Abbildung zeigt hier lediglich, dass die Modelle für jeden IT-Raum gleich aufgebaut sind



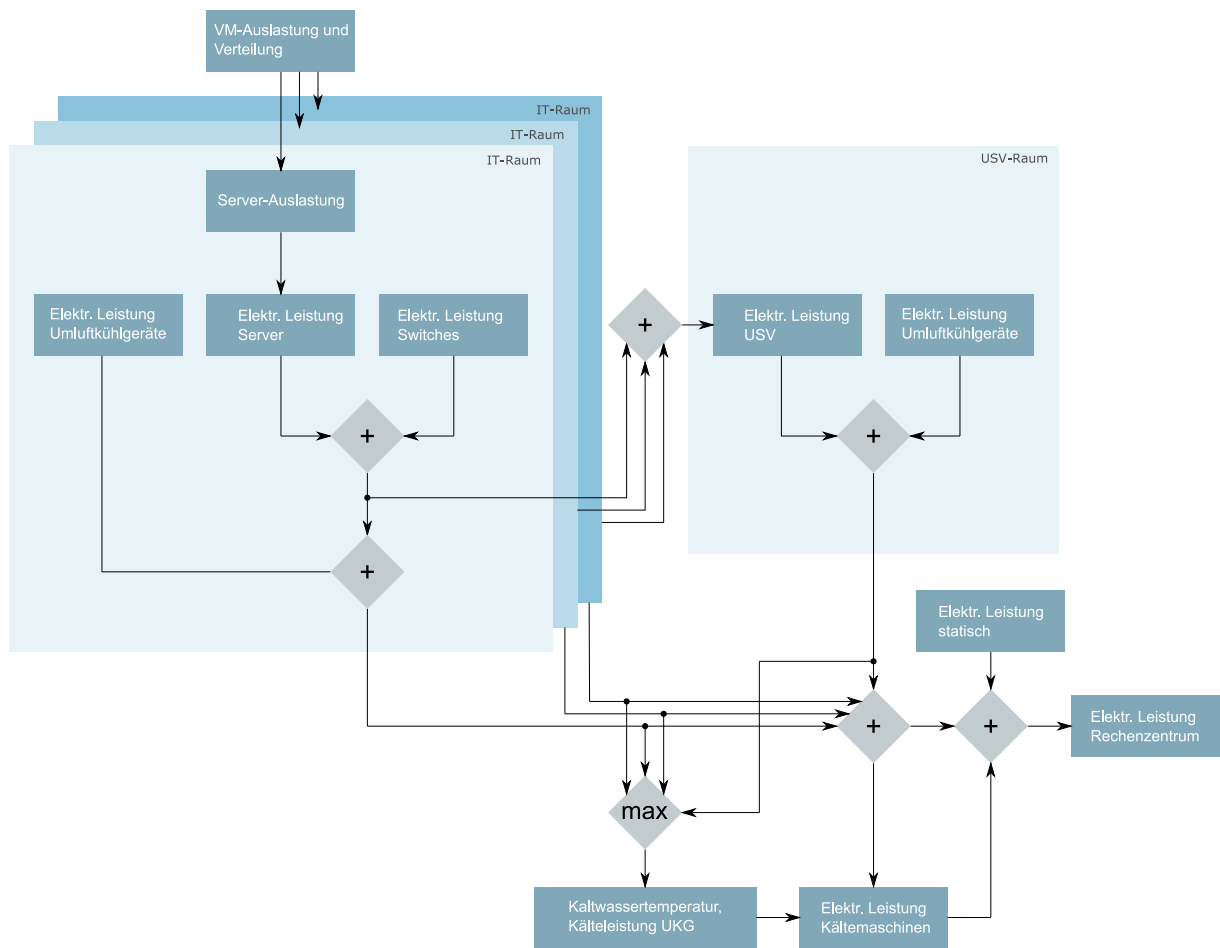


Abbildung 5.10: *Zusammenspiel der energetischen Modelle: Für die Rechenzentrumssimulation werden die gezeigten Modelle wie dargestellt kombiniert. In der Abbildung sind nur die Modelle enthalten, die für die energetische Abschätzung verwendet werden. Damit fehlen einige der beschriebenen Modelle aus diesem Kapitel, da diese primär für die Einhaltung von Grenzwerten bestimmt sind oder beispielsweise nur im Falle von möglichen Migrationen Anwendung finden.*

und die Ergebnisse aus den einzelnen IT-Räumen aufaddiert werden. Auch ist ein einzelner USV-Raum dargestellt, was für das in dieser Arbeit untersuchte Rechenzentrum tatsächlich zutrifft. Allerdings unterstützen die Modelle und deren Architektur auch hier mehrere Räume, wobei dann von einer gleichmäßigen Verteilung der Leistung auf alle USV-Geräte ausgegangen wird.

Damit sind die dynamischen Bestandteile des Energiebedarfs und die Bedarfe der IT-Räume des Rechenzentrums abgedeckt und es wird nun noch der in den Grundlagen beschriebene statische Bedarf hinzu addiert, wodurch sich der gesamte Energiebedarf des Rechenzentrums ergibt (dieser statische Bedarf entspricht dem in Abbildung 2.2 darge-

stellten Bedarf der sonstigen Infrastruktur).

## 5.9 Zusammenfassung

Dieses Kapitel hat die Modellierungen der beteiligten Geräte im Rechenzentrum vorgestellt. Der erste Teil des Kapitels beschäftigt sich mit der energetischen Modellierung von USV-Geräten, Servern und Kühlgeräten, während der zweite Teil die Auslastung und das Migrationsverhalten von virtuellen Maschinen thematisierte. Abschließend wurde das Zusammenspiel der Modelle beschrieben, was im weiteren Verlauf der Arbeit auch als Rechenzentrumssimulation bezeichnet wird.

In Kapitel 6 folgt nun die Beschreibung des DaMucA-Verfahrens, das sowohl die Modelle aus diesem Kapitel, als auch die Kriterien und Erkenntnisse aus Kapitel 4 verwendet.

# 6 DaMucA: Data Center Multi-criteria Aware Allocation

Das DaMucA-Verfahren soll den Rechenzentrumsbetrieb anhand der in Kapitel 4 identifizierten Kriterien optimieren. Dazu werden die identifizierten energetischen Freiheitsgrade und Einflüsse sowie die in den Kriterien formulierten Parameter berücksichtigt und zur Suche nach einer optimierten Allokation der virtuellen Maschinen im Rechenzentrum verwendet.

In diesem Kapitel wird zunächst die Architektur des Verfahrens grob dargestellt, dann werden die möglichen Anwendungsfälle und deren Voraussetzungen untersucht. Danach folgt die detaillierte Beschreibung des Verfahrens und eine Laufzeitbetrachtung.

## 6.1 Architektur

Das DaMucA-Verfahren besteht aus mehreren Komponenten, die in diesem Kapitel detailliert vorgestellt werden. In diesem Abschnitt wird kurz die Architektur und das Zusammenwirken dieser Komponenten beschrieben und es wird eine Einordnung des Verfahrens in den Kontext des Rechenzentrumsbetriebs vorgenommen, wobei hier insbesondere die Bedingungen des Einsatzes und die möglichen Anwendungsfälle diskutiert werden. Eine Gesamtübersicht des DaMucA-Verfahrens befindet sich in Abbildung 6.1. Das Verfahren wird in einem existierenden Rechenzentrum angewandt und benötigt Informationen über die vorhandene Infrastruktur, die ausgeführten virtuellen Maschinen und die vorhandenen Server. Diese Daten sollten beispielsweise aus einem Data Center Infrastructure Management Programm (DCIM) extrahiert werden, das mittlerweile in vielen, vor allem größeren Rechenzentren, zur Katalogisierung und Verwaltung aller Geräte angewandt wird. Aus diesen Daten wird zunächst ein modelliertes Rechenzentrum erstellt, da das DaMucA-Verfahren auf dieser Simulation arbeitet und anfallende Optimierungsentscheidungen und deren Auswirkungen zunächst simulativ untersucht.

Das DaMucA-Verfahren selbst kann grob in zwei zentrale Bestandteile zerlegt werden: die DaMucA-Metaheuristik zur Wahl der Kombination der aktiven IT-Räume und die DaMucA-Server-Heuristik, die wiederum mehrere untergeordnete Heuristiken enthält und die Platzierung der virtuellen Maschinen auf die Server vornimmt. Die genaue Begründung für diese Aufteilung und die Beschreibung der Zuständigkeiten der beiden Bereiche findet sich in Abschnitt 6.3.

Die Anwendung des DaMucA-Verfahrens fokussiert auf eine langfristige Optimierung des Rechenzentrumsbetriebs, was in der Abbildung unter „Langzeitplanung“ zusammengefasst

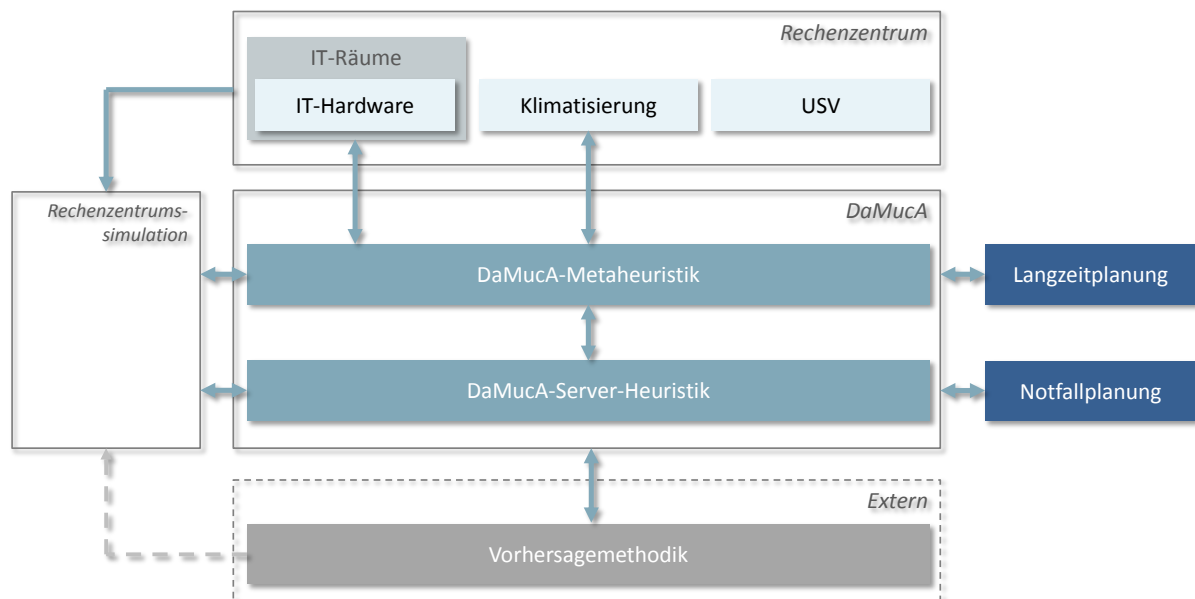


Abbildung 6.1: Architektur des DaMucA-Verfahrens sowie die Einordnung in den Kontext der Rechenzentrumskonsolidierung und Simulation

ist. Diese Langzeitplanung bezieht sich hier auf einen Zeitraum von mehreren Wochen bis wenigen Monaten. Eine genau Einordnung und Begründung dieses Zeitraums erfolgt im Abschnitt 6.2.1. Während dieses Planungszeitraumes können verschiedene Notfälle auftreten, die durch eine Evakuierung der beteiligten Server entschärft werden müssen. Auf diese Anwendungsfälle geht das nachfolgende Unterkapitel genauer ein. Schließlich greift das DaMucA-Verfahren noch auf ein externes Vorhersageverfahren aus [53] zurück, das bereits in Kapitel 3 vorgestellt wurde und zur Vorhersage der Auslastung von virtuellen Maschinen verwendet wird. Auch diese Anwendung wird im folgenden Unterkapitel thematisiert.

## 6.2 Anwendungsfälle

Das DaMucA-Verfahren nutzt gezielt die Freiheitsgrade, die durch Virtualisierung entstehen bzw. ausgenutzt werden können, um damit energetische Einsparungen auch auf Seiten der Kühltaste zu erreichen. Dazu wurde die Analyse der Einflussfaktoren auf die elektrische Leistungsaufnahme der einzelnen Geräte in der Klimakette in Kapitel 4.1.1 vorgenommen. Das Optimierungspotential durch die Anhebung der Kaltwassertemperatur  $t_{water_c}$  wurde dabei insbesondere hervorgehoben, sodass sich hier für das DaMucA-Verfahren eine Anforderung ergibt, dieses Potential auszunutzen und diese Optimierung gesondert zu betrachten. Damit allerdings eine Optimierung hinsichtlich der Kaltwassertemperatur vorgenommen werden kann, müssen zunächst einige Vorbedingungen definiert und beschrieben werden. Eine Anpassung der Kaltwassertemperatur nimmt in der Umset-

zung einige Zeit und Aufwand in Anspruch, sodass der Planungsprozess, der diesen Schritt mit einbezieht, langfristige Zeiträume betrachten muss. Hier hebt sich das DaMucA-Verfahren deutlich von den existierenden Lastmanagement- und Optimierungsverfahren ab, da diese ihre Optimierungsschritte in teils sehr kurzen zeitlichen Abständen vornehmen. Zur besseren Einordnung folgt daher eine kurze Differenzierung zwischen der Kurzzeit- und Langzeitplanung.

### 6.2.1 Kurzzeit- und Langzeitplanung

Die Kurzzeitplanung im Rechenzentrum behandelt primär die Anwendung von Server-Konsolidierung. Es können also ungenutzte Server ausgeschaltet und bei steigender Auslastung wieder aktiviert werden, aber eine Anpassung der Klimatisierung kann nicht stattfinden. Dies liegt an den notwendigen Zeiträumen, die eine Regelung beispielsweise der Kaltwassertemperatur erfordert. Hier dauert eine vollständige Anpassung ein oder mehrere Tage, was vor allem an der Trägheit im System und der Notwendigkeit der manuellen Beobachtung liegt. Denn in der Praxis werden Temperaturanpassungen beispielsweise meist nur schrittweise durchgeführt und das Ergebnis zunächst abgewartet. Ein Ein- oder Ausschalten von Servern benötigt hingegen nur wenige Minuten und ist nach dieser Zeit vollständig abgeschlossen, ein manuelles Eingreifen oder eine gesonderte Überwachung ist praktisch nur im Fehlerfall notwendig. Damit ist eine Kurzzeitplanung mehrmals täglich sinnvoll einsetzbar, sofern die Lastschwankungen der virtuellen Maschinen dies erfordern und eine entsprechende Energieersparnis zu erwarten ist.

Die Anpassung und Regelung der Klimatisierung und insbesondere der Kaltwassertemperatur sollte hingegen für deutlich längere Zeiträume erfolgen. In dieser Arbeit wird als realistischer Planungszeitraum ein Abstand von ein bis drei Monaten zwischen den vollständigen Neuplanungen, die mit dem DaMucA-Verfahren durchgeführt werden, vorgeschlagen. Dieser Zeitraum ergibt sich aus der Abwägung zwischen dem möglichen Einsparpotential und dem Aufwand und der Dauer der Temperaturregelung. Allerdings hängt der Zeitraum auch vom Rechenzentrum und der konkret eingesetzten Kühlinfrastruktur sowie der Art der Auslastung ab.

Zur Einordnung der Kurzzeit- und Langzeitplanung zeigt die Abbildung 6.2 ein Beispiel mit mehreren Planungsschritten. Zu definierten Zeitpunkten oder falls die Auslastung saisonal bedingt deutlichen Schwankungen unterliegt, wird eine vollständige Neuplanung durchgeführt. Diese Planung muss allerdings die energetischen Maxima, die zu erwarten sind, abdecken, da eine kurzfristige Neuplanung der Kombination aktiver IT-Räume und eine Anpassung der Kaltwassertemperatur aus den bereits genannten Gründen nicht einfach möglich sind. Da die Langzeitplanung also mit energetischen Maxima und damit auch mit Auslastungsmaxima kalkulieren muss, können sich zwischen den einzelnen Neuplanungen Einsparpotentiale ergeben, die durch eine zusätzliche Kurzzeitplanung realisiert werden können. Allerdings muss diese auch alle definierten Kriterien zur Stabilität sicherstellen. Die bisherigen Methoden zur kurzfristigen Optimierung des Rechenzentrumsbetriebs erfüllen diese nicht, sodass an dieser Stelle eine Erweiterung der Verfahren erfolgen muss. Dies wird in dieser Arbeit nicht mehr vorgenommen, sondern es erfolgt die Konzentration auf

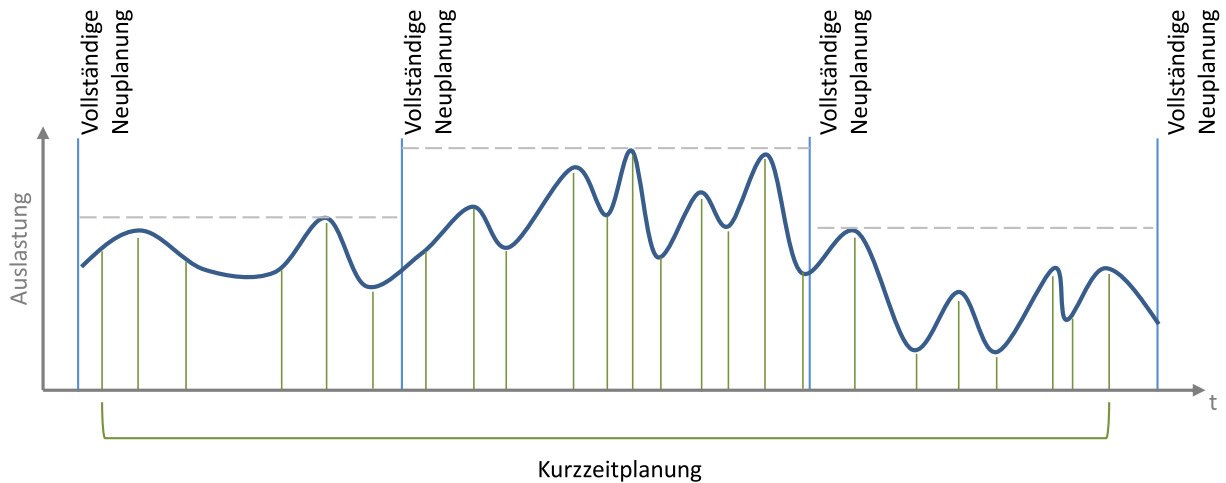


Abbildung 6.2: *Einordnung der Kurzzeit- und Langzeitplanung: Vollständige Neuplanungen beinhalten auch die Änderung der Kombination der aktiven IT-Räume, wodurch die Kühlleistung an die Auslastung angepasst werden kann. Diese Neuplanungen sollten in der Regel in größeren zeitlichen Abständen erfolgen. Dazwischen ist je nach Auslastung und Infrastruktur des Rechenzentrums auch eine Kurzzeitplanung möglich. Das DaMucA-Verfahren behandelt die Langzeitplanung, die jeweils vollständige Neuplanungen vornimmt.*

die Langzeitplanung mit dem DaMucA-Verfahren.

Tabelle 6.1: Einordnung der Kurz- und Langzeitplanung

| Merkmals                      | Kurzzeitplanung   | Langzeitplanung   |
|-------------------------------|---|---|
| <b>Zeitraum</b>               | Mehrfach am Tag   | alle ein bis drei Monate  |
| <b>Verfahren</b>              | Server-Konsolidierung in aktiven IT-Räumen                              | DaMucA-Verfahren: Vollständige Neuplanung: Kombinatorik aktiver IT-Räume zur Anpassung der Kaltwassertemperatur, Server-Konsolidierung in IT-Räumen |
| <b>Aktionen</b>               | Server ein-/ausschalten, VM-Live-Migrationen                            | Kaltwassertemperaturregelung, Server ein-/ausschalten, VM-Live-Migrationen  |
| <b>Auslastungsbetrachtung</b> | reaktive Auslastungsbetrachtung und proaktiv mittels Vorhersagemethodik | Orientierung am Maximum der VM-Auslastungen der letzten Monate bzw. am Maximum der Vorhersage für den Planungszeitraum                              |

In Tabelle 6.1 werden die Unterschiede der beiden Planungsvarianten kurz und übersichtlich

zusammengefasst. Es erfolgt ebenfalls eine Abgrenzung, welche Verfahren für die jeweiligen Planungen verwendet und welche Aktionen zur Regelung realisiert werden können.

Die weiteren Teile dieses Kapitels behandeln ausschließlich die Langzeitplanung mittels des DaMucA-Verfahrens. Dazu wird zunächst beschrieben, wie die Betrachtung der Auslastung der virtuellen Maschinen vorgenommen wird, damit die erwähnten energetischen Maxima berücksichtigt werden können.

### Auslastungsanalyse und Bildung der VM-Profile mittels Vorhersagemethodik

Das DaMucA-Verfahren muss für die Langzeitplanung eine Betrachtung der energetischen Maxima vornehmen, die durch die entstehenden Auslastungen der Anwendungen in den virtuellen Maschinen und deren Maxima erreicht werden können. Wie bei der Beschreibung der Server-Modelle dargestellt (siehe Kapitel 5.3.1), beeinflussen die CPU-Auslastungen der virtuellen Maschinen maßgeblich die elektrische Leistungsaufnahme der Server, weshalb eine Betrachtung der entstehenden CPU-Auslastungen der VM-Profile vorgenommen wird. Für die Planung wird die Auslastung  $u_v$  jeder virtuellen Maschine  $v \in V$  als ein maximaler Wert prognostiziert. Zur Berechnung diesen Werts wird sowohl die Historie als auch eine Auslastungsvorhersage herangezogen. Dieses Prinzip ist in Abbildung 6.3 beispielhaft dargestellt.

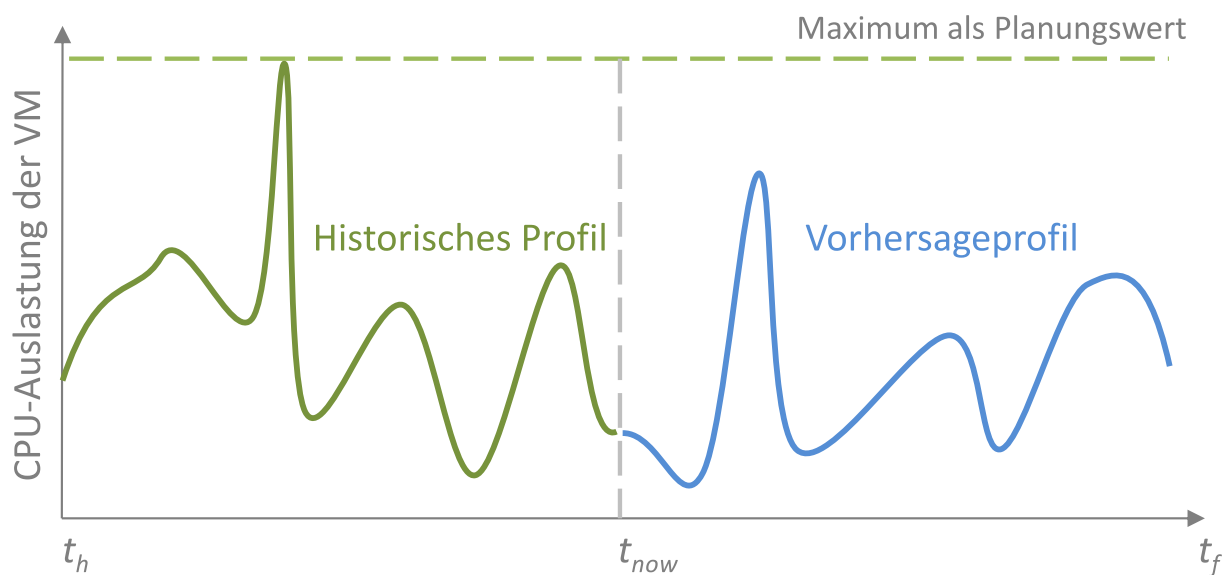


Abbildung 6.3: Auslastungsbetrachtung für die Langzeitplanung: Aus der historischen Analyse der CPU-Auslastung einer virtuellen Maschine wird ein Vorhersageprofil erzeugt. Das genutzte Vorhersageverfahren stammt aus [53]. Die Langzeitplanung muss sich nun nach dem Maximum des bekannten bzw. zu erwartenden Profilverlaufs orientieren.

Die genutzte Vorhersagemethodik wurde aus [53] entnommen und nutzt ein historisches Zeitintervall, in dem mittels Autokorrelationsanalyse periodische Schwankungen erkannt

werden. Diese lassen sich nun als Vorhersage fortsetzen. Die Methodik verwendet bereits eine Abstraktion der Auslastung und berücksichtigt explizit auftretende Maxima, sodass das Vorhersageprofil direkt zur Bestimmung dieser Maxima verwendet werden kann.

Im Zeitraum  $[t_h, t_{now}[$  liegt das historische Profil der virtuellen Maschine, während  $]t_{now}, t_f]$  eine Lastvorhersage beschreibt. Das Zeitintervall, auf das sich die Vorhersage bezieht, kann dabei unabhängig von  $[t_h, t_{now}[$  sein. Die Vorhersagemethodik kann also einen längeren historischen Zeitraum zur Analyse und Bildung des vorhergesagten Auslastungsprofils nutzen. Der Auslastungswert  $u_{v_{max}}$  für die virtuelle Maschine  $v$ , der als Grundlage für die Langzeitplanung verwendet wird, ergibt sich dann durch:

$$u_{v_{max}} = \max\left(u_v(t_h), u_v(t_{h+1}), \dots, u_v(t_{now}), u_v(t_{now+1}), \dots, u_v(t_f)\right) \quad (6.1)$$

Die Komponenten des Vektors  $\vec{u}_{V_{max}}$  enthalten nun die Auslastungsmaxima, die mittels der oben genannten Berechnung für alle virtuellen Maschinen in  $V$  bestimmt wurden:

$$\vec{u}_{V_{max}} = \left(u_{v_{max0}}, u_{v_{max1}}, \dots, u_{v_{maxn}}\right) \Big| n = |V| \quad (6.2)$$

Als zweite Ressource wird die Arbeitsspeichernutzung der virtuellen Maschinen betrachtet. Die Komponenten des Vektors  $\vec{o}_{V_{max}}$  geben hier den allokierten Arbeitsspeicher aller virtueller Maschinen in  $V$  an:

$$\vec{o}_{V_{max}} = \left(o_{v_0}, o_{v_1}, \dots, o_{v_n}\right) \Big| n = |V| \quad (6.3)$$

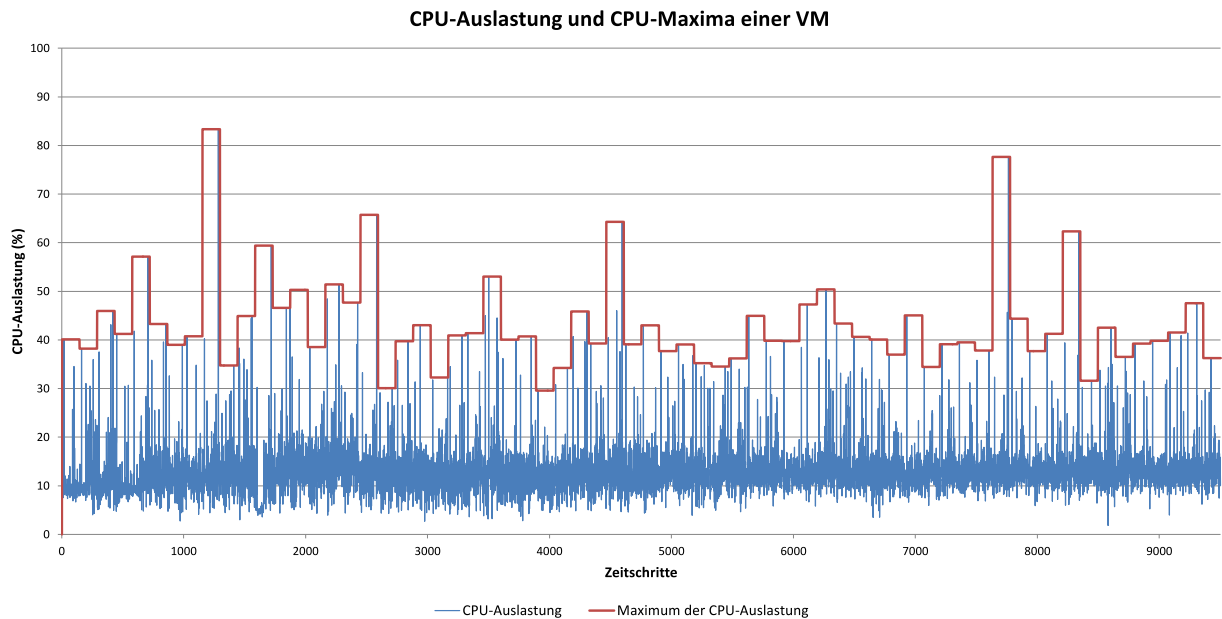
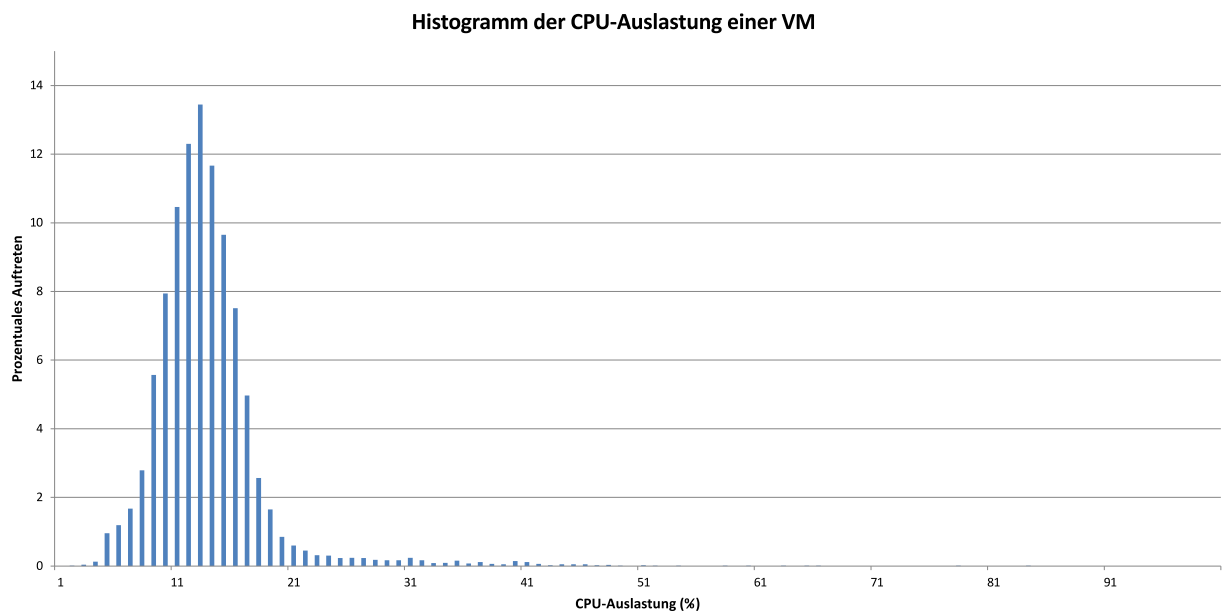
An dieser Stelle wird also kein spezielles Arbeitsspeicherprofil erstellt, sondern das tatsächlich mögliche Maximum verwendet, das beim Erstellen der virtuellen Maschine festgelegt wurde.

### Mögliche Probleme und Grenzen dieses Ansatzes

Mit Hilfe der vorgestellten Betrachtung der Auslastung der virtuellen Maschine wird eine Langzeitplanung erst ermöglicht. Allerdings hat dieser Ansatz auch mögliche Nachteile und Grenzen, die an dieser Stelle kurz diskutiert werden.

So können einzelne, nur sehr kurz auftretende Auslastungsspitzen dazu führen, dass sich die Planung an diesem kurzen Moment orientieren muss. Damit wird zwar die energetische Dimensionierung garantiert, aber es wird gleichzeitig energetisches Einsparpotential verschenkt. In Abbildung 6.4 wird dazu als Beispiel der reale Verlauf einer virtuellen Maschine über einen Zeitraum von zwei Monaten dargestellt (blau). Zusätzlich sind die Maxima eines jeden Tages markiert (rot). Es ist zu erkennen, dass sich der Algorithmus zur Langzeitplanung am Maximum der Auslastung von ca. 83% für diese virtuelle Maschine orientieren muss, obwohl dieses nur einmal und für kurze Zeit (wenige Minuten) auftritt. Das Auslastungshistogramm in Abbildung 6.5 zeigt für dieselbe virtuelle Maschine und denselben Zeitraum, dass sich die virtuelle Maschine fast immer in einem Auslastungsbereich zwischen 5% und 20% befindet. Das gefundene Maximum und der am häufigsten auftretende Auslastungsbereich unterscheiden sich hier also deutlich.



Abbildung 6.4: *CPU-Auslastung und CPU-Maxima einer Beispiel-VM*Abbildung 6.5: *Histogramm der CPU-Auslastung einer Beispiel-VM*

Zeigen nun mehrere virtuelle Maschinen im Rechenzentrum ein solches Verhalten und überlagern sich diese Auslastungsspitzen zudem, kann dies zu einer Reduktion der energetischen Einsparungen führen, da sich die Langzeitplanung an diesen Maxima ausrichten muss. Gleichzeitig darf der Algorithmus diese Spitzen nicht ignorieren, da ansonsten während dieser Zeit die ausreichende Kühlung unter Umständen nicht mehr garantiert

werden könnte.

Diese Art der Lastspitzen führen nicht nur im Optimierungsverfahren in dieser Arbeit zu Problemen, sondern sind aufgrund der aus diesen Spitzen resultierenden Probleme Gegenstand aktueller Forschung. Daher existieren auch mehrere Lösungsansätze, die dem Auftreten der Spitzen und deren Ausprägung entgegen wirken. Ein solcher Ansatz ist der Einsatz von Power-Capping (siehe dazu auch die in Kapitel 3.2 vorgestellten Arbeiten). Dabei wird der jeweilige Server, der eine Lastspitze aufzeigt, die nicht mehr durch die unterstützenden Geräte wie die Klimatisierung oder die USV abgefangen werden kann, künstlich herunter geregelt. Da diese Art der Lastspitzen nur sehr kurz auftreten [108], ist eine künstliche Begrenzung der Server-Ressourcen in diesen Fällen meist ohne nennenswerte Beeinträchtigung des Nutzers möglich. In den meisten Fällen ist davon auszugehen, dass dieser nichts von dem Vorgehen bemerkt.

In [108] wurden mehrere Dienste auf genau diese Lastspitzen untersucht. Die Forschung kommt zum Ergebnis, dass sehr verschiedenartige Dienste nur sehr geringe Auslastungen zeigen und die Lastspitzen jeweils nur seltene Extremformen sind.

Eine Anwendung von Power-Capping zur Vermeidung von kurzzeitigen Lastspitzen kann also das Potential, das durch die Ausrichtung der Langzeitplanung an diesen Maxima vergeben wird, wieder herstellen. Allerdings müssen dazu die SLA des Rechenzentrums ein solches Vorgehen erlauben. Das in diesem Kapitel beschriebene DaMucA-Verfahren orientiert sich an den tatsächlich auftretenden Maxima und setzt kein Power-Capping voraus. Eine solche Erweiterung ist außerhalb des Rahmens dieser Arbeit und gehört daher zur zukünftigen Forschung.

### 6.2.2 Reaktion auf Notfälle: Evakuierung von Servern

Das in Kapitel 4 definierte Kriterium zur Sicherstellung der Flexibilität des Betriebes beinhaltet sowohl die Bereitstellung zusätzlicher Server-Kapazitäten, als auch die Begrenzung der Dauer einer möglichen Evakuierung jedes einzelnen aktiven Servers im Rechenzentrum. Zusammengenommen soll dies dafür sorgen, dass in einem Notfall die virtuellen Maschinen der betroffenen Server innerhalb einer definierten Zeitspanne auf sichere Server migriert werden können. Als derartiger Notfall kommen mehrere Vorfälle in Frage, die hier kurz kategorisiert werden:

- *Thermischer Notfall*: Ein oder mehrere Server überhitzen plötzlich stark. Hierfür kann es verschiedene Ursachen geben, beispielsweise ausgefallene Motoren der Lüfter im Server oder der Ventilatoren im Umluftkühlgerät. Auch ist ein Versperren des Abluftkanals eines oder mehrerer Server durch Fremdgegenstände denkbar.
- *Hardwaredefekt im Server*: Dies kann ein sich andeutender Festplattendefekt oder ein Ausfall sein, bei dem das System zunächst noch weiter läuft, aber in kurzer Zeit vollständig ausfallen wird.
- *Permanent überhöhte Auslastung*: Eine ansteigende Nutzung der Anwendungen oder Fehler in der Software können zu einer permanenten Maximalauslastung von virtu-

ellen Maschinen und damit den Servern führen. Dies ist nicht mit kurz auftretenden Auslastungsspitzen zu vergleichen. Es gilt also nur dann als Notfall, wenn die Auslastung derart ansteigt, dass dies durch die lokale Flexibilität nicht abgefangen werden kann und die auftretende Auslastung nicht in der Vorhersage bekannt war.

Bei Defekten oder thermischen Notfällen müssen mehrere Server häufig vollständig ausgeschaltet werden. Bei einer ansteigenden Auslastung, die als Notfall eingestuft wird, kann unter Umständen eine Entlastung des Servers durch Migration einer oder mehrerer virtueller Maschinen ausreichen. Auch wenn ein solcher Server nicht ausgeschaltet werden muss, schlägt das DaMucA-Verfahren die gleiche Vorgehensweise vor, wie bei den anderen Notfällen.

Bei derartigen Notfällen wird davon ausgegangen, dass eine vollautomatische Behebung der Ursache nicht möglich ist und ein Mitarbeiter des Rechenzentrums eine manuelle Überwachung und Behebung des Notfalls startet. Das DaMucA-Verfahren dient lediglich dazu, die Evakuierung der betroffenen Systeme vorzunehmen, damit ein weiterer Betrieb gewährleistet wird. Die Vorgehensweise zur Reaktion auf Notfälle ist dann folgende:

1. *Ermitteln der betroffenen Server und virtuellen Maschinen:* Dies kann nur dann automatisch erfolgen, wenn ein Überwachungsprogramm, beispielsweise DCIM, die betroffenen Systeme zweifelsfrei identifizieren kann. Ansonsten wird von einer manuellen Auswahl ausgegangen. Dieser Schritt erfolgt nicht durch das DaMucA-Verfahren, sondern muss von außen erfolgen.
2. *Evakuierung:* Es folgt die Migration der betroffenen virtuellen Maschinen auf einen Server mit ausreichend freien Ressourcen, der aufgrund der globalen Flexibilität bereitgestellt wird. Stehen mehrere Server zur Auswahl, wird der Zielservers mit Hilfe der DaMucA-Server-Heuristik gewählt, die in Abschnitt 6.5 beschrieben wird. Dieser Schritt wird vom DaMucA-Verfahren vorgenommen und erfolgt auf Basis der Mitteilung der betroffenen Systeme aus Schritt 1.
3. *Ausschalten der betroffenen Server:* Ist aufgrund der Art des Notfalls ein Ausschalten der betroffenen Server notwendig, erfolgt dies nach Abschluss der Migration der letzten virtuellen Maschine des jeweiligen Servers. Die Abschaltung kann automatisch erfolgen, sollte aber normalerweise durch DCIM-Software vorgenommen werden.

Mit diesem Schritt ist nun zunächst der weitere Betrieb sichergestellt. Das weitere Vorgehen hängt nun von der Art des Notfalls ab und muss durch einen Mitarbeiter entschieden werden:

4. a) *Wiederaufnehmen des Betriebs:* Können die betroffenen Systeme wieder eingeschaltet werden, da der Fehler schnell behoben werden konnte, können die evakuierten virtuellen Maschinen wieder auf ihren ursprünglichen Server zurück migriert werden und der Betrieb geht nach ursprünglicher Planung weiter.
- b) *Vollständige Neuplanung:* Wenn eine schnelle Behebung der Ursache des Notfalls nicht möglich ist, da beispielsweise noch Reparaturen anstehen, muss ein

eventuell längerer Betrieb ohne die betroffenen Server geplant werden. In diesem Fall wird eine vollständige Neuplanung mit Hilfe des DaMucA-Verfahrens vorgenommen, wobei die betroffenen Server hier von vorne herein ausgeschlossen werden müssen. Das Verfahren muss also dieses Ausschlussprinzip unterstützen und mit den verbleibenden nutzbaren Servern einen gültigen Betriebszustand finden.

Es wird also deutlich, dass das DaMucA-Verfahren in Notfällen den weiteren, kurzfristigen Betrieb sicherstellen soll. Die eigentliche Einordnung und Behebung der Ursache des Notfalls liegt nicht im Zuständigkeitsbereich des Verfahrens und kann nur durch die Mitarbeiter erfolgen. Der Schlüssel zur erfolgreichen Bewältigung dieser Notfälle liegt also auf Seiten des DaMucA-Verfahrens in der Bereitstellung von zusätzlichen Server-Ressourcen und der Ermöglichung der Evakuierung der betroffenen Systeme.

### 6.3 Übersicht: DaMucA-Verfahren

Das DaMucA-Verfahren hat zum Ziel, einen energetisch optimierten Betriebszustand zu finden, der gleichzeitig die in Kapitel 4 definierten weiteren Kriterien erfüllt. Dazu muss eine Kombination aktiver IT-Räume gefunden werden, bei deren Nutzung einerseits in jedem IT-Raum ein energetisch optimierter Betriebszustand erreicht wird, indem wenige und möglichst effiziente Server eingeschaltet werden. Andererseits soll das Maximum der elektrischen Leistungsaufnahme und damit der Wärmeleistung aller aktiven IT-Räume gesenkt werden, wodurch eine Erhöhung der Kaltwassertemperatur erreicht werden kann. Durch diese Maßnahme ist eine signifikante Energieeinsparung bei der Kühltaste möglich, sodass eine gesonderte Betrachtung dieser Optimierung sinnvoll erscheint. Zusätzlich gelten die identifizierten Bedingungen und Einschränkungen der weiteren Kriterien, die dafür sorgen, dass der gefundene Betriebszustand aller beteiligten Geräte im Rechenzentrum betriebssicher und stabil ist und zugleich eine vom Betreiber geforderte Flexibilität erfüllt. Die Verteilung der virtuellen Maschinen auf die Server und auf die verschiedenen IT-Räume im Rechenzentrum bietet einen großen Raum möglicher Lösungskandidaten. Eine Suche nach Lösungen wird daher heuristisch angegangen, wobei das Verfahren hier schnell sinnvoll erscheinende Lösungskandidaten erzeugen und durch verschiedene Variationen eine optimierte Lösung finden soll. Damit soll das Verfahren vor allem eine hohe Skalierbarkeit bieten. Das DaMucA-Verfahren enthält nun diese Heuristiken zur Lösungsfindung und wird in zwei Bereiche unterteilt: die Meta-Heuristik zur Suche nach der Kombinatorik der aktiven IT-Räume und die deterministische Lösungssuche zur Lastverteilung auf den Servern in den aktiven IT-Räumen. Diese Aufteilung wird aus folgenden Gründen gewählt:

- Es wurde ein monotoner Zusammenhang zwischen der Wärmeleistung in den IT-Räumen und der Kaltwassertemperatur sowie zwischen der Kaltwassertemperatur und der elektrischen Leistungsaufnahme der Kältemaschinen in der Analyse herausgearbeitet. Eine Reduktion des Maximums der Wärmeleistung der IT-Räume führt damit auch zu einer Reduktion der Leistungsaufnahme der Kältemaschinen. Da hier

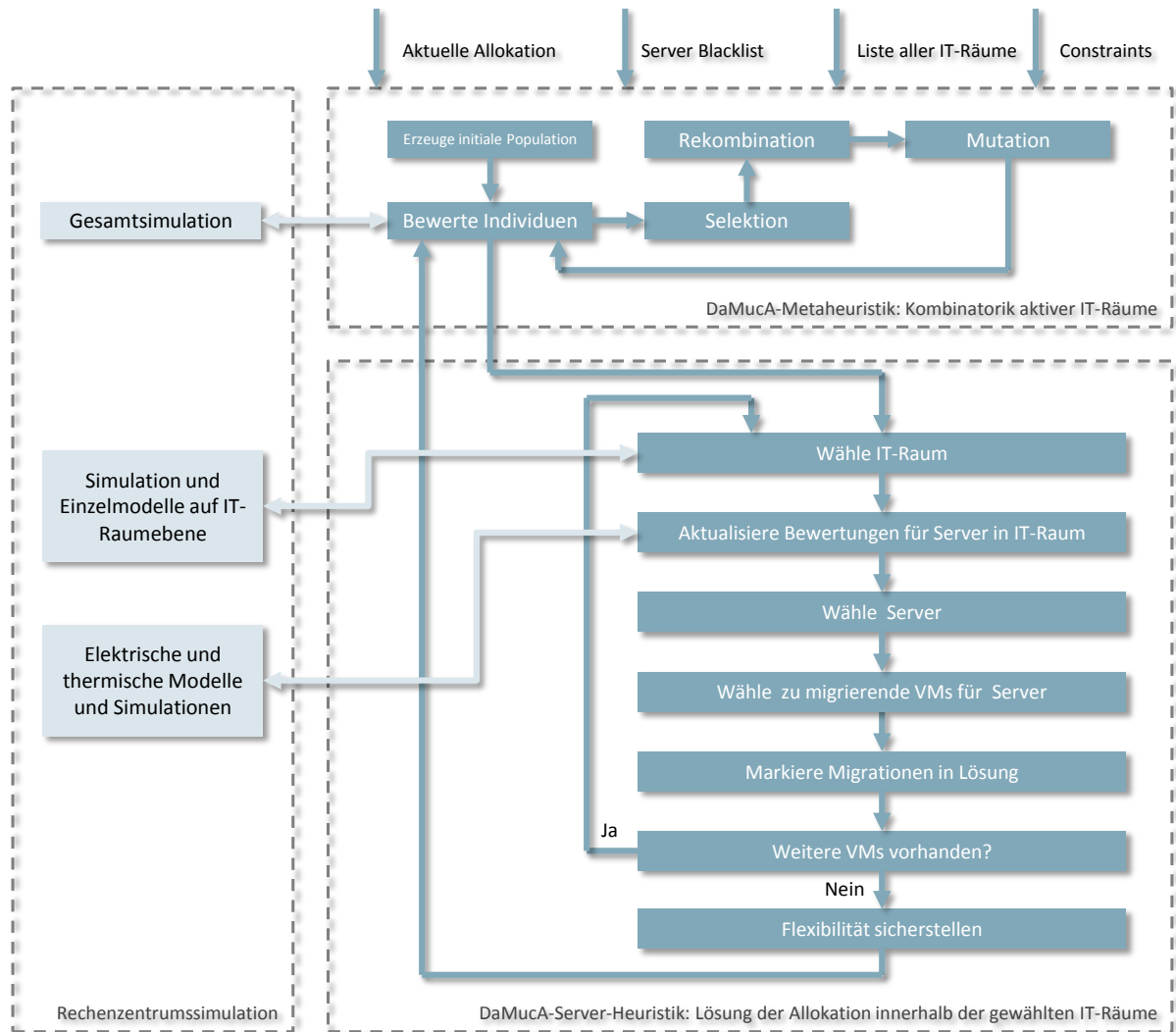


Abbildung 6.6: Übersicht DaMucA-Verfahren: Das Verfahren ist zweigeteilt und beginnt mit der DaMucA-Metaheuristik, die die Kombinatorik der aktiven IT-Räume wählt. Für diese vorgeschlagene Kombination wird dann mittels der DaMucA-Server-Heuristik eine konkrete Verteilung der virtuellen Maschinen auf die Server in diesen aktiven IT-Räumen vorgenommen.

signifikante Einsparungen erreicht werden können, erscheint es als sinnvoll, diese Optimierung gesondert zu betrachten.

- Wenn das Optimierungsproblem kombiniert betrachtet werden würde, entstünde ein derart großer Lösungsraum mit vielen Kombinationen, sodass eine Lösungssuche deutlich komplexer und langwieriger wäre. Durch die Aufteilung der Lösungssuche können kleinere Teilsuchräume erreicht werden.

- Die Trennung der beiden Suchstrategien erlaubt eine flexible Nutzung der DaMucA-Heuristiken. Soll die DaMucA-Server-Heuristik beispielsweise für zukünftige Erweiterungen in ein Kurzzeitplanungsverfahren integriert werden, kann diese einfach und effizient verwendet werden.

In Abbildung 6.6 ist das DaMucA-Verfahren und die Aufteilung in die zwei Bereiche dargestellt. Die DaMucA-Metaheuristik fokussiert die Suche nach einer optimierten Kombination aktiver IT-Räume und nutzt als Verfahren einen genetischen Algorithmus. Für eine Kombination aktiver IT-Räume muss eine vollständige Allokation gefunden werden, damit eine Lösung bewertet werden kann. Dazu wird die DaMucA-Server-Heuristik verwendet, die eine Lastverteilung innerhalb der vorausgewählten IT-Räume vornimmt. Dieses Verfahren ist deterministisch und bildet alle identifizierten Kriterien aus Kapitel 4 ab. Für die Bewertung der Lösungskandidaten werden die in Kapitel 5 beschriebenen Modelle und die vorgestellte Rechenzentrumssimulation verwendet. Dabei werden sowohl Einzelmodelle für einzelne Suchschritte genutzt, als auch ein Gesamtmodell des Rechenzentrums, damit der gesamte Energiebedarf abgeschätzt werden kann.

In den folgenden Unterkapiteln werden die beiden Verfahren zur Kombinatorik aktiver IT-Räume (Abschnitt 6.4) und zur Lastverteilung in den IT-Räumen (Abschnitt 6.5) beschrieben. Dabei werden die Algorithmen und Verfahren textuell und formal erklärt und es werden die jeweiligen wichtigsten Hauptfunktionen als Pseudo-Code dargestellt. Einfache Funktionen werden dagegen aus Gründen der Übersichtlichkeit im Fließtext beschrieben.

## 6.4 DaMucA-Metaheuristik: Kombination aktiver IT-Räume

Die DaMucA-Metaheuristik ist ausschließlich dafür verantwortlich, durch eine Kombinatorik aktiver IT-Räume das Kriterium der Energieeffizienz (vgl. Kapitel 4.2.2) zu erfüllen. Dies gilt insbesondere für die Energieeffizienz der Kältemaschinen und dementsprechend die Anhebung der Kaltwassertemperatur. Dazu wird, wie bereits erwähnt, ein genetischer Algorithmus eingesetzt. Dieser wird zunächst im Allgemeinen beschrieben und die jeweiligen Eigenschaften werden erklärt. Danach folgt eine Beschreibung des Ablaufs des Algorithmus.

### 6.4.1 Genetischer Algorithmus

Als Lösungsverfahren zur Kombinatorik der aktiven IT-Räume wird ein genetischer Algorithmus verwendet, dessen Operatoren allerdings von gängigen genetischen Algorithmen abweichen. Diese wurden modifiziert, um von vorne herein Lösungskandidaten auszuschließen, die zu keiner validen Lösung führen werden. Die Details zu diesen Modifikationen erfolgen in Abschnitt 6.4.2.

Aus folgenden Gründen wurde ein genetischer Algorithmus verwendet:

- Es ist von Beginn an nicht festzustellen, welche Kombination aktiver IT-Räume erfolgversprechend ist, eine Vorgabe bereits „sinnvoller“ Lösungskandidaten ist daher

kaum möglich. Stattdessen kann die Suche ausgehend von bekannten und zufälligen Kombinationen erfolgen. Durch Variationen dieser Kombinatorik kann so eine optimierte Lösung gefunden werden. Hier bietet der genetische Algorithmus den Vorteil, dass er in seiner Population mehrere Kandidaten gleichzeitig betrachten kann und aus diesen eine Auswahl für die erfolgversprechendsten Kandidaten verwendet werden kann.

- Mittels einfacher Modifikationen der Operatoren und der initialen Populationserstellung können Lösungskandidaten erzeugt werden, die auch mit einer höheren Wahrscheinlichkeit eine valide Lösung erzeugen. Anders herum können eindeutig illegale Lösungskandidaten ausgeschlossen werden. Dies beschleunigt die Lösungsfindung.
- Gütekriterien können anhand der Verbesserungen während der letzten Generationen definiert werden. Da die Güte der einzelnen Lösungskandidaten kontinuierlich quantifiziert werden kann, können Verbesserungen direkt bewertet und beispielsweise elitär ausgenutzt werden.
- Ein genetischer Algorithmus untersucht nicht nur eine Lösung, sondern immer mehrere Lösungskandidaten in einer Population einer Generation. Im Fall des DaMucA-Verfahrens ist die Bewertung einer Lösung eine komplexe Aufgabe, die aber aufgrund der in der Population vorhandenen Kandidaten parallel ausgeführt werden kann, sodass eine signifikante Beschleunigung der Lösungsfindung erfolgt. Außerdem kann so ein direkter Vergleich mehrerer Lösungen hinsichtlich mehrerer Kriterien in jeder Generation erfolgen.
- Die Kodierung der Chromosomen für die Kombinatorik aktiver IT-Räume kann in binärer Form erfolgen, was für einen genetischen Algorithmus besonders geeignet ist.

### Kodierung der Individuen

Die Lösungskandidaten eines genetischen Algorithmus werden als Individuen bezeichnet, jeder Wert innerhalb eines Individuums ist ein Gen. Der Wertebereich der Gene und die Zuordnung des angenommenen Werts eines Gens zur Domäne der Problemstellung werden als Kodierung bezeichnet. Ein Individuum in diesem Anwendungsfall beschreibt mittels eines Gens für jeden IT-Raum im Rechenzentrum den Nutzungszustand, also aktiv oder inaktiv. Dies wird in binärer Form kodiert. In Abbildung 6.7 ist die Kodierung formal am Beispiel eines Rechenzentrums mit zwölf IT-Räumen gezeigt. In einem Individuum  $\iota$  repräsentiert jedes Gen  $g_0, \dots, g_{11}$  die Aktivität von IT-Raum  $r_0, \dots, r_{11}$ . Die Menge der aktiven IT-Räume  $R_{active_\iota}$  unter dem Individuum  $\iota$  enthält also alle IT-Räume, deren zugeordnete Gene den Wert 1 haben. Ein konkretes Beispiel einer Kodierung ist in Abbildung 6.8 dargestellt. In diesem Beispiel sind acht aktive IT-Räume gezeigt.

Die binäre Form der Kodierung ist eine der gängigsten für genetische Algorithmen [114] und erlaubt zudem eine einfache und effiziente Implementierung der genetischen Operatoren, was unter anderem im folgenden Abschnitt thematisiert wird.

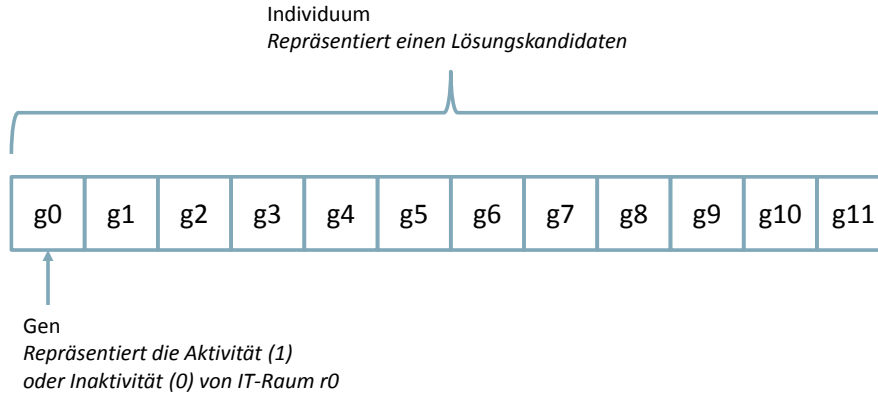


Abbildung 6.7: Kodierung eines Individuums

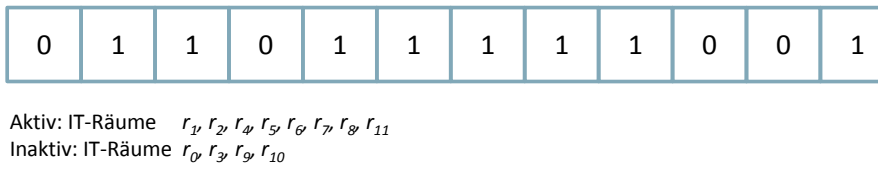


Abbildung 6.8: Beispiel der Kodierung eines Individuums

### 6.4.2 Ablauf des Algorithmus

Der Ablauf des genetischen Algorithmus entspricht der klassischen Form [114] und ist kurz in Abbildung 6.6 oben als Blockdiagramm dargestellt. Die Details des Algorithmus werden in Algorithmus 1 dargestellt. Der Algorithmus beginnt mit der Erzeugung einer initialen Population möglicher Lösungskandidaten, die in zufällig erstellten Individuen kodiert sind. Zusätzlich wird die Kodierung der aktuellen Allokation ebenfalls der Population hinzugefügt. Dadurch bezieht der Algorithmus auch immer die aktuelle Kombination aktiver IT-Räume mit ein. Ist diese bereits optimiert gewählt, führt die Wahl einer gleichen oder ähnlichen Kombinatorik zu geringerem Aufwand beim Übergang zwischen der aktuellen zur neuen Allokation.

Die Funktion  $createInitialPopulation(p_{size})$  erzeugt eine initiale Population  $P = (\iota_0, \iota_1, \dots, \iota_{p_{size}})$  der Größe  $p_{size}$ , mit der der genetische Algorithmus die Lösungssuche beginnt. Die Besonderheit bei der hier verwendeten Implementierung ist, dass jedes Individuum  $\iota \in P$  zumindest nach schnell überprüfbaren Kriterien zu einer möglichen validen Lösung führen können soll. Dazu werden die von allen virtuellen Maschinen  $V$  benötigten Ressourcen summiert und um die jeweilige geforderte Flexibilität analog zu Gleichung 4.22 aus Kapitel 4.4 ergänzt:

$$\vec{l}_{V_f} = \left( \sum_{i=1}^n (u_{v_{maxi}} \cdot (1 + f_{l_{vi}})), \sum_{i=1}^n o_{v_i} \right) | v \in V, n = |V| \quad (6.4)$$

$f_{l_v}$  ist hier der in Kapitel 4.4 definierte lokale Flexibilitätsfaktor der virtuellen Maschine  $v$ .



Die insgesamt verfügbaren Server-Ressourcen der Server in den IT-Räumen  $R_{active_i}$  werden mit  $\vec{l}_{R_i}$  bezeichnet. Damit überhaupt eine Lösungssuche mit dem Individuum  $\iota$  begonnen wird, muss gelten:

$$\vec{l}_{R_i} > \vec{l}_{V_f} \quad (6.5)$$

Wird ein Individuum neu erstellt und die genannte Bedingung trifft nicht zu, werden solange zufällig ausgewählte Gene des Individuums auf den Wert 1 gesetzt, bis die Bedingung erfüllt ist. Damit ist zwar noch nicht garantiert, dass alle Individuen in  $P$  auch zu einer validen Lösung führen, da beispielsweise für bestimmte Kombinationen von aktiven IT-Räumen der Fall auftreten kann, dass einige der definierten Kriterien nicht erfüllt werden können. Dies ist aber erst im späteren Verlauf der Lösungssuche feststellbar. Durch die genannte Bedingung werden aber offensichtlich nicht zielführende Individuen ausgeschlossen.

---

**Algorithm 1:** Genetischer Algorithmus zur Ermittlung einer optimierten Kombination aktiver IT-Räume

---

**Data:**  $p_{size}, k_{max}, n_{best}, k_{break}$   
**Result:**  $\vec{a}_{next}$

- 1  $P \leftarrow \text{createInitialPopulation}(p_{size});$
- 2  $\text{evaluate}(P);$
- 3 **for**  $k$  **to**  $k_{max}$  **step** 1 **do**
- 4  $B \leftarrow \text{selectBestIndividuals}(n_{best}, P);$
- 5 **if** *best individual in B did not change in the last  $k_{break}$  steps* **then**
- 6 |  $\text{break};$
- 7 **end**
- 8  $Z \leftarrow \text{recombineIndividuals}(B);$
- 9  $U \leftarrow \text{createNewIndividuals}(|B| + |Z|, p_{size});$
- 10  $P \leftarrow B \cup Z \cup U;$
- 11  $\vec{a}_{next} \leftarrow \text{evaluate}(P);$
- 12 **end**
- 13 **return**  $\vec{a}_{next};$

---

Die Individuen in  $P$  werden mittels des in Abschnitt 6.5 folgenden Verfahrens zur Lastverteilung in den aktiven IT-Räumen ausgewertet. Dass bedeutet also, dass die Funktion  $\text{evaluate}(P)$  für jeden Lösungskandidaten eine valide Allokation sucht. Erst bei dieser Suche kann letztendlich festgestellt werden, ob der Lösungskandidat tatsächlich zu einer möglichen Lösung führen kann, oder ob beispielsweise zu wenige aktive IT-Räume vorhanden sind, sodass bei Beachtung aller Kriterien beispielsweise nicht ausreichend freie Ressourcen zur Verfügung stehen.

Dazu wird die in Abschnitt 6.5.1 beschriebene Funktion  $\text{computeDaMucAAallocation}(\dots)$  aufgerufen, wobei die im Individuum kodierte Kombination aktiver IT-Räume der Funktion in Form der Menge der nutzbaren IT-Räume  $R_{active}$  übergeben wird. Wird durch die Funktion  $\text{computeDaMucAAallocation}(\dots)$  festgestellt, dass keine valide Lösung für dieses

Individuum gefunden werden kann, wird es entfernt und auf eine Sperrliste gesetzt, damit die Kombination in Zukunft nicht noch einmal betrachtet wird.

Der genetische Algorithmus führt nun die Suche nach verbesserten Lösungskandidaten durch, wobei dieses Vorgehen auf  $k_{max}$  Schritte bzw. Generationen begrenzt ist. Gleichzeitig gibt es ein Abbruchkriterium, wenn nach  $k_{break}$  Generationen keine Verbesserung der Individuen erfolgte. Ein Individuum  $\iota_0$  ist dann besser als ein anderes  $\iota_1$ , wenn die elektrische Leistungsaufnahme  $P_{dc}$  des gesamten Rechenzentrums unter der errechneten Allokation des Individuums geringer ist:

$$P_{dc}(\iota_0) < P_{dc}(\iota_1) \quad (6.6)$$

Aus den  $n_{best}$  besten Individuen einer jeden Generation werden durch Rekombination nun neue Individuen erstellt. Dazu wird das Verfahren *one point crossover* [62] verwendet, wobei auch hier ähnlich wie bei der Erzeugung zufälliger Individuen mittels der in Gleichung 6.5 formulierten Bedingung überprüft wird, ob eine Lösung mit dem resultierenden Individuum möglich ist.

Das *one point crossover* wählt zufällig eine Stelle innerhalb des Individuums aus, das mit einem anderen rekombiniert werden soll. Der Prozess ist in Abbildung 6.9 beispielhaft dargestellt. Als Ergebnis entstehen aus den beiden Eltern-Individuen zwei neue Kind-

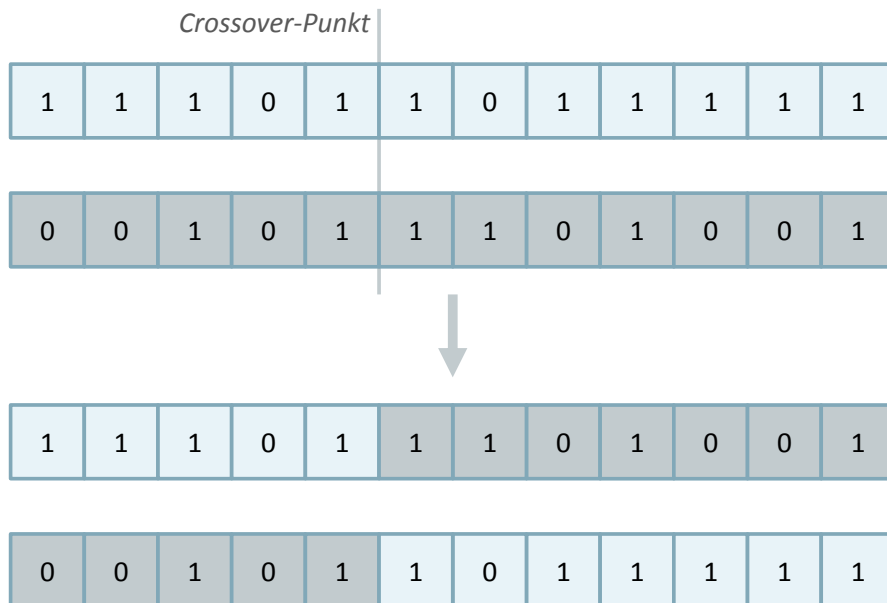


Abbildung 6.9: Beispiel der Rekombination zweier Individuen

Individuen. Nach der Rekombination findet mit einer geringen Wahrscheinlichkeit<sup>1</sup> eine Mutation der Kind-Individuen statt.

<sup>1</sup>In Experimenten haben sich Wahrscheinlichkeiten von 10% für die Mutation eines Individuums und von 3% für die Mutation eines einzelnen Gens als zielführend herausgestellt. Weitere Details zu den Parametern des genetischen Algorithmus folgen in der Evaluation in Kapitel 7.

Die Population  $P$  der nächsten Generation setzt sich jeweils aus den besten der letzten Generation ( $B$ ), den rekombinierten und eventuell mutierten Individuen ( $Z$ ) und zufällig neu erstellten ( $U$ ) zusammen.

## 6.5 DaMucA-Server-Heuristik: Lastverteilung in aktiven IT-Räumen

Der genetische Algorithmus betrachtet im Vorlauf Lösungskandidaten, die die Kombination der aktiven IT-Räume definieren. Diese beschreibt allerdings lediglich, welche IT-Räume genutzt werden dürfen und welche nicht, es wird jedoch keine Aussage darüber getätigt, welche Server in diesen IT-Räumen verwendet werden sollen. Dazu wird die DaMucA-Server-Heuristik verwendet, die für die Auswahl der Server und die VM-Verteilung zuständig ist. Diese Heuristik muss das Kriterium der Energieeffizienz auf Seiten der Server (vgl. Kapitel 4.2.1) umsetzen. Außerdem muss die thermische Stabilität im IT-Raum im Hinblick auf lokale Überhitzungen und Grenzwerte beachtet werden (vgl. Kapitel 4.3) sowie die Sicherstellung der Flexibilität der Lösung (vgl. Kapitel 4.4).

### 6.5.1 Algorithmus zur Lastverteilung auf Server in gewählten IT-Räumen

Für jede IT-Raumkombination, also die Menge der aktiven IT-Räume  $R_{active}$ , die von einem Lösungskandidaten des genetischen Algorithmus vorgegeben wird, wird eine vollständige Allokation berechnet. Erst dadurch kann festgestellt werden, ob eine valide Lösung für diese Kombination möglich ist und eine entsprechende Bewertung errechnet werden. Eine Planung findet immer ausgehend von der aktuellen Allokation  $\vec{a}_{current}$  und für die Menge aktiver IT-Räume  $R_{active}$  statt. Der Algorithmus benötigt außerdem eine Sperrmenge  $S_b$  aller Server, die nicht für eine VM-Verteilung verwendet werden dürfen. Diese Menge wird beispielsweise verwendet, wenn Server von vorne herein ausgeschlossen werden müssen, da sie von einem Notfall betroffen sind und aktuell nicht für den regulären Betrieb verwendet werden können.

Zusätzlich wird ein Constraints-Vektor  $\vec{c}$  angegeben:

$$\vec{c} = \left( f_g, \vec{f}_l, \vec{t}_{max}, \vec{\psi}_{eva} \right) \quad (6.7)$$

Dieser enthält:

- Die sicherzustellende globale Flexibilität  $f_g$  nach den Definitionen aus Kapitel 4.4.2.
- Die lokale Flexibilität  $\vec{f}_l$ . Dieser Vektor enthält für jede virtuelle Maschine  $v$  aus  $V$  den lokalen Flexibilitätsfaktor  $f_{l_v}$  nach der Definition aus Kapitel 4.4.1:

$$\vec{f}_l = \left( f_{l_{v_0}}, f_{l_{v_1}}, \dots, f_{l_{v_n}} \right) \Big| n = |V| \quad (6.8)$$

- Den Temperaturvektor  $\vec{t}_{max}$ . Dieser enthält die jeweiligen maximalen Temperaturen der warmen Abluft jedes Servers  $s$  aus  $S$ :

$$\vec{t}_{max} = \left( t_{s_{wmax_0}}, t_{s_{wmax_1}}, \dots, t_{s_{wmax_n}} \right) \Big| n = |S| \quad (6.9)$$

- Den Vektor  $\vec{\psi}_{eva}$ , der die maximal erlaubte Evakuierungszeit jedes Servers  $s$  aus  $S$  enthält:

$$\vec{\psi}_{eva} = \left( \psi_{eva_{s_0}}, \psi_{eva_{s_1}}, \dots, \psi_{eva_{s_n}} \right) \Big| n = |S| \quad (6.10)$$

Der Algorithmus in der Funktion `computeDaMucAAllocation(...)` beginnt mit dem Anlegen einer neuen, leeren Allokation  $\vec{a}_{next}$ , die dann Schrittweise mit allen virtuellen Maschinen aus  $V$  gefüllt wird. Dazu wird zunächst der IT-Raum  $r_{current} \in R_{active}$  gewählt, in dem der nächste Server für eine VM-Verteilung verwendet werden soll. Diese IT-Raumwahl mit der Funktion `selectRoom(...)` sowie die Anforderungen an den IT-Raum  $r_{current}$  werden im Detail in Abschnitt 6.5.2 beschrieben. Die gezielte Wahl des IT-Raumes dient der Erfüllung des Kriteriums der Gleichverteilung der Wärmemenge auf die IT-Räume, wie in Kapitel 4.2.2 erklärt. Sollte kein valider IT-Raum  $r_{current}$  gefunden werden, existiert für diese Kombination aus aktiven IT-Räumen  $R_{active}$  und den gegebenen Constraints  $\vec{c}$  keine Lösung. In diesem Fall wird dies dem genetischen Algorithmus mitgeteilt und dieser sperrt daraufhin diesen Lösungskandidaten.

Ansonsten fährt der Algorithmus mit der Wahl eines Servers  $s_{current}$  im gewählten IT-Raum  $r_{current}$  fort. Als Server-Kandidaten  $S_{candidates}$  werden alle Server im gewählten IT-Raum ohne die gesperrten Server  $S_b$  verwendet. Die Server-Wahl mit der Funktion `selectServer(...)` sowie die Bedingungen, unter denen ein Server verwendet werden kann, werden im Abschnitt 6.5.3 beschrieben. Auch hier gilt wieder, dass bei fehlendem validen Kandidaten  $s_{current}$  keine Lösung existiert.

Erst nachdem der Server feststeht, der als nächstes in der neuen Allokation  $\vec{a}_{next}$  verwendet werden soll, werden die virtuellen Maschinen  $V_s$  ausgewählt, die auf diesem Server ausgeführt werden sollen. Dazu wird die Funktion `selectVMs(...)` verwendet, die im Detail in Abschnitt 6.5.4 beschrieben wird. Die wichtigsten Aspekte der VM-Wahl sind die Anzahl der VMs auf dem Server und eine Vermeidung unnötiger und zusätzlicher Migrationen. Diese virtuellen Maschinen werden dann in der Allokation  $\vec{a}_{next}$  dem Server  $s_{current}$  zugewiesen und aus der Menge der noch zu verteilenden VMs  $V$  entfernt.

Nachdem alle virtuellen Maschinen verteilt wurden, wird die globale Flexibilität mit der Funktion `assureGlobalFlexibility(...)` überprüft und gegebenenfalls behoben. Die globale Flexibilität  $f_g$  beschreibt, wie in Kapitel 4.4 definiert, die Anzahl der Server, die in einer beliebigen Kombination ausgeschaltet werden können, ohne dass dazu weitere Server eingeschaltet werden müssen. Die Einhaltung dieses Werts ist damit für die Möglichkeit der verzögerungsfreien Reaktion auf Notfälle verantwortlich. Da der Wert eine absolute Anzahl von Servern angibt und erst nach der vollständigen Verteilung der virtuellen Maschinen die tatsächliche Anzahl aktiver Server bekannt ist, kann die globale Flexibilität erst nach der Erstellung der Allokation  $\vec{a}_{next}$  sichergestellt werden. Die Details der Abschätzung der

---

**Procedure** computeDaMucAAllocation

---

```

input :  $\vec{a}_{current}$ ,  $R_{active}$ ,  $S_b$ ,  $V$ ,  $\vec{c}$ 
output:  $\vec{a}_{next}$ 
1  $\vec{a}_{next} \leftarrow$  empty allocation;
2 while  $V \neq \emptyset$  do
3    $r_{current} \leftarrow$  selectRoom( $\vec{a}_{next}$ ,  $R_{active}$ ,  $S_b$ ,  $\vec{c}$ );
4   if has no  $r_{current}$  then
5     /* No solution possible with active room combination  $R_{active}$  and
6       given constraints  $\vec{c}$  */
7     return false;
8   end
9    $S_{candidates} \leftarrow S_{r_{current}} \setminus S_b$ ;
10   $s_{current} \leftarrow$  selectServer( $\vec{a}_{next}$ ,  $S_{candidates}$ ,  $\vec{c}$ );
11  if has no  $s_{current}$  then
12    /* No solution possible with active room combination  $R_{active}$  and
13      given constraints  $\vec{c}$  */
14    return false;
15  end
16   $V_s \leftarrow$  selectVMs( $V$ ,  $\vec{a}_{current}$ ,  $\vec{a}_{next}$ ,  $s_{current}$ );
17  foreach  $v \in V_s$  do
18    setVM( $\vec{a}_{next}$ ,  $v$ ,  $s_{current}$ );
19     $V \leftarrow V \setminus \{v\}$ ;
20  end
21 end
22 assureGlobalFlexibility( $\vec{a}_{next}$ );
23 return  $\vec{a}_{next}$ ;

```

---

tatsächlichen globalen Flexibilität  $f_g$  und der möglichen Korrektur werden in Abschnitt 6.5.5 beschrieben.

Das Ergebnis des Algorithmus ist die neue Allokation  $\vec{a}_{next}$  für die aktiven IT-Räume  $R_{active}$ . Diese Allokation wird mit dem von genetischen Algorithmus generierten Lösungskandidaten verknüpft und hinsichtlich der elektrischen Leistungsaufnahme des Rechenzentrums unter dieser Lösung bewertet. Dazu wird die Rechenzentrumssimulation verwendet, die die elektrische Leistungsaufnahme unter der gefundenen Allokation und mit der VM-Auslastung als durchschnittlichen Wert abschätzt.

Die weiteren Kriterien werden dann nicht mehr überprüft, da der Algorithmus zur Berechnung der neuen Allokation in *computeDaMucAAllocation(...)* diese Kriterien nach den gegebenen Constraints  $\vec{c}$  garantiert und bei Nichterfüllung keine valide Lösung erzeugt.

### 6.5.2 IT-Raumwahl

Der Algorithmus zur IT-Raumwahl löst primär das Problem der gleichmäßigen Verteilung der Leistungsaufnahme auf die IT-Räume und der Absenkung des Maximums dieser nach den Gleichungen 4.11 - 4.13. Dazu werden IT-Räume, die im aktuellen Status der Allokation  $\vec{a}_{next}$  eine relativ niedrige elektrische Leistungsaufnahme haben, in der Wahl bevorzugt. Zur Ermittlung der Leistungsaufnahme nach IT-Räumen wird die Rechenzentrumssimulation verwendet, wobei der aktuelle Stand der Verteilung der Allokation  $\vec{a}_{next}$  und die Auslastung der virtuellen Maschinen nach der Beschreibung in Abschnitt 6.2.1 betrachtet werden. An dieser Stelle muss also, wie im genannten Abschnitt 6.2.1 beschrieben, das Maximum der Auslastung betrachtet werden, da die hier vorgenommene Verteilung auf die Anhebung der Kaltwassertemperatur abzielt.

Im schlechtesten Fall besitzt ein Rechenzentrum eine Menge an IT-Räumen, in denen sich sehr ineffiziente (und meist ältere) Server befinden, während sich in den restlichen IT-Räumen die effizienten (und meist neueren) Server befinden. Unabhängig davon, dass dies ein Planungsmangel des Rechenzentrums wäre, kann die folgende Funktion diesen Mangel nicht erkennen und kompensieren. Stattdessen würde dies im übergeordneten genetischen Algorithmus zur Festlegung der aktiven IT-Räume passieren. Dieser würde dann die Lösungen mit der Nutzung der effizienteren IT-Räume bevorzugen. Somit kann die folgende Funktion deutlich vereinfacht und der benötigte Rechenaufwand damit deutlich reduziert werden. Es erfolgt eine Greedy-Suche nach dem jeweils nächsten passenden IT-Raum.

---

**Procedure** selectRoom

---

```
input :  $\vec{a}_{next}$ ,  $R_{active}$ ,  $S_b$ ,  $\vec{c}$ 
output:  $r_{current}$ 
1  $R_{candidates} \leftarrow$  sorted list of  $R_{active}$  by power consumption (ascending) under  $\vec{a}_{next}$ ;
2 foreach  $r_{current} \in R_{candidates}$  do
3    $S_{candidates} \leftarrow S_{inactive_{r_{current}}} \setminus S_b$ ;
4   if  $\exists s \in S_{candidates}$  that can be switched on according to constraints  $\vec{c}$  then
5     return  $r_{current}$ ;
6   end
7 end
```

---

Die Überprüfung in der Funktion  $selectRoom(...)$  in Zeile 4, ob ein Server  $s$  im IT-Raum  $r$  eingeschaltet werden kann, bezieht sich primär auf die Wärmeverteilung und Grenztemperaturen im IT-Raum  $r$  und der dort vorhandenen Server, sowie auf die grundsätzliche Verfügbarkeit eines Servers, der noch nicht eingeschaltet wurde. Die konkrete Wahl des als nächstes zu verwendenden Servers erfolgt später. An dieser Stelle wird also nur sichergestellt, dass der IT-Raum, der als nächstes verwendet werden soll, mindestens einen potentiellen Server-Kandidaten hat. Ist dies nicht der Fall, muss der IT-Raum nicht weiter in der Suche betrachtet werden.

In Abbildung 6.10 werden beispielhaft die Auswirkungen dieser IT-Raumwahl auf die durchschnittliche Leistungsaufnahme in den einzelnen IT-Räumen dargestellt. Wie zu er-

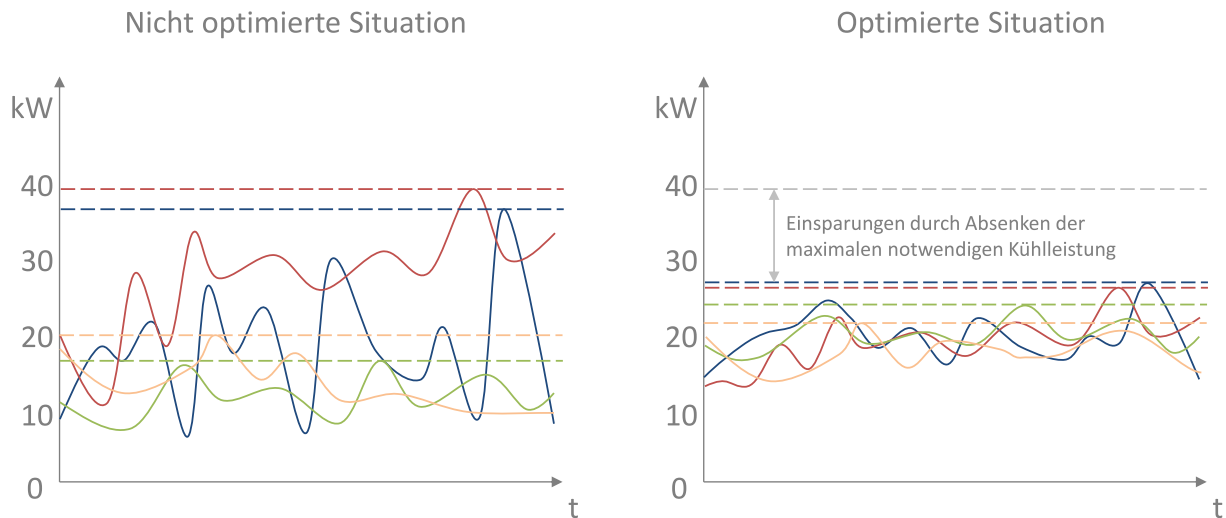


Abbildung 6.10: Anhand von beispielhaften Leistungskurven von vier verschiedenen IT-Räumen zeigt die Abbildung den Einfluss der IT-Raumwahl. Die Leistungsaufnahmen der IT-Räume liegen im optimierten Fall näher beieinander, die Maxima sind abgesenkt, sodass sich der Kühlkreislauf an niedrigeren Maxima der Leistungsaufnahmen orientieren kann. Diese Absenkungen der Maxima ergeben genau Einsparungen bei der Kühlkette. Die Grafik zeigt aus Gründen der Übersichtlichkeit vereinfachte Kurven, die Ergebnisse entsprechen aber einer Simulation der IT-Räume mit realen Lastdaten, sodass auch das genannte Einsparungsbeispiel in der Größenordnung realistisch ist.

kennen ist, sorgt die beschriebene IT-Raumwahl für eine Angleichung der Leistungsaufnahmen und einer gleichzeitigen Senkung des Maximums, indem niedriger ausgelastete IT-Räume priorisiert und die am höchsten ausgelasteten IT-Räume entlastet werden.

### 6.5.3 Server-Wahl

Der Algorithmus zur Server-Wahl muss mehrere Kriterien berücksichtigen. So müssen bereits grundsätzlich die effizientesten Server-Typen bevorzugt werden. Dies erfolgt unabhängig von der aktuellen Allokation  $\vec{a}_{current}$  oder dem momentanen Zwischenstand der neuen Allokation  $\vec{a}_{next}$ . Außerdem müssen die thermischen Bedingungen aus Kapitel 4.3, die sich auf das Innere eines IT-Raumes beziehen, durch die Server-Wahl abgedeckt werden. Dies ist insbesondere die thermische Gleichverteilung im IT-Raum und die Einhaltung der Temperaturgrenzwerte.

Das Verfahren zur Server-Wahl bildet diese Kriterien ab, wobei einige als feste Bedingungen formuliert sind, und andere Kriterien durch Bewertungen und relative Verglei-

che untereinander umgesetzt werden. Zu Beginn der Funktion  $selectServer(...)$  werden alle möglichen Server-Kandidaten in der Menge  $S_{candidates}$  zusammengefasst. Diese sind zunächst alle ausgeschalteten Server. Allerdings wird danach eine Überprüfung vorgenom-

---

**Procedure** selectServer

---

**input** :  $\vec{a}_{next}$ ,  $S_{candidates}$ ,  $\vec{c}$   
**output**:  $s$

- 1  $S_{candidates} \leftarrow S_{candidates} \setminus \text{active servers};$
- 2  $v_{min} \leftarrow \text{choose virtual machine from } V \text{ with smallest } u_v;$
- 3 **foreach**  $s \in S_{candidates}$  **do**
- 4     **if** *simulate  $v_{min}$  additionally on  $s$ , if  $s$  violates any temperature or evacuation time constraint in  $\vec{c}$*  **then**
- 5          $S_{candidates} \leftarrow S_{candidates} \setminus \{s\};$
- 6     **end**
- 7 **end**
- 8 **if**  $S_{candidates} = \emptyset$  **then**
- 9     **return** false;
- 10 **end**
- 11  $\vec{\phi}_{efficiency} \leftarrow \text{normalized efficiency score of each server } s \in S_{candidates};$
- 12  $\vec{\phi}_{fg} \leftarrow \text{normalized global flexibility score of each server } s \in S_{candidates};$
- 13  $\vec{\phi}_{tdist} \leftarrow \text{normalized thermal distribution scores for each server } s \in S_{candidates} \text{ under current state of } \vec{a}_{next};$
- 14  $\vec{\phi}_S \leftarrow \vec{\phi}_{efficiency} + \vec{\phi}_{fg} + \vec{\phi}_{tdist};$
- 15 **return**  $s \in S_{candidates}$  with highest rating in  $\vec{\phi}_S;$

---

men, ob alle Server aus  $S_{candidates}$  mindestens eine virtuelle Maschine aufnehmen könnten, ohne eine der festen Bedingungen zu verletzen. Für diese Überprüfung werden das Server-Temperaturmodell und die Modelle zur Abschätzung der Migrationszeit verwendet. Kann ein Server  $s \in S_{candidates}$  die virtuelle Maschine mit der geringsten Auslastung, die noch zu verteilen ist, nicht ohne eine solche Verletzung aufnehmen, wird er aus der Kandidatenmenge entfernt. Bei der späteren Verteilung der virtuellen Maschinen müssen diese festen Bedingungen ebenfalls überprüft werden, allerdings kann durch Ausschluss an dieser Stelle Rechenzeit bei der Bewertungsberechnung gespart werden. Denn für die Kriterien, die nicht durch feste Bedingungen definiert sind, muss eine Bewertung aller Server in  $S_{candidates}$  erfolgen. Dies betrifft die Energieeffizienzbewertung, die Flexibilitätsbewertung und die Bewertung zur thermischen Leistungsverteilung.

Die Komponenten des Vektors  $\vec{\phi}_{efficiency}$  enthalten für jeden Server  $s \in S_{candidates}$  die Effizienzbewertung, die unabhängig von der Allokation  $\vec{a}_{next}$  ermittelt wird:

$$\vec{\phi}_{efficiency} = \left( \phi_{e_0}, \phi_{e_1}, \dots, \phi_{e_n} \right) \Big| n = |S_{candidates}| \quad (6.11)$$

Die einzelnen Bewertungen  $\phi_{e_i}$  der Server ergeben sich dabei aus einer normalisierten Form



der Server-Energieeffizienz, die in Kapitel 4.2.1 beschrieben wurde. Dazu wird eine lineare Abbildung der einzelnen absoluten Server-Effizienzwerte  $e_{s_i}$  (nach Gleichung 4.6) aller Server aus dem Intervall  $[e_{s_{min}}, e_{s_{max}}]$  nach  $[0, 1]$  vorgenommen:

$$\phi_{e_i} = \frac{(e_{s_i} - e_{s_{min}})}{(e_{s_{max}} - e_{s_{min}})} \quad (6.12)$$

Dabei ist  $e_{s_{min}}$  das Minimum und  $e_{s_{max}}$  das Maximum aller Effizienzwerte der Server aus  $S_{candidates}$ .

In Abbildung 6.11 wird der Vektor  $\vec{\phi}_{efficiency}$  für alle Server-Typen aus dem frei verfügbaren *SPECpower\_ssj 2008 benchmark* [124] in Kombination mit der absoluten Leistungsaufnahme dargestellt. Die roten Balken zeigen die Leistungsaufnahme der Server und die grüne

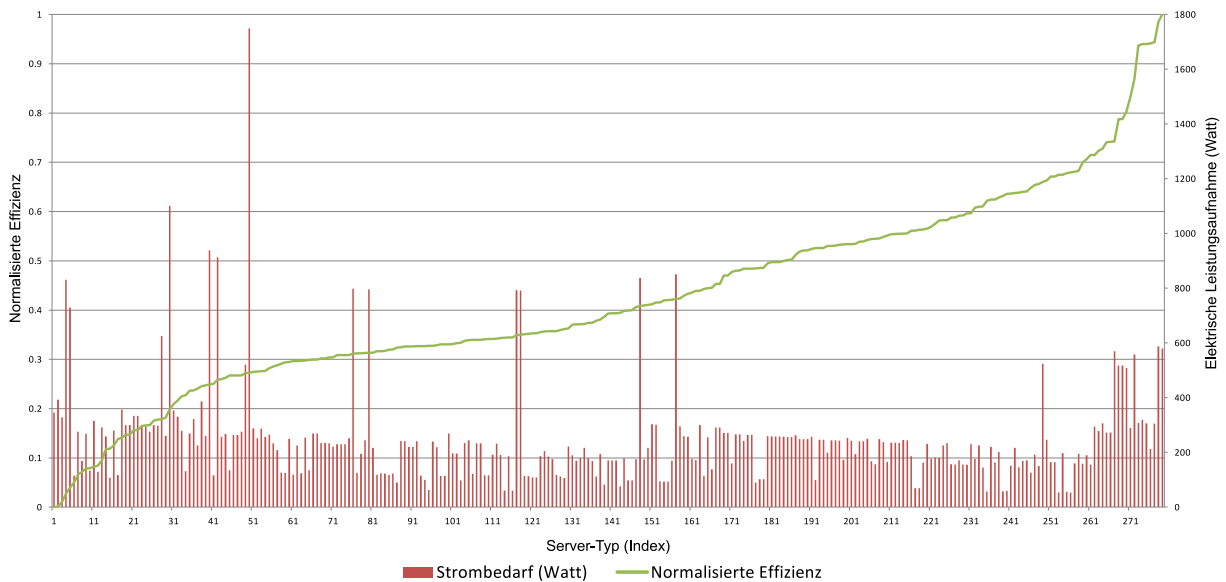


Abbildung 6.11: *Effizienzbewertung und Leistungsaufnahme aller Server-Typen, die sich im frei verfügbaren SPECpower\_ssj 2008 [124] befinden. Die grüne Linie zeigt die Effizienz, während die roten Balken die jeweilige Leistungsaufnahme der Server am effizientesten Punkt darstellen.*

Linie zeigt die Effizienz der einzelnen Server-Typen. Auffällig ist, dass die mittlere elektrische Leistungsaufnahme aller Server sehr ähnlich ist. Effizientere Server können also meist mit der ungefähr gleichen Energie mehr leisten, was in diesem Fall Kapazität für mehr virtuelle Maschinen bedeutet. Allerdings gibt es auch einige Ausnahmen, sodass einige effiziente Server sehr wenig Energie benötigen, aber im Gegenzug auch nur geringe Kapazitäten bereitstellen. Dabei ist es fragwürdig, ob es immer sinnvoll ist, auch diese Server bevorzugt zu verwenden. Damit die zusätzliche Betrachtung der Kapazität von Server-Typen mit einbezogen werden kann, wird die Flexibilitätsbewertung für jeden Server-Typ verwendet.

Die Bewertungen der globalen Flexibilität eines jeden Server-Typs folgt einem ähnlichen Prinzip wie die Effizienzbewertung. Zunächst enthält der Vektor  $\vec{\phi}_{fg}$  die in das Intervall  $[0, 1]$  normalisierten Einzelbewertungen:

$$\vec{\phi}_{fg} = \left( \phi_{fg_0}, \phi_{fg_1}, \dots, \phi_{fg_n} \right) \Big| n = |S_{candidates}| \quad (6.13)$$

Jede Einzelbewertung der globalen Flexibilität muss die prognostizierte Veränderung der globalen Flexibilität  $f_g$  ausdrücken, die sich ergibt, wenn der entsprechende Server als nächstes verwendet und damit in der Allokation  $\vec{a}_{next}$  eingeschaltet wird. Anders ausgedrückt muss die Bewertung der globalen Flexibilität  $\vec{\phi}_{fg}$  diejenigen Server bevorzugen, die bei Ausführung ihrer zugewiesenen virtuellen Maschinen proportional mehr freie Ressourcen zur Verfügung stellen. In einer naiven Herangehensweise würde also die globale Flexibilitätswertung an dieser Stelle tatsächlich ausschließlich die absoluten verfügbaren Ressourcen eines Servers betrachten. Allerdings gibt der Zielwert  $f_g$  wie in Kapitel 4.4.2, definiert die absolute Anzahl der gleichzeitig ausschaltbaren Server an. Dies hängt direkt von den eingesetzten virtuellen Maschinen und deren Lastprofilen ab. Daher wird die Flexibilitätswertung  $\vec{\phi}_{fg_i}$  in Abhängigkeit der im Rechenzentrum vorhandenen VM-Profile bestimmt. Die Wertung bevorzugt also den Server  $s$ , der eine größere Anzahl virtueller Maschinen  $n_{vs}$  aufnehmen kann. Dieser Wert wird zunächst für jeden Server  $s$  bestimmt:

$$n_{vs} = \min \left( \left\lfloor \frac{u_s}{\text{avg}(\vec{u}_{V_{max}})} \right\rfloor, \left\lfloor \frac{o_s}{\text{avg}(\vec{o}_{V_{max}})} \right\rfloor \right) \quad (6.14)$$

Dabei entspricht  $\vec{u}_{V_{max}}$  dem in Gleichung 6.2 und  $\vec{o}_{V_{max}}$  dem in Gleichung 6.3 definierten Auslastungsvektor. Diese einzelnen Werte für jeden Server werden nun wiederum mittels des Minimums und Maximums in das Intervall  $[0, 1]$  normalisiert:

$$\phi_{fg_i} = \frac{(n_{vs_i} - n_{vs_{min}})}{(n_{vs_{max}} - n_{vs_{min}})} \quad (6.15)$$

Somit ergeben sich die normalisierten globalen Flexibilitätswertungen.

Die globale Flexibilitätswertung kann durch Addition mit der Server-Effizienzbewertung verknüpft werden:

$$\vec{\phi}_{fe} = \vec{\phi}_{fg} + \vec{\phi}_{efficiency} \quad (6.16)$$

Das Ergebnis ist in Abbildung 6.12 dargestellt. Die blauen Balken zeigen die Server-Effizienzbewertung  $\vec{\phi}_{efficiency}$ , die roten Balken die globale Flexibilitätswertung  $\vec{\phi}_{fg}$  und die grünen Balken die Kombination  $\vec{\phi}_{fe}$ . Es ist zu erkennen, dass grundsätzlich effiziente Server gewählt werden, allerdings wirkt  $\vec{\phi}_{fg}$  korrigierend, sodass beispielsweise bei Index 41 und 46 ineffizientere Server bevorzugt werden, da diese mehr virtuelle Maschinen aufnehmen können, als vergleichbar effiziente Server.

Die Bewertung der Wärmeverteilung im IT-Raum erfolgt analog zu dem in Kapitel 4.3.2 beschriebenen Verfahren auf Rack-Ebene. Dies bedeutet, dass alle Server, die sich in einem Rack befinden, die gleiche Bewertung bekommen.

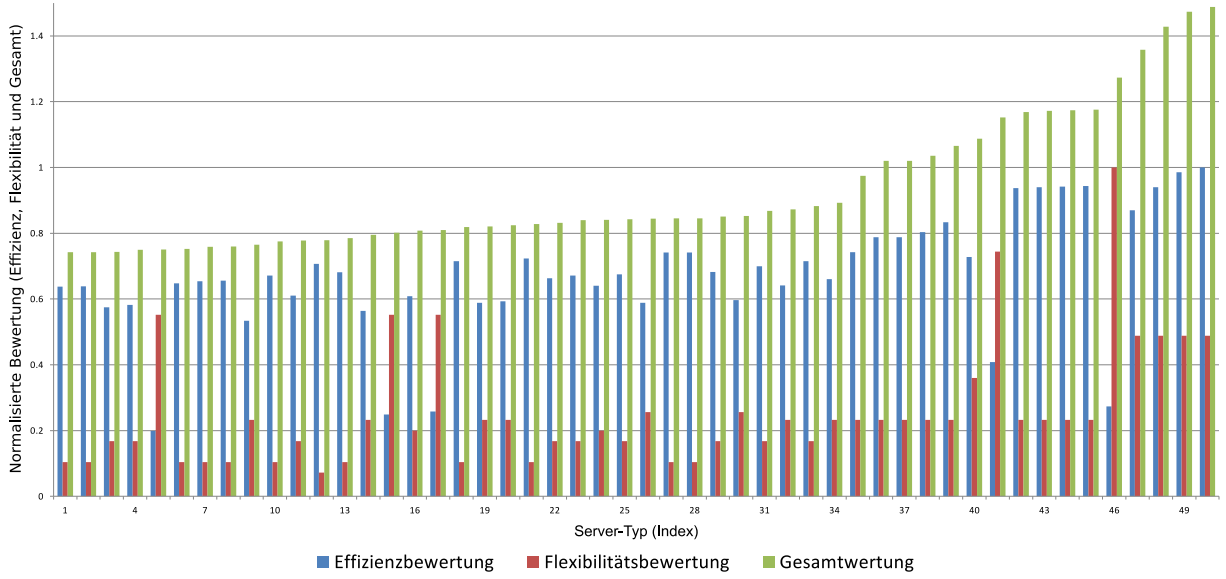


Abbildung 6.12: *Kombinierte Bewertung der Server-Effizienz und Flexibilität: Grundsätzlich werden die effizientesten Server priorisiert, aber die Flexibilitätsbewertung sorgt dafür, dass einige Server, die die globale Flexibilität überproportional anheben, eine ebenfalls insgesamt höhere Bewertung bekommen. Die dargestellten Bewertungen entsprechen  $\vec{\phi}_{efficiency}$  (blau),  $\vec{\phi}_{fg}$  (rot) und  $\vec{\phi}_{fe}$  (grün).*

Die thermische Leistungsverteilung wird nach Gleichung 2.2 anhand der elektrischen Leistungsverteilung bewertet. Die Bewertung  $\phi_{tdist_i} \in \mathbb{R} | 0 \leq \phi_{tdist_i} \leq 1$  für den Server  $s_i$  erfolgt über die elektrische Leistungsaufnahme des Racks  $P_{k_i}$ , in dem sich der Server befindet:

$$\phi_{tdist_i} = \frac{(P_{k_i} - P_{k_{min}})}{(P_{k_{max}} - P_{k_{min}})} \quad (6.17)$$

Die Leistungsaufnahme ergibt sich durch eine Simulation der Server im Rack unter dem bisherigen Stand der Allokation  $\vec{a}_{next}$ . Damit wird der Vektor  $\vec{\phi}_{tdist}$  wie folgt definiert:

$$\vec{\phi}_{tdist} = \left( \phi_{tdist_0}, \phi_{tdist_1}, \dots, \phi_{tdist_n} \right) \Big| n = |S_{candidates}| \quad (6.18)$$

Da alle Einzelbewertungen  $\vec{\phi}_{efficiency}$ ,  $\vec{\phi}_{fg}$  und  $\vec{\phi}_{tdist}$  normalisiert sind, können diese nun durch Addition verknüpft werden. Damit werden die genannten Kriterien in einer Gesamtwertung für alle Server kombiniert und es wird der Server gewählt, dem die höchste Wertung zugeordnet ist.

#### 6.5.4 Wahl der zu migrierenden virtuellen Maschinen

Wenn nun der nächste zu verwendende Server  $s_{target}$  im gewählten IT-Raum gefunden wurde, wird eine Liste von virtuellen Maschinen  $V_s$  ausgewählt, die in der neuen Allokation

$\vec{a}_{next}$  auf diesem Server ausgeführt werden sollen. Dazu wird die aufgeführte Funktion *select-VMs(...)* verwendet. Die Funktion bevorzugt zunächst die virtuellen Maschinen, die sich in der aktuellen Allokation  $\vec{a}_{current}$  auf dem Server befinden. Falls solche virtuelle Maschinen existieren und für diese ausreichend Ressourcen inklusive der zu berücksichtigenden lokalen Flexibilität auf dem Server vorhanden sind, werden zusätzlich noch die individuellen Grenzwerte, die für den Server gelten, berücksichtigt. Dies sind die Grenzwerte der Server-Temperatur, des effizientesten Auslastungsbereichs und der maximalen Evakuierungszeit. An dieser Stelle werden, wie auch an anderen Stellen in diesem Kapitel erwähnt, die entsprechenden Einzelmodelle aus Kapitel 5 verwendet, um diese Grenzwerte zu überprüfen. Durch die Bevorzugung der virtuellen Maschinen, die sich bereits auf dem Server befinden, werden unnötige Migrationen vermieden. Dieser erste Teil des Algorithmus, der in den Zeilen 2-12 beschrieben wird, kann bereits zu einer vollständigen Befüllung des Servers führen, beispielsweise in dem Fall, wenn die Constraints  $\vec{c}$  verschärft wurden. Dann stoppt die Funktion und gibt die Menge der identifizierten virtuellen Maschinen  $V_s$  als Ergebnis in Zeile 9 zurück.

Sind die Grenzwerte des Servers  $s_{target}$  noch nicht erreicht, werden nach den gleichen Maßstäben virtuelle Maschinen aus der bisher noch nicht verteilten Menge  $V_{unallocated}$  verwendet.

Alle virtuellen Maschinen, die unter der Allokation  $\vec{a}_{next}$  auf dem Server  $s_{target}$  ausgeführt werden sollen, werden der Menge  $V_s$  zugefügt. Die Funktion gibt diese als Ergebnis zurück. Die Zuordnung der virtuellen Maschinen aus  $V_s$  erfolgt dann wiederum in der bereits beschriebenen Funktion *computeDaMucAAllocation(...)*. Befindet sich eine virtuelle Maschine  $v$  in Allokation  $\vec{a}_{current}$  auf Server  $s_0$  und in der neuen Allokation  $\vec{a}_{next}$  auf Server  $s_1$ , so bedeutet dies automatisch und implizit eine Migration der virtuellen Maschine  $v$  von  $s_0$  nach  $s_1$ .

### 6.5.5 Sicherstellen der globalen Flexibilität

Erst nachdem eine nach den bisher beschriebenen Kriterien valide Allokation  $\vec{a}_{next}$  gefunden wurde, kann die tatsächliche globale Flexibilität  $f_g$  bestimmt werden. Hierzu muss, wie bereits in Kapitel 4.4 beschrieben, heuristisch vorgegangen werden, da eine vollständige Prüfung aller möglichen Kombinationen auszuschaltender Server nicht in realistischer Zeit möglich ist.

Stattdessen wird eine Heuristik verwendet, die aufgrund der von den virtuellen Maschinen benötigten und den von den Servern bereitgestellten Ressourcen eine Abschätzung für den Ist-Wert von  $f_g$  vornimmt. Grob zusammengefasst wird davon ausgegangen, dass wenn die ersten  $f_g$  Server mit den meisten verfügbaren Ressourcen ausgeschaltet werden können, dies auch für eine beliebige Kombination von  $f_g$  Servern gilt. Das konkrete heuristische Vorgehen ist dabei wie folgt:

1. *Bestimmen einer Referenz-VM*: Auf Basis der durchschnittlichen Auslastungen der virtuellen Maschinen wird eine Referenz-VM gebildet. Die CPU-Auslastung  $u_{ref}$  wird aus dem arithmetischen Mittel der identifizierten Maxima nach der Methode aus

---

```

Procedure selectVMs


---


  input :  $\vec{a}_{current}, \vec{a}_{next}, V_{unallocated}, s_{target}, \vec{c}$ 
  output:  $V_s$ 
  1  $V_s \leftarrow \emptyset$ ;
  2 /* First part of the algorithm: chose fitting VMs that are already on
     server  $s_{target}$  in  $\vec{a}_{current}$  */
  3 if  $s_{target}$  is active in  $\vec{a}_{current}$  then
  4   foreach  $v$  on server  $s_{target}$  in  $\vec{a}_{current}$  do
  5     if  $V_s \cup \{v\}$  fit on  $s$  in  $\vec{a}_{next}$  according to temperature models, efficiency
       metric, evacuation time and resources with local flexibility then
  6        $V_s \leftarrow V_s \cup \{v\}$ ;
  7        $V_{unallocated} \leftarrow V_{unallocated} \setminus \{v\}$ ;
  8     else
  9       return  $V_s$ ;
  10    end
  11  end
  12 end
  13 /* Second part of the algorithm: chose additional VMs from  $V_{unallocated}$ 
     for  $s_{target}$  if fitting */
  14 foreach  $v \in V_{unallocated}$  do
  15   if  $V_s \cup \{v\}$  fit on  $s$  in  $\vec{a}_{next}$  according to temperature models, efficiency metric,
       evacuation time and resources with local flexibility then
  16      $V_s \leftarrow V_s \cup \{v\}$ ;
  17      $V_{unallocated} \leftarrow V_{unallocated} \setminus \{v\}$ ;
  18   else
  19     return  $V_s$ ;
  20   end
  21 end
  22 return  $V_s$ 

```

---

Kapitel 6.2.1 berechnet. Die Arbeitsspeicherbelegung  $o_{ref}$  wird ebenfalls aus dem arithmetischen Mittel des allokierten Arbeitsspeichers aller virtuelle Maschinen aus  $V$  berechnet.

2. *Bestimmen des Ressourcenbedarfs*: Der Ressourcenbedarf  $\vec{l}_{V_f}$  aller virtueller Maschinen inklusive der akkumulierten lokalen Flexibilität wird nach Gleichung 6.4 berechnet.
3. *Bestimmen der Ressourcenverfügbarkeit*: Die Verfügbarkeit  $\vec{l}_{S_{active}(\vec{a}_{next})}$  aller Ressourcen der Server  $S_{active}$  in der Allokation  $\vec{a}_{next}$  wird berechnet.
4. *Sortieren der Server*: Alle Server aus  $S_{active}$  werden nach der verfügbaren Ressourcengröße sortiert. Die Sortierung erfolgt hier nach der Anzahl Referenz-VMs, die auf

den Servern ausgeführt werden können. Der Server mit den meisten Ressourcen und damit der größten Anzahl Referenz-VMs, die ausgeführt werden können, steht als erster in der sortierten Liste.

5. *Bestimmen von  $f_g$*

- a) *Start*: Beginne mit  $f_g = 1$ .
- b) *Wiederhole*: Wähle die ersten  $f_g$  aktiven Server aus der sortierten Liste und entferne diese aus  $S_{active}$ . Wenn  $\vec{l}_{S_{active}(\vec{a}_{next})} \geq \vec{l}_{V_f}$ , dann erhöhe  $f_g$  um eins und wiederhole die Überprüfung, ansonsten stop.

Damit gibt  $f_g | f_g \in \mathbb{N}$  die heuristisch geschätzte globale Flexibilität der Allokation  $\vec{a}_{next}$  an. Grundlage der Heuristik ist die Annahme, dass jede virtuelle Maschine im Rechenzentrum auf jedem beliebigen Server ausgeführt werden könnte. Trifft dies für eine Teilmenge virtueller Maschinen nicht zu, da diese beispielsweise an bestimmte Server gebunden sind, müssen diese aus der Flexibilitätsbetrachtung ausgeschlossen und gesondert behandelt werden. In diesem Fall sollte eine definierte Teilmenge der Server speziell für diese virtuellen Maschinen reserviert werden und die globale Flexibilitätsbetrachtung erfolgt dann für die restlichen Systeme.

Sowohl die heuristische Bestimmung von  $f_g$  der Allokation  $\vec{a}_{next}$  als auch die möglicherweise notwendige Korrektur erfolgt nun mit der Funktion *assureGlobalFlexibility(...)*. Die darin aufgerufene Funktion *heuristicGlobalFlexibility(...)* (Zeilen 5 und 10) nimmt genau die oben beschriebene Abschätzung von  $f_g$  vor. Der erste Teil der Funktion nimmt die Zuschaltung weiterer Server vor, bis der geforderte Wert für  $f_g$  erreicht wurde (Zeilen 1-11). Die Auswahl der Server erfolgt hier wiederum mittels der bereits beschriebenen Kriterien und der DaMucA-Server-Heuristik.

Der zweite Teil der Funktion (Zeilen 12-20) nimmt nun eine Gleichverteilung der virtuellen Maschinen über die bereits eingeschalteten und die neu hinzu geschalteten Server vor. Hier gelten ebenfalls die bereits beschriebenen Grenzwerte und Bedingungen für die Überprüfung, ob eine virtuelle Maschine  $v$  auf dem Server  $s_n$  ausgeführt werden darf. Zusätzlich wird überprüft, ob die elektrische Leistungsaufnahme der beiden Server sinken würde. Dies führt dazu, dass die Verteilung der Auslastung nur soweit vollzogen wird, bis beide Server in ihrem jeweils effizientesten Auslastungsbereich betrieben werden.

## 6.6 Laufzeit und Komplexität

In diesem Abschnitt wird eine Laufzeit- und Komplexitätsbetrachtung des DaMucA-Verfahrens vorgenommen. Zunächst werden die Metaheuristik und die Server-Heuristik getrennt analysiert und abschließend wird eine gemeinsame Betrachtung vorgenommen.

**Procedure** assureGlobalFlexibility

---

```

input :  $\vec{a}_{next}, R_{active}, S_b, f_{gtarget}$ 
1 /* First part of the algorithm: find additional servers ( $S_{flex}$ ) to
   fulfill global flexibility target  $f_{gtarget}$  */
2  $S_{active} \leftarrow$  set with all active servers under allocation  $\vec{a}_{next}$ ;
3  $S_{candidates} \leftarrow S_{R_{active}} \setminus (S_{active} \cup S_b)$ ;
4  $S_{flex} \leftarrow \emptyset$ ;
5  $f_g \leftarrow$  heuristicGlobalFlexibility( $S_{active}$ );
6 while  $f_g < f_{gtarget}$  do
7    $s \leftarrow$  selectServer( $S_{candidates}, R_{active}$ );
8    $S_{flex} \leftarrow S_{flex} \cup \{s\}$ ;
9    $S_{candidates} \leftarrow S_{candidates} \setminus (S_{flex} \cup S_b)$ ;
10   $f_g \leftarrow$  heuristicGlobalFlexibility( $S_{active}$ );
11 end
12 /* Second part of the algorithm: balance virtual machines over
   additional servers as long as no criteria gets worse */
13 foreach  $s_n \in S_{flex}$  do
14   foreach  $s_r \in S_{active_r}$  do
15      $v \leftarrow$  VM on  $s_r$  under  $\vec{a}_{next}$ ;
16     if  $v$  fits on  $s_n$  and  $(P_{s_n} + P_{s_r})$  decreases when  $v$  is on  $s_n$  then
17       | setVM( $\vec{a}_{next}, v, s_n$ );
18     end
19   end
20 end

```

---

**6.6.1 Laufzeit der DaMucA-Metaheuristik**

Die grundlegende Problemklasse der Kombinatorik der aktiven IT-Räume würde bei vollständiger Überprüfung auf eine Komplexität von  $\mathcal{O}(2^n) | n = |R|$  hinauslaufen. Für eine realistische Anzahl verschiedener IT-Räume in einem mittelgroßen bis großen Rechenzentrum ist eine solche Kombinatorik in realistischer Zeit nicht überprüfbar. Daher wurde eine Metaheuristik in Form eines genetischen Algorithmus angewandt, um die Laufzeit zu verringern. Die Laufzeit des genetischen Algorithmus hängt maßgeblich von der Anzahl der benötigten Generationen bis zur Findung einer optimierten Lösung ab. Diese Kennzahl wird speziell in der Evaluation in Kapitel 7 genauer untersucht.

**6.6.2 Laufzeit der DaMucA-Server-Heuristik**

Die DaMucA-Server-Heuristik beginnt mit der Verteilung der virtuellen Maschinen in der Funktion `computeDaMucAllocation(...)`. Hier wird im schlechtesten Fall ein Schleifendurchgang für jede virtuelle Maschine  $v \in V$  durchgeführt. Damit ergibt sich hier zunächst die lineare Komplexität  $\mathcal{O}(|V|)$ .

Die Raumwahl iteriert im schlechtesten Fall über jeden Server  $s \in S$ , da in der Raumwahl ein IT-Raum mit mindestens einem validen Server-Kandidaten gesucht wird. Dieser Fall tritt dann auf, wenn alle IT-Räume mögliche Kandidaten sind und erst der letzte IT-Raum als tatsächlicher IT-Raum  $r_{current}$  verwendet werden kann. Hier gilt dann ebenfalls eine lineare Komplexität von  $\mathcal{O}(|S|)$ . Dasselbe gilt für die Wahl des als nächstes zu verwendenden Servers, wobei hier nur noch die Teilmenge  $S_{candidates} \in S$  in Betrachtung gezogen wird. Damit ist hier wiederum die lineare Komplexität  $\mathcal{O}(|S_{candidates}|)$  gegeben. Die Wahl der virtuellen Maschinen kann nun im schlechtesten Fall wiederum zur Betrachtung aller  $v \in V$  führen, was für die einzelne Funktion die lineare Komplexität  $\mathcal{O}(|V|)$  ergibt. Kombiniert ergibt sich damit für die Funktion  $computeDaMucAAllocation(\dots)$  eine Komplexität von:

$$\begin{aligned} & \mathcal{O}(|V|) \cdot \left( \mathcal{O}(|S|) + \mathcal{O}(|S_{candidates}|) + \mathcal{O}(|V|) \right) \\ & = \mathcal{O}(|V|^2) + \mathcal{O}(|V|) \cdot \mathcal{O}(|S|) + \mathcal{O}(|V|) \cdot \mathcal{O}(|S_{candidates}|) \end{aligned} \quad (6.19)$$

Damit ergibt sich als maßgeblicher Faktor der Komplexität ein quadratischer Einfluss der Anzahl der virtuellen Maschinen.

Die reale Laufzeit des gesamten DaMucA-Verfahrens liegt jedoch aufgrund mehrerer Optimierungen deutlich unter den Laufzeiten des schlechtesten Falls. So können beispielsweise mehrere Berechnungen zwischengespeichert oder mittels vorberechneter Hash-Tabellen direkt zugreifbar gemacht werden. Außerdem können mittels zusätzlicher Kandidatenanalysen die Mengen potentieller zu untersuchender Objekte (primär Server und virtuelle Maschinen) deutlich verkleinert werden. Damit kann in der Realität eine vollständige Lösungsfindung inklusive Simulation der Lösung für ein Rechenzentrum mit mehr als 8.000 Servern und über 64.000 virtuellen Maschinen auf einem Desktop-PC in ein bis zwei Stunden erreicht werden.

Dieses Ergebnis wurde mit einer Implementierung erreicht, die auf Stabilität und nicht auf Geschwindigkeit hin optimiert wurde. So wurde beispielsweise auf die Implementierung von Parallelisierung verzichtet. Dies alleine kann bereits zu einer zusätzlichen signifikanten Reduktion der Laufzeit führen.

## 6.7 Zusammenfassung

Kapitel 6 hat die Beschreibung des DaMucA-Verfahrens sowie die Voraussetzungen für dessen Anwendung vorgenommen. Dazu wurde zunächst eine Einordnung des Verfahrens in den Kontext der Langzeitplanung des Rechenzentrums beschrieben. Hier ist die Bildung der VM-Profile auf Basis der historischen Auslastung und der Vorhersagemethodik ein ausschlaggebender Faktor für das Optimierungsverfahren. Die Möglichkeiten und Grenzen dieses Vorgehens wurden diskutiert. Danach wurde das DaMucA-Verfahren im Detail vorgestellt, wobei die Unterscheidung zwischen der DaMucA-Metaheuristik, die aus einem genetischen Algorithmus besteht, und der DaMucA-Server-Heuristik, die die Verteilung



der IT-Last in den IT-Räumen realisiert, vorgenommen wurde. Abschließend wurde eine Laufzeitanalyse des DaMucA-Verfahrens vorgestellt, wobei dargelegt wurde, dass die vollständige Planung eines großen Rechenzentrums in realistischer Zeit durchgeführt werden kann.



# 7 Evaluation

In den folgenden Abschnitten werden Auswertungen der Simulationen des DaMucA-Verfahrens dargestellt. Dabei wird das Ergebnis mit dem nicht optimierten Zustand verglichen. Dieser Zustand entspricht dem in Rechenzentren im Normalfall vorzufindenden Betriebsmodus, in dem alle Server eingeschaltet und die virtuellen Maschinen gleichmäßig auf diese verteilt sind.

Außerdem wird das DaMucA-Ergebnis mit einem Ergebnis des Optimierungsverfahrens verglichen, das das Allokations-Problem als klassisches Bin-Packing betrachtet und als Lösungsstrategie die First-Fit-Decreasing-Heuristik (FFD) einsetzt. Dieses Verfahren ist in der aktuellen Forschung eines der am meisten eingesetzten Verfahren für diese Problemstellung und die Güte des Verfahrens wurde bereits mehrfach untersucht. Allerdings ist die durch FFD gefundene Lösung nach den in Kapitel 4.1.2 definierten Kriterien praktisch immer illegal, da eine Betrachtung der thermischen Grenzen nicht stattfindet und gleichzeitig eine maximale Packungsdichte der virtuellen Maschinen auf den Servern zu erreichen versucht wird. Dadurch werden lokale Hot-Spots und globale Kühlungsineffizienzen erzeugt. Der Vergleich findet aber aufgrund der großen Popularität der FFD-Heuristik in dieser Domäne und des Referenzcharakters trotzdem statt. In den einzelnen Auswertungen wird gesondert auf die thermischen Bedingungen und die Verbesserungen durch das DaMucA-Verfahren eingegangen.

Tabelle 7.1: Details des simulierten Rechenzentrums

|  |   |
|--|---|
| <b>Anzahl IT-Räume</b>                       | 12  |
| <b>Anzahl Server</b>                         | 8.111                                     |
| <b>Anzahl Racks</b>                          | 576 (48 pro IT-Raum)                      |
| <b>Anzahl virtueller Maschinen</b>           | 64.506                                    |
| <b>Ressourcen aller Server (akkumuliert)</b> | CPU: 225.534,688 GHz; RAM: 265.781,248 GB |

## 7.1 Evaluationsszenario

Zur Evaluation des DaMucA-Verfahrens wurden die Räumlichkeiten des in der Realität untersuchten Rechenzentrums in einer Simulation nachgebildet. Es wurden sowohl vermessene Server-Typen als auch welche aus dem SPEC-Benchmark verwendet, da so eine größere

Menge unterschiedlicher Typen zur Verfügung steht. Die Auslastungsprofile der virtuellen Maschinen stammen ausschließlich von realen Geschäftsanwendungen, die in virtuellen Maschinen ausgeführt werden. Die Auslastungsdaten lagen in minütlicher Auflösung vor. Die Tabelle 7.1 zeigt die Details des evaluierten Szenarios. Das Rechenzentrum hat insgesamt zwölf IT-Räume mit insgesamt 8.111 Servern, die in jeweils 48 Racks pro IT-Raum untergebracht sind. Die Simulation wurde mit 64.506 virtuellen Maschinen durchgeführt. Das DaMucA-Verfahren und der Algorithmus zur FFD-Optimierung wurden prototypisch mittels der Programmiersprache Java implementiert. Auf dieser Basis wurde dann die in diesem Kapitel dargestellte Auswertung durchgeführt.

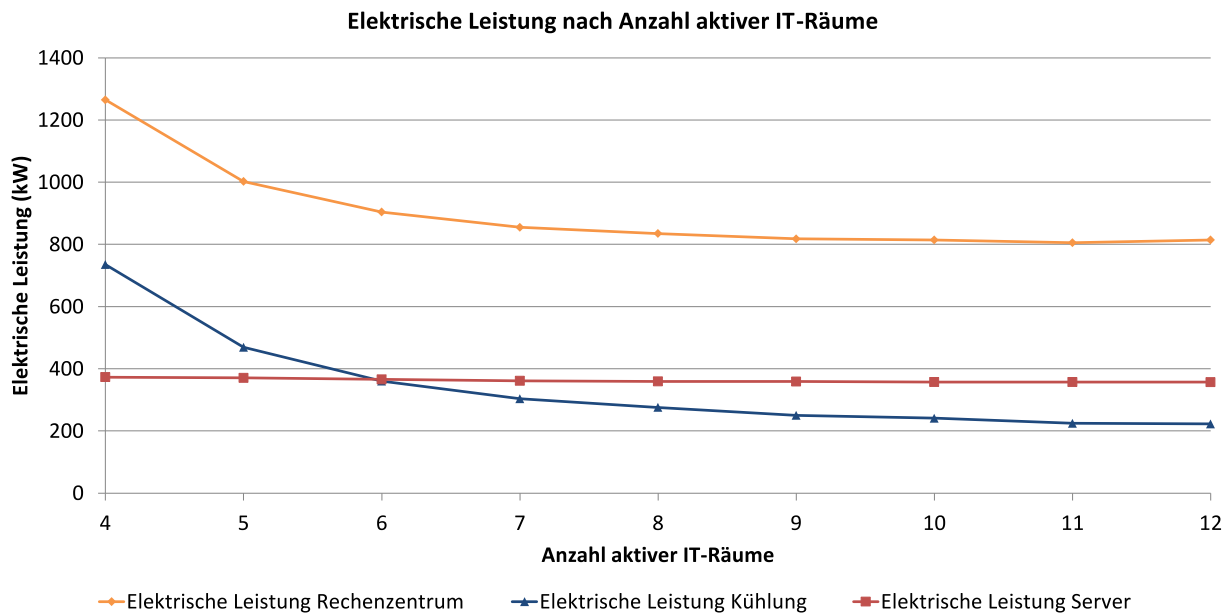


Abbildung 7.1: In Abhängigkeit von der Anzahl der aktiven IT-Räume zeigt die Abbildung die elektrische Leistung des gesamten Rechenzentrums (orange), der Kühlung (blau) und der Server (rot). Mit zunehmender Anzahl aktiver IT-Räume sinkt die benötigte Leistung für die Kühlung signifikant. Die gesamte elektrische Leistungsaufnahme des Rechenzentrums hat ihr Minimum bei elf aktiven IT-Räumen. Dieses Minimum ergibt sich aus Einsparungen, die bei den weiteren Geräten im IT-Raum erreicht werden und nicht in den gezeigten Kurven zu erkennen sind.

## 7.2 DaMucA-Metaheuristik: IT-Raumwahl mittels genetischem Algorithmus

Der genetische Algorithmus wählt die Kombination aktiver IT-Räume, wobei für diese Kombination eine valide Allokation mittels der DaMucA-Server-Heuristik erstellt wird. Das Bewertungskriterium für die Güte einer Kombination ist die elektrische Leistungs-

aufnahme des Rechenzentrums unter der Allokation, die zu der gefundenen Kombination gehört. Alle weiteren Kriterien werden durch die DaMucA-Server-Heuristik zur Allokationssuche abgebildet. Da die DaMucA-Metaheuristik explizit die energetische Optimierung der Kühlung durch Anheben der Kaltwassertemperatur fokussiert, wird in diesem Teil der Evaluation auch eben dieser Einfluss dargestellt. Die Betrachtung der weiteren Kriterien erfolgt in der Evaluation in Kapitel 7.3.

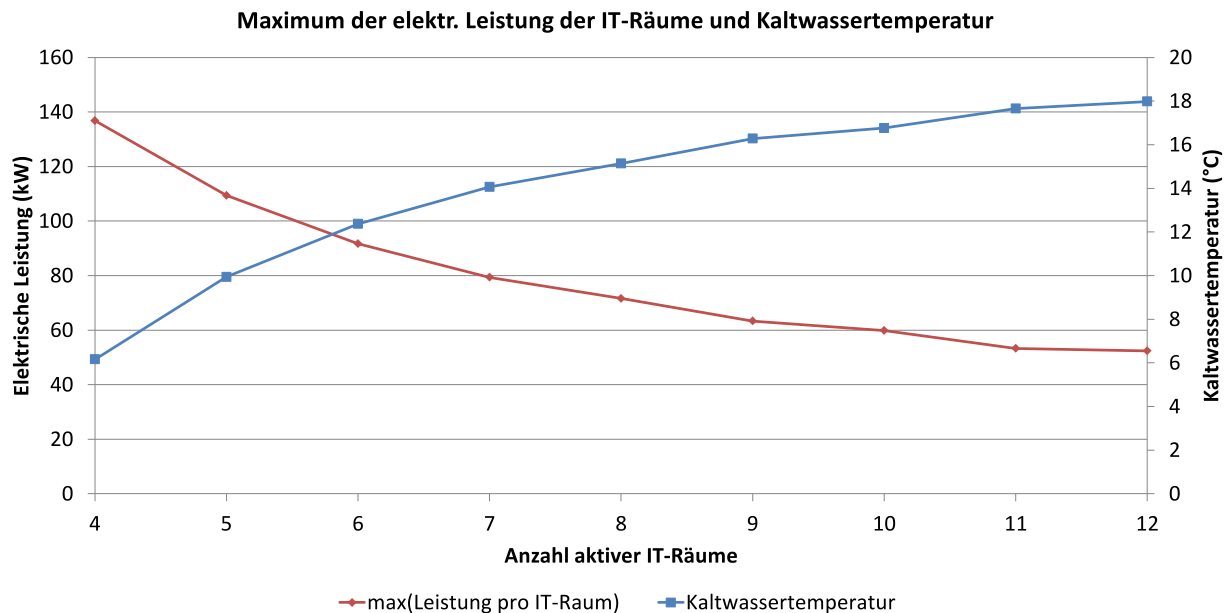


Abbildung 7.2: Die Abbildung zeigt in Abhängigkeit der Anzahl der aktiven IT-Räume das jeweilige Maximum der elektr. Leistung pro IT-Raum und die sich daraus ergebende Kaltwassertemperatur. Werden beispielsweise für die IT-Last im Rechenzentrum nur vier IT-Räume verwendet, führt dies zu einer maximalen elektrischen Leistung von ca. 138kW pro IT-Raum. Damit diese abgeführt werden kann, wird eine Kaltwassertemperatur von ca. 6°C benötigt. Wird die IT-Last dagegen auf zwölf IT-Räume verteilt, sinkt das Maximum auf ca. 50kW. Hier ist eine Kaltwassertemperatur von ca. 18°C ausreichend.

### 7.2.1 Einfluss der Anzahl und Kombination aktiver IT-Räume

Die folgenden Grafiken und Auswertungen zeigen die Auswirkungen der Anzahl der aktiven IT-Räume auf die Leistungsaufnahme des Rechenzentrums und der wichtigsten Geräteklassen. Für eine gegebene Anzahl an aktiven IT-Räumen existieren mehrere Kombinationen valider Lösungskandidaten. Für die folgenden Auswertungen wurden während der Lösungsfindung die jeweils besten Lösungskandidaten für die jeweils angegebene Anzahl aktiver IT-Räume ausgewertet. Die Güte der Lösungskandidaten bezieht sich hierbei immer auf eine möglichst geringe Leistungsaufnahme des gesamten Rechenzentrums.

In Abbildung 7.1 ist die elektrische Leistungsaufnahme des gesamten Rechenzentrums

(orange), der Kühlung (blau) und der Server (rot) in Abhängigkeit von der Anzahl der aktiven IT-Räume dargestellt. Es ist deutlich zu erkennen, dass die Leistungsaufnahme des Rechenzentrums mit zunehmender Anzahl aktiver IT-Räume signifikant sinkt. Das Minimum liegt bei der Nutzung von elf aktiven IT-Räumen. Gleichzeitig ist zu erkennen, dass die Leistungsaufnahme der Server annähernd konstant ist. Das Absinken der gesamten Leistungsaufnahme liegt hier also an der sinkenden Leistungsaufnahme der Kühlung. Die Vermutung liegt hier nahe, dass bei einer steigenden Anzahl aktiver IT-Räume die IT-Last gleichmäßig auf diese verteilt werden kann, sodass das Maximum der Leistungsaufnahme aller IT-Räume (siehe Gleichung 4.3) gesenkt werden kann. Genau dieses Optimierungsziel wurde in den Kapiteln 4.1.1 und 4.2.2 formuliert. Erst ab einem bestimmten Punkt lohnt sich das Zuschalten weiterer IT-Räume nicht mehr, da hier die zusätzliche Grundlast eines IT-Raumes (Umluftkühlgeräte, Netzwerkgeräte, ...) die Ersparnisse übersteigt. Zur genaueren Untersuchung dieses Vorgangs ist in Abbildung 7.2 das jeweilige Maximum der Leistungsaufnahme aller IT-Räume (rot) und die sich daraus ergebende notwendige Kaltwassertemperatur (blau) dargestellt. Hier bestätigt sich die oben formulierte Vermutung. Es ist gut zu erkennen, dass bei Zuschaltung des zwölften aktiven IT-Raumes nur noch geringe Unterschiede bei beiden Kurven auftreten.

Aufgrund der Skalierung in Abbildung 7.1 ist die Leistungsaufnahme der Server nicht differenziert zu erkennen. Daher wird diese in Abbildung 7.3 dediziert betrachtet. Es zeigt sich, dass zwar Schwankungen sowohl bei der Anzahl aktiver Server (grüne Kurve) als auch bei deren Leistungsaufnahme (rote Kurve) auftreten, allerdings sind diese eher gering. Trotzdem sinkt die Leistungsaufnahme aller Server mit zunehmender Anzahl aktiver IT-Räume. Dies ist damit zu erklären, dass die DaMucA-Server-Heuristik, die die Verteilung der virtuellen Maschinen auf die Server vornimmt, bei einer höheren Anzahl aktiver IT-Räume mehr Spielraum bei der Wahl der Server hat. So können primär die effizientesten Server aus allen IT-Räumen gewählt werden, was zu zusätzlichen energetischen Einsparungen führt. Es ist aber zu erkennen, dass diese mit einer Varianz von ca. 17kW oder 4,5% eher gering ausfällt.

Die Anzahl der aktiven Server schwankt ebenfalls nur geringfügig, bis auf einen Ausreißer bei vier aktiven IT-Räumen. Hier liegt die Erklärung darin, dass in diesem Fall fast alle Server in diesen IT-Räumen eingeschaltet werden müssen, also auch tendenziell ältere, weniger effiziente Server mit geringen Kapazitäten. Damit alle virtuellen Maschinen untergebracht werden können, müssen diese verwendet werden, steht jedoch eine größere Auswahl von Servern zur Verfügung, vermeidet der Algorithmus eine Nutzung dieser Server.

Die bisherigen Darstellungen zeigten Abhängigkeiten zur Anzahl der aktiven IT-Räume. Es können bei einer gegebenen Anzahl aber auch verschiedene Kombinationen von aktiven IT-Räumen genutzt werden. Der genetische Algorithmus untersucht während des Prozesses der Lösungsfindung solche verschiedenen Kombinationen. In Abbildung 7.4 ist dazu zu jeder Anzahl von aktiven IT-Räumen die jeweilige Variation der Lösungsgüte je nach Kombination dargestellt. Die Grafik zeigt die Differenz der mittleren elektrischen Leistungsaufnahme des gesamten Rechenzentrums unter der Kombination aktiver IT-Räume mit der geringsten und der höchsten Leistungsaufnahme. Beispielsweise gibt es für die Anzahl von fünf aktiven IT-Räumen mehrere verschiedene Kombinationen, die sich durch ca.

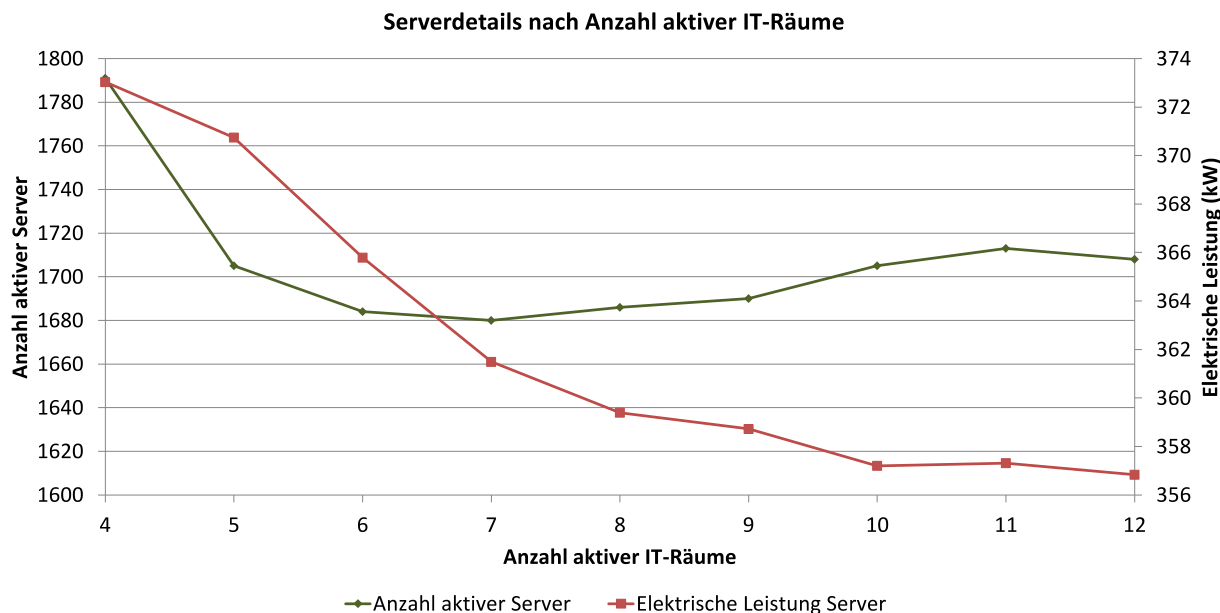


Abbildung 7.3: Anzahl aktiver Server und deren elektr. Leistungsaufnahme in Abhängigkeit der Anzahl aktiver IT-Räume. Die Anzahl aktiver Server schwankt zwischen 1680 und 1790. Die Zahl hängt direkt von der Art der Server in den genutzten IT-Räumen ab. Da mit steigender Zahl aktiver IT-Räume das Auswahlangebot der Server steigt, sinkt hier auch ebenfalls die elektrische Leistungsaufnahme der Server. Der Algorithmus kann also hier effizientere Server auswählen. Allerdings ist die Varianz hier eher gering.

30kW mittlerer Leistungsaufnahme unterscheiden.

Es ist zu erkennen, dass die Differenz bei weniger aktiven IT-Räumen tendenziell größer ist, als bei der Nutzung von fast allen IT-Räumen. Hier gilt die gleiche Begründung, die bereits oben erwähnt wurde: Bei steigender Anzahl aktiver IT-Räume steigt ebenfalls die Auswahlmenge der Server, wodurch die DaMucA-Server-Heuristik eine effizientere Allokation erzeugen kann. Die Kombination der aktiven IT-Räume spielt dann nur eine geringe Rolle, da die DaMucA-Server-Heuristik hier die primäre Optimierung vornimmt. Können allerdings aufgrund der IT-Auslastung im Rechenzentrum nur wenige aktive IT-Räume für einen effizienten Betrieb verwendet werden, ist die Kombination der aktiven IT-Räume ein relevanter Einflussfaktor. Die DaMucA-Metaheuristik untersucht daher in jedem Fall sowohl die Anzahl der aktiven IT-Räume als auch mögliche Kombinationen.

Die Suche nach einer Kombination aktiver IT-Räume wurde bereits in der Analyse mit dem Ziel beschrieben, primär den Energiebedarf der Kühlung zu senken, während die anderen Kriterien durch die DaMucA-Server-Heuristik abgedeckt werden. Die bisher dargestellten Ergebnisse der Evaluation der DaMucA-Metaheuristik zeigen, dass die aktiven IT-Räume einen signifikanten Einfluss auf den Energiebedarf der Kühlung haben und der Algorithmus dieses Potential nutzen kann.

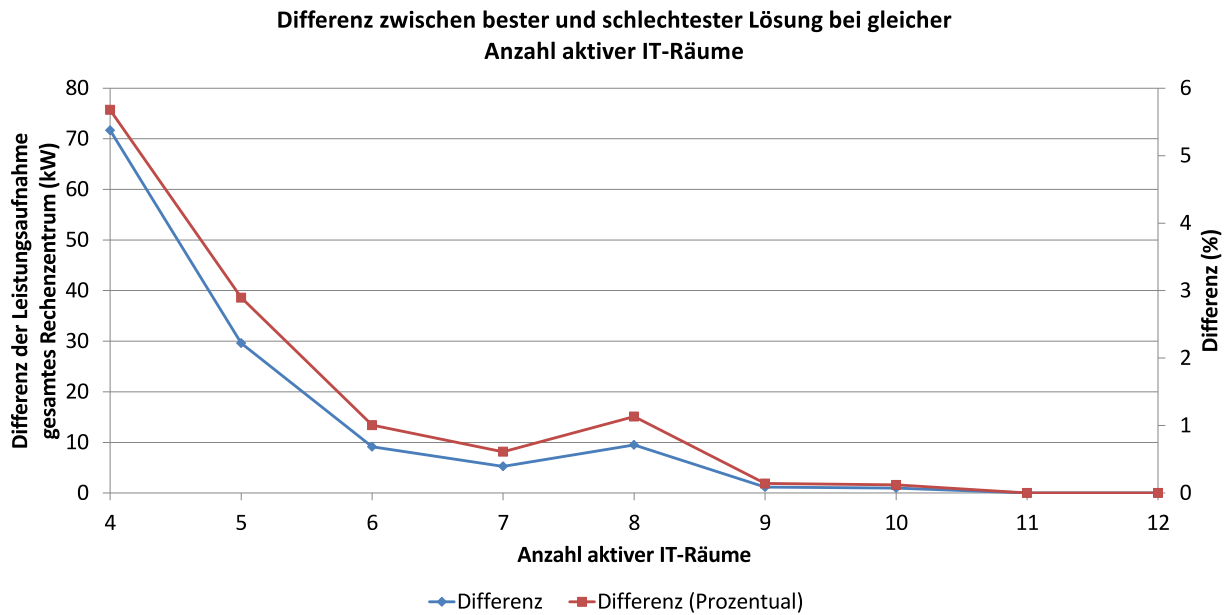


Abbildung 7.4: Die Abbildung zeigt die Differenz der besten und schlechtesten Lösung in Abhängigkeit der Anzahl der aktiven IT-Räume als absoluten (blau) und prozentualen (rot) Wert. Damit kann erkannt werden, wie groß der Einfluss der Kombination der IT-Räume bei gleicher Anzahl auf die elektrische Leistungsaufnahme des Rechenzentrums ist. Bei wenigen aktiven IT-Räumen kommen tendenziell größere Differenzen zustande.

Die Kombinatorik der aktiven IT-Räume ist primär bei wenigen aktiven IT-Räumen signifikant, wie die Auswertung in Abbildung 7.4 gezeigt hat. Dies ist beispielsweise dann relevant, wenn ein Rechenzentrum nur eine sehr geringe Auslastung hat und damit nur wenige aktive IT-Räume nutzen muss. Steigt dagegen die Anzahl der aktiven IT-Räume, die mindestens aufgrund der Auslastung benötigt werden, ist der Einfluss der Kombinatorik geringer. Hier ist es unter Umständen möglich, auf die Untersuchung der Kombinatorik zu verzichten und damit das Verfahren zu beschleunigen. Dies ist vor allem dann sinnvoll, falls zukünftig eine dynamischere Steuerung der Klimatisierung möglich wird und damit häufigere Optimierungen durch das DaMucA-Verfahren notwendig werden.

### 7.2.2 Lösungsfindung: Generationen bis zum Abbruchkriterium

Zur Evaluation des genetischen Algorithmus wird ebenfalls die Anzahl der Schritte bzw. Generationen betrachtet, die benötigt werden, um eine Lösung zu finden, die einem Gütekriterium entspricht. In diesem Fall ist das Gütekriterium, dass keine bessere Lösung nach  $k_{break}$  Generation gefunden werden konnte. Zur Bestimmung von  $k_{break}$  wurden zunächst mehrere unbeschränkte Durchgänge des genetischen Algorithmus durchgeführt. Hier wurde auch die Populationsgröße  $p_{size}$  untersucht und festgelegt. Zunächst wurde mit  $p_{size} = 20$  der Parameter auf einen Wert gesetzt, der nach Erfahrungen aus ähnlichen Anwendungsge-



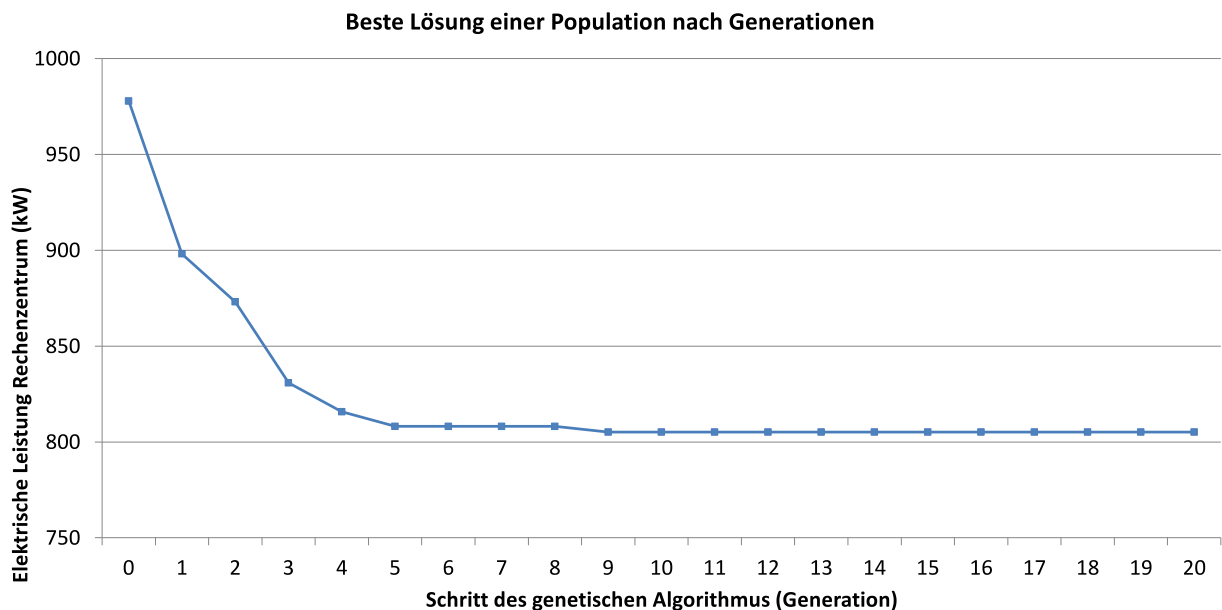


Abbildung 7.5: Innerhalb der ersten 20 Generationen wurde eine Lösung gefunden, die nicht mehr verbessert werden konnte. Besonders während den ersten Generationen finden signifikante Verbesserungen der Lösungen statt. Bereits in der neunten Generation wurde dann die Lösung gefunden, die bis zum Abbruch als beste Lösung beibehalten wurde. Zur Untersuchung der Verbesserung der Lösungen wurde das Abbruchkriterium des genetischen Algorithmus deaktiviert und es wurden die ersten 100 Generationen ausgewertet. Zur besseren Übersichtlichkeit zeigt die Abbildung lediglich die ersten 20 Generationen, da im weiteren Verlauf keine Veränderung mehr stattfand.

bieten von genetischen Algorithmen als sinnvoll angesehen wird. Verschiedene Durchgänge haben dann gezeigt, dass mit  $p_{size} = 10$  ein schnellerer Durchlauf des Algorithmus ohne Verschlechterung der Lösung erreicht werden konnte, weshalb dieser Wert beibehalten wurde.

Die Durchläufe des genetischen Algorithmus ohne Abbruchkriterium wurden für die ersten 100 Generationen durchgeführt. In Abbildung 7.5 ist ein solcher Durchgang dargestellt, wobei aus Gründen der Lesbarkeit nur die ersten 20 Generationen in der Abbildung vorhanden sind. Bereits nach der neunten Generation war keine weitere Veränderung der besten Lösung mehr zu beobachten. Aufgrund dieser Beobachtungen wurde das Abbruchkriterium  $k_{break} = 5$  festgesetzt.

Interessant ist zudem die Betrachtung der Entwicklung der Population während der ersten zehn Generationen, da hier die zu verwendende Lösung gefunden wurde. In Abbildung 7.6 wird dazu der jeweils beste (blau) und schlechteste (rot) Lösungskandidat einer jeden Generation dargestellt. Die Schwankungen der schlechtesten Lösungen entstehen durch Mutationen und die zufällige Erzeugung neuer Lösungskandidaten.

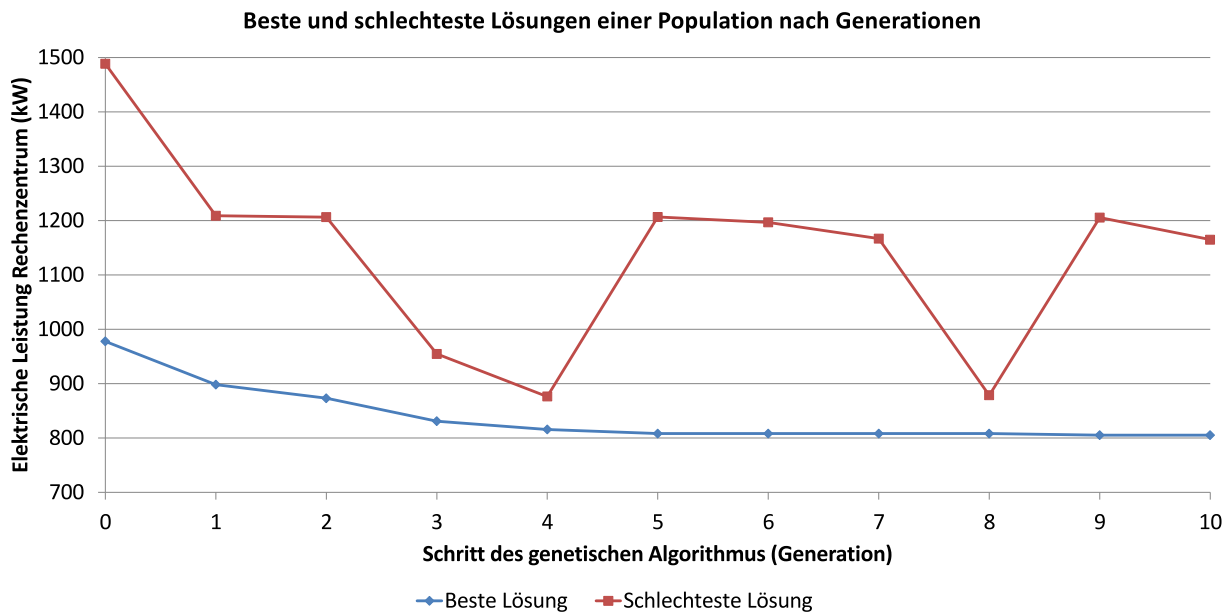


Abbildung 7.6: *Wiederum zur besseren Übersichtlichkeit zeigt die Darstellung die ersten zehn Generationen. In diesen wurde auch die letztendlich verwendete Lösung gefunden. Die Abbildung zeigt nun die jeweils beste und schlechteste Lösung in jeder Generation. Analog zu Abbildung 7.5 ist zu Beginn eine stetige Verbesserung der jeweils besten Lösung zu erkennen. Gleichzeitig schwankt die schlechteste Lösung stark. Dies ist auf die zufällige Mutation und Erzeugung von zufälligen neuen Lösungskandidaten zurückzuführen. Diese führen unter Umständen zu neuen und besseren Lösungskandidaten, können aber auch deutlich schlechtere Lösungen vorschlagen.*

## 7.3 DaMucA-Server-Heuristik: Optimierung in den aktiven IT-Räumen

In diesem Unterkapitel wird die Verteilung der IT-Last in den gewählten aktiven IT-Räumen untersucht und ausgewertet. Hier werden insbesondere die identifizierten Kriterien mit Ausnahme der Optimierung der Kaltwassertemperatur abgebildet. Alle Auswertungen beziehen sich auf den besten Lösungskandidaten, der von der DaMucA-Metaheuristik gefunden wurde und der die Kombination der aktiven IT-Räume vorgibt. Dieser beinhaltet elf aktive IT-Räume, es sind die IT-Räume 2-12 aktiv, IT-Raum 1 ist inaktiv.

### 7.3.1 Elektrische Leistung auf Ebene der IT-Räume

Zunächst wird die mittlere elektrische Leistung pro aktivem IT-Raum untersucht. Wie bereits zu Beginn dieses Kapitels 7 erwähnt, erfolgt bei den folgenden Auswertungen jeweils der Vergleich zwischen dem nicht optimierten Betrieb, einem FFD-optimierten und einem DaMucA-optimierten Betriebszustand. Der nicht optimierte Betrieb entspricht jeweils dem tatsächlichen Stand des untersuchten Rechenzentrums.

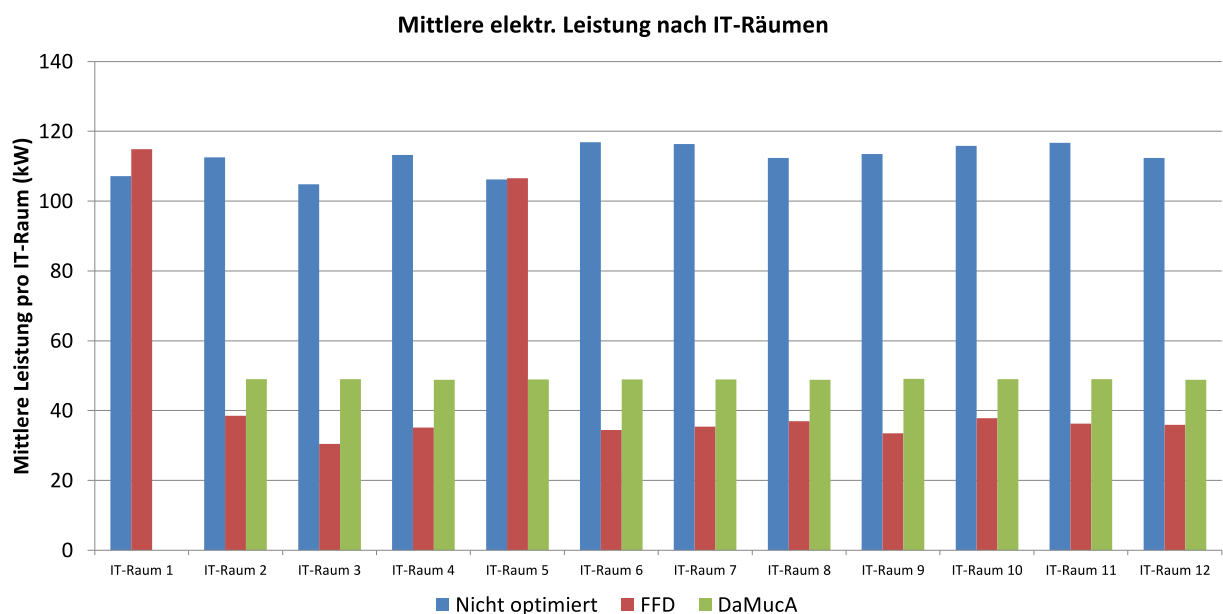


Abbildung 7.7: Vergleich der mittleren elektrischen Leistung nach IT-Räumen für den nicht optimierten (blau), einen FFD- (rot) und einen DaMucA-optimierten (grün) Betriebszustand. Der nicht optimierte Betriebszustand zeigt zwar eine relativ ausgeglichene Leistung über die IT-Räume hinweg, dafür ist die Leistung insgesamt sehr hoch. Der FFD-optimierte Betriebszustand senkt zwar die Leistung insgesamt, zeigt aber in den IT-Räumen 1 und 5 Leistungsspitzen. DaMucA hat zwar den IT-Raum 1 abgeschaltet, lastet dafür aber alle anderen IT-Räume gleichmäßig aus.

In Abbildung 7.7 wird der Vergleich der mittleren elektrischen Leistungsaufnahme pro IT-Raum und Betriebszustand dargestellt. Es ist zu erkennen, dass im nicht optimierten Betrieb (blau) alle IT-Räume aktiv sind und Leistungsaufnahmen über 100 kW zeigen. Da in diesem Zustand alle Server aktiv sind, wird die tatsächliche elektrische Leistungsaufnahme nur durch die IT-Auslastung bestimmt. Der FFD-optimierte Betriebszustand (rot) nutzt ebenfalls alle IT-Räume, zeigt aber signifikante Unterschiede in der jeweiligen Leistungsaufnahme. Die IT-Räume 1 und 5 sind stark ausgelastet, während alle anderen IT-Räume unter 40 kW mittlerer Leistung liegen. Dies liegt an der Verteilung der Server im simulierten Rechenzentrum: IT-Räume 1 und 5 beherbergen mehr Server mit großer Kapazität als die anderen IT-Räume. Da das FFD-Verfahren die Verortung ignoriert, ist die entsprechende Ungleichverteilung das Ergebnis.

Tabelle 7.2: Standardabweichung in kW der mittleren elektrischen Leistungsaufnahme

| IT-Raum | Nicht optimiert | FFD  | DaMucA |
|---------|-----------------|------|--------|
| 1       | 0,79            | 4,44 | -      |
| 2       | 1,01            | 1,0  | 0,86   |
| 3       | 0,95            | 0,74 | 0,76   |
| 4       | 1,0             | 0,79 | 0,8    |
| 5       | 0,86            | 4,23 | 0,83   |
| 6       | 1,05            | 0,94 | 0,9    |
| 7       | 1,06            | 0,92 | 0,86   |
| 8       | 1,01            | 1,03 | 0,82   |
| 9       | 1,06            | 0,87 | 0,84   |
| 10      | 1,03            | 1,03 | 0,85   |
| 11      | 1,07            | 0,89 | 0,83   |
| 12      | 1,06            | 0,92 | 0,87   |

Das DaMucA-Verfahren hingegen nutzt die IT-Räume 2-12 und lastet diese annähernd gleich aus, was dem formulierten Optimierungsziel aus Kapitel 4.2.2 entspricht. Auch ist zu erkennen, dass DaMucA die IT-Räume 2-4 und 6-12 energetisch etwas höher auslastet, als FFD. Im Gegenzug kann allerdings das Maximum gesenkt werden, wodurch die Einsparungen in der Kühlung erreicht werden können. Die DaMucA-Server-Heuristik setzt also an dieser Stelle die Gleichverteilung der energetischen Auslastung über die IT-Räume um, die die DaMucA-Metaheuristik zuvor als aktiv gewählt hat.

Tabelle 7.2 zeigt die Standardabweichung der mittleren elektrischen Leistungsaufnahme der einzelnen IT-Räume aus Abbildung 7.7. Im nicht optimierten sowie im DaMucA-optimierten Betrieb ist die jeweilige Standardabweichung niedrig. Dies kann mit der geringeren Verdichtung der virtuellen Maschinen auf den Servern erklärt werden. Während des stark

verdichteten FFD-optimierten Betriebs ist eine höhere Standardabweichung (vor allem in den IT-Räumen 1 und 5) zu erkennen.

Insgesamt sind die Abweichungen aber bei allen Verfahren relativ gering. Die Schwankungen der Auslastungen der virtuellen Maschinen heben sich teilweise gegenseitig auf oder sind dementsprechend gering, so dass der beobachtete Effekt eintritt.

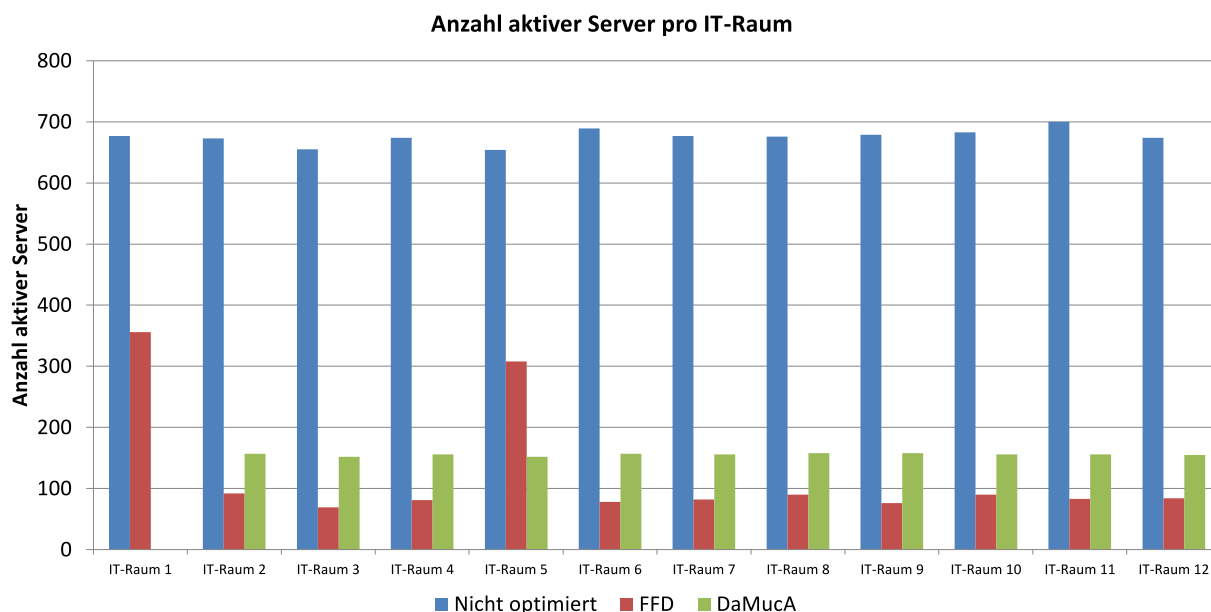


Abbildung 7.8: Analog zu dem Vergleich aus Abbildung 7.7 wird hier die Anzahl der aktiven Server pro IT-Raum dargestellt. Der nicht optimierte Betriebszustand nutzt alle Server. FFD führt zu einer signifikanten Konzentration der aktiven Server in den IT-Räumen 1 und 5, schaltet aber in den weiteren IT-Räumen lediglich ungefähr die Hälfte der Server im Vergleich zum DaMucA-Verfahren ein. In Relation zu der mittleren Leistung pro IT-Raum zeigt sich hier, dass das DaMucA-Verfahren deutlich mehr Server einschaltet, aber in Relation dazu weniger Mehrleistung pro IT-Raum benötigt.

In Abbildung 7.8 wird nun analog zu Abbildung 7.7 die Anzahl der aktiven Server pro IT-Raum dargestellt. Wie bereits beschrieben sind alle Server im nicht optimierten Betrieb aktiv. FFD hingegen optimiert auf eine Minimierung der Anzahl der aktiven Server hin, während DaMucA teils deutlich mehr aktive Server in den IT-Räumen verwendet. Relativ dazu gesehen wird aber weniger zusätzliche Energie dafür benötigt. Dies wird in Abbildung 7.9 ersichtlich. Auch wenn DaMucA zu erhöhtem Leistungsbedarf in einem IT-Raum im Vergleich zu FFD führt, so schaltet DaMucA in jedem Fall prozentual mehr Server ein. Die Anzahl der aktiven Server ist jedoch nicht unbedingt die maßgebliche Kennzahl zum Vergleich der IT-Kapazität. Stattdessen muss die konkrete Ressourcenkapazität der aktiven Server betrachtet werden. Dies erfolgt in den folgenden beiden Auswertungen. Zum besseren Verständnis und Übersichtlichkeit werden die freien Ressourcen wiederum in Form

der Anzahl der Referenz-VMs angegeben (analog zu der Definition in Kapitel 6.5.5), die auf den Servern in Abhängigkeit der freien Ressourcen ausgeführt werden könnten. Die ermittelten Werte für die Referenz-VM betragen in dieser Evaluation ca. 2 GHz für die CPU-Auslastung und ca. 2 GB allozierter Arbeitsspeicher.

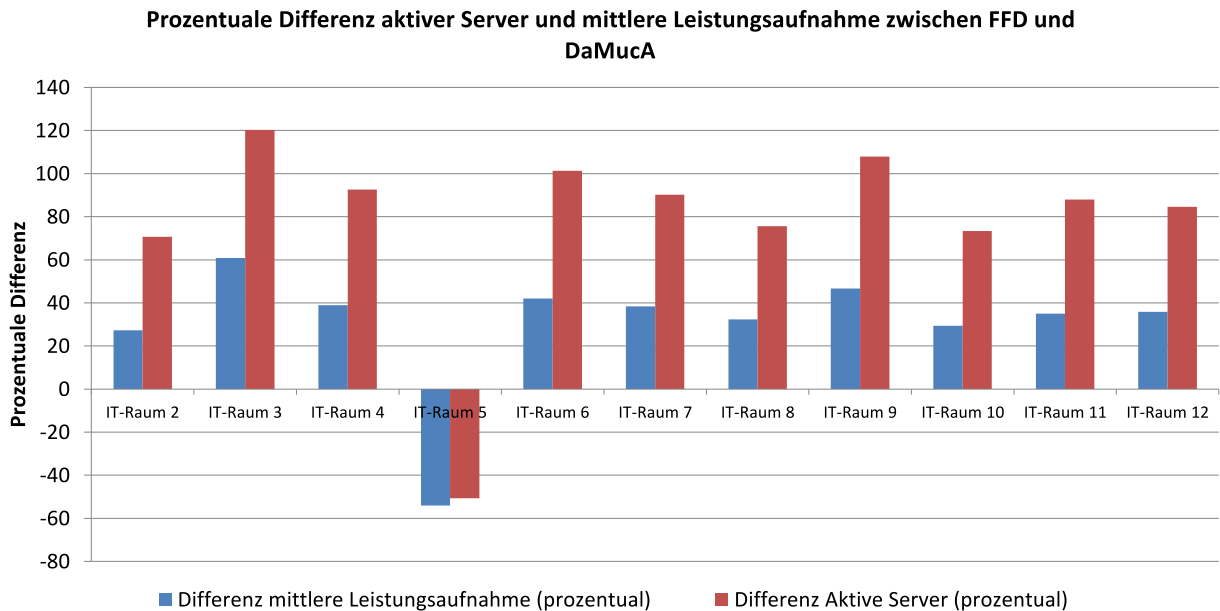


Abbildung 7.9: Die Abbildung zeigt die prozentuale Differenz zwischen der Anzahl aktiver Server (rot) und der mittleren Leistungsaufnahme (blau) pro IT-Raum im Vergleich FFD und DaMucA. Beispielsweise benötigt die DaMucA-Lösung ca. 23% mehr Leistung in IT-Raum 2 als die FFD-Lösung, DaMucA schaltet aber ca. 68% mehr Server ein.

Die Abbildungen 7.10 und 7.11 zeigen nun die Anzahl der Referenz-VMs pro IT-Raum, die mit den jeweiligen freien Server-Ressourcen ausgeführt werden könnten. Abbildung 7.10 vergleicht alle drei Betriebszustände. Da hier die Werte für FFD nicht zu erkennen sind, stellt Abbildung 7.11 diese noch einmal im Vergleich zu DaMucA dar. Grundsätzlich stellt der nicht optimierte Betrieb die allermeisten freien Ressourcen zur Verfügung. Dies ist nicht verwunderlich, da hier alle Server aktiv und jeweils nur geringfügig ausgelastet sind. Im direkten Vergleich zwischen FFD und DaMucA wird aber deutlich, dass DaMucA eine signifikant höhere Anzahl freier Ressourcen vorhält. Wird nun wiederum der prozentuale Vergleich aus Abbildung 7.9 betrachtet, kann an dieser Stelle zusammengefasst werden, dass DaMucA mit nur geringer zusätzlicher Leistungsaufnahme signifikant mehr freie Ressourcen zur Verfügung stellen kann.

Die Ergebnisse dieser Evaluation wurden mittels einer Simulation über 60 Tage erreicht. An dieser Stelle werden nun noch einige absolute Kennzahlen genannt, die in den Abbildungen nicht zu erkennen sind. Außerdem wird die benötigte Energie über die 60 Tage genannt. Die Werte sind in der folgenden Tabelle 7.3 zusammenfassend für die DaMucA-Allokation

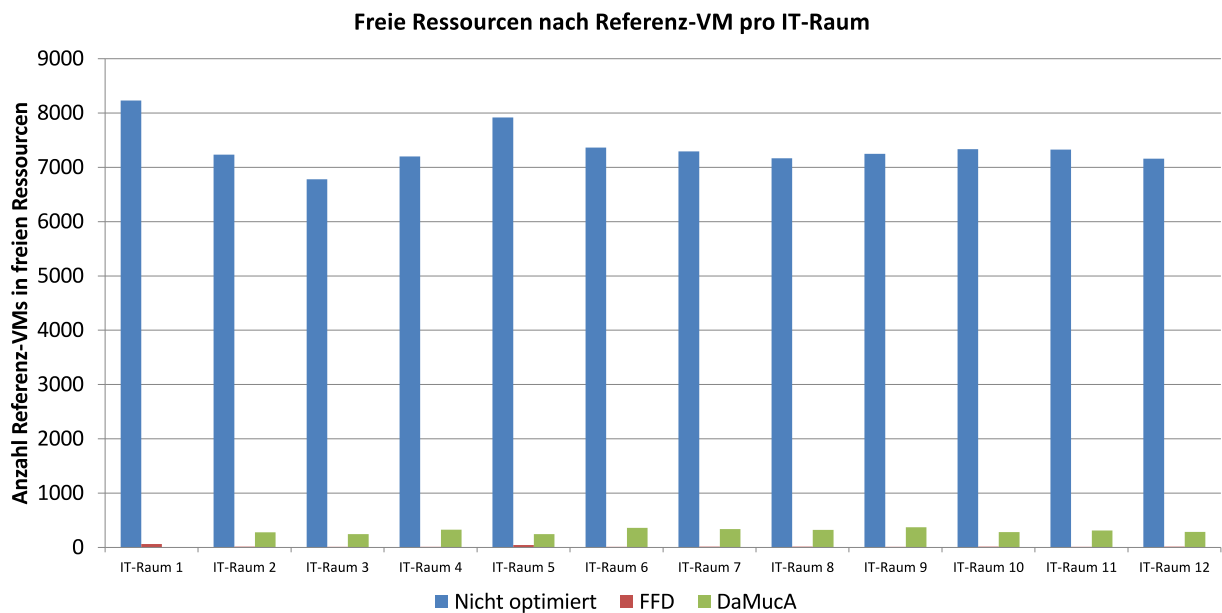


Abbildung 7.10: *Freie Ressourcen nach Referenz-VM pro IT-Raum: Werden alle Betriebszustände verglichen, ist deutlich zu erkennen, dass der nicht optimierte Zustand die mit Abstand meisten freien Ressourcen vorhält. Dies ist nicht verwunderlich, da hier alle Server eingeschaltet sind. Die freien Ressourcen des DaMucA- und FFD-Verfahrens sind im Vergleich dazu sehr gering.*

dargestellt.

Die berechneten Maxima der elektrischen Leistung definieren die notwendige Kaltwassertemperatur und liegen eng beieinander. Der ausschlaggebende Wert beträgt 53,33 kW in IT-Raum 8. In den Abbildungen 7.12 und 7.13 ist der Simulationsverlauf der elektrischen Leistung in den IT-Räumen dargestellt. Damit einzelne Kurven besser zu erkennen sind, wurde die Darstellung auf zwei Abbildungen aufgeteilt. Insgesamt ist zu erkennen, dass die elektrischen Leistungen der IT-Räume sehr ähnlich sind und jeweils unter dem berechneten Maximum liegen.

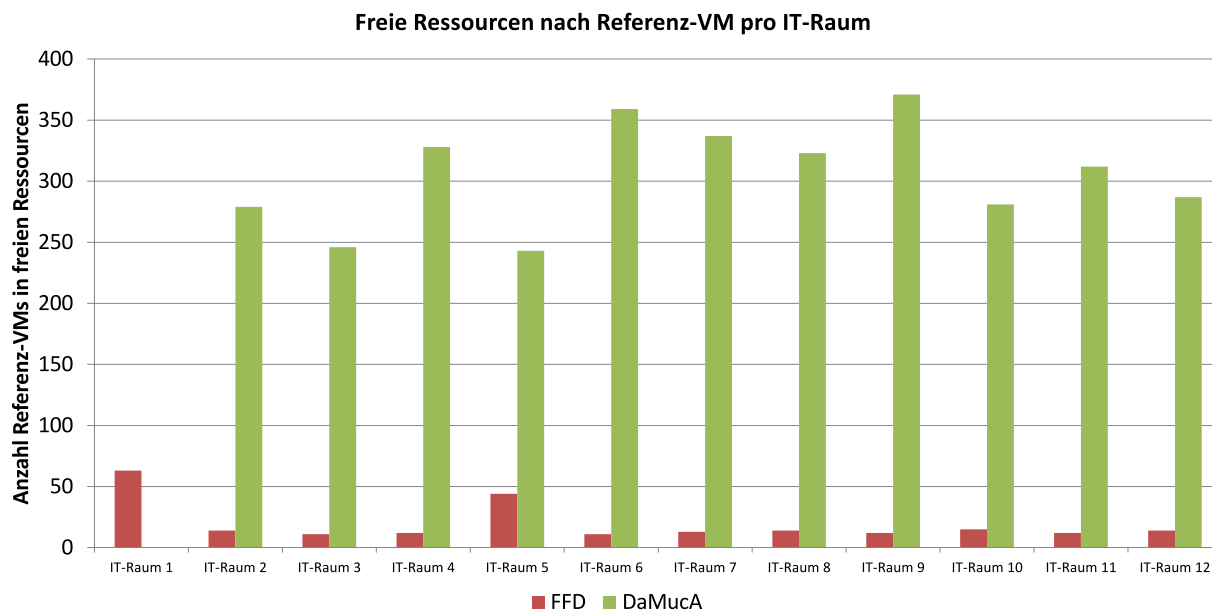


Abbildung 7.11: *Freie Ressourcen nach Referenz-VM pro IT-Raum (nur FFD und DaMucA)*

Tabelle 7.3: Übersicht der Ergebnisse der 60-Tage-Simulation (DaMucA-Allokation)

| IT-Raum | Berechnetes Maximum elektr. Leistung (kW) | aktive Server | benötigte Energie (kWh) |
|---------|---|---------------|-------------------------|
| 1       | 0   | 0             | 0                       |
| 2       | 50,44                                     | 157           | 62606,02                |
| 3       | 52,65                                     | 152           | 63661,12                |
| 4       | 53,18                                     | 156           | 68317,96                |
| 5       | 53,3                                      | 152           | 65259,71                |
| 6       | 52,93                                     | 157           | 67442,65                |
| 7       | 53,17                                     | 156           | 66838,11                |
| 8       | <b>53,33</b>                              | 158           | 67232,96                |
| 9       | 52,97                                     | 158           | 67708,10                |
| 10      | 53,19                                     | 156           | 68756,63                |
| 11      | 53,19                                     | 156           | 72478,94                |
| 12      | 52,94                                     | 155           | 68347,05                |



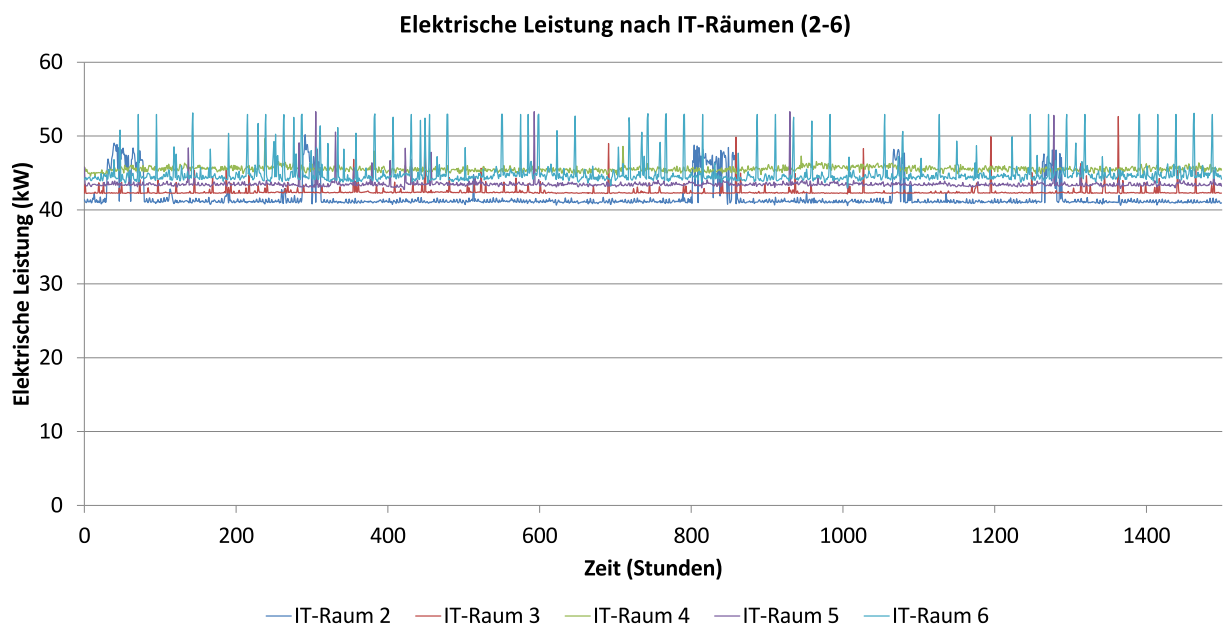


Abbildung 7.12: Die Abbildung zeigt den elektrischen Leistungsverlauf in den IT-Räumen 2-6 der Simulation über 60 Tage unter der Allokation, die mit dem DaMucA-Verfahren gefunden wurde. IT-Raum 1 ist deaktiviert und daher nicht in der Grafik enthalten. Es ist deutlich zu erkennen, dass die Leistungen der einzelnen IT-Räume dicht beieinander liegen und insgesamt unter dem berechneten Maximum bleiben. Damit einzelne Kurven der IT-Räume erkannt werden können, wurde die Auswertung auf zwei Grafiken aufgeteilt, die Darstellung der Leistungen der IT-Räume 7-12 ist in der folgenden Abbildung 7.13 zu finden.

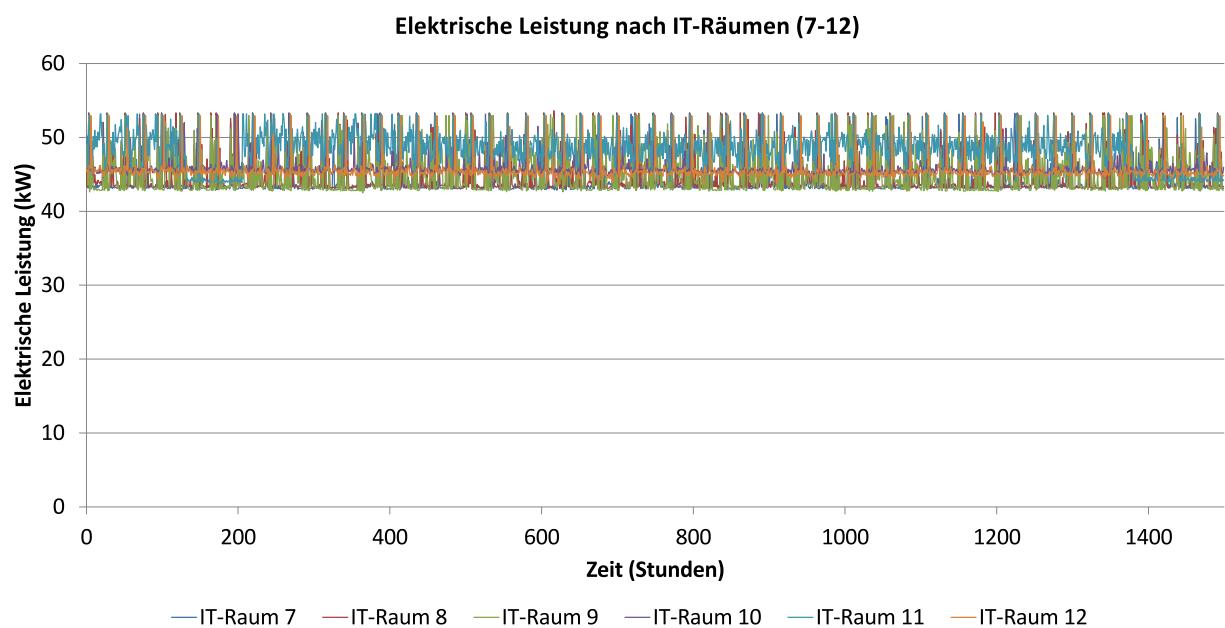


Abbildung 7.13: *Die Abbildung zeigt den elektrischen Leistungsverlauf in den IT-Räumen 7-12 der Simulation über 60 Tage unter der Allokation, die mit dem Da-MucA-Verfahren gefunden wurde. Auch hier liegen die Leistungen der einzelnen IT-Räume dicht beieinander und insgesamt unter dem berechneten Maximum. In einigen dieser IT-Räume sind primär stark periodisch schwankende Anwendungen vorhanden, was deutlich in der Abbildung zu erkennen ist. Auswirkungen auf die Lösung hat dies nicht, alle Grenzwerte werden eingehalten.*

### 7.3.2 Thermische Stabilität und Wärmeverteilung in IT-Räumen

Ein wichtiges Kriterium für die Betriebsstabilität des Rechenzentrums ist die Wärmeverteilung und die Einhaltung der Grenzwerte, wie in Kapitel 4.3 identifiziert und definiert. In diesem Abschnitt erfolgt die Evaluation des DaMucA-Verfahrens hinsichtlich dieser thermischen Parameter. Wie im vorherigen Abschnitt erfolgt auch hier der Vergleich mit dem nicht optimierten und dem FFD-optimierten Betriebszustand. Dieser Abschnitt betrachtet ausschließlich die thermische Situation innerhalb der einzelnen IT-Räume.

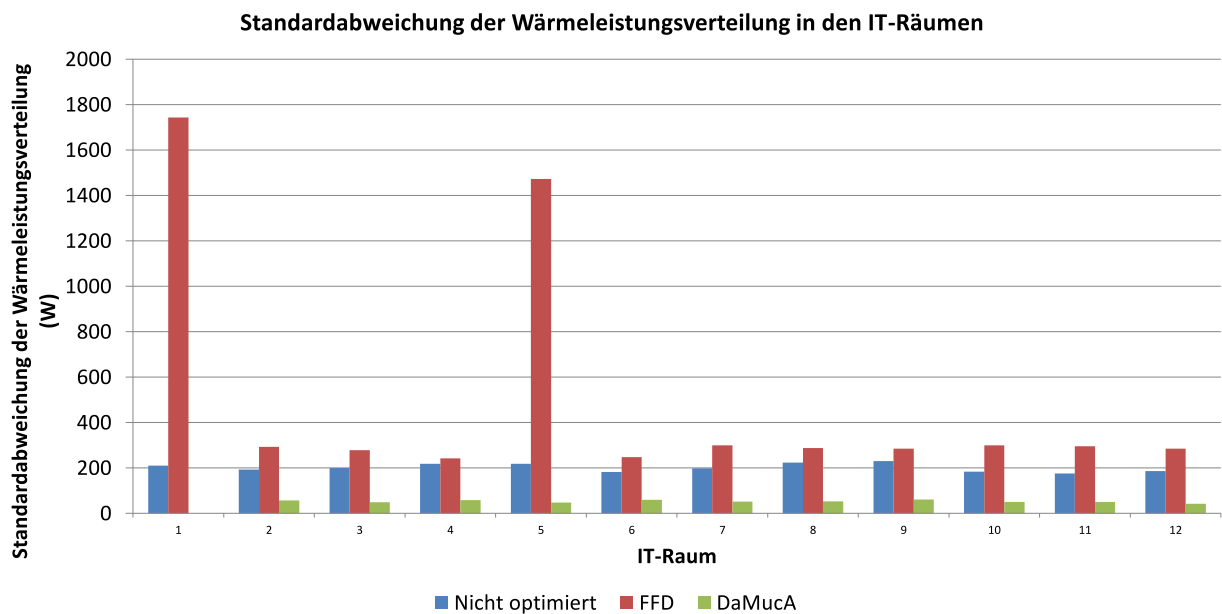


Abbildung 7.14: *Standardabweichung der Wärmeleistungsverteilung in den IT-Räumen nach Gleichung 4.20 aus Kapitel 4.3.2: DaMucA zeigt die geringsten Abweichungen, gefolgt vom nicht optimierten Betriebszustand. FFD hingegen zeigt signifikante Abweichungen, vor allem in den IT-Räumen 1 und 5.*

Zunächst wird in Abbildung 7.14 ein Vergleich der Wärmeleistungsverteilung mittels der Standardabweichung  $\sigma_K$  wie in Kapitel 4.3 beschrieben dargestellt. Die Abbildung bezieht sich auf die Wärmeleistungsverteilung in den einzelnen IT-Räumen und der Bewertung nach Gleichung 4.20. In IT-Räumen 2-4 und 6-12 ist die Verteilung im nicht optimierten und im FFD-optimierten Zustand ähnlich. FFD zeigt jedoch deutliche Ausreißer in den IT-Räumen 1 und 5. Diese IT-Räume waren bereits in den vorherigen Auswertungen aufgefallen. Die effizienten Server mit großer Kapazität in diesen IT-Räumen werden übermäßig ausgelastet, was auch in der thermischen Betrachtung zu signifikanten lokalen Wärmeschwerpunkten führt. Die Auswertung in Abbildung 7.14 auf Basis der Standardabweichung zeigt sofort, dass die IT-Räume 1 und 5 thermische Besonderheiten aufweisen, allerdings ist der Ort in den jeweiligen IT-Räumen hier nicht zu erkennen. Dies wird erst in den nachfolgenden Auswertungen genauer dargestellt.

Der DaMucA-optimierte Betriebszustand hat nach Abbildung 7.14 die geringsten Abweichungen der thermischen Wärmeleistungsverteilung vorzuweisen. Damit wurde auf den ersten Blick das Optimierungsziel der Gleichverteilung aus Kapitel 4.3 erfüllt. In den folgenden Auswertungen wird nun ein genauerer Fokus auf die örtliche Wärmeverteilung in den IT-Räumen und die zu erwartenden Temperaturen gelegt.

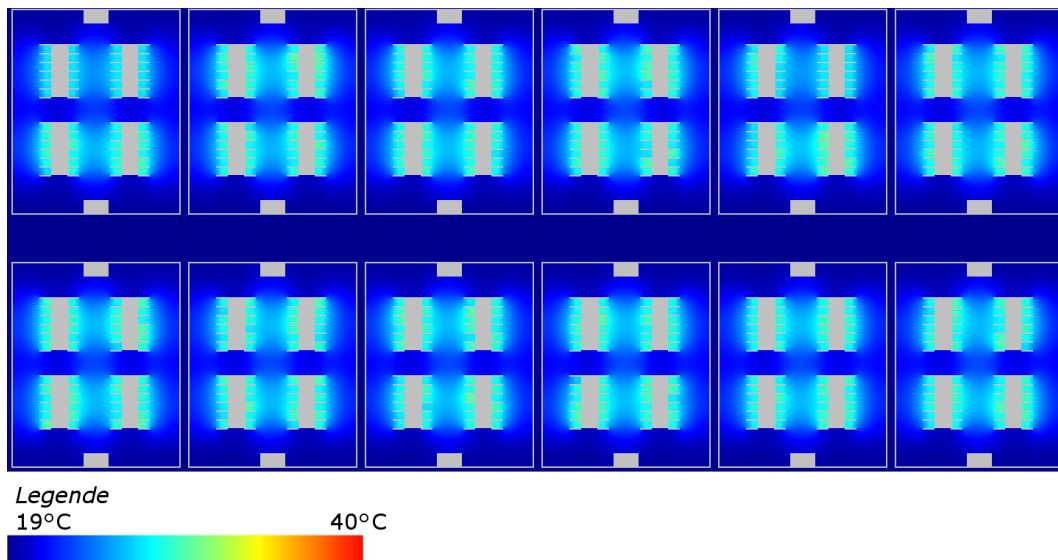


Abbildung 7.15: *Wärmeleistungsverteilung im nicht optimierten Betriebszustand: Alle IT-Räume und Server sind aktiv, es tritt eine relativ gleichmäßige Wärmeleistungsverteilung auf. Geringere Abweichungen entstehen hauptsächlich aufgrund der unterschiedlichen Server-Typen. Aufgrund der Nutzung aller Server entsteht insgesamt eine moderate Wärmeentwicklung, die durch die Klimatisierung abgeführt werden muss. Alle Temperaturgrenzwerte werden aber eingehalten.*

Die Abbildungen 7.15, 7.16 und 7.17 zeigen für diese Betrachtung die simulierten Temperaturen des nicht optimierten, FFD-optimierten und DaMucA-optimierten Betriebszustandes. Dazu wurde die Wärmeleistungsverteilung, die in Abbildung 7.14 mit Hilfe der Standardabweichung bewertet wurde, als Temperaturkarte dargestellt. Die dargestellten Temperaturen beschreiben die Temperatur der warmen Luft, die auf der Rückseite der Racks austritt. Jedes Feld eines Racks beschreibt also die Temperatur, die durch die Wärmeentwicklung aller Server im jeweiligen Rack entsteht. Zur Ermittlung der Temperaturen wurde die Rechenzentrumssimulation verwendet, wobei eine physikalische Simulation der Temperaturerhöhung und Temperaturverteilung in den IT-Räumen vorgenommen wurde. Es sei angemerkt, dass die Abbildungen explizit ausschließlich die Wärmeleistungsverteilung zeigen, Strömungsverhalten und weitere mögliche thermodynamische Effekte sind hier außen vorgelassen worden.

Sowohl im nicht optimierten Fall als auch bei der DaMucA-Lösung wurden alle maximalen

Temperaturgrenzen der Server eingehalten. FFD hingegen konsolidiert die virtuellen Maschinen auf möglichst wenige aktive Server mit Priorisierung effizienter Server, wodurch signifikante lokale Temperaturüberschreitungen bei diesen Servern auftreten. In Abbildung 7.16 kann nun im Detail nachvollzogen werden, was Abbildung 7.14 bereits angedeutet hat: In den IT-Räumen 1 und 5 tritt eine Konzentration der Auslastung auf die Rack-Reihen, die diese Server beinhalten, auf, wodurch hier eine übermäßige Wärmebelastung entsteht. Die geforderten Temperaturgrenzen können hier nicht eingehalten werden, die Gefahr von Hot-Spots steigt hier signifikant an. Die restlichen IT-Räume dagegen beinhalten nur geringe Last und bleiben dementsprechend kühl.

In Abbildung 7.17 kann erkannt werden, dass das DaMucA-Verfahren sowohl die Temperaturgrenzen eingehalten, als auch eine Gleichverteilung der Wärmeleistung in den IT-Räumen und über die IT-Räume hinweg erreicht hat.

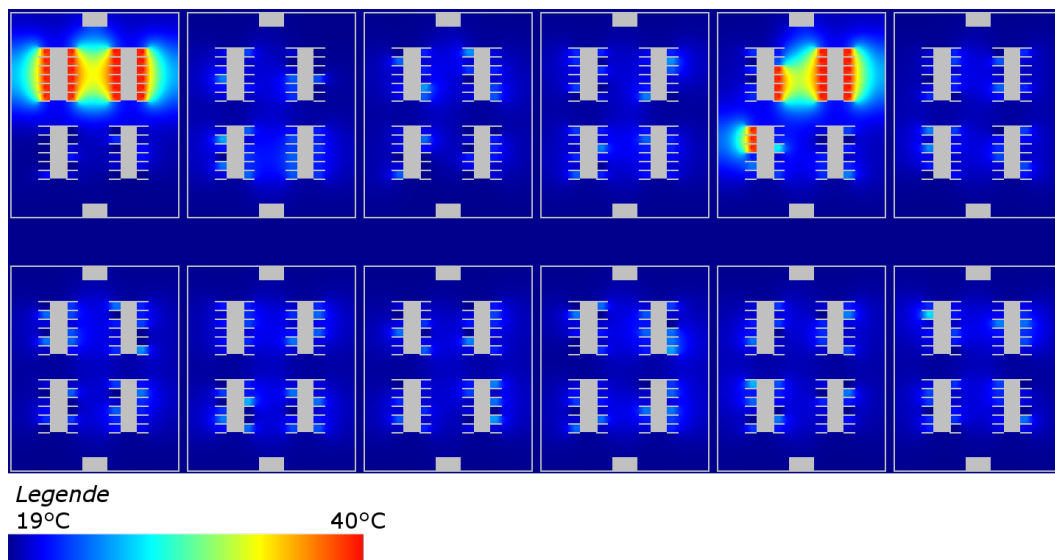


Abbildung 7.16: *Wärmeleistungsverteilung im FFD-optimierten Betriebszustand: Die Wahl der genutzten Server ist unter dieser Allokation vollkommen unabhängig von der Verortung, was deutlich in der Wärmeleistungsverteilung zu erkennen ist. In den IT-Räumen 1 und 5 gibt es extreme Konzentrationen in einzelnen Racks. Die vorgeschriebenen Grenzwerte können hier nicht eingehalten werden. Die weiteren IT-Räume hingegen zeigen eine eher niedrige Wärmeleistung.*

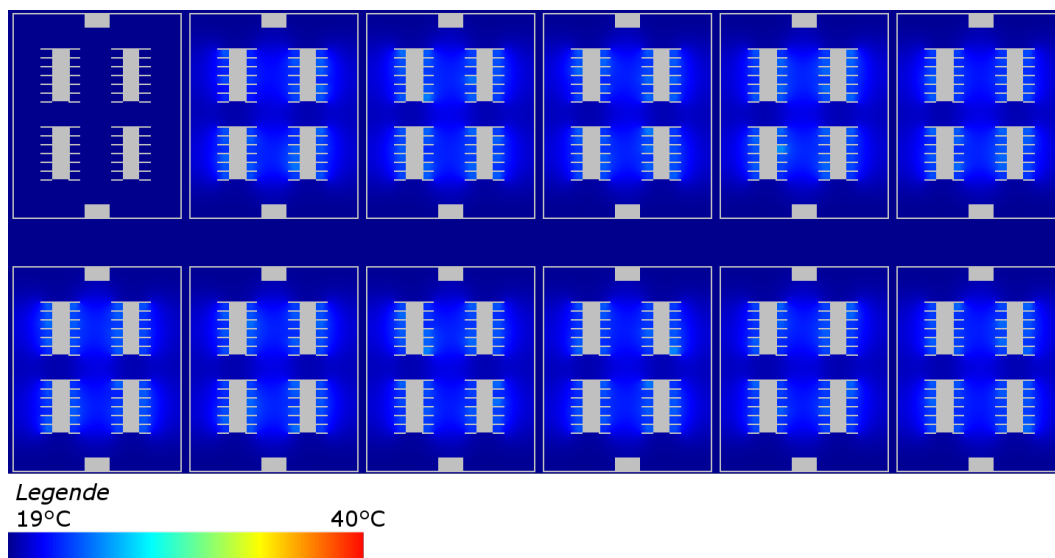


Abbildung 7.17: *Wärmeleistungsverteilung im DaMucA-optimierten Betriebszustand: Hier werden nur die IT-Räume 2-12 verwendet. In diesen IT-Räumen zeigt sich eine gleichmäßige Wärmeleistungsverteilung, es sind nur geringe Abweichungen innerhalb der IT-Räume zu erkennen. Insgesamt ist die Wärmeleistung niedrig, was im Vergleich zum nicht optimierten Betriebszustand vor allem durch die Reduktion der Anzahl aktiver Server und deren effizienter Nutzung erreicht wird. In allen IT-Räumen können alle Grenzwerte eingehalten werden.*

### 7.3.3 Evakuierungszeit

Die Evakuierungszeit jedes Servers wird, wie in Kapitel 4.4.3 definiert, als fester einzuhaltender Grenzwert betrachtet. Die tatsächlich benötigte Zeit für die Evakuierung eines Servers wird ausschließlich durch die Migrationsdauer der virtuellen Maschinen auf dem Server bestimmt. In Tabelle 7.4 wird die Evakuierungszeit der drei untersuchten Betriebszustände nicht optimiert, FFD-optimiert und DaMucA-optimiert dargestellt.

Tabelle 7.4: Evakuierungszeit der unterschiedlichen Betriebszustände

| Wert                        | Nicht optimiert | FFD   | DaMucA       |
|-----------------------------|-----------------|-------|--------------|
| Minimum der Zeit (Sekunden) | 8,5             | 19,8  | <b>70,5</b>  |
| Maximum der Zeit (Sekunden) | 30,8            | 132,3 | <b>118,8</b> |
| Mittelwert (Sekunden)       | 22,5            | 122,4 | <b>109,8</b> |

Die Tabelle zeigt jeweils das Minimum, Maximum und den Mittelwert der Evakuierungszeit aller aktiven Server unter dem jeweiligen Betriebszustand. Der nicht optimierte und der FFD-optimierte Betriebszustand beachten keine Begrenzung der Evakuierungszeit. Für das DaMucA-Verfahren wurde eine Begrenzung auf 120 Sekunden<sup>1</sup> festgelegt. Es ist zu erkennen, dass DaMucA diese Werte einhalten kann. Allerdings ist die Evakuierungszeit im nicht optimierten Betrieb signifikant geringer als in den beiden optimierten Betriebszuständen. An dieser Stelle muss darauf hingewiesen werden, dass eine energetische Optimierung mittels Konsolidierungsstrategien diesen Nachteil nicht eliminieren kann. Die geringe Evakuierungszeit im nicht optimierten Zustand wird deshalb erreicht, weil die Server nur geringfügig ausgelastet sind. Das DaMucA-Verfahren bietet hier aber die Begrenzung auf einen festen Wert an, sodass eine maximale Zeit garantiert werden kann.

Zur genaueren Betrachtung dieses Sachverhalts wird die Veränderung der elektrischen Leistungsaufnahme des Rechenzentrums in Abhängigkeit der maximal erlaubten Evakuierungszeit in Abbildung 7.18 dargestellt. Die gezeigte Auswertung bezieht sich ausschließlich auf das DaMucA-Verfahren. Bei geringer werdender maximaler Evakuierungszeit eines jeden Servers steigt der Bedarf zusätzliche Server und die mittlere elektrische Leistungsaufnahme des Rechenzentrums signifikant an.

### 7.3.4 Energieeinsparung gegenüber Flexibilität

Die globale Flexibilität steht, wie auch in Kapitel 4 identifiziert, der Optimierung des Energiebedarfs des Rechenzentrums direkt entgegen. Wird eine höhere Flexibilität gefordert, bedeutet dies auch eine Steigerung der elektrischen Leistung des Rechenzentrums, da dies zur Nutzung zusätzlicher Server-Ressourcen und gleichzeitig zu einer ineffizienteren

<sup>1</sup>In Kapitel 4.3.1 wurde von erlaubten Evakuierungszeiten von 200-220 Sekunden berichtet. Diese sind jedoch Maximalwerte, daher werden hier etwas niedrigere Werte angenommen, die zudem einen Einfluss auf die Lösungsfindung haben.

Auslastung dieser führt. Denn bei steigender Flexibilität werden zusätzliche freie Server-Ressourcen benötigt, die Server arbeiten dann also in einem niedrigeren Auslastungsbereich und damit meist auch Wirkungsgrad.

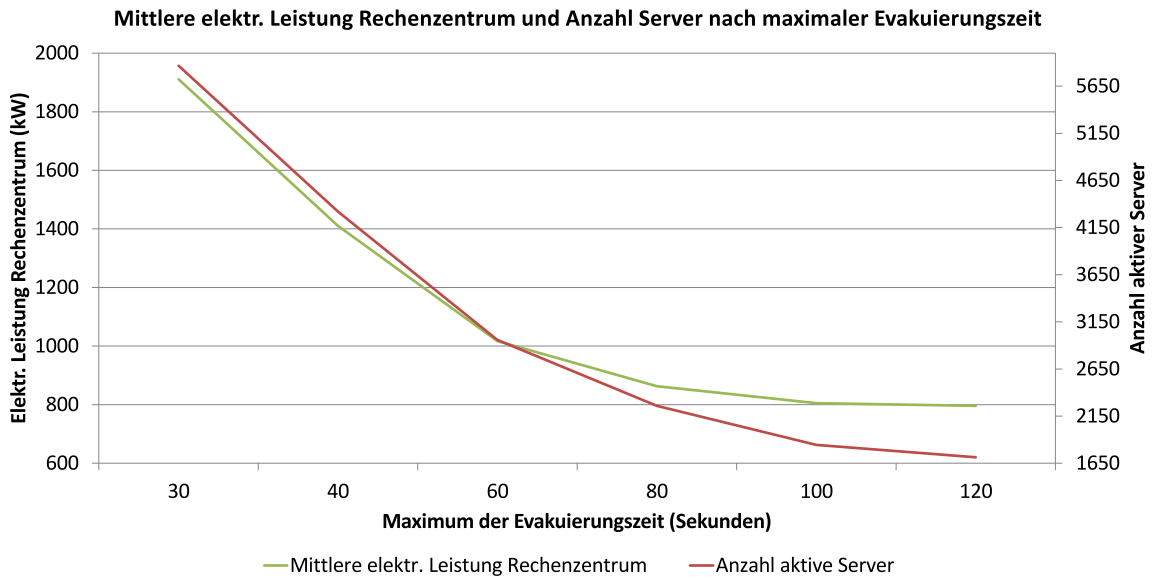


Abbildung 7.18: *Abhängig vom festgelegten Maximum der Evakuierungszeit verändert sich die Anzahl der notwendigen Server und die mittlere elektrische Leistung des Rechenzentrums signifikant.*

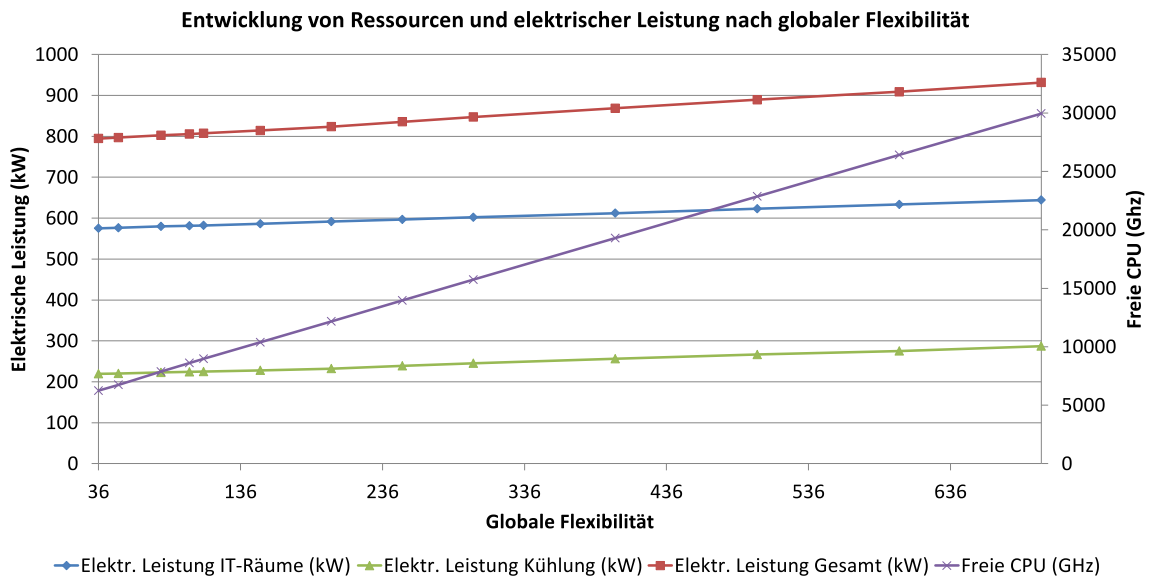


Abbildung 7.19: *Mit steigender globaler Flexibilität werden zusätzliche freie Server-Ressourcen benötigt. Diese zusätzlichen Ressourcen sorgen für einen Anstieg der mittleren elektrischen Leistung der Server und der Kühlgeräte.*



Dieser Zusammenhang wird in Abbildung 7.19 mittels mehrerer Kenngrößen dargestellt. In Abhängigkeit der geforderten globalen Flexibilität  $f_g$  werden die freien Ressourcen (CPU, violette Kurve) und die mittlere elektrische Leistung des Rechenzentrums gezeigt. Allerdings führt das DaMucA-Verfahren dazu, dass die zusätzlichen freien Ressourcen in einem höheren Maße ansteigen als die dazu benötigte zusätzliche elektrische Leistung. Der prozentuale Vergleich in Abbildung 7.20 bestätigt dies. Werden beispielsweise bei einer globalen

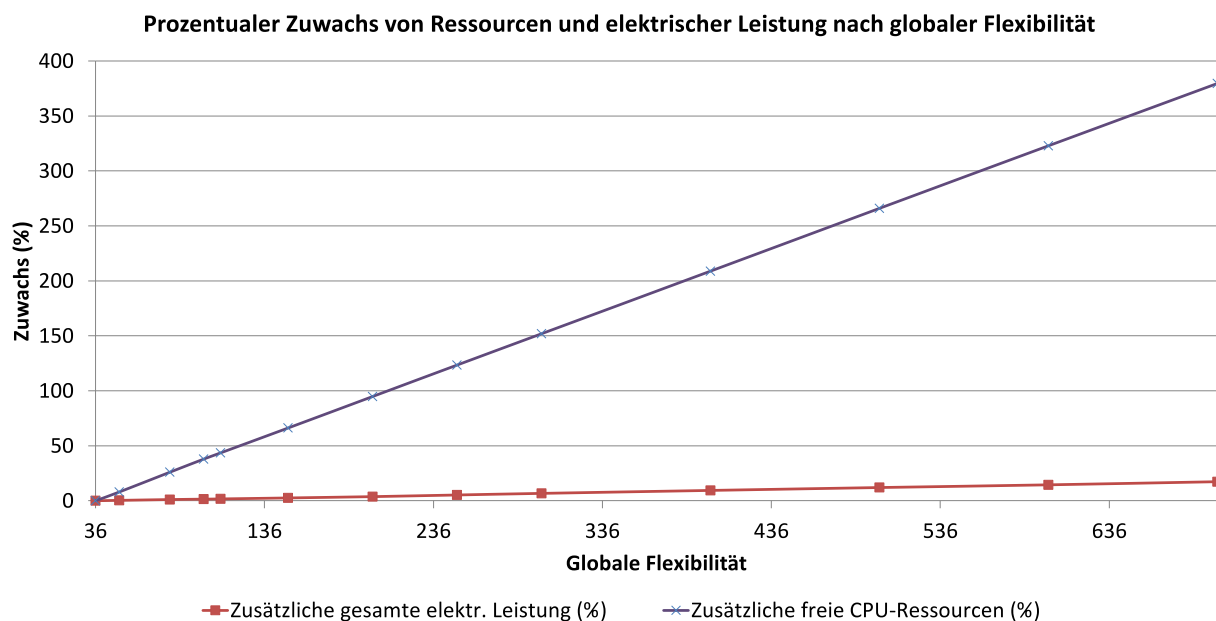


Abbildung 7.20: *Mit steigender globaler Flexibilität ist eine signifikante Erhöhung der freien Server-Ressourcen erreichbar. Diese werden im gezeigten Beispiel mehr als verdreifacht. Im Gegensatz dazu steigt die dafür benötigte mittlere elektrische Leistung des Rechenzentrums nur moderat an. Das DaMucA-Verfahren sorgt also mit der globalen Flexibilität für eine Erhöhung der verfügbaren Server-Ressourcen, sucht aber jeweils einen energetisch optimierten Betriebszustand.*

Flexibilität von 300 ca. 150% mehr freie Ressourcen als im primär energetisch optimierten Zustand<sup>2</sup> vorgehalten, steigt die elektrische Leistung lediglich um ca. 10%. Damit kann auch in einem energetisch optimierten Betriebszustand eine hohe Flexibilität erreicht werden. Auch ist eine energetische Regelbarkeit anhand der globalen Flexibilität möglich, sodass ein Rechenzentrumsbetreiber hier bewusst zwischen energetischer Einsparung und flexiblem Betriebszustand abwägen kann.

<sup>2</sup>Die minimale globale Flexibilität  $f_g$  beginnt in der Abbildung bei 36. Dies liegt daran, dass das DaMucA-Verfahren bei der energetischen Optimierung den effizientesten Betriebspunkt der Server nutzt. Da dieser nicht immer bei 100% Auslastung liegt, werden einige Server niedriger ausgelastet, was zu freien Ressourcen führt. Bei Berechnung der globalen Flexibilität werden auch diese Ressourcen berücksichtigt.

### 7.3.5 Energieeinsparung bei der Kühlung

Energieeinsparungen bei der Kühlkette und damit bei den Kältemaschinen werden, wie in Kapitel 4.2.2 beschrieben, durch Anheben der Kaltwassertemperatur erreicht, was durch Absenken des Maximums der Leistungsaufnahme aller IT-Räume erreicht wird. Der Zusammenhang zwischen der Kaltwassertemperatur des Kühlwassers und dem Maximum der Leistungsaufnahme der IT-Räume wurde für die Lösung nach dem DaMucA-Verfahren in Abbildung 7.2 bereits dargestellt. In den weiteren Auswertungen wurde zudem belegt, dass das DaMucA-Verfahren die geforderten Grenzwerte und die thermische Stabilität einhalten kann, sodass die energetischen Optimierungen der Kühlung auch tatsächlich im stabilen Betrieb vorgenommen werden können. Tabelle 7.5 zeigt nun den Vergleich der für die Optimierung der Kühlung relevanten Parameter.

Tabelle 7.5: Optimierung der Kühlleistung im Vergleich

| Wert  | Nicht optimiert | FFD    | DaMucA        |
|---|-----------------|--------|---------------|
| Standardabweichung der mittleren elektr. Leistungsverteilung (kW) | 3,986           | 28,166 | <b>0,783</b>  |
| Variationskoeffizient der mittleren elektr. Leistungsverteilung   | 0,0107          | 0,1769 | <b>0,005</b>  |
| Maximum (kW)  | 116,86          | 114,86 | <b>53,33</b>  |
| Notwendige Kaltwassertemperatur (°C)                              | 8,91            | 9,18   | <b>17,66</b>  |
| Mittlere Elektr. Leistung Kältemaschinen (kW)                     | 1341,86         | 554,47 | <b>224,28</b> |

Zunächst ist zu erkennen, dass die Standardabweichung (und zusätzlich die daraus normalisierte Verteilung) der elektrischen Leistungsverteilung und damit auch der Wärmeleistungsverteilung bei der DaMucA-Lösung deutlich geringer ist, als beim nicht optimierten und FFD-optimierten Betrieb. Dies geht einher mit einer Absenkung des Maximums der elektrischen Leistung der IT-Räume auf 53,33 kW, was zu einer notwendigen Kaltwassertemperatur von 17,66°C führt. Kombiniert mit der zusätzlichen Optimierung der Leistungsaufnahme der Server führt dies zu einer Reduktion der mittleren elektrischen Leistung der Kühlung auf 224,28 kW.

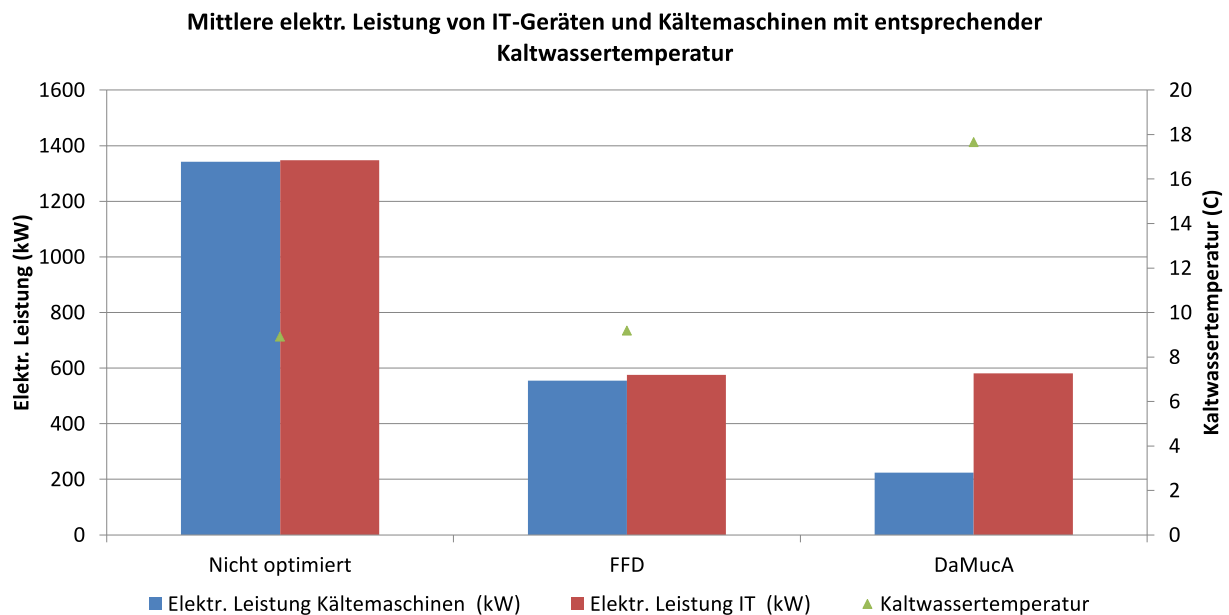


Abbildung 7.21: Zusammenfassender Vergleich der Einsparungen: Der nicht optimierte Betrieb und FFD benötigen ähnliche Kaltwassertemperaturen. FFD erzielt die Einsparungen sowohl bei Servern als auch bei der Kühlung durch die signifikante Reduzierung der Anzahl der aktiven Server. DaMucA erzielt im Vergleich zu FFD keine weiteren Einsparungen bei den Servern, sorgt aber für die konsequente Umsetzung der identifizierten Kriterien für einen stabilen Betriebszustand und ermöglicht außerdem die Anhebung der Kaltwassertemperatur. Dadurch werden die zusätzlichen Einsparungen im Vergleich zu FFD erreicht. Standardabweichungen der elektrischen Leistungsaufnahmen: nicht optimiert: 13,8kW, 13,97kW; FFD: 9,54kW, 10,96kW; DaMucA: 7,62kW, 8,85kW

In Abbildung 7.21 wird dieser Effekt zusätzlich zur mittleren elektrischen Leistung der IT-Geräte dargestellt. Hier ist deutlich zu erkennen, dass FFD und DaMucA eine fast gleiche elektrische Leistung der IT-Geräte, also insbesondere der Server erreichen. Durch die Konsolidierung kann diese Leistung mehr als halbiert werden. Zusätzliche Einsparungen erreicht DaMucA im Vergleich zu FFD hingegen ausschließlich durch die Optimierung der Kühlung.

Dieser Effekt kann ebenfalls mit der Power Usage Effectiveness-Metrik (PUE) [13] beschrieben werden, die aktuell als Quasi-Standard dazu verwendet wird, die Effizienz von Rechenzentren zu bewerten und untereinander zu vergleichen [116]. Im nicht optimierten Betrieb beträgt dieser ca. 2, was als ineffizient bewertet wird. Die DaMucA-Optimierung hingegen führt zu einem PUE von ca. 1,3, was eine signifikante Verbesserung darstellt.

## 7.4 Gesamtenergieeinsparung

In diesem Abschnitt wird nun abschließend die gesamte Energieeinsparung des Rechenzentrums durch die DaMucA-Optimierung betrachtet. In Abbildung 7.22 wird die für ein Jahr simulierte Ersparnis des DaMucA-Verfahrens im Vergleich zum nicht optimierten und FFD-optimierten Betrieb dargestellt. Die Ergebnisse aus Abbildung 7.21 zeigen sich auch hier. Wird nun die prozentuale Einsparung in Abbildung 7.23 hinzugezogen, so zeigt sich

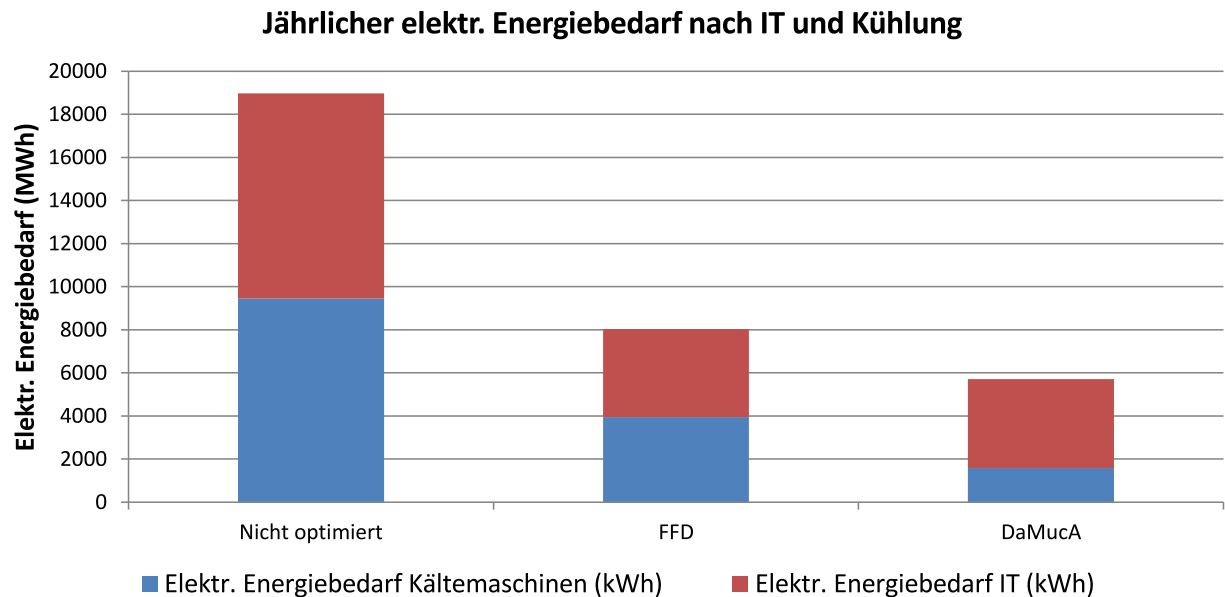


Abbildung 7.22: Die in Abbildung 7.21 beschriebene Analyse zeigt sich auch im jährlichen Vergleich. Hier erreicht das DaMucA-Verfahren eine signifikante Senkung des Energiebedarfs des gesamten Rechenzentrums.

ein energetisches Einsparpotential von bis zu 70% gegenüber dem nicht optimierten Betriebszustand, das durch das DaMucA-Verfahren realisiert werden kann.

## 7.5 Einfluss des Szenarios auf Leistungsfähigkeit des Verfahrens

Die Leistungsfähigkeit des DaMucA-Verfahrens wird zusätzlich zu den bereits untersuchten Parametern durch mehrere Parameter des Szenarios, also des Rechenzentrums und der anfallenden Anwendungslast, bestimmt. Je nach Aufbau des Rechenzentrums und Verteilung der Server in den IT-Räumen können Situationen auftreten, bei denen das DaMucA-Verfahren kaum noch Optimierungen vornehmen kann. Die dafür primär verantwortlichen Parameter werden hier kurz aufgelistet und es wird das jeweilige Problem erklärt.

- *Anzahl verfügbarer IT-Räume:* Wenn ein Rechenzentrum nur wenige IT-Räume besitzt und diese (fast) alle verwendet werden müssen, damit überhaupt die Anwendun-

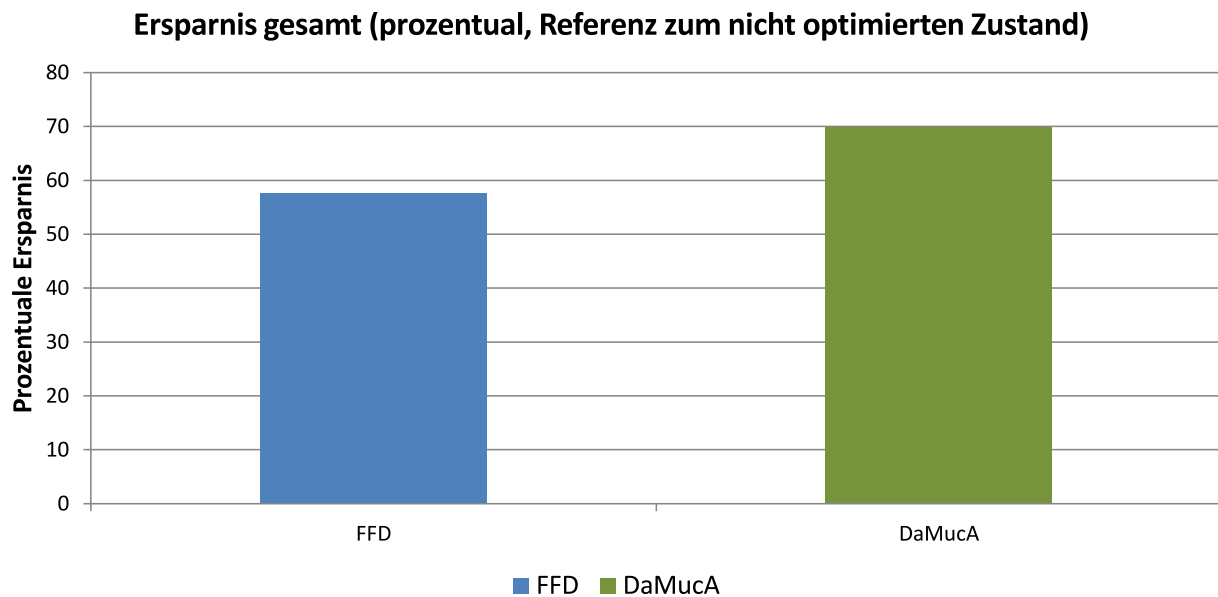


Abbildung 7.23: Die Aussagen aus Abbildung 7.22 werden hier als prozentuale Ersparnis dargestellt. Im Vergleich zum nicht optimierten Betriebszustand erreicht das DaMucA-Verfahren bis zu 70% Energieeinsparung. Der State of the Art in Form der FFD-Optimierung liegt hier bei knapp 60% Einsparung. Im Vergleich zu FFD kann DaMucA ca. 28% zusätzliche Einsparungen erreichen.

gen auf den Servern stabil ausgeführt werden können, ist die Kombinatorik der aktiven IT-Räume kaum noch relevant für die Lösungsfindung. In diesem Fall kann die DaMucA-Metaheuristik übersprungen werden. Die DaMucA-Server-Heuristik wird dann nur zur Verteilung der Auslastung innerhalb der IT-Räume verwendet. In diesem Fall hängt die mögliche Optimierung der Kühlung allein von der tatsächlich anfallenden Anwendungslast ab. Eine gesonderte Optimierung durch das DaMucA-Verfahren ist dann nicht möglich.

- *Verteilung der Server in IT-Räumen:* Wenn die Server in mindestens einem IT-Raum, der genutzt werden muss, so verteilt sind, dass sie sehr eng beieinander stehen und weite Teile des IT-Raumes leer stehen, kann das Verfahren hier kaum Optimierungen vornehmen. Wenn die Anwendungsauslastung es erlaubt, viele Server auszuschalten, kann dies zwar helfen, eine thermisch stabile Optimierung in diesem IT-Raum vorzunehmen. Aber je nach Anzahl benötigter Server kann eine solche Situation in einem IT-Raum zu einer insgesamt nicht optimierbaren Lage führen. Denn in einem solchen Fall muss die Kälteleistung angehoben werden, damit die thermische Stabilität in diesem IT-Raum gewährleistet werden kann. Dies führt dann wiederum zur Erhöhung der Kälteleistung in allen IT-Räumen, auch wenn diese dort nicht benötigt wird.
- *Streuung der Server-Leistung:* Die Streuung der Server-Leistungen hingegen hat kaum

Auswirkungen auf das DaMucA-Verfahren, da hier jeweils die Abwägung zwischen den passendsten Servern vorgenommen wird. Ineffizientere und schwächere Server werden damit nur bei absoluter Notwendigkeit verwendet. Allerdings gilt hier, dass die Streuung wiederum in Kombination mit der räumlichen Platzierung ein wichtiger Aspekt ist. Unterschiedlich effiziente Server sollten nebeneinander in den IT-Räumen platziert werden, damit das Verfahren eine möglichst große Auswahl der Kombination aktiver Server hat.

- *Schwankungen der Anwendungsauslastung:* Das DaMucA-Verfahren schöpft, wie in der Evaluation dargelegt, ein großes Optimierungspotential aus der Anpassung der Kälteleistung. Dabei müssen jedoch immer die beschriebenen zeitlichen Bedingungen der Kaltwassertemperaturregelung beachtet werden. Die Evaluation mit realen Anwendungen hat gezeigt, dass dies möglich ist und die beschriebenen Optimierungen erreichbar sind. Bei speziellen Anwendungen kann es allerdings zu signifikanten Leistungsspitzen kommen, die mehrere Server eines einzelnen IT-Raumes betreffen und das Verfahren zwingen, die Kühlung dementsprechend anzupassen. Je nach Periode der Spitzen ist damit eine Optimierung der Kühlung zumindest unter den aktuellen Möglichkeiten der Klimaregelung kaum noch möglich.

Insgesamt kann das DaMucA-Verfahren auch bei Rechenzentren, deren Infrastruktur „gewachsen“ ist und dementsprechend eine breite Streuung der Server-Typen aufweist, die beschriebenen Optimierungen erreichen. Die oben dargelegten Aspekte stellen vor allem Sonderfälle dar.

## 7.6 Zusammenfassung

Dieses Kapitel behandelte die Evaluation des DaMucA-Verfahrens. Es wurde jeweils zwischen der Optimierung durch die DaMucA-Metaheuristik und die DaMucA-Server-Heuristik unterschieden und die jeweiligen Einflüsse dargestellt. Insbesondere wurde gezeigt, dass das DaMucA-Verfahren die geforderten Grenzwerte und identifizierten Kriterien einhalten konnte, während die in der Forschung aktuell populäre FFD-Optimierung dies nicht berücksichtigt und dementsprechend zu ungültigen Lösungen führt. Trotzdem wurde der weitere Vergleich mit FFD vorgenommen, da dieses Verfahren den aktuellen Stand der Forschung repräsentiert.

Das DaMucA-Verfahren konnte zudem durch die Gleichverteilung der Wärmeleistungen der IT-Räume und dem gleichzeitigen Absenken des Maximums eine signifikante Anhebung der Kaltwassertemperatur erreichen. Dies führt zu signifikanten energetischen Einsparungen bei der Kühlung, was in Kombination mit der angewandten Server-Konsolidierung zu deutlichen energetischen Einsparungen beim Betrieb des Rechenzentrums führt. Hier ist ein Einsparungspotential von 70% im Vergleich zum nicht optimierten Betrieb möglich, während FFD hier nur knapp 60% erreicht - wohlgermerkt ohne die weiteren Kriterien zu betrachten.

Zusätzlich kann DaMucA die geforderte Flexibilität sicherstellen, wobei eine Erhöhung der globalen Flexibilität nur in geringerem Maße zu einem Ansteigen der benötigten elektrischen Leistung führt.





## 8 Zusammenfassung und Ausblick

Die vorliegende Dissertation behandelte die energetische Analyse des Rechenzentrums und eine darauf aufbauende Optimierung, die durch zusätzliche Kriterien erweitert wurde. Der Bedarf einer energetischen Optimierung wurde mit dem steigenden Energiebedarf von Rechenzentren weltweit begründet. Gleichzeitig ist durch das steigende Bewusstsein der Rechenzentrumsbetreiber für einen energieeffizienten Betrieb die realistische Anwendungsmöglichkeit gegeben. Damit allerdings Stabilitätsbedenken der Betreiber entgegnet werden kann, wurden von vorne herein Kriterien zur Sicherstellung der Betriebsstabilität und der Flexibilität mit in die Betrachtung einbezogen.

Damit eine ganzheitliche energetische Optimierung des Rechenzentrums vorgenommen werden konnte, wurden zunächst Freiheitsgrade einzelner Geräte, energetische Einflüsse und Wechselwirkungen zwischen verschiedenen Geräteklassen im Rechenzentrum untersucht. Als maßgebliche Einflussfaktoren wurde die Verteilung der virtuellen Maschinen auf die Server und die Leistungsverteilung auf die IT-Räume identifiziert. Aufgrund einer zentralisierten Konzeption der USV-Geräte ist hier eine direkte Regelbarkeit nicht möglich. Stattdessen sollten diese Geräte bereits bei der Planung des Rechenzentrums insofern berücksichtigt werden, dass diese möglichst modular und flexibel betrieben werden können. Dadurch wird bei schwankender elektrischer Leistung automatisch ein möglichst effizienter Betriebszustand eingenommen.

Für die eigentliche Optimierung wurden drei generelle Kriterien identifiziert: Optimierung der elektrischen Leistungsaufnahme, Wahrung der thermischen Stabilität und Sicherstellen einer definierten Flexibilität. Diese Kriterien wurden jeweils in weitere Unter Aspekte unterteilt, die direkt durch das Optimierungsverfahren DaMucA umgesetzt werden sollten. Als besonderer Einfluss auf die elektrische Leistungsaufnahme des Rechenzentrums wurde die Verteilung der IT-Auslastung und damit der elektrischen Leistung auf die IT-Räume herausgearbeitet, da durch diese die Kaltwassertemperatur der Kühlung angepasst werden kann. Durch Ausnutzung dieses Freiheitsgrades konnte das DaMucA-Verfahren signifikante Energieeinsparungen bei der Kühlung erreichen, die die Einsparungen von aktuell in der Forschung publizierten Verfahren wie FFD übertreffen. So kann das DaMucA-Verfahren insgesamt Einsparungen von 70% im Vergleich zum nicht optimierten Betrieb erreichen. Allerdings bezieht sich die DaMucA-Optimierung aufgrund der Trägheit der Regelbarkeit des Kühlsystems ausschließlich auf längere Zeiträume.

Durch die konsequente Einbeziehung der thermischen Situation im IT-Raum konnte das DaMucA-Verfahren zudem die Einhaltung der Temperaturgrenzwerte sicherstellen und zu einer gleichmäßigen Verteilung der Wärmeleistung beitragen. Dies senkt die Wahrscheinlichkeit des Auftretens von Hot-Spots.

Zusätzlich zu den energetischen Optimierungen wurde mit der Definition der globalen

Flexibilität eine Metrik geschaffen, die auf die Vorhaltung von Reserve-Ressourcen abzielt und im Notfall zu schnellen Evakuierungen betroffener Server führen kann. Auch wurde in der Evaluation gezeigt, dass auch bei einer Steigerung der globalen Flexibilität immer noch signifikante Energieeinsparungen möglich sind.

### Zukünftige Möglichkeiten der Vertiefung

Mehrere Bereiche des Optimierungsverfahrens und der Modellierung können noch erweitert oder mit weiteren Ergebnissen existierender Forschung kombiniert werden. Diese werden hier kurz aufgeführt.

- *Definition der globalen Flexibilität:* Die Definition der globalen Flexibilität und die Bewertung und Abschätzung dieser beinhalten ein enormes Potential. Durch diese kann eine Abwägung der Nutzung von Server-Konsolidierung nach definierten Maßstäben erfolgen. Der in dieser Arbeit vorgestellte heuristische Ansatz zur Abschätzung der tatsächlichen globalen Flexibilität ist pessimistisch. Allerdings ist eine genauere Ermittlung aufwendig und beinhaltet mehrere weitere kombinatorische Probleme, die in dieser Arbeit nicht mehr aufgegriffen wurden.
- *Thermale Flexibilitätmetrik:* Die Flexibilitätmetrik kann um thermale Zonen erweitert werden. Dafür ist allerdings eine CFD-Simulation der IT-Räume erforderlich. Die Metrik wird dann so erweitert, dass die Auswirkungen des Ausfalls eines Umluftkühlgerätes auf die einzelnen Server-Standorte analysiert und einbezogen werden. Dies hat einerseits eine zeitliche Dimension, es kann also analysiert werden, welche Server als erste evakuiert werden müssten. Andererseits können die Server in thermale Zonen eingeteilt werden, wobei die Vorhaltung von Reserveservern jeweils redundant in verschiedenen thermalen Zonen erfolgen muss. Die eigentliche Metrik muss dafür unter Umständen nicht geändert werden, sondern es kann dann lediglich die Anzahl der gleichzeitig zu evakuierenden und auszuschaltenden Server an die Anzahl der Server in den thermischen Zonen im Rechenzentrum angepasst werden.
- *Kurzzeit- und Langzeitplanung:* Das DaMucA-Verfahren konzentriert sich auf eine Langzeitplanung, was vor allem aufgrund der Trägheit der Regelung der Klimatisierung erfolgt. Hier wurde gezeigt, dass wesentliche Energieeinsparungen erreicht werden können. Grundsätzlich kann die Kombination des DaMucA-Verfahrens mit einem existierenden Kurzzeitplanungsverfahren erfolgen, wodurch je nach Lasttyp und Rechenzentrum weitere Einsparungen möglich sind. Allerdings muss dieses Kurzzeitplanungsverfahren die Kriterien der DaMucA-Heuristik ebenfalls unterstützen, existierende Verfahren müssten also teils deutlich erweitert werden und vor allem so abgeändert werden, dass sie zur Verteilung der Last die Heuristiken des DaMucA-Verfahrens verwenden. Insbesondere die räumliche Verteilung der IT-Auslastung ist bei bisherigen Kurzzeitplanungsverfahren kaum thematisiert worden, sodass hier ein relevanter zusätzlicher Forschungsbedarf erscheint.

- 
- *Dynamische Klimaregelung:* Wenn mehr bewährte Erfahrungen aus der Praxis hinsichtlich der Regelung der Kaltwassertemperatur der Klimatisierung vorliegen, sollte die dynamischere und damit häufigere Anpassung der Kaltwassertemperatur überprüft und gegebenenfalls vorgenommen werden. Hier können sich je nach Auslastungsprofilen und den darin auftretenden Schwankungen zusätzliche Einsparpotentiale ergeben. Durch den modularen Aufbau des DaMucA-Verfahrens ist eine häufigere Optimierung effizient umsetzbar. Beispielsweise kann je nach Beschaffenheit des Rechenzentrums und der Art der Anwendungsauslastung auf eine regelmäßige Suche nach der Kombinatorik der aktiven IT-Räume verzichtet werden (wie bereits kurz in Kapitel 7.2.1 beschrieben). Die Optimierung wird in diesem Fall nur durch die DaMucA-Server-Heuristik vorgenommen.
  - *Automatische Temperaturmodelle der Server:* Die Server-Temperaturen sollten im Rechenzentrum generell permanent analysiert und überwacht werden. Aus diesen Daten können Modelle abgeleitet werden, die Abweichungen oder Anomalien frühzeitig erkennen und so dem DaMucA-Verfahren frühzeitig eine notwendige Umverteilung der Auslastung mitteilen. Über einen längeren Zeitraum kann dann eine lernende Komponente entwickelt werden, die thermale Zonen oder einzelne Server, die zu Überhitzungen neigen, grundsätzlich niedriger priorisiert.



# Literaturverzeichnis

- [1] Z. Abbasi, G. Varsamopoulos, and S. K. S. Gupta. Thermal aware server provisioning and workload distribution for internet data centers. In *Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing, HPDC '10*, pages 130–141, New York, NY, USA, 2010. ACM.
- [2] A. Adelin, P. Owezarski, and T. Gayraud. On the impact of monitoring router energy consumption for greening the internet. In *Grid Computing (GRID), 2010 11th IEEE/ACM International Conference on*, pages 298–304, 2010.
- [3] D. Aikema, R. Simmonds, and H. Zareipour. Data centres in the ancillary services market. In *Green Computing Conference (IGCC), 2012 International*, pages 1–10, June 2012.
- [4] A. Al-Qawasmeh, S. Pasricha, A. Maciejewski, and H. Siegel. Thermal-aware performance optimization in power constrained heterogenous data centers. In *Parallel and Distributed Processing Symposium Workshops PhD Forum (IPDPSW), 2012 IEEE 26th International*, pages 27–40, May 2012.
- [5] M. Ammann, C. Briceno, K. Connell, and S. Rungta. Upgrading data center network architecture to 10 gigabit ethernet. *Intel White Paper, IT Best Practices - Data Centers*, 2011.
- [6] Apachefoundation. Apache HTTP Project. <https://httpd.apache.org/>, 2016.
- [7] Apachefoundation. Apache JMeter. <http://jmeter.apache.org/>, 2016.
- [8] APC. Understanding delta conversion online "the difference". [http://www.apcmedia.com/salestools/TDOY-5UQVDS/TDOY-5UQVDS\\_R1\\_EN.pdf](http://www.apcmedia.com/salestools/TDOY-5UQVDS/TDOY-5UQVDS_R1_EN.pdf), 2004.
- [9] ASHRAE. 2011 thermal guidelines for data processing environments – expanded data center classes and usage guidance. *Whitepaper prepared by ASHRAE Technical Committee (TC) 9.9, Mission Critical Facilities, Technology Spaces, and Electronic Equipment*, 2011.
- [10] N. Bansal, A. Lodi, and M. Sviridenko. A tale of two dimensional bin packing. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pages 657–666, Oct 2005.
- [11] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield. Xen and the art of virtualization. In *Proceedings of the nineteenth ACM symposium on Operating systems principles*, 2003.
- [12] L. A. Barroso and U. Hölzle. The case for energy-proportional computing. *IEEE Computer*, 40, 2007.
- [13] C. Belady, A. Rawson, J. Pflueger, and T. Cader. Green grid data center power efficiency metrics: Pue and dcie. *the Green Grid white paper*, 2008.
- [14] T. L. Bergman, A. S. Lavine, F. P. Incropera, and D. P. DeWitt. Fundamentals of heat and mass transfer. *7th Edition*, 2006.

- [15] A. Berl, S. Klingert, M. Beck, and H. de Meer. Integrating data centres into demand-response management: A local case study. In *Industrial Electronics Society, IECON 2013 - 39th Annual Conference of the IEEE*, pages 4762–4767, Nov 2013.
- [16] BITKOM. Energieeffizienz im rechenzentrum - ein leitfaden zur planung, zur modernisierung und zum betrieb von rechenzentren, band 2. *Schriftenreihe Umwelt und Energie*, 2010.
- [17] N. Bobroff, A. Kochut, and K. Beaty. Dynamic placement of virtual machines for managing sla violations. In *2007 10th IFIP/IEEE International Symposium on Integrated Network Management*, pages 119–128, May 2007.
- [18] T. Breen, E. Walsh, J. Punch, A. Shah, and C. Bash. From chip to cooling tower data center modeling: Part i influence of server inlet temperature and temperature rise across cabinet. In *Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), 2010 12th IEEE Intersociety Conference on*, pages 1–10, June 2010.
- [19] A. Capozzoli, M. Chinnici, M. Perino, and G. Serale. *Energy Efficient Data Centers: Third International Workshop, E2DC 2014, Cambridge, UK, June 10, 2014, Revised Selected Papers*, chapter Review on Performance Metrics for Energy Efficiency in Data Center: The Role of Thermal Management, pages 135–151. Springer International Publishing, Cham, 2015.
- [20] T. Carli, S. Henriot, J. Cohen, and J. Tomasik. A packing problem approach to energy-aware load distribution in clouds. In *arXiv:1403.0493*, 2014.
- [21] Y. A. Cengel, R. H. Turner, and J. M. Cimbala. Fundamentals of thermal-fluid sciences. *4th edition*, London: McGraw Hill Higher Education, 2012.
- [22] G. Chen, W. He, J. Liu, S. Nath, L. Rigas, L. Xiao, and F. Zhao. Energy-aware server provisioning and load dispatching for connection-intensive internet services. In *Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation*, NSDI'08, pages 337–350, Berkeley, CA, USA, 2008. USENIX Association.
- [23] H. Chen, M. Caramanis, and A. Coskun. The data center as a grid load stabilizer. In *Proceedings of the Asia and South Pacific Design Automation Conference (ASP-DAC)*, 2014.
- [24] H. Chen, A. Coskun, and M. Caramanis. Real-time power control of data centers for providing regulation service. In *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, pages 4314–4321, Dec 2013.
- [25] H. Chen, H. Kang, G. Jiang, and Y. Zhang. Coordinating virtual machine migrations in enterprise data centers and clouds. In *IFIP/IEEE International Symposium on Integrated Network Management (IM 2013)*, 2013.
- [26] J. Choi, Y. Kim, A. Sivasubramaniam, J. Srebric, Q. Wang, and J. Lee. A cfd-based tool for studying temperature in rack-mounted servers. *Computers, IEEE Transactions on*, 57(8):1129–1142, Aug 2008.
- [27] Cisco. Cisco global cloud index: Forecast and methodology, 2014–2019. [http://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/Cloud\\_Index\\_White\\_Paper.pdf](http://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/Cloud_Index_White_Paper.pdf), 2015.
- [28] K. Clark, C. and Fraser, S. Hand, J. G. Hansen, E. Jul, C. Limpach, I. Pratt, and A. Warfield. Live migration of virtual machines. In *Proceedings of the 2Nd Conference on Symposium on Networked Systems Design & Implementation - Volume 2*, NSDI'05, pages 273–286, Berkeley, CA, USA, 2005. USENIX Association.

- 
- [29] J. E. G. Coffman, M. R. Garey, and D. S. Johnson. Approximation algorithms for bin packing: A survey. In D. S. Hochbaum, editor, *Approximation Algorithms for NP-hard Problems*, pages 46–93. PWS Publishing Co., Boston, MA, USA, 1997.
- [30] C. Cremer, W. Eichhammer, M. Friedewald, P. Georgieff, S. Rieth-Hoerst, B. Schlomann, P. Zoche, B. Aebischer, and A. Huser. Energy consumption of information and communication technology (ict) in germany up to 2010, summary of the final report to the german federal ministry of economics and labour. *Karlsruhe/Zurich*, 2003.
- [31] A. Dalvandi, M. Gurusamy, and K. C. Chua. Time-aware vm-placement and routing with bandwidth guarantees in green cloud data centers. In *IEEE 5th International Conference on Cloud Computing Technology and Science (CloudCom), vol.1*, pages 212–217, 2013.
- [32] G. Drenkelfort, L. Goschin, and F. Behrendt. Ein quasi-stationäres modell zur prognose des infrastruktur-energiebedarfs von rechenzentren. In M. Horbach, editor, *GI-Jahrestagung*, volume 220 of *LNI*, pages 864–875. GI, 2013.
- [33] D. Economou, S. Rivoire, and C. Kozyrakis. Full-system power analysis and modeling for server environments. In *In Workshop on Modeling Benchmarking and Simulation (MOBS, 2006)*.
- [34] Emerson Network Power. Data center users’ group special report - energy efficiency and capacity concerns increase. *Data Center Users’ Group (DCUG)*, 2012.
- [35] E. Falkenauer and A. Delchambre. A genetic algorithm for bin packing and line balancing. In *Robotics and Automation, 1992. Proceedings., 1992 IEEE International Conference on*, pages 1186–1192 vol.2, May 1992.
- [36] X. Fan, W.-D. Weber, and L. A. Barroso. Power provisioning for a warehouse-sized computer. In *The 34th ACM International Symposium on Computer Architecture*, 2007.
- [37] M. H. Ferdous, M. Murshed, R. N. Calheiros, and R. Buyya. Network-aware virtual machine placement and migration in cloud data centers. In *Emerging Research in Cloud Distributed Computing Systems. Ed. Susmit Bagchi. Hershey: IGI Global*, pages 42–91, 2015.
- [38] K. Fichter. Zukunftsmarkt energieeffiziente rechenzentren. *Fallstudie im Auftrag des Bundesministeriums für Umwelt, Naturschutz und Reaktorsicherheit (BMU)*, 2007.
- [39] V. Ganti and G. Ghatikar. Smart grid as a driver for energy-intensive industries: A data center case study. In *Grid-Interop 2012*, 2012.
- [40] M. Ghamkhari and H. Mohsenian-Rad. Data centers to offer ancillary services. In *SmartGrid-Comm’12*, pages 436–441, 2012.
- [41] M. Ghamkhari and H. Mohsenian-Rad. Optimal integration of renewable energy resources in data centers with behind-the-meter renewable generator. In *IEEE International Conference on Communications (ICC)*, 2012.
- [42] G. Ghatikar, V. Ganti, N. Matson, and M. Piette. Demand response opportunities and enabling technologies for data centers: Findings from field studies. *Lawrence Berkeley National Laboratory*, 2012.
- [43] G. Ghatikar, M. Piette, S. Fujita, A. McKane, J. Dudley, and A. Radspieler. Demand response and open automated demand response opportunities for data centers. *Lawrence Berkeley National Laboratory*, 2010.

- [44] S. Govindan, D. Wang, L. Chen, A. Sivasubramaniam, and B. Urgaonkar. Modeling and analysis of availability of datacenter power infrastructure. *Technical Report CSE-10-006, The Pennsylvania State University, Philadelphia, PA, USA*, 2010.
- [45] S. Govindan, D. Wang, A. Sivasubramaniam, and B. Urgaonkar. Leveraging stored energy for handling power emergencies in aggressively provisioned datacenters. *SIGARCH Comput. Archit. News*, 40(1):75–86, March 2012.
- [46] A. Greenberg, N. Jain, S. Kandula, C. Kim, P. Lahiri, D. A. Maltz, P. Patel, and S. Sengupta. VL2: A scalable and flexible data center network. In *SIGCOMM*. Association for Computing Machinery, Inc., August 2009.
- [47] M. Gupta and S. Singh. Greening of the internet. In *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, SIGCOMM '03, pages 19–26, New York, NY, USA, 2003. ACM.
- [48] S. Gurumurthi, A. Sivasubramaniam, M. J. Irwin, N. Vijaykrishnan, M. Kandemir, T. Li, and L. K. John. Using complete machine simulation for software power estimation: The softwatt approach. In *Proceedings of the 8th International Symposium on High-Performance Computer Architecture*, HPCA '02, pages 141–, Washington, DC, USA, 2002. IEEE Computer Society.
- [49] C. Hankendi, S. Reda, and A. Coskun. vcap: Adaptive power capping for virtualized servers. In *Low Power Electronics and Design (ISLPED), 2013 IEEE International Symposium on*, pages 415–420, Sept 2013.
- [50] T. Harvey, M. Patterson, and J. Bean. Updated air-side free cooling maps: The impact of ashrae 2011 allowable ranges. *The Green Grid*, 2012.
- [51] R. Hintemann. Die trends in rechenzentren bis 2015. *Ergebnisse einer neuen Studie im Auftrag des Umweltbundesamtes*, 2010.
- [52] R. Hintemann, K. Fichter, and L. Stobbe. Materialbestand der rechenzentren in deutschland. <http://www.uba.de/uba-info-medien/4037.html>, 2010.
- [53] M. Hoyer, K. Schroeder, D. Schlitt, and W. Nebel. Proactive dynamic resource management in virtualized data centers. In *Proceedings of the 2nd International Conference on Energy-Efficient Computing and Networking*, 2011.
- [54] C. Hyser, B. McKee, R. Gardner, and B. Watson. Autonomic virtual machine placement in the data center. In *HPL-2007-189*, 2008.
- [55] C. Isci, J. Liu, B. Abali, J. Kephart, and J. Kouloheris. Improving server utilization using fast virtual machine migration. *IBM Journal of Research and Development*, 55(6):4:1–4:12, Nov 2011.
- [56] S. Janacek, K. Schroeder, G. Schomaker, W. Nebel, M. Rueuschen, and G. Pistor. Modeling and approaching a cost transparent, specific data center power consumption. In *2012 International Conference on Energy Aware Computing*, pages 68–73, december 2012.
- [57] C. Jo, E. Gustafsson, J. Son, and B. Egger. Efficient live migration of virtual machines using shared storage. *SIGPLAN Not.*, 48(7):41–50, Mar. 2013.
- [58] I. T. Jolliffe. Principal component analysis. *Springer*, 2002.



- [59] M. Jonas, R. Gilbert, J. Ferguson, G. Varsamopoulos, and S. K. S. Gupta. A transient model for data center thermal prediction. In *Green Computing Conference (IGCC), 2012 International*, pages 1–10, 2012.
- [60] A. Kansal, F. Zhao, J. Liu, N. Kothari, and A. A. Bhattacharya. Virtual machine power metering and provisioning. In *Proceedings of the 1st ACM Symposium on Cloud Computing*, SoCC '10, pages 39–50, New York, NY, USA, 2010. ACM.
- [61] J. M. Kaplan, W. Forrest, and N. Kindler. Revolutionizing data center energy efficiency. [http://www.mckinsey.com/client-service/bto/pointofview/pdf/revolutionizing\\_data\\_center\\_efficiency.pdf](http://www.mckinsey.com/client-service/bto/pointofview/pdf/revolutionizing_data_center_efficiency.pdf), McKinsey & Company, 2008.
- [62] Y. Kaya, M. Uyar, and R. Tekin. A novel crossover operator for genetic algorithms: Ring crossover. *CoRR*, abs/1105.0355, 2011.
- [63] G. Keller, M. Tighe, H. Lutfiyya, and M. Bauer. An analysis of first fit heuristics for the virtual machine relocation problem. In *Network and service management (cnsm), 2012 8th international conference and 2012 workshop on systems virtualization management (svm)*, pages 406–413, Oct 2012.
- [64] R. Khanna, J. Jaiber, and T. Rangarajan. Phase-aware predictive thermal modeling for proactive load-balancing of compute clusters. In *Energy Aware Computing, 2012 International Conference on*, pages 1–6, Dec 2012.
- [65] Y. Koh, R. Knauerhase, P. Brett, M. Bowman, Z. Wen, and C. Pu. An analysis of performance interference effects in virtual environments. In *Performance Analysis of Systems Software, 2007. ISPASS 2007. IEEE International Symposium on*, pages 200–209, April 2007.
- [66] J. Koomey and J. Taylor. New data supports finding that 30 percent of servers are 'comatose', indicating that nearly a third of capital in enterprise data centers is wasted. [http://anthesisgroup.com/wp-content/uploads/2015/06/Case-Study\\_DataSupports30PercentComatoseEstimate-FINAL\\_06032015.pdf](http://anthesisgroup.com/wp-content/uploads/2015/06/Case-Study_DataSupports30PercentComatoseEstimate-FINAL_06032015.pdf), Steyer-Taylor Center for Energy Policy and Finance, Stanford University and Anthesis Group, 2015.
- [67] A. Krioukov, C. Goebel, S. Alspaugh, Y. Chen, D. Culler, and R. Katz. Integrating renewable energy using data analytics systems: Challenges and opportunities. *IEEE Data Eng. Bull.*, 34(1):3–11, 2011.
- [68] M. L., A. Wierman, L. Andrew, and E. Thereska. Dynamic right-sizing for power-proportional data centers. In *INFOCOM, 2011 Proceedings IEEE*, pages 1098–1106, April 2011.
- [69] A. Lewis, S. Ghosh, and N.-F. Tzeng. Run-time energy consumption estimation based on workload in server systems. In *Proceedings of the 2008 Conference on Power Aware Computing and Systems, HotPower'08*, pages 4–4, Berkeley, CA, USA, 2008. USENIX Association.
- [70] J. Li and J. F. Martinez. Dynamic power-performance adaptation of parallel computation on chip multiprocessors. In *High-Performance Computer Architecture, 2006. The Twelfth International Symposium on*, pages 77–87, Feb 2006.
- [71] S. Li, M. Brocanelli, W. Zhang, and X. Wang. Data center power control for frequency regulation. In *Power and Energy Society General Meeting (PES), 2013 IEEE*, pages 1–5, July 2013.
- [72] H. Liu, C.-Z. Xu, H. Jin, J. Gong, and X. Liao. Performance and energy modeling for live migration of virtual machines. In *Proceedings of the 20th International Symposium on High Performance Distributed Computing, HPDC '11*, pages 171–182, New York, NY, USA, 2011. ACM.

- [73] Z. Liu, Y. Chen, C. Bash, A. Wierman, D. Gmach, Z. Wang, M. Marwah, and C. Hyser. Renewable and cooling aware workload management for sustainable data centers. In *Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems*, SIGMETRICS '12, pages 175–186, New York, NY, USA, 2012. ACM.
- [74] Z. Liu, A. Wierman, Y. Chen, B. Razon, and N. Chen. Data center demand response: avoiding the coincident peak via workload shifting and local generation. In *Proceedings of the ACM SIGMETRICS/international conference on Measurement and modeling of computer systems*, SIGMETRICS '13, pages 341–342, New York, NY, USA, 2013. ACM.
- [75] P. Mahadevan, P. Sharma, S. Banerjee, and P. Ranganathan. Energy aware network operations. In *Proceedings of the 28th IEEE International Conference on Computer Communications Workshops*, INFOCOM'09, pages 25–30, Piscataway, NJ, USA, 2009. IEEE Press.
- [76] P. Mahadevan, P. Sharma, S. Banerjee, and P. Ranganathan. A power benchmarking framework for network devices. In *Proceedings of the 8th International IFIP-TC 6 Networking Conference*, NETWORKING '09, pages 795–808, Berlin, Heidelberg, 2009. Springer-Verlag.
- [77] V. Mann, A. Gupta, P. Dutta, A. Vishnoi, P. Bhattacharya, R. Poddar, and A. Iyer. Remedy: Network-aware steady state vm management for data centers. In R. Bestak, L. Kencl, L. Li, J. Widmer, and H. Yin, editors, *NETWORKING 2012*, volume 7289 of *Lecture Notes in Computer Science*, pages 190–204. Springer Berlin Heidelberg, 2012.
- [78] M. Marwah, R. Sharma, R. Shih, C. Patel, V. Bhatia, M. Mekanapurath, R. Velumani, and S. Velayudhan. Data analysis, visualization and knowledge discovery in sustainable data centers. In *Proceedings of the 2Nd Bangalore Annual Compute Conference*, COMPUTE '09, pages 2:1–2:8, New York, NY, USA, 2009. ACM.
- [79] A. Mazer. The state of data center evolution. <https://www.downloads.siemens.com/download-center/Download.aspx?pos=download&fct=getasset&id1=A6V10595703>, 2013.
- [80] H. Maziku and S. Shetty. Towards a network aware vm migration: Evaluating the cost of vm migration in cloud data centers. In *Cloud Networking (CloudNet), 2014 IEEE 3rd International Conference on*, pages 114–119, Oct 2014.
- [81] D. Meisner and T. F. Wenisch. Stochastic queuing simulation for data center workloads. *Exascale Evaluation and Research Techniques Workshop*, 2010.
- [82] A. Memari, J. Vornberger, J. M. Gómez, and W. Nebel. A data center simulation framework based on an ontological foundation. In *Proceedings of the 28th EnviroInfo 2014 Conference, Oldenburg, Germany*, pages 461–468, September 2014.
- [83] X. Meng, C. Isci, J. Kephart, L. Zhang, E. Bouillet, and D. Pendarakis. Efficient resource provisioning in compute clouds via vm multiplexing. In *Proceedings of the 7th International Conference on Autonomic Computing*, ICAC '10, pages 11–20, New York, NY, USA, 2010. ACM.
- [84] A. Menon, J. R. Santos, Y. Turner, G. Janakiraman, and W. Zwaenepoel. Diagnosing performance overheads in the xen virtual machine environment. In *Proceedings of the 1st ACM/USENIX International Conference on Virtual Execution Environments*, VEE '05, pages 13–23, New York, NY, USA, 2005. ACM.
- [85] H. Mohsenian-Rad and A. Leon-Garcia. Coordination of cloud computing and smart power grids. In *Proc. of IEEE Smart Grid Communications Conference, Gaithersburg, Maryland*, 2010.

- 
- [86] J. Moore, J. Chase, and P. Ranganathan. Consil: Lowcost thermal mapping of data centers. In *Proceedings of the Workshop on Tackling Computer Systems Problems with Machine Learning Techniques (SysML)*, 2006.
- [87] J. Moore, J. Chase, P. Ranganathan, and R. Sharma. Making scheduling „cool”: Temperature-aware workload placement in data centers. In *Proceedings of the Annual Conference on USENIX Annual Technical Conference, ATEC '05*, pages 5–5, Berkeley, CA, USA, 2005. USENIX Association.
- [88] T. Mukherjee, A. Banerjee, G. Varsamopoulos, S. K. S. Gupta, and S. Rungta. Spatio-temporal thermal-aware job scheduling to minimize energy consumption in virtualized heterogeneous data centers. In *Comput. Netw.*, 2009.
- [89] B. Mumei, J. Tang, and S. Hashimoto. Enabling green networking with a power down approach. In *Communications (ICC), 2012 IEEE International Conference on*, pages 2867–2871, June 2012.
- [90] R. Nathuji and K. Schwan. Vpm tokens: Virtual machine-aware power budgeting in datacenters. In *Proceedings of the 17th International Symposium on High Performance Distributed Computing, HPDC '08*, pages 119–128, New York, NY, USA, 2008. ACM.
- [91] S. Nedeveschi, L. Popa, G. Iannaccone, S. Ratnasamy, and D. Wetherall. Reducing network energy consumption via sleeping and rate-adaptation. In *Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation, NSDI'08*, pages 323–336, Berkeley, CA, USA, 2008. USENIX Association.
- [92] J. Niemann, K. Brown, and V. Avelar. Hot-aisle vs. cold-aisle containment for data centers. *APC White Paper*, 2010.
- [93] NRDC. Data center efficiency assessment. <https://www.nrdc.org/sites/default/files/data-center-efficiency-assessment-IP.pdf>, 2014.
- [94] E. Pakbaznia, M. Ghasemazar, and M. Pedram. Temperature-aware dynamic resource provisioning in a power-optimized datacenter. In *Design, Automation Test in Europe Conference Exhibition (DATE), 2010*, pages 124–129, March 2010.
- [95] E. Pakbaznia and M. Pedram. Minimizing data center cooling and server power costs. In *Proceedings of the 14th ACM/IEEE international symposium on Low power electronics and design, ISLPED '09*, pages 145–150, New York, NY, USA, 2009. ACM.
- [96] L. Parolini, B. Sinopoli, B. Krogh, and Z. Wang. A cyber-physical systems approach to data center modeling and control for energy efficiency. *Proceedings of the IEEE*, 100(1):254–268, Jan 2012.
- [97] L. Parolini, B. Sinopoli, and B. H. Krogh. Reducing data center energy consumption via coordinated cooling and load management. In *Proceedings of the 2008 Conference on Power Aware Computing and Systems, HotPower'08*, pages 14–14, Berkeley, CA, USA, 2008. USENIX Association.
- [98] C. Patel, R. Sharma, C. Bash, and A. Beitelmal. Thermal considerations in cooling large scale high compute density data centers. In *Thermal and Thermomechanical Phenomena in Electronic Systems, 2002. ITherm 2002. The Eighth Intersociety Conference on*, pages 767–776, 2002.
- [99] C. D. Patel, C. E. Bash, C. Belady, L. Stahl, and D. Sullivan. Computational Fluid Dynamics Modeling of High Compute Density Data Centers to Assure System Inlet Air Specifications. In *Proceedings of IPACK'01, The Pacific Rim/ASME International Electronic Packaging Technical Conference and Exhibition*, 2001.

- [100] C. D. Patel, C. E. Bash, R. Sharma, M. Beitelmal, and R. Friedrich. *Smart cooling of data centers*, volume 2, pages 129–137. ASME, 12 2003.
- [101] P. D. Patel, M. Karamta, M. D. Bhavsar, and M. B. Potdar. Live virtual machine migration techniques in cloud computing: A survey. *International Journal of Computer Applications*, 86(16):18–21, January 2014.
- [102] M. Pedram and I. Hwang. Power and performance modeling in a virtualized server system. In *Proceedings of the 2010 39th International Conference on Parallel Processing Workshops*, 2010.
- [103] S. Pelley, D. Meisner, T. Wenisch, and J. Vangilder. Understanding and abstracting total data center power. *Proc. of the 2009 Workshop on Energy Efficient Design (WEED)*, 2009.
- [104] T. Pering, T. Burd, and R. Brodersen. Dynamic voltage scaling and the design of a low-power microprocessor system. In *In Power Driven Microarchitecture Workshop, attached to ISCA98*, 1998.
- [105] B. Petschke. State of the art energy efficient data centre air conditioning - dynamic free cooling. *STULZ White Paper - Dynamic Free Cooling, Rel. 2*, 2008.
- [106] P. Pulvermüller. Bestandsaufnahme effiziente rechenzentren in deutschland. *eco – Verband der deutschen Internetwirtschaft e.V.*, 2008.
- [107] A. Qouneh, C. Li, and T. Li. A quantitative analysis of cooling power in container-based data centers. In *Workload Characterization (IISWC), 2011 IEEE International Symposium on*, pages 61–71, Nov 2011.
- [108] P. Ranganathan, P. Leech, D. Irwin, and J. Chase. Ensemble-level power management for dense blade servers. In *Proceedings of the 33rd Annual International Symposium on Computer Architecture*, ISCA '06, pages 66–77, Washington, DC, USA, 2006. IEEE Computer Society.
- [109] L. Rao, X. Liu, L. Xie, and W. Liu. Minimizing electricity cost: Optimization of distributed internet data centers in a multi-electricity-market environment. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9, March 2010.
- [110] Raritan. Px intelligent rack pdus. [http://ram.raritanassets.com/resources/data\\_sheets/raritan-ds-px.pdf](http://ram.raritanassets.com/resources/data_sheets/raritan-ds-px.pdf), 2015.
- [111] N. Rasmussen. Electrical efficiency modeling for data centers. In *White Paper 113, Revision 2*, Schneider Electric Data Center Science Center, 2011.
- [112] P. V. V. Reddy and L. Rajamani. Evaluation of different hypervisors performance in the private cloud with sigar framework. In *International Journal of Advanced Computer Science and Applications(IJACSA)*, volume 5, 2014.
- [113] S. Rivoire, P. Ranganathan, and C. Kozyrakis. A comparison of high-level full-system power models. In *Proceedings of the 2008 Conference on Power Aware Computing and Systems*, HotPower'08, pages 3–3, Berkeley, CA, USA, 2008. USENIX Association.
- [114] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Pearson Education, 2 edition, 2003.
- [115] K. Rybina, A. Patni, and A. Schill. Analysing the migration time of live migration of multiple virtual machines. In *CLOSER 2014 - Proceedings of the 4th International Conference on Cloud Computing and Services Science*, pages 590–597, 2014.

- 
- [116] D. Schlitt, G. Schomaker, and W. Nebel. Gain more from pue: Assessing data center infrastructure power adaptability. In S. Klingert, M. Chinnici, and M. Rey Porto, editors, *Energy Efficient Data Centers*, volume 8945 of *Lecture Notes in Computer Science*, pages 152–166. Springer International Publishing, 2015.
- [117] K. Schroeder and W. Nebel. Demand response and site management for cloud based services. In *Energy Aware Computing, 2012 International Conference on*, pages 1–6, Dec 2012.
- [118] K. Schroeder and W. Nebel. Behavioral model for cloud aware load and power management. In *Proceedings of the 2013 International Workshop on Hot Topics in Cloud Services*, HotTopiCS '13, pages 19–26, New York, NY, USA, 2013. ACM.
- [119] K. Schroeder and W. Nebel. Inter cloud capable dynamic resource management with model of behavior. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, SAC '13, pages 408–410, New York, NY, USA, 2013. ACM.
- [120] V. Schwalfenberg and M. Schmidt. Infrastruktur im energieeffizienten rechenzentrum. *RITTAL White Paper, V10*, 2008.
- [121] G. Semeraro, G. Magklis, R. Balasubramonian, D. H. Albonesi, S. Dwarkadas, and M. L. Scott. Energy-efficient processor design using multiple clock domains with dynamic voltage and frequency scaling. In *High-Performance Computer Architecture, 2002. Proceedings. Eighth International Symposium on*, pages 29–40, Feb 2002.
- [122] R. Sharma, C. Bash, C. Patel, R. Friedrich, and J. Chase. Balance of power: dynamic thermal management for internet data centers. *Internet Computing, IEEE*, 9(1):42–49, Jan 2005.
- [123] V. Soteriou and L.-S. Peh. Dynamic power management for power optimization of interconnection networks using on/off links. In *Proceedings of the 11th Symposium on High Performance Interconnects*, page 15–20, 2003.
- [124] S. P. E. C. (SPEC). Power and performance benchmark methodology v2.1. [https://www.spec.org/power/docs/SPEC-Power\\_and\\_Performance\\_Methodology.pdf](https://www.spec.org/power/docs/SPEC-Power_and_Performance_Methodology.pdf), 2012.
- [125] A. Stage and T. Setzer. Network-aware migration control and scheduling of differentiated virtual machine workloads. In *Software Engineering Challenges of Cloud Computing, 2009. CLOUD '09. ICSE Workshop on*, pages 9–14, May 2009.
- [126] G. Staples. Torque resource manager. In *SC '06: Proceedings of the 2006 ACM/IEEE conference on Supercomputing*, page 8, 2006.
- [127] A. Strunk and W. Dargie. Does live migration of virtual machines cost energy? In *Advanced Information Networking and Applications (AINA), 2013 IEEE 27th International Conference on*, pages 514–521, March 2013.
- [128] Q. Tang, S. Gupta, and G. Varsamopoulos. Thermal-aware task scheduling for data centers through minimizing heat recirculation. In *Cluster Computing, 2007 IEEE International Conference on*, pages 129–138, Sept 2007.
- [129] Q. Tang, S. Gupta, and G. Varsamopoulos. Energy-efficient thermal-aware task scheduling for homogeneous high-performance computing data centers: A cyber-physical approach. *Parallel and Distributed Systems, IEEE Transactions on*, 19(11):1458–1472, Nov 2008.

- [130] Q. Tang, S. K. S. Gupta, D. Stanzione, and P. Cayton. Thermal-aware task scheduling to minimize energy usage of blade server based datacenters. In *Dependable, Autonomic and Secure Computing, 2nd IEEE International Symposium on*, pages 195–202, Sept 2006.
- [131] Q. Tang, T. Mukherjee, S. K. S. Gupta, and P. Cayton. Sensor-based fast thermal evaluation model for energy efficient high-performance datacenters. In *Proceedings - 4th International Conference on Intelligent Sensing and Information Processing*, 2006.
- [132] M. Tighe, G. Keller, M. Bauer, and H. Lutfiyya. A distributed approach to dynamic vm management. In *Network and Service Management (CNSM), 2013 9th International Conference on*, pages 166–170, Oct 2013.
- [133] N. Tolia, Z. Wang, M. Marwah, C. Bash, P. Ranganathan, and X. Zhu. Delivering energy proportionality with non energy-proportional systems: Optimizing the ensemble. In *Proceedings of the 2008 Conference on Power Aware Computing and Systems, HotPower'08*, pages 2–2, Berkeley, CA, USA, 2008. USENIX Association.
- [134] M. Ton and B. Fortenburry. High performance buildings: Data centers uninterruptible power supplies (ups). *LBNL*, 2005.
- [135] Transparency Market Research. Green data center market. <http://www.transparencymarketresearch.com/green-data-center.html>, 2015.
- [136] M. S. Vasudevan, B. R. Mohan, and D. K. Damodaran. Performance measuring and comparison of virtualbox and vmware. In *International Proceedings of Computer Science & Information Tech, 2012, Vol. 27*, page 42, 2012.
- [137] VDE. *Feuchte Luft - h,x-Diagramm - Praktische Anwendungs- und Arbeitshilfen*. VDE Verlag, 2012.
- [138] VDI. *VDI-Wärmeatlas. Berechnungsunterlagen für Druckverlust, Wärme- und Stoffübergang*. Springer, Berlin, 10., bearb. u. erw. a. edition, 2006.
- [139] A. Verma, P. Ahuja, and A. Neogi. Power-aware dynamic placement of hpc applications. In *Proceedings of the 22Nd Annual International Conference on Supercomputing, ICS '08*, pages 175–184, New York, NY, USA, 2008. ACM.
- [140] H. Wang, J. Huang, X. Lin, and H. Mohsenian-Rad. Exploring smart grid and data center interactions for electric power load balancing. *SIGMETRICS Perform. Eval. Rev.*, 41(3):89–94, Jan. 2014.
- [141] R. Wang, N. Kandasamy, C. Nwankpa, and D. R. Kaeli. Datacenters as controllable load resources in the electricity market. In *Proceedings of the 2013 IEEE 33rd International Conference on Distributed Computing Systems, ICDCS '13*, pages 176–185, Washington, DC, USA, 2013. IEEE Computer Society.
- [142] X. Wang, M. Chen, C. Lefurgy, and T. Keller. Ship: A scalable hierarchical power control architecture for large-scale data centers. *Parallel and Distributed Systems, IEEE Transactions on*, 23(1):168–176, Jan 2012.
- [143] C. Weber, A. Saiyed, M. Kamani, and P. Sivalingam. A scientific review on virtual machine performance. *Faculty of Business and IT*, 2(2.335):2–466, 2013.
- [144] F. Xu, F. Liu, H. Jin, and A. Vasilakos. Managing performance overhead of virtual machines in cloud computing: A survey, state of the art, and future directions. *Proceedings of the IEEE*, 102(1):11–31, Jan 2014.

- [145] H. Xu, C. Feng, and B. Li. Temperature aware workload management in geo-distributed data centers. *Parallel and Distributed Systems, IEEE Transactions on*, 26(6):1743–1753, June 2015.
- [146] J. Xu and J. Fortes. Multi-objective virtual machine placement in virtualized data center environments. In *Green Computing and Communications (GreenCom), 2010 IEEE/ACM Int'l Conference on Int'l Conference on Cyber, Physical and Social Computing (CPSCom)*, pages 179–188, Dec 2010.
- [147] J. Xu and J. Fortes. A multi-objective approach to virtual machine management in datacenters. In *Proceedings of the 8th ACM International Conference on Autonomic Computing, ICAC '11*, pages 225–234, New York, NY, USA, 2011. ACM.
- [148] Yole Developpement. New technologies and architectures for efficient data center report. [http://yole.fr/Datacenter\\_Silicon\\_Photonics.aspx](http://yole.fr/Datacenter_Silicon_Photonics.aspx), 2015.
- [149] X. Zhan and S. Reda. Techniques for energy-efficient power budgeting in data centers. In *Proceedings of the 50th Annual Design Automation Conference, DAC '13*, pages 176:1–176:7, New York, NY, USA, 2013. ACM.
- [150] Y. Zhang, Y. Wang, and X. Wang. Greenware: Greening cloud-scale data centers to maximize the use of renewable energy. In F. Kon and A.-M. Kermarrec, editors, *Middleware 2011*, volume 7049 of *Lecture Notes in Computer Science*, pages 143–164. Springer Berlin Heidelberg, 2011.
- [151] W. Zheng, A. P. Centeno, F. Chong, and R. Bianchini. Logstore: Toward energy-proportional storage servers. In *Proceedings of the 2012 ACM/IEEE International Symposium on Low Power Electronics and Design, ISLPED '12*, pages 273–278, New York, NY, USA, 2012. ACM.
- [152] R. Zhou, Z. Wang, A. McReynolds, C. E. Bash, T. W. Christian, and R. Shih. Optimization and control of cooling microgrids for data centers. In *Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), 2012 13th IEEE Intersociety Conference on*, pages 338–343, May 2012.





# Abbildungsverzeichnis

|      |   |     |
|------|---|-----|
| 1.1  | Energiebedarf der Rechenzentren weltweit . . . . .                              | 2   |
| 1.2  | Marktentwicklung von „grünen Rechenzentren“ . . . . .                           | 3   |
| 2.1  | Rechenzentrumsaufteilung (vereinfacht) . . . . .                                | 13  |
| 2.2  | Elektrische Leistungsaufnahme nach Gewerken im Rechenzentrum . . . . .          | 14  |
| 2.3  | Stromversorgungsarchitektur im Rechenzentrum . . . . .                          | 17  |
| 2.4  | Kühlkette im Rechenzentrum . . . . .  | 20  |
| 2.5  | Schematische Darstellung einer Kaltgangeinhausung . . . . .                     | 23  |
| 2.6  | Netzwerktopologie VL2 . . . . .   | 25  |
| 2.7  | Grundlagen der Virtualisierung . . . . .  | 27  |
| 2.8  | Konzept der Live-Migrationen . . . . .  | 28  |
| 2.9  | Ressourcenbelegung eines Servers durch VMs . . . . .                            | 31  |
| 2.10 | Beispiel einer Server-Konsolidierung . . . . .                                  | 35  |
| 3.1  | Vorhersagemethodik von Hoyer <i>et al.</i> . . . . .                            | 38  |
| 3.2  | Architektur des autonomen Ressourcenmanagements von Xu <i>et al.</i> . . . . .  | 41  |
| 3.3  | Kontrollfluss der multi-kriteriellen Optimierung von Xu <i>et al.</i> . . . . . | 42  |
| 4.1  | Unterschiedliche Allokationen der gleichen virtuellen Maschinen . . . . .       | 50  |
| 4.2  | Energetische Einflussfaktoren im Rechenzentrum . . . . .                        | 52  |
| 4.3  | Details der Klimakette . . . . .  | 55  |
| 4.4  | Vergleich der Server-Energieeffizienz . . . . .                                 | 63  |
| 4.5  | Verteilung der Wärmeleistung der IT-Räume . . . . .                             | 66  |
| 4.6  | Verteilung der Wärmeleistung im IT-Raum . . . . .                               | 72  |
| 4.7  | Globale Flexibilität . . . . .  | 75  |
| 4.8  | Kombinierter Lösungsansatz . . . . .  | 82  |
| 5.1  | USV-Effizienzverlauf . . . . .  | 87  |
| 5.2  | Server-Leistungsaufnahme: Modellgenauigkeit . . . . .                           | 89  |
| 5.3  | Logarithmisches Mittel der Temperaturdifferenz . . . . .                        | 93  |
| 5.4  | Diagramm des Korrekturfaktors für Kreuzwärmeübertrager . . . . .                | 95  |
| 5.5  | Zusammenhang der Kälteenergie und Temperaturen . . . . .                        | 98  |
| 5.6  | Coefficient of Performance einer Kältemaschine . . . . .                        | 99  |
| 5.7  | Auslastungsvergleich: prozentual . . . . .                                      | 101 |
| 5.8  | Auslastungsvergleich: absolut . . . . .   | 101 |
| 5.9  | VM-Migrationszeiten nach CPU-Auslastung . . . . .                               | 103 |
| 5.10 | Zusammenspiel der Modelle . . . . .   | 107 |
| 6.1  | DaMucA-Architektur . . . . .  | 110 |
| 6.2  | Einordnung der Kurzzeit- und Langzeitplanung . . . . .                          | 112 |
| 6.3  | Auslastungsbetrachtung für die Langzeitplanung . . . . .                        | 113 |
| 6.4  | CPU-Auslastung und CPU-Maxima einer VM . . . . .                                | 115 |
| 6.5  | Histogramm der CPU-Auslastung einer VM . . . . .                                | 115 |
| 6.6  | Übersicht DaMucA-Verfahren . . . . .  | 119 |
| 6.7  | Kodierung eines Individuums . . . . .   | 122 |

|      |  |     |
|------|--|-----|
| 6.8  | Beispiel der Kodierung eines Individuums . . . . .   | 122 |
| 6.9  | Beispiel der Rekombination zweier Individuen . . . . .                                     | 124 |
| 6.10 | Einfluss der IT-Raumwahl auf das Ergebnis . . . . .  | 129 |
| 6.11 | Server-Effizienz im Vergleich . . . . .  | 131 |
| 6.12 | Kombinierte Bewertung Server-Effizienz und Flexibilität . . . . .                          | 133 |
|      |  |     |
| 7.1  | Elektrische Leistung nach Anzahl aktiver IT-Räume . . . . .                                | 142 |
| 7.2  | Maximum der elektr. Leistung der IT-Räume und Kaltwassertemperatur . . . . .               | 143 |
| 7.3  | Server-Details nach Anzahl aktiver IT-Räume . . . . .                                      | 145 |
| 7.4  | Differenz zwischen bester und schlechtester Lösung bei gleicher Anzahl aktiver IT-Räume .  | 146 |
| 7.5  | Beste Lösung einer Population nach Generationen . . . . .                                  | 147 |
| 7.6  | Beste und schlechteste Lösungen einer Population nach Generationen . . . . .               | 148 |
| 7.7  | Mittlere elektrische Leistung nach IT-Räumen . . . . .                                     | 149 |
| 7.8  | Anzahl aktiver Server pro IT-Raum . . . . .  | 151 |
| 7.9  | Prozentuale Differenz aktive Server und mittlere Leistungsaufnahme (FFD und DaMucA) .      | 152 |
| 7.10 | Freie Ressourcen nach Referenz-VM pro IT-Raum . . . . .                                    | 153 |
| 7.11 | Freie Ressourcen nach Referenz-VM pro IT-Raum (nur FFD und DaMucA) . . . . .               | 154 |
| 7.12 | Elektrische Leistung nach IT-Räumen (2-6) . . . . .  | 155 |
| 7.13 | Elektrische Leistung nach IT-Räumen (7-12) . . . . .                                       | 156 |
| 7.14 | Standardabweichung der Wärmeleistungsverteilung in den IT-Räumen . . . . .                 | 157 |
| 7.15 | Wärmeleistungsverteilung im nicht optimierten Betriebszustand . . . . .                    | 158 |
| 7.16 | Wärmeleistungsverteilung im FFD-optimierten Betriebszustand . . . . .                      | 159 |
| 7.17 | Wärmeleistungsverteilung im DaMucA-optimierten Betriebszustand . . . . .                   | 160 |
| 7.18 | Mittlere elektr. Leistung Rechenzentrum und Anzahl Server nach maximaler Evakuierungszeit  | 162 |
| 7.19 | Entwicklung von Ressourcen und elektrischer Leistung nach globaler Flexibilität . . . . .  | 162 |
| 7.20 | Prozentualer Zuwachs von Ressourcen und elektrischer Leistung nach globaler Flexibilität . | 163 |
| 7.21 | Mittlere elektr. Leistung von IT-Geräten und Kältemaschinen mit Kaltwassertemperatur . .   | 165 |
| 7.22 | Jährlicher elektrischer Energiebedarf nach IT und Kühlung . . . . .                        | 166 |
| 7.23 | Ersparnis gesamt (prozentual) . . . . .  | 167 |

# Tabellenverzeichnis

|     |   |     |
|-----|---|-----|
| 2.1 | USV-Topologien . . . . .  | 18  |
| 3.1 | Vergleich der Lastmanagementverfahren . . . . .                                 | 44  |
| 4.1 | Vergleich der Einflussfaktoren . . . . .  | 58  |
| 5.1 | Benötigte Messwerte für die Charakterisierung des Umluftkühlgerätes . . . . .   | 94  |
| 6.1 | Einordnung der Kurz- und Langzeitplanung . . . . .                              | 112 |
| 7.1 | Simuliertes Rechenzentrum . . . . .   | 141 |
| 7.2 | Standardabweichung in kW der mittleren elektrischen Leistungsaufnahme . . . . . | 150 |
| 7.3 | Übersicht der Ergebnisse der 60-Tage-Simulation (DaMucA-Allokation) . . . . .   | 154 |
| 7.4 | Evakuierungszeit der unterschiedlichen Betriebszustände . . . . .               | 161 |
| 7.5 | Optimierung der Kühlleistung im Vergleich . . . . .                             | 164 |



# Symbolverzeichnis

|                       |   |
|-----------------------|---|
| $\vec{a}$             | Allokation der virtuellen Maschinen auf die aktiven Server im Rechenzentrum |
| $\vec{a}_{current}$   | Aktuelle Allokation im Rechenzentrum  |
| $\vec{a}_{next}$      | Neue Allokation, die durch das DaMucA-Verfahren erstellt wird               |
| $COP_{chiller}$       | Coefficient of Performance der Kältemaschinen                               |
| $c_{pair}$            | Spezifische Wärmekapazität der Luft   |
| $c_{pwater}$          | Spezifische Wärmekapazität des Kühlwassers (Wasser-Glykol-Gemisch)          |
| $\vec{c}$             | Vektor der Constraints  |
| $e_s$                 | Energieeffizienz des Servers $s$  |
| $e_{ups}$             | Energieeffizienz einer USV  |
| $\varepsilon_{NTU}$   | Effektivität des Umluftkühlgeräts   |
| $\vec{\varepsilon}_s$ | Ressourcenpuffer des Servers $s$  |
| $f$                   | Korrekturfaktor für Kreuzwärmeübertrager                                    |
| $f_g$                 | Globale Flexibilität einer Allokation                                       |
| $f_{l_v}$             | Lokale Flexibilität der virtuellen Maschine $v$                             |
| $\vec{f}_l$           | Vektor mit der lokalen Flexibilität aller virtueller Maschinen              |
| $K_r$                 | Menge aller Racks im IT-Raum $r$  |
| $k$                   | Bezeichnung eines spezifischen Racks  |
| $kpm(\dots)$          | Modellfunktion, die die elektrische Leistung der Kältemaschinen abschätzt   |
| $L$                   | Menge aller Ressourcenbedarfe aller virtueller Maschinen                    |
| $\vec{l}_R$           | Ressourcenkapazität aller Server unter der Menge aktiver IT-Räume $R$       |
| $\vec{l}_s$           | Ressourcenbelegung des Servers $s$  |
| $\vec{l}_{scap}$      | Ressourcenkapazität des Servers $s$   |

|                     |   |
|---------------------|---|
| $\vec{l}_v$         | Ressourcenbedarf der virtuellen Maschine $v$  |
| $\vec{l}_{V_f}$     | Summierter Ressourcenbedarf aller virtueller Maschinen inklusive der geforderten lokalen Flexibilität |
| $\dot{m}_{air}$     | Massenstrom der Luft  |
| $\dot{m}_{air_s}$   | Massenstrom der Luft in Server $s$  |
| $\dot{m}_{water}$   | Massenstrom des Kühlwassers   |
| $NTU$               | Einheitenlose Kennzahl der NTU-Methode  |
| $o_\varepsilon$     | RAM-Belegungspuffer   |
| $o_{ref}$           | Allozierter RAM einer Referenz-VM   |
| $o_s$               | RAM-Belegung des Servers $s$  |
| $o_{s_{cap}}$       | RAM-Kapazität des Servers $s$   |
| $o_v$               | Allozierter RAM der virtuellen Maschine $v$   |
| $\vec{o}_{V_{max}}$ | Vektor mit dem allokierten RAM aller virtuellen Maschinen $v \in V$                                   |
| $P_{base}$          | Elektrische Grundleistung Rechenzentrum   |
| $P_{c_r}$           | Statische elektrische Leistung des IT-Raumes $r$  |
| $P_{chiller}$       | Elektrische Leistung der Kältemaschinen   |
| $P_{cooling}$       | Elektrische Leistung Kühlung gesamt   |
| $P_{crah_r}$        | Elektrische Leistung der Umluftkühlgeräte in IT-Raum $r$  |
| $P_{dc}$            | Elektrische Leistung Rechenzentrum gesamt   |
| $P_{el}$            | Elektrische Leistung allgemein  |
| $P_k$               | Elektrische Leistung aller Server im Rack $k$   |
| $\vec{P}_{K_r}$     | Vektor, der die elektrische Leistung aller Racks im IT-Raum $r$ beinhaltet                            |
| $P_r$               | Elektrische Leistung des IT-Raums $r$   |
| $P_s$               | Elektrische Leistung des Servers $s$  |
| $P_{s_{max}}$       | Maximal mögliche elektrische Leistung des Servers $s$   |
| $P_{server}$        | Elektrische Leistung aller Server im Rechenzentrum  |
| $P_{ups}$           | Elektrische Leistung inklusive Eigenbedarf einer USV  |
| $P_{ups_{cap}}$     | Elektrische Anschlusskapazität einer USV  |
| $P_{ups_{load}}$    | Aktuelle anliegende elektrische Leistung einer USV (der zu versorgenden Geräte)                       |

|                    |  |
|--------------------|--|
| $\psi_{eva_{max}}$ | Maximal erlaubte Evakuierungszeit aller Server   |
| $\psi_{eva_s}$     | Evakuierungszeit des Servers $s$   |
| $\psi_{mig_v}$     | Migrationszeit der virtuellen Maschine $v$   |
| $\vec{\psi}_{eva}$ | Vektor, der die maximal erlaubte Evakuierungszeit aller Server beinhaltet  |
| $\dot{Q}_{crah}$   | Wärmeleistung, die ein einzelnes Umluftkühlgerät abtransportieren kann   |
| $\dot{Q}_{crah_r}$ | Wärmeleistung, die alle Umluftkühlgeräte im IT-Raum $r$ abtransportieren können                                    |
| $\dot{Q}_r$        | Wärmeleistung im IT-Raum $r$   |
| $\vec{Q}_R$        | Vektor der einzelnen Wärmeleistungen aller IT-Räume  |
| $\dot{Q}_s$        | Wärmeleistung des Servers $s$  |
| $\dot{Q}_{therm}$  | Wärmeleistung allgemein  |
| $\dot{Q}_{total}$  | Gesamte anfallende Wärmeleistung aller IT-Räume  |
| $R$                | Menge aller IT-Räume   |
| $R_{active}$       | Menge aller aktiven IT-Räume in einer Allokation   |
| $r$                | Bezeichnung eines spezifischen IT-Raums  |
| $\rho_K$           | Variationskoeffizient von $\sigma_K$ (normalisierte Gleichverteilung der Leistungen in den Racks in einem IT-Raum) |
| $\rho_Q$           | Variationskoeffizient von $\sigma_Q$ (normalisierte Gleichverteilung der Wärmeleistungen der IT-Räume)             |
| $S$                | Menge aller Server im Rechenzentrum  |
| $S_k$              | Menge aller Server im Rack $k$   |
| $S_r$              | Menge aller Server im IT-Raum $r$  |
| $\vec{S}$          | Vektor aller Server im Rechenzentrum, sortiert nach Energieeffizienz   |
| $s$                | Bezeichnung eines spezifischen Servers   |
| $spm_s(\dots)$     | Modellfunktion, die die elektrische Leistung des Servers $s$ abschätzt   |
| $stm_s(\dots)$     | Modellfunktion, die die Temperatur der warmen ausgehenden Luft des Servers $s$ abschätzt                           |
| $\sigma_K$         | Standardabweichung der Gleichverteilung der Leistung der Racks in einem IT-Raum                                    |
| $\sigma_Q$         | Standardabweichung der Gleichverteilung der Wärmeleistungen aus $\vec{Q}_R$  |

|                       |   |
|-----------------------|---|
| $t_{air_c}$           | Kaltlufttemperatur im Kaltgang der IT-Räume   |
| $t_{air_w}$           | Warmlufttemperatur in den IT-Räumen   |
| $t_{s_{cold}}$        | Eingehende (kalte) Lufttemperatur des Servers $s$   |
| $t_{s_{cmax}}$        | Erlaubtes Maximum der eingehenden (kalten) Lufttemperatur des Servers $s$                             |
| $t_{s_{warm}}$        | Ausgehende (warme) Lufttemperatur des Servers $s$   |
| $t_{s_{wmax}}$        | Erlaubtes Maximum der ausgehenden (warmen) Lufttemperatur des Servers $s$                             |
| $t_{water_c}$         | Kaltwassertemperatur im Kühlkreislauf   |
| $t_{water_w}$         | Warmwassertemperatur im Kühlkreislauf   |
| $\vec{t}_{max}$       | Vektor der die maximalen Ablufttemperaturen aller Server beinhaltet                                   |
| $\overline{\Delta T}$ | Logarithmisches Mittel der Temperaturdifferenz  |
| $u_\epsilon$          | CPU-Auslastungspuffer   |
| $u_{ref}$             | CPU-Auslastung einer Referenz-VM  |
| $u_s$                 | CPU-Auslastung des Servers $s$  |
| $u_{scap}$            | CPU-Kapazität des Servers $s$   |
| $u_v$                 | CPU-Auslastung der virtuellen Maschine $v$  |
| $u_{vmax}$            | Maximum der CPU-Auslastung der virtuellen Maschine $v$ nach Auslastungsanalyse und Vorhersagemethodik |
| $\vec{u}_{Vmax}$      | Vektor mit den Komponenten des Maximums von $u_{vmax}$ für alle $v \in V$                             |
| $ukm(\dots)$          | Modellfunktion, die die mögliche abtransportierbare Wärmeleistung eines Umluftkühlgerätes abschätzt   |
| $U \cdot A$           | Charakterisierung der Wärmeübertragerfläche (Betrachtung erfolgt nur gemeinsam)                       |
| $V$                   | Menge aller virtueller Maschinen im Rechenzentrum   |
| $v$                   | Bezeichnung einer spezifischen virtuellen Maschine  |
| $X$                   | Menge aller Ressourcenkapazitäten aller Server  |



# Abkürzungsverzeichnis

|        |  |
|--------|--|
| CFD    | Computational Fluid Dynamics, oder auch numerische Strömungsmechanik   |
| COP    | Coefficient Of Performance   |
| CPU    | Central Processing Unit  |
| DaMucA | Data Center Multi-criteria Aware Allocation, Name des entwickelten Optimierungsverfahrens                                      |
| DCIM   | Data Center Infrastructure Management  |
| DVFS   | Dynamic Voltage and Frequency Scaling  |
| FF     | First Fit, Klasse der First-Fit-Heuristiken  |
| FFD    | First Fit Decreasing, Heuristik zur Lösung des Bin-Packing-Problems  |
| GA     | Genetischer Algorithmus  |
| GGA    | Gruppierender genetischer Algorithmus  |
| HPC    | High Performance Computing   |
| IT     | Informationstechnik, hier meist als Oberbegriff für die technischen Geräte (insbesondere Server und Netzwerk) im Rechenzentrum |
| LPM    | Last- und Powermanagement  |
| MB     | Megabyte   |
| MHz    | Megahertz  |
| NTU    | Number of Transfer Units   |
| PDU    | Power Distribution Unit  |
| RAM    | Random Access Memory   |
| SLA    | Service Level Agreement  |
| UKG    | Umluftkühlgerät  |
| USV    | Unterbrechungsfreie Stromversorgung  |
| VM     | Virtuelle Maschine   |