



FAKULTÄT II – INFORMATIK, WIRTSCHAFTS- UND
RECHTSWISSENSCHAFTEN
DEPARTMENT FÜR INFORMATIK

Entwicklung einer auslastungsabhängigen Energieeffizienzmetrik für Rechenzentren

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 Daniel Schlitt

geboren am 24.01.1984 in Leer

Gutachter

Prof. Dr.-Ing. Wolfgang Nebel
Apl. Prof. Dr.-Ing. habil. Jürgen Sauer

Tag der Disputation: 17. Mai 2016

Zusammenfassung

Der Ausbau von Rechenzentren (RZ) durch die vermehrte Nutzung von Internetdiensten sowie durch das stetig zunehmende Datenaufkommen ist verantwortlich für steigende Betriebskosten der RZ. Da die Energiekosten einen maßgeblichen Anteil der Betriebskosten darstellen, liegt eine Anhebung der Energieeffizienz im Interesse der RZ-Betreiber. Als Grundvoraussetzung für Optimierungsmaßnahmen muss allerdings zunächst der Ist-Zustand der Energieeffizienz bekannt sein, welcher mit Hilfe von Energieeffizienzmetriken ermittelt werden kann.

Im Rahmen dieser Arbeit wurde mit der Load Dependent Energy Efficiency (LDEE) eine Energieeffizienzmetrik für Rechenzentren entwickelt, welche die wichtigsten Fragestellungen von RZ-Verantwortlichen beantwortet. Neben dem allgemeinen Überblick über die Energieeffizienz (1) umfassen diese einen Vergleich mit anderen Rechenzentren (2), die Identifikation der Quellen für Ineffizienzen (3) sowie die Abschätzung der Einsparungen durch potentielle Optimierungsmaßnahmen (4). Während gängige Metriken nur einen allgemeinen Überblick liefern, ist die LDEE aufgrund ihres Konzepts bei allen vier Fragestellungen anwendbar.

Die LDEE basiert auf auslastungsabhängigen Performance- und Powermodellen einzelner RZ-Komponenten sowie einem IT-Lastmodell zur Abschätzung der Verteilung der Systemauslastungen und fasst diese zur Betrachtung des gesamten Rechenzentrums geeignet zusammen. Die Modelle setzen sich dabei aus Parametern zusammen, die aus öffentlich verfügbaren Datenbanken oder Datenblättern gespeist werden können, so dass eine Modellcharakterisierung ohne Messungen im Produktivbetrieb durchführbar ist. Für den Fall, dass auf der Infrastrukturseite detaillierte Daten fehlen, wurde zudem mit der Infrastructure Power Adaptability (IPA) eine Ersatzkennzahl entwickelt, die den dynamischen Infrastrukturoverhead aus dem historischen Verlauf der Power Usage Effectiveness (PUE) ermittelt. Die IPA kann in Kombination mit der PUE dazu genutzt werden, die Leistungsaufnahme der Infrastruktur in Abhängigkeit zur Auslastung abzuschätzen.

Die Einzelmodelle sowie die Metrik als Ganzes wurden in verschiedenen Evaluationen untersucht. Das IT-Lastmodell und die Powermodelle wurden direkt mit verfügbaren Messwerten verglichen und besitzen abgesehen von der Freiluftkühlung jeweils einen relativen Fehler von unter 5%. Die Performancemodelle sowie die Gesamtmetriken wurden hingegen mit alternativen Verfahren verglichen und den von den Fragestellungen abgeleiteten Anforderungen gegenübergestellt. In weiteren Teilevaluationen wurde diesbezüglich die breitere Anwendbarkeit der LDEE gegenüber den Alternativen demonstriert.

Abstract

Data centers' total operational expenditures (OpEx) are rising due to the need for data center extensions caused by the increasing data center service usage as well as continuously growing data volumes. In order to save energy costs, which represent a significant share of the OpEx, the data center operator's focus is shifting towards energy efficiency improvements. However, before optimizations can be implemented, one has to gather information about the actual state of the data center which can be done with the aid of energy efficiency metrics.

Within the scope of this thesis, the Load Dependent Energy Efficiency (LDEE) – an energy efficiency metric for data centers which is applicable on a broad spectrum of use cases – has been developed. Among providing a general overview of energy efficiency (1) further use cases are fair comparisons of data centers' energy efficiencies (2), the identification of sources causing inefficiencies (3), and estimations of achievable savings by potential optimization measures (4). While common metrics may only provide a general overview on energy efficiency, the LDEE may also be used for supporting the optimization processes.

The LDEE makes use of load dependent performance and power models for single data center components as well as an IT load model abstracting the distribution of system utilization on all IT systems. These models are then combined to represent the total data center performance and power in relation to estimated usage. The component models base on publicly available data sources such as benchmark results in databases or operating parameters in data sheets. Thus, there is no need in actual measurements which may disrupt a data center's operation or at least constitute a high effort. In the case of missing models or data for infrastructure components, the Infrastructure Power Adaptability (IPA) metric, which also has been developed in the context of this work, may be applied. The IPA uses traces from the Power Usage Effectiveness (PUE) to derive the actual dependency of infrastructure power to IT power. By combining PUE with IPA the load dependent infrastructure power can be estimated.

The component models and the LDEE as a whole were analyzed in several evaluations. The IT load model and the power models were compared directly to available measured data and have relative errors of less than 5%, except for the free cooling model. In contrast, the performance models as well as the whole metric were applied on synthetic data center data and compared with alternate approaches. In a discussion, the results are examined regarding the requirements for energy efficiency metrics, which have been derived from the aforementioned use cases. Further evaluations demonstrate the broader application possibilities of the LDEE compared to the alternative approaches.

Inhaltsverzeichnis

1	Einleitung	11
1.1	Wissenschaftlicher Beitrag	15
1.2	Aufbau des Dokuments	16
2	Grundlagen	19
2.1	Definition Rechenzentrum	19
2.1.1	Ausrichtung von Rechenzentren	19
2.1.2	Verfügbarkeitsklassen	21
2.1.3	Service-Level-Agreement und Quality of Service	22
2.1.4	Architektur	22
2.2	RZ-Infrastruktur	24
2.2.1	Stromversorgung	24
2.2.2	Klimatisierung	25
2.3	IT-Systeme	28
2.3.1	Server	28
2.3.2	Storage	29
2.3.3	Netzwerk	29
2.4	Software	31
2.4.1	Anwendungen	31
2.4.2	Systemsoftware	31
2.4.3	Managementsysteme	32
2.5	Energieeffizienz	35
2.5.1	Definition der Energieeffizienz	35
2.5.2	Definition im RZ-Kontext	36
2.5.3	Abgrenzung	37
2.5.4	Parameter	39
2.6	Verwandte Arbeiten	40
2.6.1	Energieeffizienzmetriken für Rechenzentren	41
2.6.2	Energiebenchmarks für IT-Systeme	44
2.6.3	Powermodelle	45
2.6.4	Performancemodelle	47
3	Definition der LDEE Metrik	51
3.1	Anwendungsfälle für Energieeffizienzmetriken	51
3.1.1	Bestimmung der Energieeffizienz	52
3.1.2	Vergleich der Energieeffizienz	52

Inhaltsverzeichnis

3.1.3	Identifizierung von Ineffizienzen	52
3.1.4	Bewertung von Optimierungsmaßnahmen	52
3.1.5	Exploration von potentiellen Modifikationen	53
3.1.6	Auswirkung auf Kosten	53
3.2	Anforderungen	53
3.2.1	Anmerkung zu den Anforderungen	55
3.3	Vergleich von Energieeffizienzmetriken	56
3.3.1	Zusammenfassung	61
3.4	Definition der LDEE	61
3.4.1	Definition der Hauptfunktion	63
3.4.2	Funktionen für Performance und Power	63
3.4.3	Auslastungsabhängigkeit	64
3.4.4	Eingabeparameter	64
3.4.5	RZ-Klassen	65
3.4.6	Definition des Resultats	67
4	Modellbildung	73
4.1	Beispielszenario	73
4.2	Aufbau der RZ-Modelle	75
4.3	IT-Lastmodell	76
4.3.1	Anforderungen	77
4.3.2	Modellansatz	77
4.3.3	Konvertierung der Auslastung	79
4.3.4	Bestimmung der aktiven Server	81
4.3.5	Bestimmung der Auslastung	83
4.4	Performancemodelle	85
4.4.1	Anforderungen	86
4.4.2	Bestimmung der Performance	86
4.4.3	Aufbau der Performancemodelle	88
4.4.4	Server	93
4.4.5	Storage	96
4.4.6	Netzwerk	97
4.4.7	Fehlende Performancewerte	99
4.4.8	Kombiniertes RZ-Performancemodell	100
4.5	Powermodelle	105
4.5.1	Anforderungen	105
4.5.2	Bestimmung der Leistungsaufnahme	105
4.5.3	Server	106
4.5.4	Storage	109
4.5.5	Netzwerk	112
4.5.6	USV	113
4.5.7	Klimatisierung	114
4.5.8	Kombiniertes RZ-Powermodell	121

4.6	Evaluation	124
4.6.1	IT-Lastmodell	124
4.6.2	RZ-Performancemodell	128
4.6.3	RZ-Powermodell	135
5	Definition der IPA Metrik	137
5.1	Motivation	137
5.2	Power Variability	138
5.2.1	Konzeption	139
5.2.2	Beispiel	140
5.3	Infrastructure Power Adaptability	142
5.3.1	Konzeption	143
5.3.2	Beispiel	145
5.4	Anwendung als Ersatzkennzahl	147
5.5	Evaluation	149
5.5.1	Vergleich IPA mit PUE Scalability	149
5.5.2	Genauigkeit bei Nutzung als Ersatzkennzahl	154
6	Prozessintegration	157
6.1	Umsetzung	157
6.1.1	Gesamtkonzept	158
6.1.2	Anbindung an DCIM	159
6.1.3	Anbindung an Cloud Computing Manager	163
6.1.4	Zugriff auf Benchmarks	164
6.1.5	Verknüpfung von Komponenten und Modellen	164
6.1.6	Ablauf der Anwendung	166
6.2	Anwendungsbeispiel	167
6.2.1	RZ-Hardware	168
6.2.2	RZ-Software	170
6.2.3	Datenbasis	171
6.2.4	Evaluation	172
7	Fazit	187
	Literaturverzeichnis	191
	Abbildungsverzeichnis	207
	Tabellenverzeichnis	209
	Abkürzungsverzeichnis	211

1 Einleitung

Rechenzentren (RZ) besitzen vielfältige Anwendungen in den unterschiedlichsten Branchen für sowohl geschäftliche als auch private Zielgruppen [HC14]. Typische Anwendungsfelder sind dabei unter anderem die Implementierung von Geschäftsprozessen in Unternehmen, das Angebot dynamischer Webapplikationen und Speicherdienste über das Internet (Cloudservices), die Bereitstellung frei konfigurierbarer Server- sowie Storagehardware oder die massiv-parallele Verarbeitung von großen Datenmengen. Rechenzentren in Kombination mit den Kommunikationsnetzen bilden somit sprichwörtlich das Rückgrat unserer modernen Informationsgesellschaft.

Wie Untersuchungen der letzten Jahre zeigen, wird der Trend einer steigenden RZ-Nutzung auch zukünftig spürbar sein. Beispielsweise erwarten Analysten von Arthur D. Little und dem eco-Verband in [KSS⁺13] eine durchschnittliche Wachstumsrate von durchschnittlich 11% über alle Segmente des deutschen Internetmarkts für die Jahre 2012 bis 2016. Dabei wird für internetbezogene Dienste und Anwendungen wie Cloudservices, Hosting oder Content-Delivery-Netzwerke mit 32,7% pro Jahr ein besonders starkes Wachstum erwartet. Ein ähnliches Ergebnis liefert auch die Studie von 451 Research, welche ein jährliches Wachstum des Cloud-Computing-Markts von 36% prognostizieren¹. Ebenso wächst nach einer Studie [Cis15] der Cisco Systems der IP-Datenverkehr über das Internet mit 23% jährlich. Da diese Daten in der Regel von oder zu Rechenzentren übertragen werden, müssen entsprechende Kapazitäten in den Standorten vorhanden sein bzw. ausgebaut werden.

Der Trend der steigenden Nutzung wirkt sich somit auf die Größe und Anzahl der Rechenzentren aus. Nach Hintemann et al. [HC14] stieg die Anzahl der physikalischen Server in deutschen Rechenzentren zwischen 2003 und 2013 um ca. 50% auf 1,6 Mio. Die Anzahl aller Server inklusive virtueller Maschinen (VM) habe sich im gleichen Zeitraum sogar mehr als vervierfacht. Zudem sei eine Konzentration der Server in größeren Rechenzentren erkennbar, da die Anzahl der Rechenzentren mit einer Fläche von $\geq 100 \text{ m}^2$ deutlich anstiegen (ca. 25%). Auch eine Marktanalyse der Serververkäufe, die von der International Data Corporation (IDC) Anfang 2015 durchgeführt wurde, demonstriert weiterhin steigende Absatzzahlen². Die Wachstumsrate der Einnahmen betrug hierbei 17,2% und die der Absatzzahlen 8,4% im Vorjahresvergleich.

Diese Entwicklung der Rechenzentren hat naturgemäß auch Auswirkungen auf

¹<http://www.firstpost.com/business/cloud-computing-market-revenue-approach-20-billion-end-2016-2074901.html> (letzter Zugriff: 26.11.2015)

²<http://www.idc.com/getdoc.jsp?containerId=prUS25643215> (letzter Zugriff: 26.11.2015)

1 Einleitung

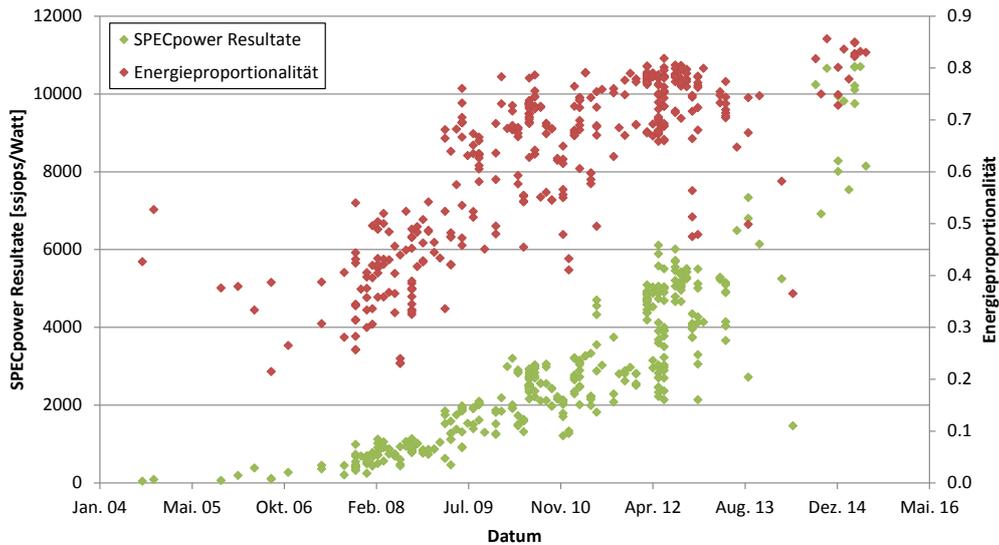


Abbildung 1.1: Entwicklung der Energieeffizienz und -proportionalität nach SPECpower_ssj2008

den Energiebedarf. Einzelne Serversysteme sind zwar im Laufe der Zeit deutlich energieeffizienter geworden, wie es auch in Abbildung 1.1 anhand der Benchmarkergebnisse³ des SPECpower_ssj2008-Benchmarks ersichtlich ist. Ebenso stieg die Energieproportionalität der Server an, welche den Anteil des dynamischen Bereichs der Leistungsaufnahme (Differenz zwischen minimaler und maximaler Leistungsaufnahme) im Verhältnis zur maximalen Leistungsaufnahme beschreibt (vgl. [RPE11]). Dennoch bedürfen Rechenzentren aufgrund ihres Wachstums immer mehr elektrische Energie. So benötigten laut Koomey [Koo11] zwischen 2005 und 2010 Rechenzentren weltweit 56% mehr Energie als in den Jahren 2000 bis 2005. Insgesamt entsprach der Energiebedarf der Rechenzentren ca. 1,3% der weltweiten Energieaufnahme im Jahr 2010. In Deutschland haben Rechenzentren mit ca. 10 TWh einen Anteil von 1,8% am gesamten Bedarf elektrischer Energie, vgl. [HC14]. Einen steigenden Energiebedarf von Rechenzentren hat auch eine globale Befragung durch die DatacenterDynamics im Jahr 2012 ergeben⁴. So stieg die durchschnittliche Leistungsaufnahme zwischen 2011 und 2012 von 24 GW auf 38 GW, was einem Wachstum von 63% entspricht.

Durch den Ausbau der RZ-Komponenten und dem damit steigenden Energiebedarf nehmen die Energiekosten einen wesentlichen Anteil der Gesamtkosten (TCO – Total Cost of Ownership) eines Rechenzentrums ein, welcher aufgrund sinkender Hardwarebeschaffungskosten noch weiter ansteigt [Gen14]. Der An-

³https://www.spec.org/power_ssj2008/results/power_ssj2008.html
(letzter Zugriff: 26.11.2015)

⁴<http://www.computerweekly.com/news/2240164589/Datacentre-power-demand-grew-63-in-2012-Global-datacentre-census> (letzter Zugriff: 26.11.2015)

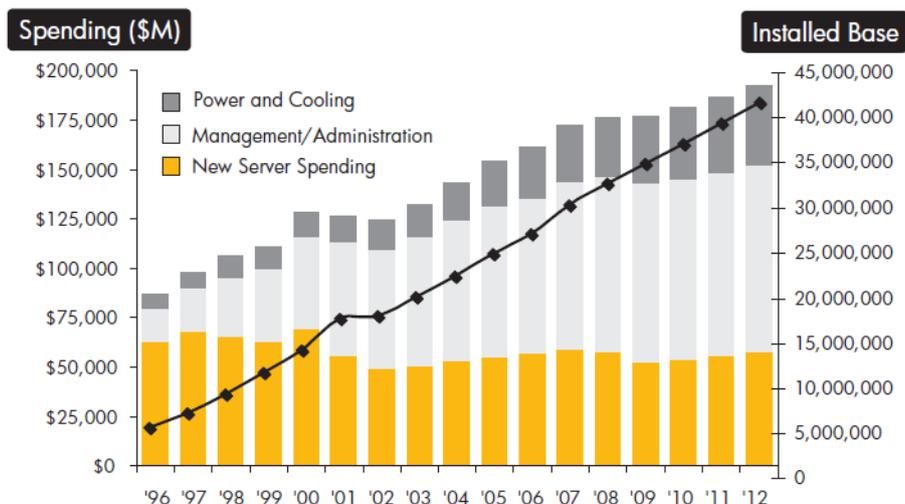


Abbildung 1.2: Aufteilung der Kosten von Rechenzentren – Grafik unverändert aus [Hew09] zitiert

teil beläuft sich je nach Rechenzentrum auf zwischen 15% [GHMP08, HA14] und 30%⁵ der TCO. Diese Zahlen decken sich auch mit den in Abbildung 1.2 dargestellten Abschätzungen, die einen steten Anstieg der Energiekosten aufzeigen [Hew09].

Um die Betriebskosten (OpEx – Operational Expenditure) eines Rechenzentrums zu verbessern, ist ein möglicher Ansatz die Reduktion des Energiebedarfs. Typische Methoden sind hierbei unter anderem der Austausch ineffizienter Komponenten durch (zumeist modernere) effizientere Komponenten oder die Optimierung des Betriebs der vorhandenen Komponenten durch Anpassung der Konfiguration oder durch Einsatz eines übergreifenden Managements. Doch um über den besten Optimierungsansatz entscheiden zu können, sind Informationen über den Ist-Zustand des Rechenzentrums notwendig. An dieser Stelle setzen Energieeffizienzmetriken an. Diese haben das Ziel die Betriebsdaten, die z.B. durch Messungen ermittelt werden, geeignet zusammenzufassen, um somit einen Überblick über die aktuelle Effizienz hinsichtlich der sinnvollen Nutzung der aufgenommenen Energie zu liefern.

Die Power Usage Effectiveness (PUE) ist die De-Facto-Standardmetrik zur Bewertung und zum Vergleich der Energieeffizienz von Rechenzentren [BKST13]. Sie erfährt eine breite Anwendung von Cloudrechenzentren über Konzerne (z.B. veröffentlichen Google und Facebook regelmäßig ihre Ergebnisse⁶) bis hin zu mittelständischen Unternehmen und wird auch zur Anforderungsdefinition geplan-

⁵<http://www.zdnet.com/article/energy-costs-mean-tough-decisions-for-datacenter-owners/> (letzter Zugriff: 26.11.2015)

⁶<http://www.techrepublic.com/article/data-centers-focus-on-pue-in-their-quest-to-use-electricity-efficiently/> (letzter Zugriff: 26.11.2015)

1 Einleitung

ter Rechenzentren eingesetzt⁷. Die PUE wurde von The Green Grid allerdings nie zum Zweck der Energieeffizienzbewertung entwickelt, sondern sollte nur den energetischen Overhead durch die Infrastruktur darstellen. Entsprechend berücksichtigt die PUE nur die Energiebedarfe und nicht die getätigte Arbeit. Dadurch kann sie beispielsweise nicht zur Beurteilung von Optimierungen beim Serverbetrieb wie z.B. durch Virtualisierung eingesetzt werden, vgl. [ND08].

Zu diesem Zweck gibt es weitere Metriken wie die Data Center Energy Productivity [HMP⁺09] oder die Data Center Performance per Energy [Gre12], die neben der Energie auch die Arbeit bzw. Produktivität des Rechenzentrums betrachten. Diese wird aus Messungen an den Produktivsystemen von vordefinierten Kennzahlen wie die Anzahl abgearbeiteter Aufgaben (Tasks) oder durchgeführter Operationen pro Sekunde ermittelt. Ein großer Nachteil entsprechender Metriken ist der nicht zu vernachlässigende Aufwand durch die Vermessung aller Produktivsysteme, sofern diese überhaupt vollumfänglich realisiert werden kann. Zudem verhindert die Abhängigkeit von konkreten Messungen eine Anwendung auf abstrakten RZ-Betrachtungen z.B. zur Planung von Erweiterungen.

Im Rahmen dieser Arbeit wurde daher mit der Load Dependent Energy Efficiency (LDEE) eine Energieeffizienzmetrik entwickelt, die das Verhältnis zwischen der Leistungsfähigkeit (Performance) und der Leistungsaufnahme (Power) eines Rechenzentrums in Abhängigkeit zur Auslastung betrachtet. Die Grundidee ist dabei, die Gesamtperformance sowie -power des Rechenzentrums durch die Kombination einzelner, auslastungsabhängiger Komponentenmodelle zu ermitteln. Ein einzelnes Modell liefert dabei eine Abschätzung der Performance oder Power einer RZ-Komponente für eine gegebene Auslastung. Die Komponentenmodelle sind dabei so ausgelegt, dass eine Anwendung keine aktuellen Messdaten erfordert. Eine Anpassung bzw. Charakterisierung der Modelle auf die konkret eingesetzten Komponenten wird mit öffentlich verfügbaren Daten wie Benchmarkresultaten oder Datenblättern durchgeführt.

Durch dieses Vorgehen ergeben sich diverse Vorteile der LDEE gegenüber bislang bekannten Energieeffizienzmetriken. So kann die Energieeffizienz für beliebige Arbeitslasten bzw. Auslastungen und nicht nur für die tatsächlich anliegende ermittelt werden. Vergleiche zwischen Rechenzentren können damit unabhängig der jeweiligen Arbeitslasten durchgeführt werden. Dies ermöglicht einen fairen Vergleich der RZ-Infrastrukturen. Ebenso kann mit der LDEE bestimmt werden, welches Rechenzentrum für eine bestimmte Arbeitslast effizienter ist. Die Unabhängigkeit von realen Messdaten ermöglicht zudem Analysen potentieller Anpassungen des Rechenzentrums, bevor diese angewandt werden. Dazu gehört der Austausch von HW-Komponenten ebenso wie die Änderung von HW- und SW-Konfigurationen. Voraussetzung hierfür ist, dass etwaige Konfigurationsänderungen durch die vorhandenen Modelle abgedeckt werden und dass Modelle der ausgetauschten Komponenten verfügbar sind. Des Weiteren ermöglicht der

⁷z.B. <http://www.commscope.com/Solutions/Modular-Data-Center-Solutions/>
(letzter Zugriff: 26.11.2015)

Ansatz auf Basis von Komponentenmodellen neben einer allgemeinen Abschätzung der Energieeffizienz eines gesamten Rechenzentrums auch die Betrachtung einzelner Komponenten. Dadurch können Ineffizienzen nicht nur identifiziert, sondern auch die Quelle bis auf Komponentenebene zurückverfolgt werden.

Die LDEE stellt durch die größere Funktionalität eine deutliche Verbesserung gegenüber den bislang eingesetzten Energieeffizienzmetriken dar.

1.1 Wissenschaftlicher Beitrag

Im Rahmen dieser Arbeit entstand mit der LDEE eine Energieeffizienzmetrik, die nicht nur analog zu anderen Metriken einen Überblick über die Gesamteffizienz liefert, sondern ebenso Vergleiche zwischen nahezu frei konfigurierbaren RZ-Konstellationen (beschränkt durch den Umfang und die Verfügbarkeit von Komponentenmodellen) sowie Effizienzbetrachtungen einzelner Komponenten ermöglicht. Während der Entwicklung der Energieeffizienzmetrik sind zudem weitere Ergebnisse entstanden, die einen Beitrag für das Forschungsumfeld der Energieeffizienz in Rechenzentren (Green IT) darstellen.

Anforderungsanalyse: Auf Basis typischer Fragestellungen bzgl. der Anwendung von Energieeffizienzmetriken durch RZ-Verantwortliche wurden die Anforderungen an eine sinnvolle, multifunktionale Metrik erarbeitet. Außerdem wurde die Energieeffizienz im Kontext von Rechenzentren definiert und mögliche Einflussparameter diskutiert. Auf Grundlage dieser Analysen wurde der prinzipielle, modellbasierte Ansatz der LDEE entwickelt und weitere Anforderungen für die jeweiligen Modelle abgeleitet.

IT-Lastmodell: Ein Teilergebnis der LDEE-Entwicklung ist ein IT-Lastmodell zur Abschätzung von Lastverteilungen auf Server- und Stagesysteme. Das Modell berücksichtigt dabei potentiell eingesetzte Lastverteilungsalgorithmen und schätzt sowohl die aktiven, genutzten Systeme als auch den jeweils konkreten Grad der Nutzung bzw. die Auslastung ab.

Performancemodelle: Zur Beurteilung der Leistungsfähigkeit bzgl. der Datenverarbeitung wurden Performancemodelle für Server- und Stagesysteme entwickelt. Im Gegensatz zu verbreiteten Modellen werden diese mit öffentlich verfügbaren Benchmarkergebnissen trainiert bzw. charakterisiert, so dass keine Messungen im Produktivbetrieb notwendig sind. Des Weiteren wurde ein Konzept zur Zusammenfassung der Einzelresultate zu einer gesamten RZ-Performance definiert und hierbei ebenso die Netzwerkperformance inkludiert.

Powermodelle: Die Powermodelle der Server- und Stagesysteme unterscheiden sich von bekannten Modellen aus dem Forschungsumfeld in der Hinsicht, dass

1 Einleitung

diese ebenso wie die Performancemodelle auf öffentlich verfügbaren Daten aufsetzen, so dass keine Messungen im Produktivbetrieb notwendig sind. Die weiteren genutzten Powermodelle basieren auf bekannten Arbeiten oder sind im Rahmen des AC4DC-Projekts [SJS⁺14] entstanden. Ein relevanter Beitrag findet sich allerdings im Ansatz zur Kombination der Modelle zur Ermittlung der gesamten RZ-Power. Dieser Ansatz berücksichtigt die Abhängigkeiten zwischen den einzelnen Komponenten, den Betriebszustand der Klimatisierung (Freiluftkühlung) und den Einfluss der Eingangsparameter (Auslastung, Außentemperaturen).

Form des Resultats: Ein erklärtes Ziel in der Entwicklung der Metrik ist die Darstellung der Ergebnisse in einem geeigneten, verständlichen Format. Hierzu wurde zur Repräsentation der Performance eine Referenzeinheit definiert, die mit der Auslastung skaliert und mit der Leistungsaufnahme in Relation gesetzt wird, um letztlich einen relativen Energieeffizienzwert zu erhalten. Zudem wurden unterschiedliche Darstellungsformen für diverse Anwendungszwecke der LDEE definiert und diskutiert.

Kennzahl für Adaptivität der Infrastruktur: Neben der LDEE zur Bewertung der Energieeffizienz wurde mit der Infrastructure Power Adaptability (IPA) eine Metrik zur Darstellung der Anpassungsfähigkeit der Infrastruktur gegenüber den IT-Systemen hinsichtlich ihrer Leistungsaufnahmen entwickelt. Die IPA stellt eine Erweiterung der Power Usage Effectiveness bzgl. der Dynamik der Leistungsaufnahmen (z.B. aufgrund von Schwankungen in der Systemauslastung) dar. Sind keine Powermodelle für Infrastrukturkomponenten vorhanden, kann sie zudem als Ersatzkennzahl zur Abschätzung der Infrastrukturleistungsaufnahme innerhalb der LDEE dienen.

1.2 Aufbau des Dokuments

Nach der Motivation der Arbeit in diesem Kapitel folgt in Kapitel 2 zunächst eine Beschreibung des Aufbaus von Rechenzentren und der enthaltenen Komponenten sowie die Definition der Energieeffizienz im aufgezeigten Kontext. Zudem wird der aktuelle Stand der Technik, speziell im Hinblick auf die im Rahmen dieser Arbeit genutzten Verfahren und Techniken, erläutert. Kapitel 3 befasst sich dann mit der Definition der LDEE, wobei zuvor Anwendungsfälle und Anforderungen von Energieeffizienzmetriken definiert und verbreitete Metriken bzgl. dieser untersucht werden. Aufbauend auf der Definition der LDEE werden im darauffolgenden Kapitel 4 die einzelnen Bestandteile der Metrik – also die RZ-Komponentenmodelle – im Detail beschrieben. Entsprechend wird die Modellbildung der Performance, der Power sowie der IT-Auslastung sowohl für RZ-Komponenten als auch für ein ganzes Rechenzentrum erläutert und um eine Evaluation der (kombinierten) Modelle ergänzt. Kapitel 5 behandelt die Definition und Evaluation einer Metrik zur Bewertung der Adaptivität der Leistungsauf-

nahme von Infrastrukturkomponenten bei eventuell vorliegenden Schwankungen in der IT-Leistungsaufnahme. Diese Metrik kann als Ersatz für fehlende Modelle der RZ-Infrastruktur dienen, was ebenso in Kapitel 5 thematisiert wird. In Kapitel 6 wird eine beispielhafte Umsetzung der LDEE als Erweiterung von verfügbaren Managementsystemen für Rechenzentren zur Integration in die Unternehmensprozesse zum RZ-Management vorgestellt. Zudem wird die Gesamtmetrikt auf Beispielrechenzentren angewandt und die Ergebnisse hinsichtlich der Anforderungen diskutiert. Das Dokument schließt anschließend in Kapitel 7 mit einem Fazit inklusive eines Ausblicks.

2 Grundlagen

Dieses Kapitel behandelt die Energieeffizienz im Kontext von Rechenzentren und die Bestimmung dieser durch Metriken. Dazu wird zunächst der Aufbau und die Funktionsweise von Rechenzentren als Zielplattform beschrieben. Die Beschreibung beginnt mit einer grundsätzlichen Definition in Abschnitt 2.1 und unterteilt sich dann in die drei Themenbereiche Infrastruktur (2.2), IT-Systeme (2.3) sowie die darauf betriebene Software (2.4). Anschließend wird eine Definition der Energieeffizienz jeweils allgemein und im Kontext von Rechenzentren in Abschnitt 2.5 geliefert. Im abschließenden Abschnitt 2.6 wird dann auf die bisherigen Ergebnisse und Fortschritte in der Industrie und Forschung bzgl. der Bewertung der Energieeffizienz im RZ-Kontext eingegangen.

2.1 Definition Rechenzentrum

Ein Rechenzentrum bezeichnet ein Gebäude oder eine Räumlichkeit, in dem zentralisiert die Rechenkapazitäten (Server, Storage, Netzwerk) einer oder mehrerer Firmen oder Organisationen und die zum Betrieb notwendige Infrastruktur (Stromversorgung, Klimatisierung) verortet sind [HFS10]. Der Zweck eines RZ ist die Verarbeitung, Speicherung und Verteilung von Daten. Zur Realisierung eines reibungslosen Betriebs von Rechenzentren werden fünf Designkriterien verfolgt [AP03]. Eine hohe Verfügbarkeit steht für die möglichst durchgehende Bereitstellung der im RZ betriebenen Dienste und Anwendungen. Die Skalierbarkeit sorgt dafür, dass die Rechenkapazitäten dem Bedarf entsprechend bereitgestellt werden. Eine hohe Performance (Leistungsfähigkeit) ist maßgeblich entscheidend, dass überhaupt genug Ressourcen für eine zweckmäßige Skalierung vorhanden sind. Die Sicherheit soll zum einen physikalisch durch Zugangsbeschränkungen und zum anderen virtuell durch den Einsatz entsprechender Maßnahmen (z.B. Firewalls) vor böswilligen Eingriffen schützen. Bei alledem muss die Handhabbarkeit (Manageability) durch die betreuenden Systemadministratoren weiterhin gegeben sein, um auf Anfragen oder bei Problemen akkurat und kurzfristig reagieren zu können.

2.1.1 Ausrichtung von Rechenzentren

Rechenzentren lassen sich ihrer Ausrichtung bzw. ihrem Zweck nach typologieren. Die Typen richten sich dabei häufig nach dem angewandten Geschäftsmodell, basieren zum Teil aber auch auf den konkret betriebenen Arbeitslasten. Die typischen Rechenzentrumsklassen werden nachfolgend beschrieben:

2 Grundlagen

- **Enterprise:** Diese Rechenzentren sind einem Unternehmen (Enterprise) zugeordnet und bieten interne Dienste wie Email, ERP oder CRM an. Sie übernehmen außerdem die Speicherung und Verwaltung der Nutzerdaten und betreiben zum Teil die kompletten Systemumgebungen der Nutzer z.B. als virtualisierte Maschinen. Als Bestandteil eines Unternehmens oder einer Organisation werden entsprechende Rechenzentren durch diese finanziert.
- **Hosting:** Hosting-RZ bieten Kunden die Nutzung ihrer Server-, Storage- und Netzwerkkomponenten gegen Bezahlung an (IaaS – Infrastructure-as-a-Service). Die Nutzer können dabei selbst entscheiden, welche Software sie auf den HW-Systemen betreiben, müssen dafür allerdings auch alles selbst einrichten. Durch den Einsatz von Virtualisierung werden inzwischen vermehrt virtuelle Server angeboten, was einen Schritt in Richtung der im weiteren Verlauf beschriebenen Cloud-RZ darstellt.
- **Co-Location:** Rechenzentren mit der Ausrichtung Co-Location haben ein analoges Geschäftsmodell zu den Hosting-RZ mit dem Unterschied, dass die Kunden ihre Hardware selbst kaufen und einsetzen. Die Kunden bezahlen somit praktisch für den Stellplatz inkl. Stromversorgung, Klimatisierung und Netzanbindung unter Berücksichtigung der ausgehandelten Verfügbarkeit.
- **Cloud:** Cloud-RZ stellen das genaue Gegenteil von Enterprise-RZ dar. In Cloud-RZ werden einer Vielzahl von Kunden unterschiedliche, dedizierte SW-Dienste angeboten (SaaS – Software-as-a-Service), deren Kosten anhand des Nutzungsgrads abgerechnet werden. Die Bandbreite der angebotenen Dienste reicht von Office-Anwendungen über Shopsysteme bis hin zu CRM-Systemen. Mit Platform-as-a-Service als offeneres SW-System zur Anwendungsentwicklung und Cloud-Storage als reine Speicherplattform gibt es noch weitere mögliche Ausrichtungen von Cloud-RZ.
- **High-Performance-Computing (HPC):** HPC-Rechenzentren bieten – im Gegensatz zu den anderen Typen – Nutzern keine Dienste an, sondern dienen der Bearbeitung rechen- oder speicherintensiver, hochparalleler Aufgaben. HPC-Rechenzentren sind vor allem im wissenschaftlichen Bereich vertreten, werden aber auch gelegentlich in der Industrie z.B. bei der Produktentwicklung genutzt. Finanziert werden HPC-Rechenzentren jeweils durch die betreibenden Einrichtungen.

Pulvermüller hat in [Pul08] eine ähnliche Klassifizierung der Rechenzentren vorgenommen und die Anteile der Klassen anhand von Umfragen ermittelt. Dabei fällt der größte Anteil mit 41% auf die 'Interne IT' (Enterprise), gefolgt von Hosting (34%) und Co-Location (18%). Den geringsten Anteil veranschlagen HPC-Rechenzentren mit 7%. Cloud-Rechenzentren wurden nicht separat betrachtet.

2.1.2 Verfügbarkeitsklassen

Der Zweck eines Rechenzentrums ist die Bereitstellung von Rechenressourcen bzw. den darauf betriebenen Anwendungen für Nutzer. Die wichtigste Eigenschaft vor allen anderen ist entsprechend die Verfügbarkeit der angebotenen Dienste. Kann diese nicht sichergestellt werden, kann der Zweck des Rechenzentrums nicht erfüllt werden.

Die Gründe für Ausfälle der angebotenen Dienste sind vielfältig, da der Betrieb von vielen Faktoren abhängt. So kann z.B. die IT-Hardware direkt durch Defekte betroffen sein, die Klimatisierung ausfallen, so dass die IT heruntergefahren werden muss oder der Dienst aufgrund unterbrochener Netzwerkverbindungen nicht mehr erreichbar sein. Damit diese potentiellen Fehlerquellen nicht zu einem Ausfall führen, werden die Komponenten und Strukturen des Rechenzentrums in der Regel redundant ausgelegt. Der Umfang der Redundanzen kann sich dabei stark unterscheiden. Zur besseren Einordnung des Redundanzgrads hat das Uptime Institute daher eine Kategorisierung in die vier nachfolgend beschriebenen Verfügbarkeitsklassen [TSRB06] – die sogenannten Tier I bis IV – eingeführt.

- **Tier I:** Die Verfügbarkeitsklasse Tier I beschreibt die Bereitstellung der Basiskapazitäten ohne explizite Redundanzen. Es muss zwar eine unterbrechungsfreie Stromversorgung (USV) für die IT-Komponenten sowie eine geeignete Klimatisierung vorhanden sein, allerdings sind diese Komponenten und alle Pfade nicht-redundant, so dass ein einzelner Defekt zu einem Ausfall führen kann (Single-Point-of-Failure). Die Verfügbarkeit des Tier I wird mit 99,671% angegeben. Der Aufbau wird beispielhaft in Abbildung 2.1 dargestellt, wobei die jeweiligen Komponenten (und Abkürzungen) in Abschnitt 2.1.4 näher erläutert werden.
- **Tier II:** In Tier II werden weiterhin einfache Versorgungspfade (Strom, Kälte) genutzt, die Komponenten dieser Pfade allerdings redundant ausgelegt. Neben einer etwas höheren Ausfallsicherheit von 99,749% verbessert sich auch die Wartbarkeit der Infrastrukturkomponenten aufgrund der Redundanz.
- **Tier III:** Die Verfügbarkeitsklasse Tier III ergänzt die Redundanzen aus Tier II um jeweils einen zusätzlichen passiven Pfad zur Energie- und Kälteversorgung, die bei einem Ausfall des aktiven Pfades diesen ersetzt. Durch den passiven Versorgungspfad steigt die Verfügbarkeit auf 99,982% und die Wartung oder ein Wechsel von Komponenten ist effektiv ohne Einfluss auf den IT-Betrieb durchführbar.
- **Tier IV:** Tier IV stellt die höchste Verfügbarkeitsklasse nach Definition des Uptime Institutes dar. Es wird ein komplett fehlertoleranter (d.h. kein einziger verbleibender Single-Point-of-Failure) Betrieb mit jeweils zwei gleichzeitig aktiven Versorgungspfaden und eine zusätzliche n+1-Redun-

2 Grundlagen

danz aller Komponenten eines Pfads vorausgesetzt. Die Verfügbarkeit steigt damit auf 99,991%, was einer jährlichen Ausfallszeit von 0,8 h entspricht.

Da die Verfügbarkeit eines Rechenzentrums ein relevantes Nutzungskriterium für Anwender ist, bietet das Uptime Institute einen Zertifizierungsprozess für Rechenzentren an¹. Mit dem resultierenden Zertifikat können die RZ-Betreiber ihre Standorte entsprechend bewerben, da somit potentiellen Kunden eine bestimmte Verfügbarkeit zugesichert werden kann.

Den gleichen Ansatz wie das Uptime Institute verfolgt das deutsche Bundesamt für Sicherheit in der Informationstechnik (BSI) mit einer eigenen Definition von fünf Verfügbarkeitsklassen, welche in [Her13a] beschrieben werden. Die einzelnen Klassen sind ähnlich denen des Uptime Institutes aufgebaut, betrachten insgesamt aber eine noch höhere Verfügbarkeit von bis zu 99,999% (ca. 5min./a Ausfallzeit). Auf die genaue Definition wird in dieser Arbeit nicht näher eingegangen, da im weiteren Verlauf nur auf die Klassifizierung des Uptime Institutes Bezug genommen wird.

2.1.3 Service-Level-Agreement und Quality of Service

Durch die Ausrichtung eines großen Teils der Rechenzentren als Dienstleister wie z.B. bei Hosting- oder Cloud-Rechenzentren (vgl. Abschnitt 2.1.1) ist es üblich die für einen Auftraggeber zu vollbringenden Leistungen in Service-Level-Agreements (SLA, dt. Dienstgütevereinbarung) festzuhalten [BYV⁺09, Mah14]. Typische Inhalte solcher SLA sind unter anderem der Zweck, die Leistungsbeschreibung, die Verfügbarkeit des Dienstes, die Preisgestaltung sowie Vertragsstrafen bei Nichteinhaltung der Abmachungen. Neben den kundenbezogenen SLA zwischen Auftragnehmer und -geber gibt es auch dienstbezogene SLA, die übergreifend für alle Nutzer gelten. Durch die verteilte Ressourcennutzung von Cloud-Diensten sind kundenbezogene SLA in Cloud-Rechenzentren beispielsweise schwierig umsetzbar.

Die Quality of Service (QoS) bzw. Dienstgüte repräsentiert die Anforderungen an einen Dienst, die in den SLA vertraglich festgehalten werden [BYV⁺09, Mah14]. Die Anforderungen manifestieren sich in Leistungskennzahlen wie z.B. Datendurchsatz, Fehlerrate, Latenzen oder Verfügbarkeit, die einen definierten Grenzwert nicht überschreiten dürfen. Damit Kunden die Einhaltung der SLA kontrollieren und damit Vertrauen bzgl. des Anbieters aufbauen können, bieten Rechenzentren typischerweise Mechanismen zur Überprüfung der relevanten QoS-Parameter an.

2.1.4 Architektur

Der grundlegende Aufbau eines Rechenzentrums wird in Abbildung 2.1 dargestellt. Das Schaubild repräsentiert ein RZ der Verfügbarkeitsklasse Tier I (vgl.

¹<https://journal.uptimeinstitute.com/explaining-uptime-institutes-tier-classification-system/> (letzter Zugriff: 26.11.2015)

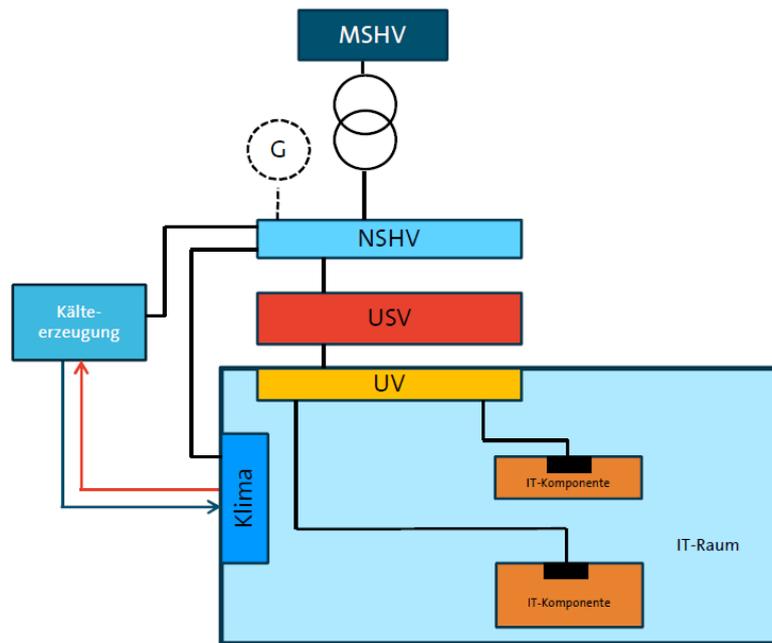


Abbildung 2.1: Tier I Rechenzentrum – Grafik unverändert aus [Her13b] zitiert

Abschnitt 2.1.2) und besitzt dementsprechend keine redundant ausgelegten Komponenten. Der Energieversorger liefert den elektrischen Strom über die Mittelspannungshauptverteilung (MSHV) an die Niederspannungshauptverteilung (NSHV), welche die im RZ lokalisierte Verbindungskomponente zum Energieversorger darstellt. Optional besitzt das Rechenzentrum zusätzlich eine Netzersatzanlage (NEA) bzw. einen Generator (G) zum Ausgleich etwaiger Stromausfälle. Über die unterbrechungsfreie Stromversorgung (USV) und die Stromunterverteilung (UV) werden die IT-Komponenten mit elektrischem Strom versorgt. Die Klimageräte und Kälteerzeugung werden hingegen direkt von der NSHV gespeist.

Wie bereits in Abschnitt 2.1.2 erwähnt wurde, hat die Sicherstellung der Verfügbarkeit eines Dienstes maßgeblichen Einfluss auf die Redundanz in der RZ-Architektur. Auch spezielle, definierte Anforderungen in den SLA können sich auf die Architektur auswirken, z.B. wenn die Nutzung konkreter Kühlmechanismen wie Wasserkühlung zur Zusicherung definierter Betriebstemperaturen in einem Co-Location-RZ verlangt wird. In Abbildung 2.2 wird der Aufbau eines Tier IV Rechenzentrums schematisch dargestellt. Im Vergleich zu Abbildung 2.1 sind die Energieversorgungspfade über zwei unterschiedliche Energieversorger ausgelegt und werden zusätzlich untereinander gekoppelt. Zudem kann die Klimatisierung über einen einzelnen Pfad gewährleistet werden, so dass auch hier eine doppelte Absicherung besteht. Die einzelnen Komponenten in der Stromversorgung und Klimatisierung sind zusätzlich redundant ausgelegt. Damit werden nicht nur die Single-Point-of-Failures eliminiert, sondern selbst bei zwei unabhängigen Ausfä-

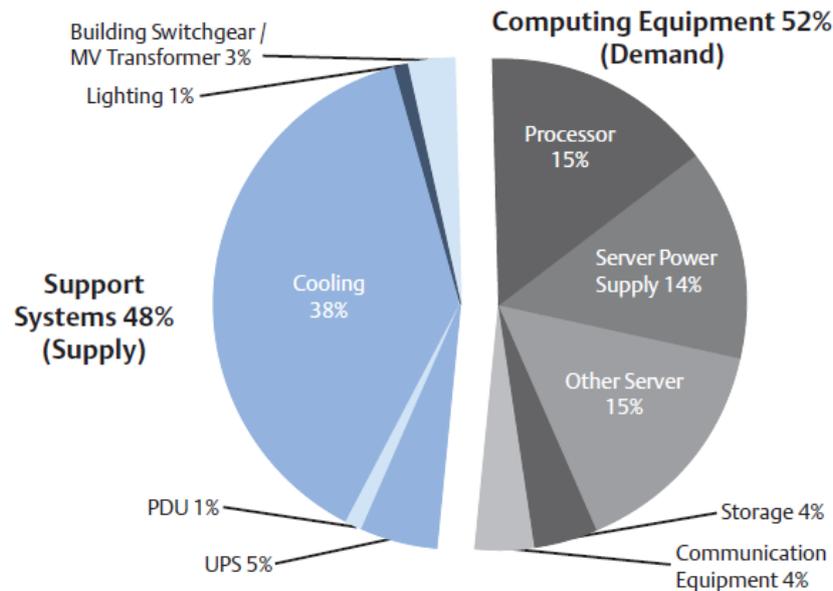


Abbildung 2.3: Aufteilung der Energie auf die RZ-Komponenten – Grafik unverändert aus [Eme09] zitiert

des Dieselgenerators zu überbrücken, ist die IT-Hardware über eine unterbrechungsfreie Stromversorgung (USV) mit der NSHV verbunden. Diese filtert zum einen Netzschwankungen der Spannung oder Frequenz heraus, um negative Auswirkungen auf die Servernetzteile zu verhindern. Zum anderen schaltet sie im Falle eines Totalausfalls auf einen Batteriebetrieb um, der die Zeit bis zum Umschalten auf die NEA überbrückt. Über die Stromunterverteilung wird der von der USV aufbereitete Strom an die IT-Hardware weitergeleitet.

Die Spannungswandlung, die Netzfilterung und das Laden der Batterien sind jeweils mit einem Verlust in der Leistung behaftet. Typische Werte für die Hauptverteilung (Building Switchgear), die USV (UPS) sowie die Unterverteilung (PDU) werden in Abbildung 2.3 angegeben. Aufgrund fehlender Messdaten zur Modellierung werden die NSHV sowie die Stromunterverteilung in dieser Arbeit nicht näher betrachtet. Durch den kleinen Anteil an der Gesamtleistungsaufnahme eines Rechenzentrums ist der Einfluss dieser Komponenten auf die Energieeffizienz allerdings gering.

2.2.2 Klimatisierung

Die Klimatisierung eines Rechenzentrums setzt sich grundsätzlich aus den drei Bereichen der Kälteverteilung, dem Kältetransport und der Kälteerzeugung zusammen. Die Kälteverteilung sorgt dafür, dass die durch die IT-Hardware erwärmte Luft in den Serverräumen herabgekühlt und neu verteilt wird. Die Wärme wird dann über das Kühlmedium von den Serverräumen aus dem Gebäude

2 Grundlagen

transportiert. Die Kälteerzeugung kühlt das Kühlmedium anschließend wieder ab, indem es die Abwärme an die Umgebung überträgt. Das abgekühlte Kühlmedium wird schließlich zurück zur Kälteverteilung der Serverräume transportiert.

Die häufigsten Klimallösungen nutzen die Luft als wärmeführendes Medium im Serverraum. Die Möglichkeiten sind dabei sehr variantenreich [DR14]. Noch immer weit verbreitet sind am Rand der Serverräume platzierte Umluftkühlgeräte bzw. Computer Room Air Handler (CRAH), die die Luft im Raum mittels Lüftern ansaugen, über einen Luft-Wasser-Wärmetauscher mit gekühltem Wasser abkühlen und danach mit weiteren Lüftern wieder in den Raum befördern. Die Effizienz dieser Variante hängt direkt von der Luftführung [Ras12, Her13a] ab, ob die Reihen der Serverschränke (Racks) beispielsweise in Warm- und Kaltgänge eingeteilt werden und ob diese zur Vermeidung einer Vermischung von Kalt- und Warmluft eingehaust sind. Andere Varianten befördern das gekühlte Wasser direkt zu den Racks, wo über Luft-Wasser-Wärmetauscher, die in oder über der Rackreihe, im Unterboden oder an den Serverauslässen platziert sind, die warme Luft abgekühlt wird. Die Positionierung der Kühlgeräte an den Wärmeerzeugern erlaubt eine höhere Server- bzw. Energiedichte pro Rack und ermöglicht insgesamt einen effizienteren Betrieb [DR14]. Alternativ zu luftorientierten Klimallösungen ist auch eine Wasserkühlung², bei der das Kühlwasser direkt über die Mikrochips der Server geleitet wird, möglich. Die Führung von Kältemitteln innerhalb der Serverräume erhöht allerdings auch die Gefahr von Schäden an der IT bei einer Leckage.

Zur möglichst optimalen Einstellung der Betriebsparameter im Serverraum hinsichtlich eines zuverlässigen und energieeffizienten Betriebs veröffentlicht der amerikanische Berufsverband American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) im Abstand weniger Jahre aktualisierte Richtlinien [ASH12]. Die möglichen Einstellungen sind dabei in vier Klassen unterteilt, deren aktuelle Definition in [ASH12] zu finden ist. Die maßgeblichen Klimaparameter sind die Trockentemperatur, welche der üblichen Definition der Außentemperatur entspricht, die Kühlgrenztemperatur, die tiefste durch Verdunstung erreichbare und somit von der relativen Feuchtigkeit abhängige Temperatur, und die relative Luftfeuchtigkeit jeweils im Kaltgang vor dem IT-Equipment. Höhere Kaltgangtemperaturen ermöglichen dabei grundsätzlich einen kostengünstigeren Betrieb, da die Kälteerzeugung durch die höhere Differenz der Wasserrücklauftemperatur zur Außentemperatur effizienter arbeitet. Die im Serverraum realisierbaren Einstellungen richten sich nach der Hardware mit den stringentesten Anforderungen. Während moderne IT-Hardware bereits auf höhere Temperaturen ausgelegt wird, fallen ältere Komponenten häufig noch in die Klassen mit stringenteren Anforderungen, so dass hierdurch die möglichen Betriebsparameter eingeschränkt werden. Neben den erlaubten Parametern gibt ASHRAE für alle vier Klassen eine identische, sehr eingeschränkte, empfohlene

²<http://www.hpcwire.com/2013/12/09/reality-check-liquid-cooling-data-center/>
(letzter Zugriff: 26.11.2015)

Konfiguration an, die zwar für jede IT-Hardware zulässig ist, damit allerdings auch die geringsten Optimierungsmöglichkeiten bietet.

Die aus dem Serverraum an das Kühlmedium übertragene Wärme wird über den Kühlkreislauf zur Kälteerzeugung transportiert. Im Regelfall ist das eingesetzte Kühlmedium Wasser (ggf. mit Glykol versetzt) und der Transport wird entsprechend von Pumpen vorgenommen. Die Leitungen sind als Kreislauf aufgebaut, so dass das Kühlmedium nach dem Abkühlen wieder zu den Klimageräten in den Serverräumen transportiert wird. Zur Kälteerzeugung werden häufig Kompressionskälteanlagen (Chiller) eingesetzt, die durch Verdichtung und Expansion das im Rechneraum erwärmte Kältemittel herabkühlen und die Wärmeenergie an einen weiteren Kreislauf, der die Wärme nach außen leitet, abgibt [Eva12]. Die Wärme wird dann letztlich über Kühltürme, Trockenkühler oder Kondensatoren an die Außenluft abgegeben. Da der Verdichtungsprozess viel Energie benötigt, werden vermehrt energieeffizientere Kühllösungen angewandt. Dazu zählt vor allem die Nutzung der Freiluftkühlung, indem beispielsweise die Chiller bei geeigneten Temperaturen kurzgeschlossen und nur die angeschlossenen Kühltürme zur Abkühlung des Kühlmediums genutzt werden [Her13a]. Eine Alternative ist z.B. die Adsorptionskühlung, die über die Abwärme eines Blockheizkraftwerks (BHKW) betrieben werden kann [DH09].

Die Außenluft kann auch direkt zur Kühlung genutzt werden, wobei die Luft allerdings gefiltert und die Luftfeuchtigkeit ggf. geregelt werden muss, bevor sie in den Rechneraum geleitet wird [Her13a]. Eine weitere Möglichkeit ist die indirekte Nutzung über einen Luft-Luft-Wärmetauscher, bei der durch die Trennung der Kreisläufe keine Filterung notwendig ist [Tor13]. In beiden Fällen werden als einzige aktive Komponenten Lüfter benötigt, so dass der Energiebedarf sehr niedrig ist. Eine intelligente Ausnutzung der Thermik erhöht die Effizienz noch weiter. Aufgrund der Abhängigkeit der Freiluftkühlung von den klimatischen Bedingungen sind allerdings für viele Standorte zusätzliche Klimälösungen notwendig, die bei zu hohen Außentemperaturen die Kühlung übernehmen. Der Anteil der Freiluftkühlung kann jedoch durch eine Anhebung der Servereingangstemperaturen nach ASHRAE [ASH12] erhöht werden. Die Effektivität lässt sich weiterhin durch eine zusätzliche Ausnutzung des Verdunstungseffekts steigern, indem der Wärmetauscher bzw. ein Gitter in der Luftführung mit Wasser bestäubt wird [Tor13, Eme15].

Analog zur Luftkühlung ist ebenso eine Klimatisierung, die vollständig auf Wasser als einziges Kühlmedium setzt, möglich. Das Kühlwasser aus den Rechneräumen wird dabei mit kaltem Wasser aus der Umwelt, z.B. dem Meer, einem Brunnen oder Wärmetauschbecken, abgekühlt³. Auch die Weiternutzung des Warmwassers in Bürogebäuden ist denkbar und zudem leichter realisierbar als bei warmer Abluft⁴.

³www.networkworld.com/article/2166377/data-center/swedish-data-center-saves--1-million-a-year-using-seawater-for-cooling.html (letzter Zugriff: 26.11.2015)

⁴<http://www.datacenterknowledge.com/data-centers-that-recycle-waste-heat/> (letzter Zugriff: 26.11.2015)

Im Rahmen dieser Arbeit wird der typische Fall der RZ-Klimatisierung bestehend aus CRAH, Pumpen sowie Chiller inkl. Freiluftkühlung betrachtet. Zur näheren Betrachtung der weiteren beschriebenen Klimatisierungsoptionen lagen zu wenige Daten vor. Von abweichenden Klimalösungen kann jedoch mit dem in Kapitel 5 beschriebenen Verfahren abstrahiert werden.

2.3 IT-Systeme

Die Informationstechnik (IT) eines Rechenzentrums bildet das Kernstück, um welches herum die Infrastruktur und damit zusammengenommen das Gesamtsystem konzipiert und aufgebaut wird. Die RZ-IT besteht grundsätzlich aus Server-, Storage- und Netzwerkkomponenten. Die Server sind dabei für die Verarbeitung von Daten und die Storage-Systeme für die Speicherung dieser verantwortlich, während das RZ-Netzwerk die Kommunikation der IT untereinander und nach außen ermöglicht.

2.3.1 Server

Ein Server bezeichnet zum einen eine Software, die Clients eine (über ein Netzwerk) zugreifbare Anwendung bzw. Dienst bereitstellt, und zum anderen den Computer (Host), auf dem eine solche Software ausgeführt wird. Im Rahmen dieser Arbeit bezieht sich der Begriff Server nur auf die Hardware, die zum Betrieb einer Server-Client-Architektur genutzt wird.

HW-Server lassen sich in Klassen basierend auf ihren Eigenschaften wie z.B. der Bauart oder dem Anwendungszweck einteilen. Eine Einteilung nach Bauart mit den Kategorien Towerserver, Rackserver, Bladeserver, Unixserver und Mainframes wird beispielsweise von Hintemann et al. in [HFS10] genutzt. Während Towerserver einzeln in eigenen Gehäusen betrieben werden, werden Rackserver in Serverschränke (Racks) und Bladeserver in Bladeenclosures, die wiederum selbst in Racks verortet sind, eingeschoben. Unixserver sind nach Hintemann et al. mittelgroße Server mit einem Unixderivat als Betriebssystem und Mainframes bezeichnen auf Zuverlässigkeit und hohen Datendurchsatz ausgelegte Großrechner. Eine ähnliche Klassifizierung ist die Einteilung in Volume-, Midrange- und High-End-Server. Das Marktforschungsunternehmen IDC teilt die Server dabei entsprechend ihrer Marktpreise⁵ in die Kategorien ein, während Koomey in [Koo07] die Leistungsaufnahme nutzt. Nach [HFS10] lassen sich Tower-, Rack- und Bladeserver den Volumeservern, Unixserver den Midrangeservern und Mainframes den High-End-Servern zuordnen.

Neben der Kategorisierung nach HW-Aspekten wird auch häufig eine Einordnung nach der betriebenen Anwendung vorgenommen, wie z.B. Webserver, Emailserver, Dateiserver oder allgemein Applikationsserver. Dies macht Sinn, wenn einem Server genau eine Anwendung dediziert zugeordnet ist. Für den

⁵<http://www.idc.com/getdoc.jsp?containerId=prUS25643215> (letzter Zugriff: 26.11.2015)

Fall, dass eine Anwendung auf mehreren Servern betrieben wird, spricht man hingegen von einem Servercluster bzw. Cluster Computing. Mit dem Einzug der Virtualisierung gibt es auch den umgekehrten Fall des Betriebs mehrerer Serveranwendungen auf einer physikalischen Maschine. Jede dieser Anwendungen wird dann als ein virtueller Server bzw. eine virtuelle Maschine (VM) betitelt.

2.3.2 Storage

Storagesysteme dienen der permanenten Speicherung sowie der Bereitstellung von digitalen Daten in einem Netzwerk. Im Prinzip bestehen Storagesysteme aus Serverhardware, an die Komponenten zur Festspeicherung angeschlossen sind. Als Primärspeicher werden hierzu Flashspeicher (z.B. SSD – Solid State Drives) oder Festplatten (HDD – Hard Disk Drives) genutzt. Des Weiteren wird ein geeigneter Netzwerkanschluss für hohe Datenübertragungsraten (z.B. FC – Fibre Channel) benötigt. Die Serverhardware der Storagesysteme dient dann als Controller und übernimmt die eigenständige Verwaltung der Datenspeicherung bzgl. der Redundanz (RAID) und des Zugriffs.

Es gibt grundsätzlich zwei unterschiedliche Umsetzungen von Storageumgebungen: Network-Attached Storage (NAS) und Storage Area Networks (SAN). Beide Implementierungen stellen Nutzern Speicherbereiche im Netzwerk bereit – die Unterschiede liegen dabei in der Art der Bereitstellung. Während ein NAS dateibasiert ist und das Dateisystem der Daten somit vorgibt, können Nutzer für den über ein SAN bereitgestellten blockbasierten Speicher selbst über das Dateisystem entscheiden [GVM00]. Ein NAS wird über dateibasierte Protokolle wie das Network File System (NFS) oder den Server Message Block (SMB) im allgemeinen Netzwerk (LAN – Local Area Network) des Rechenzentrums realisiert. Ein SAN hingegen erfordert eine Abtrennung zum allgemeinen RZ-LAN, die z.B. virtuell über iSCSI oder physikalisch per Fibre Channel umgesetzt werden kann.

Zusätzlich kann Speichervirtualisierung [Sku12] in einer Storageumgebung implementiert werden, die durch Abstraktion des nutzbaren Speicherplatzes von den konkreten Speichermedien eine höhere Nutzung ermöglicht. Zudem erhöht sie die Möglichkeiten des Datenmanagements durch die Zentralisierung des virtualisierten Speichers und indem Migrationen permanenter Daten während des Datenzugriffs ermöglicht werden.

2.3.3 Netzwerk

Das RZ-Netzwerk verbindet die Server- und Storagesysteme untereinander sowie nach außen mit dem Unternehmensnetzwerk und dem Internet und ist damit von zentraler Bedeutung für die Realisierung einer Server-Client-Architektur. Die wichtigste Komponente zum Aufbau eines Netzwerks ist dabei der Switch, über den die IT-Systeme an das Netzwerk angeschlossen werden. Ein Switch stellt eine Netzwerkbrücke mit multiplen Anschlüssen (Ports) dar, welche Datenpakete von einem Quell- zu einem oder mehreren Zielsystemen leitet. Im Gegensatz zu einem

Hub, der Datenpakete an alle angeschlossenen Systeme weiterleitet, werden die Pakete in einem Switch auf Basis der Hardwareadressen zielgerichtet zugestellt. Die meisten größeren und für Rechenzentren ausgelegten Switches lassen sich modular über einsteckbare Linecards um zusätzliche Ports erweitern. Neben den am weitesten verbreiteten Switches für Ethernet (LAN) gibt es auch Modelle für Fibre Channel, InfiniBand und weitere Netzwerktechnologien.

Um einen reibungslosen Netzwerkbetrieb zu gewährleisten, können die QoS-Parameter überprüft und die erlaubten Grenzwerte zum Beispiel im Rahmen von SLA [SKL14] festgehalten werden, vgl. Abschnitt 2.1.3. Typische Parameter im Kontext der Netzwerkperformance sind beispielsweise der Datendurchsatz oder die Paketverluste, vgl. [KSL⁺10]. Werden die Grenzwerte überschritten, ist dies ein Anzeichen für vorhandene Performanceengpässe und somit negative Auswirkungen auf die das Netzwerk verwendende Dienste. Eine der wichtigsten Eigenschaften für einen reibungslosen Betrieb sind die Skalierbarkeit des Netzwerks [AFLV08] mit der Anzahl verbundener Knoten (Server, Storage) sowie die realisierbaren Verbindungsgeschwindigkeiten zwischen den Knoten und nach außen. Diese hängen maßgeblich von der genutzten Netzwerkarchitektur ab. Die bekanntesten Architekturen sind dabei Three-Tier, Fat-Tree und DCell.

Die Three-Tier-Architektur ist als Baum aufgebaut, der – wie der Name bereits andeutet – in die drei Ebenen Access, Aggregate und Core [BKZ⁺13] unterteilt wird. Die Server bilden die Blätter und sind an den Access-Switches in der untersten Ebene angeschlossen. Darüber verbinden Aggregate-Switches jeweils mehrere Access-Switches miteinander und in der höchsten Ebene (direkt unter der Wurzel) werden letztlich alle Aggregate-Switches über die Core-Switches miteinander verbunden. Da der Großteil des internen Netzwerkverkehrs über die Core-Switches läuft, gibt es dort eine sehr hohe Überbuchung, die letztlich Ursache für eine schlechte Bandbreite zwischen den Access- und Aggregate-Switches unterschiedlicher Sektionen ist [AFLV08].

RZ-Netzwerke im Fat-Tree-Design [AFLV08] sollen dem beschriebenen Problem der Three-Tier-Architektur entgegenwirken, indem die beiden unteren Ebenen in sogenannte Pods der Größe k mit jeweils $(k/2)^2$ Knoten und $k/2$ Access- und Aggregate-Switches unterteilt werden. Die $(k/2)^2$ Core-Switches sind in dieser Architektur nur noch mit einem Aggregate-Switch pro Pod verbunden. Während mit dem Fat-Tree eine Überbuchung unterbunden und die volle Bandbreite zwischen den Pods gewährleistet wird, ist die Skalierbarkeit durch die Anzahl der Ports pro Switch begrenzt [AFLV08].

In der DCell-Architektur werden die Knoten nicht nur über Switches sondern zum Teil auch direkt miteinander verbunden [GWT⁺08]. Die Hierarchie ist in diesem Fall nicht in Ebenen sondern in rekursiv aufgebauten Zellen angeordnet, so dass eine $DCell_k$ aus mehreren $DCell_{k-1}$ besteht. Eine $DCell_0$ ist damit die Basiskomponente der Architektur und besteht aus n Knoten, die mit einem Switch verbunden sind. Eine übergeordnete $DCell_1$ besteht dann aus $n + 1$ $DCell_0$, wobei jeweils einer der n Server eine Verbindung mit einer anderen $DCell_0$ herstellt, so dass sich eine komplette Verknüpfung aller $DCell_0$ ergibt.

Die Zusammensetzung weiterer, übergeordneter Zellen ist dann analog. Durch den rekursiven Aufbau skaliert die DCell-Architektur zwar gut mit der Anzahl Knoten, allerdings gibt es nach Bilal et al. [BKZ⁺13] bei großen Netzwerken eine hohe Überbuchung und eine geringe Bandbreite zwischen entfernten Knoten.

2.4 Software

Der Nutzen des Rechenzentrums ergibt sich durch die auf der IT-Hardware betriebene Software. Diese unterteilt sich in Anwendungssoftware bzw. Applikationen und Systemsoftware. Während die Anwendungen bestimmte Funktionen, Aufgaben oder Aktivität für die Nutzer erfüllen sollen und häufig im Kontext von Geschäftsprozessen und Workflows eingesetzt werden, stellt die Systemsoftware die Plattform zum Betrieb der Applikationen inkl. Möglichkeiten zur Kontrolle und Verwaltung dar. Die unterschiedlichen Formen von Software im Rechenzentrum werden nachfolgend beschrieben.

2.4.1 Anwendungen

Eine Anwendung ist eine Menge auf einer Systemsoftware (Betriebssystem) ausführbarer Programme, die eine oder mehrere Funktionen, Aufgaben oder Aktivitäten für den Nutzer der Anwendung erfüllen und über eine entsprechende Nutzerschnittstelle verfügen [Die05]. Die auf den IT-Systemen des Rechenzentrums betriebenen Anwendungen besitzen zumeist den Fokus der Unterstützung der Abläufe von Geschäftsprozessen und Workflows des Unternehmens [Wes12]. Die Anwendungen sind damit ein wichtiger Bestandteil für einen reibungslosen Unternehmensbetrieb. Typische Anwendungen eines Enterpriserechenzentrums sind z.B. Enterprise-Resource-Planning (ERP) zum Management der Unternehmensressourcen wie Kapital, Personal sowie Materialien oder Customer-Relationship-Management (CRM) zur Verwaltung von Kunden/Kontakten. Solche Anwendungen werden nicht nur in Enterprise-RZ betrieben, sondern auch vermehrt ausgelagert (Outsourcing). Viele der Anwendungen in Hosting-, Co-Location- und Cloud-RZ sind damit ebenso Bestandteil der Geschäftsprozesse der mietenden Unternehmen. Ebenso lassen sich typische HPC-Anwendungen wie Simulationen oder Videorendering als Workflow zur Umsetzung von Geschäftsprozessen definieren.

2.4.2 Systemsoftware

Als Systemsoftware werden Programme bezeichnet, die die Hardware betreiben und steuern. Sie bieten Anwendungen eine Plattform, auf der diese ausgeführt werden können. Die Systemsoftware unterteilt sich dabei in die Kategorien Betriebssysteme und spezifische Systemdienste [Die05]. Betriebssysteme organisieren die grundsätzliche Zusammenarbeit der Hardwarekomponenten und ermöglichen Anwendungen auf der Hardware ausgeführt zu werden. Systemdiens-

2 Grundlagen

te verbessern den Betrieb der Hardware, indem sie Möglichkeiten zur Analyse, Konfiguration oder Optimierung anbieten. Beispiele hierfür sind Werkzeuge zur Speicherverwaltung oder Systemmonitore. In die Gruppe der Systemsoftware fallen auch typische Managementsysteme für Rechenzentren, wie sie nachfolgend beschrieben werden.

2.4.3 Managementsysteme

Es gibt Managementsysteme für jeden Bereich des Rechenzentrums – von der Infrastruktur über die IT-Hardware bis zum Betrieb der Software. Ihr Ziel ist die Unterstützung bei der Überprüfung, Verwaltung und Optimierung des Betriebs der betrachteten Komponenten. Entsprechend werden die Managementsysteme in die Kategorie der Systemsoftware eingeordnet. Nachfolgend werden die für diese Arbeit relevanten Managementsysteme beschrieben.

Data Center Infrastructure Management

Data Center Infrastructure Management (DCIM) ist eine Managementsoftware zur Überwachung (Monitoring) und Verwaltung (Controlling) der RZ-Infrastrukturen. Das Ziel von DCIM ist die Verbindung des zuvor getrennten IT- und Gebäudemanagements in einer einzigen Managementumgebung, um eine übergreifende Betrachtung und Verwaltung zu realisieren [Gen14]. Typische Anwendungsfälle für DCIM sind Inventarisierung, Kapazitätsplanung, Lebenszykluskostenanalyse und die Überwachung der Systemkomponenten über integrierte oder zusätzlich installierte Sensoren sowie die Erstellung von Berichten auf Basis der Messdaten. Die Software wird in der Regel direkt im Rechenzentrum betrieben und liest die vorhandenen, bekannten Komponenten sowie Sensoren über das Netzwerk aus, z.B. über das Simple Network Management Protocol (SNMP). Während die IT-Hardware bereits über das DCIM überwacht und gesteuert werden kann, mangelt es noch an Möglichkeiten zur Konfiguration und Verwaltung der Anwendungssoftware. Die unterschiedlichen Aufgabengebiete zwischen DCIM und IT Service Management werden in Abbildung 2.4 dargestellt. DCIM übernehmen dabei alle die Hardware betreffenden Aufgabengebiete wie Datenmessung und -speicherung, Berichterstattung oder Kapazitätsplanung und -optimierung. Das IT Service Management hingegen verwaltet den Betrieb der Anwendungen unter Berücksichtigung von Einflussfaktoren wie Virtualisierung oder Systemressourcennutzung.

Virtualisierung

Die Virtualisierung bezeichnet ein Verfahren zur Abstraktion eines Anwenders von konkreten Ressourcen. Im Kontext der Rechenzentren bezeichnet die Virtualisierung in der Regel die Abstraktion eines Serverbetriebssystems (Anwender) von der physikalischen Hardware (Ressource) in Form einer virtuellen Maschine (VM) [BDF⁺03]. Dem Betriebssystem kann über die VM die exklusive Nutzung

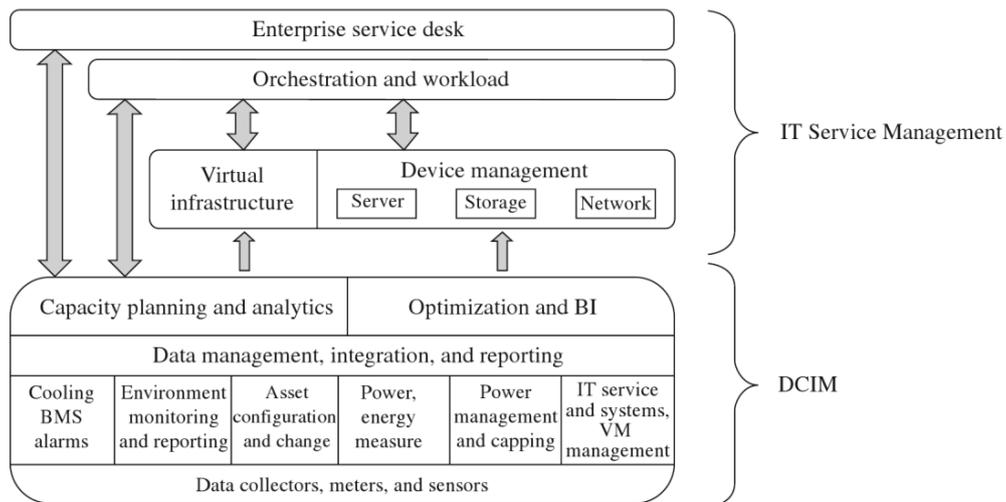


Abbildung 2.4: Umfang von DCIM gegenüber IT Service Management (Cloud Manager) – Grafik unverändert aus [Gen14] zitiert

einer konfigurierbaren Hardware vorgetauscht werden. Dazu wird auf dem physikalischen Server der Hypervisor, eine Kontrollsoftware zum Betrieb und zur Verwaltung der virtuellen Maschinen, installiert. Durch das Konzept der Abstraktion von der realen Hardware und der Abschirmung von VM untereinander ist es möglich, mehrere VM bzw. die darin enthaltenen Anwendungen parallel und unabhängig voneinander auf einer physikalischen Maschine zu betreiben. Dadurch können die Hardwareressourcen besser ausgenutzt werden. Insgesamt werden zum Betrieb der gleichen Menge Anwendungen also weniger Server benötigt, so dass die Virtualisierung zur energetischen Optimierung eines Rechenzentrums beiträgt [LWL⁺09].

Ein weiterer nennenswerter Vorteil ist die Möglichkeit zur unterbrechungsfreien Verschiebung der VM zwischen den Servern per Live-Migration [CFH⁺05]. Diese Funktion wurde ursprünglich für den Anwendungsfall der Wartung implementiert, um Server aus dem Produktivbetrieb nehmen zu können ohne die Anwendungen oder Dienste zu unterbrechen. Live-Migration kann allerdings auch zur Ausfallsicherung (Failover) und zur Wiederherstellung bei Ausfällen (Disaster Recovery) eingesetzt werden. Ein weiterer Anwendungszweck ist die dynamische, energetische Optimierung des IT-Betriebs, wie sie von Last- und Powermanagementverfahren für Server und VM umgesetzt wird.

Last- und Powermanagement

Die meisten Last- und Powermanagementverfahren (LPM) für Server und den darauf betriebenen Diensten setzen auf Virtualisierung auf und nutzen dort die Funktion zum dynamischen, unterbrechungsfreien Verschieben von VM zwischen

den Servern (Live-Migration). Beispiele hierfür sind GreenCloud [LWL⁺09], p-Mapper [VAN08] oder AC4DC [SJS⁺14]. Die Grundidee einer dynamischen Verteilung der VM auf Basis von Auslastungsanalysen ist die gleiche, die Umsetzung allerdings zum Teil sehr unterschiedlich. Unterschiede gibt es vor allem in der Konzeption der Analysen und in den Verteilungsalgorithmen. Da im Rahmen dieser Arbeit auf die RZ-Simulation des AC4DC-Projekts zurückgegriffen wird, wird an dieser Stelle das AC4DC-LPM [SJS⁺14] näher vorgestellt.

Das Grundprinzip des AC4DC-LPM ist die Minimierung der Anzahl aktiver Server entsprechend des vorliegenden Ressourcenbedarfs um durch das Abschalten nicht genutzter Server Energie einzusparen. Das LPM nutzt dabei eine statische Verteilung der VM auf die Server als Ausgangszustand, wobei die Server in dieser Verteilung die maximalen Ressourcenbedarfe aller VM gleichzeitig bewältigen können. Von dieser Verteilung ausgehend werden die VM per Live-Migration verschoben, so dass möglichst wenige Server stärker ausgelastet sind und die restlichen deaktiviert werden können.

Aufgrund der zeitlichen Verzögerungen durch das Reaktivieren von Serversystemen reicht es nicht aus, auf Änderungen im Ressourcenbedarf zu reagieren. Um bei steigendem Ressourcenbedarf Engpässe durch eine Überbuchung der Ressourcen eines Servers zu vermeiden, muss die Verteilung proaktiv angepasst werden. Das AC4DC-LPM untersucht entsprechend nicht nur die vorliegenden Bedarfe, sondern auch die zukünftig eintretenden. Dazu abstrahiert das LPM über eine Vorhersagekomponente die zukünftigen Bedarfe aus den historischen Werten [HSSN11] und verwendet diese zur Ermittlung zukünftig eintretender Ressourcenengpässe. Wird ein solcher erkannt, kann die Verteilung früh genug angepasst werden, so dass dieser nicht mehr eintritt. Das LPM erlaubt dabei standardmäßig keine parallelen (d.h. gleichzeitige) Migrationen, da das Verfahren nicht feststellen kann, ob die Performance bei parallelen Migrationen ausreicht. Der Grad der parallelen Migrationen hat direkten Einfluss auf den Planungszeitraum, für den das LPM den zukünftigen Ressourcenbedarf überprüft, da dieser direkt von den notwendigen Migrationen zur Wiederherstellung der statischen Ausgangsverteilung abhängt.

Cloud Manager

Als Cloud Manager werden Plattformen bezeichnet, die die Virtualisierungsumgebung, die Server- und Stagesysteme, das Netzwerk sowie die Anwendungen und Dienste unter einer Managementsoftware zusammenfassen. In Abbildung 2.4 werden Cloud Manager in Abgrenzung zu DCIM-Systemen unter der Bezeichnung IT Service Management dargestellt. Die Ausrichtung fokussiert sich dabei auf das Thema Cloud Computing, so dass Cloud Manager den Zusammenschluss der Hard- und Software zu einer privaten, öffentlichen oder hybriden Cloud zum Ziel haben. Typische Funktionen von Cloud Managern [VLDWF12] sind die Koordination der Virtualisierungsverfahren unterschiedlicher Anbieter (z.B. Xen, VMware, KVM), die Verwaltung der Replikation und Integrität der Datenspei-

cherung, das Management von Netzwerken (z.B. LAN, Virtual LAN, Software-Defined Networks) und korrespondierende IP-Adressen sowie die Darstellung von Informationen und Berichten in einem Dashboard. Zudem bieten Cloud Manager standardmäßig diverse Nutzerschnittstellen an, um den kombinierten Betrieb mit bestehender (eigener) Software zu ermöglichen.

Die bekanntesten und verbreitetsten Cloud-Managementplattformen sind OpenStack⁶, OpenNebula⁷, Nimbus⁸ und Eucalyptus⁹. Alle vier Cloud Manager werden als Open-Source-Software zur freien Verfügung angeboten. Vergleiche der Cloud Manager finden sich in [VLDWF12] und [DJ13].

2.5 Energieeffizienz

Die im Rahmen dieser Arbeit entwickelte Metrik bewertet die Energieeffizienz eines Rechenzentrums. Um die Ergebnisse der Metrik korrekt einordnen zu können, wird in diesem Abschnitt die Energieeffizienz im Kontext von Rechenzentren definiert. Zudem wird die Energieeffizienz von weiteren Begriffen mit ähnlicher Bedeutung abgegrenzt und es werden unterschiedliche nutzbare Parameter diskutiert.

2.5.1 Definition der Energieeffizienz

Effizienz beschreibt im Allgemeinen mit möglichst wenig Aufwand ein bestimmtes Ergebnis zu erzielen. Sie ist damit die Grundlage des ökonomischen Prinzips, welches durch die Maximierung des Ergebnisses (Maximumprinzip) oder die Minimierung des Aufwands (Minimumprinzip) eine Nutzenmaximierung verfolgt [Peh10]. Beispiele für mögliche Aufwände sind dabei Geld, Materialien, Zeit oder Energie. Die Form des Ergebnisses hängt vom Einsatzgebiet ab und kann ein konkretes, materielles Produkt darstellen oder eine vollbrachte, sinnvolle Arbeit bzw. einen erzielten Nutzen. Die Effizienz (*Eff*) ist also definiert als Relation eines erreichten, sinnvollen Ergebnisses (*UsefulOutput*) zum dafür benötigten Aufwand (*Input*), wie in der nachfolgenden Gleichung dargestellt wird.

$$\text{Eff} = \frac{\text{Useful Output}}{\text{Input}} \quad (2.1)$$

Die energiewirtschaftliche Betrachtung der Effizienz nennt sich Energieeffizienz (EE) und bedeutet den Einsatz der Energie in ein System zur Erreichung eines Ergebnisses zu reduzieren [Peh10]. Als Input wird demnach exklusiv die Aufnahme der Energie des untersuchten Systems berücksichtigt. Der Output hingegen bleibt weiterhin systemspezifisch und kann daher von der Anzahl gefertigter Bauteile durch eine Fräsmaschine über die Kühlung von Lebensmitteln

⁶<https://www.openstack.org/> (letzter Zugriff: 26.11.2015)

⁷<http://openebula.org/> (letzter Zugriff: 26.11.2015)

⁸<http://www.nimbusproject.org/> (letzter Zugriff: 26.11.2015)

⁹<http://open.eucalyptus.com/> (letzter Zugriff: 26.11.2015)

in einem Kühlschrank bis hin zur Wettervorhersage eines HPC-Clusters reichen. Die Gleichung der Effizienz wird entsprechend nur im Nenner angepasst.

$$EE = \frac{\text{Useful Output}}{\text{Energy Used}} \quad (2.2)$$

Wenn der Output ebenso wie der Input die Energie oder die Leistung betrachtet, entspricht die Effizienz dem Wirkungsgrad η der Energieübertragung innerhalb eines Systems [Peh10]. Dies funktioniert mit beliebigen Kombinationen von Energieformen – die Wahl des Outputs richtet sich dabei nach dem sinnvollen Nutzen, der wiederum vom Einsatzzweck abhängt. Zum Beispiel kann für eine Lampe der Wirkungsgrad der Umwandlung von elektrischer Energie zum einen in Lichtenergie (Leuchtmittel) und zum anderen in Wärmeenergie (Wärmelampe) betrachtet werden.

2.5.2 Definition im RZ-Kontext

Die allgemeine Definition der Energieeffizienz berücksichtigt zwar alle Energieformen (elektrische, mechanische, thermische, chemische, ...), im Kontext der Rechenzentren wird die Energieeffizienz allerdings grundsätzlich auf die elektrische Energie reduziert [Ras07b, Her15]. Dies hat mehrere Gründe:

- Alle Komponenten des Rechenzentrums werden mit elektrischer Energie betrieben.
- Die elektrische Energie ist für die meisten Rechenzentren die einzige zugeführte Energieform.
- Die insgesamt vom RZ aufgenommene Energie kann mit geeignetem Equipment relativ einfach und zuverlässig an der Einspeisung gemessen werden.

Selbst wenn ein Rechenzentrum über ein Dieselaggregat (z.B. über ein BHKW) betrieben wird, wird die chemische Energie im Energieträger (Diesel) in elektrische umgewandelt, welche wiederum an der Einspeisung gemessen werden kann. Eine Ausnahme betrifft die Unterstützung der Klimatisierung durch extern zugeführte Kälte über ein Fernkältenetz. Die so zugeführte Wärmeenergie wird in typischen Betrachtungen der Energieeffizienz für Rechenzentren nicht berücksichtigt, so dass die dadurch reduzierte Energieaufnahme der lokalen Klimatisierung die Energieeffizienzbewertung verfälscht. Allerdings betreiben nur sehr wenige Rechenzentren ihre Klimatisierung über ein Fernkältenetz, da eine entsprechende Infrastruktur und ein Anbieter in direkter Nähe verfügbar sein müssen¹⁰. Zudem kann die korrekte Effizienz ermittelt werden, wenn die finanziellen Aufwände für die Strom- und Kälteversorgung anstelle der Energie betrachtet werden.

¹⁰www.greendatacenternews.org/articles/774040/district-heating-means-cooling-as-a-service-for-da/ (letzter Zugriff: 26.11.2015)

Die elektrische Energie, die das Rechenzentrum aufnimmt, wird letztlich sowohl durch die IT-Hardware als auch die Infrastruktur komplett in Wärmeenergie umgewandelt [Ras07a]. Wird diese vollständig durch die RZ-Klimatisierung abgeführt, reicht es die Energieeffizienz auf Grundlage der elektrischen Energie an der Einspeisung zu beurteilen. Wird allerdings ein Teil der Wärme weitergenutzt, um beispielsweise Büros zu beheizen, steigt dadurch die Effizienz des Gesamtsystems (RZ und Büros). Die Energieeffizienz im RZ-Kontext wird daher auf das Rechenzentrum selbst begrenzt und weitere angeschlossene Systeme/Abnehmer nicht mehr berücksichtigt [AAF12]. Extern weitergenutzte Abwärme erhöht damit nicht die Energieeffizienz des RZ, sondern wird dem abnehmenden System angerechnet oder über eine separate Metrik wie die Energy Reuse Effectiveness [TVCA10] bewertet. Wird Wärmeenergie allerdings im Rahmen des RZ-Betriebs weitergenutzt – z.B. für eine Absorptionskühlung oder zum Vorheizen eines Notstromaggregats – fließt diese Weiternutzung implizit in die Energieeffizienz mit ein, da für diese Systeme weniger elektrische Energie benötigt würde.

Der sinnvolle Output (also der Zähler in Gleichung 2.2) eines Rechenzentrums ist im Gegensatz zur aufgenommenen, elektrischen Energie nicht eindeutig bestimmbar. Im Prinzip hängt der Output direkt von der im Rechenzentrum betriebenen Software (vgl. Abschnitt 2.4) und den dadurch realisierten Workflows bzw. Prozessen ab. Allerdings ist der Einfluss der Software auf die Gesamtproduktivität zum einen schwierig zu messen und zum anderen schlecht untereinander vergleichbar, so dass zur Ermittlung der Energieeffizienz häufig Ersatzkennzahlen für den sinnvollen Output genutzt werden [HMP⁺09]. Mögliche Parameter werden dabei im Abschnitt 2.5.4 beschrieben.

2.5.3 Abgrenzung

Es gibt einige Begriffe und Kennzahlen, deren Bedeutungen der Effizienz sehr ähnlich sind und die aufgrund dessen häufig verwechselt werden. Die Wichtigsten werden nachfolgend in Abgrenzung zur Effizienz erläutert.

Effektivität

Die Effektivität bezeichnet im allgemeinen das Verhältnis eines erreichten zu einem zuvor definierten Ziel [Peh10]. Im Gegensatz zur Effizienz spielt der Mitteleinsatz bzw. der Aufwand hierbei keine Rolle, sondern nur die Qualität des erreichten Ziels bzw. der Output ist entscheidend. Im Prinzip beschreibt die Effektivität damit die Wirksamkeit eines Vorgehens während die Effizienz die Wirtschaftlichkeit repräsentiert. Die Effektivität ist also implizit in der Effizienz enthalten.

Produktivität

Die Produktivität ist ebenso wie die Effizienz durch die Relation des Outputs zum Input definiert. Nach Cantner et al. [CKH07] kann man die Produktivität als spezifische, inhaltliche Ausprägung der Effizienz ansehen. Während die Effizienz dem Zweck diene, das ökonomische Prinzip (vgl. Abschnitt 2.5.1) anzuwenden, betrachte die Produktivität im Allgemeinen reale Outputs und Inputs eines Produktionsprozesses. Theoretisch kann die Produktivität allerdings für beliebige Formen von Inputs und Outputs angewandt werden, womit sie zur Effizienz identische Aussagen liefert. Daher werden im Rahmen dieser Arbeit Effizienz und Produktivität analog behandelt.

Coefficient of Performance

Der Coefficient of Performance (COP) beschreibt prinzipiell die Energieeffizienz einer Wärmepumpe und kann somit als spezielle Ausprägung der Effizienz angesehen werden. Der COP bildet die Effizienz über das Verhältnis der beförderten Wärmeenergie gegenüber der dafür benötigten elektrischen Energie der Wärmepumpe ab. Je nach Einsatzzweck der Wärmepumpe, ob sie zur Kühlung oder zur Heizung eingesetzt wird, wird dabei die abgeführte oder zugeführte Wärmeenergie betrachtet. Der COP ist somit entsprechend Gleichung 2.3 definiert [MSBB10].

$$COP = \frac{Q}{W} \quad (2.3)$$

Q repräsentiert dabei die abgeführte/zugeführte Wärmeenergie und W die benötigte Arbeit (in der Regel die aufgenommene elektrische Energie E). Da zwei Formen von Energie miteinander in Relation gesetzt werden, ist der COP einheitenlos.

Energy Efficiency Ratio

Die Energy Efficiency Ratio (EER) ist eine US-amerikanische Kennzahl für die Kühleffizienz von Klimageräten. Sie ist definiert über die Relation der Kühlleistung in British Thermal Units Per Hour (BTU/h) zur aufgenommenen elektrischen Leistung in Watt des Klimageräts [AHR08]. Zudem wird die EER für vordefinierte Betriebstemperaturen ermittelt um unterschiedliche Klimageräte – zur Erleichterung der Produktauswahl – direkt miteinander vergleichen zu können. Im Gegensatz zum COP ist die EER für alle Klimageräte definiert und nicht auf Wärmepumpen begrenzt. Ihre Definition sieht allerdings keine Anwendung auf gemessene Daten zur Bewertung des tatsächlichen Betriebs vor, auch wenn dies theoretisch möglich ist. Neben der EER wurde außerdem die Seasonal Energy Efficiency Ratio (SEER) definiert. Diese bewertet die Kühleffizienz für einen festgelegten, typischen Jahrestemperaturverlauf.

2.5.4 Parameter

Die Energieeffizienz setzt sich nach Gleichung 2.2 aus zwei Parametern zusammen. Der Nenner ist dabei im Kontext von Rechenzentren – wie in Abschnitt 2.5.2 bereits erläutert wurde – auf die zum Betrieb benötigte elektrische Energie festgelegt. Für den Zähler, der mit dem Output ein Maß der erzielten Ergebnisse darstellt, ist die Wahl des geeigneten Parameters variabel. Nachfolgend werden mögliche, nutzbare Parameter vorgestellt.

Elektrische Leistung/Energie

Je nachdem ob die Energieeffizienz für einen Zeitraum oder für einen einzelnen Zeitpunkt ermittelt wird, wird entweder die elektrische Energie (E) oder die elektrische Leistung (P) genutzt. Die Energie entspricht dem Integral der Leistung über den betrachteten Zeitraum und die Leistung ist die Energie pro Zeit. Im Rahmen dieser Arbeit werden die Begriffe elektrische Leistung, Leistungsaufnahme und Power synonym benutzt, wobei die Leistungsaufnahme den Aspekt der von Geräten aufgenommenen elektrischen Leistung hervorhebt und Power vornehmlich im Kontext der Modellierung von Leistungsaufnahmen elektrischer Geräte genutzt wird. Der allgemeinere Begriff der Leistung wird im Rahmen dieser Arbeit hingegen möglichst vermieden. Der Grund hierfür ist die mehrfache Belegung des Begriffs der Leistung, welche z.B. im Bereich der Informatik auch für die Verarbeitungsgeschwindigkeit oder Leistungsfähigkeit bzw. Performance von Hard- und Software steht.

Arbeit

Der sinnvolle Output eines Rechenzentrums ist die getätigte Arbeit, die einen Beitrag in der Wertschöpfungskette des betreibenden Unternehmens leistet. Wie bereits in Abschnitt 2.4 andiskutiert wurde, ist dieser durch die Vielfalt der Anwendungen sowie ihre unterschiedlichen Einflüsse auf Workflows und Prozesse nur sehr schwierig bewertbar. Die konkrete Arbeit kann zwar in Form von Tasks mit einer definierbaren Aufgabe wie z.B. die Bearbeitung einer Datenbankabfrage oder die Beantwortung eines Seitenaufrufs gemessen werden, die Einordnung der jeweiligen Sinnhaftigkeit – vor allem im Vergleich untereinander – muss bislang allerdings manuell vorgenommen werden. Eine vollständige Abdeckung sowie korrekte Einordnung bzgl. des Nutzens aller Anwendungen bzw. Tasks ist somit enorm aufwändig.

Performance

Im Informatikbereich bezeichnet Performance die Verarbeitungsgeschwindigkeit oder Leistungsfähigkeit von Hardware- und Softwaresystemen. Die Performance wird dabei immer auf den Nutzen des betreffenden Systems bezogen. Für Server

2 Grundlagen

heißt dies beispielsweise, dass die Performance sich an den Arbeitslasten der Anwendungen, die auf dem Server ausgeführt werden, orientiert. Die Performance zweier Systeme ist damit nur vergleichbar, wenn beide Systeme die identische Arbeit verrichten. Zur Ermittlung der Performance wird der erzielte Nutzen bzw. die erledigte Arbeit typischerweise der dafür benötigten Zeit gegenübergestellt. Die Performance stellt damit eine Ausprägung der Effizienz mit dem Fokus auf den zeitlichen Aufwand dar und ist daher ein potentieller Parameter zur Bestimmung der Energieeffizienz.

Über Benchmarks lässt sich die Performance eines Systems mit standardisierten Arbeitslasten ermitteln, so dass man mit anderen Systemen vergleichbare Resultate erhält. In der Regel wird dazu die Zeit der Ausführung der Arbeitslast oder eine Rate der abgeschlossenen Tasks pro Zeiteinheit gemessen.

Auslastung

Die Auslastung eines IT-Systems beschreibt den Anteil der Systemressourcen, der mit Aufgaben wie Berechnungen oder der Speicherung von Daten belegt ist. Da manche Subsysteme keine Teilbelastung ermöglichen, sondern nur zwischen Bearbeitung und Leerlauf wechseln (z.B. CPU-Kerne, wobei diese ihre Bearbeitungsgeschwindigkeit über Performance States skalieren können), wird die Auslastung anhand des Durchschnitts über dem Intervall der zeitlichen Auflösung berechnet. Da eine höhere Ressourcennutzung einen höheren generierten Output impliziert, kann die Auslastung der Ressourcen als Anzeichen dafür genutzt werden, wie viel Arbeit das System leistet. Somit ist die Systemauslastung ein möglicher Parameter für die Energieeffizienz.

Die Auslastung muss allerdings zusätzlich in Relation zu den ausgelasteten Systemressourcen gesetzt werden, denn die an einem System anliegende und für die Auslastung verantwortliche Arbeitslast (Workload) kann auf unterschiedlichen IT-Systemen unterschiedliche Auslastungen verursachen [PMWV09]. Diese hängt direkt mit der Performance des IT-Systems bzgl. der Arbeitslast zusammen. So benötigt beispielsweise die gleiche Arbeitslast bei einem performanteren System einen geringeren Anteil an Systemressourcen und damit eine geringere Auslastung als bei einem langsameren System.

2.6 Verwandte Arbeiten

Die in dieser Arbeit beschriebene Entwicklung der auslastungsabhängigen Energieeffizienzmetrik LDEE bezieht sich auf die Themenfelder der Modellierung und Bewertung von Rechenzentren und ihre IT-Komponenten. Im Folgenden wird dazu der aktuelle Stand der Technik näher erläutert. Zunächst werden bekannte Energieeffizienzmetriken vorgestellt, mit denen die LDEE im weiteren Verlauf der Arbeit verglichen wird. Des Weiteren wird auf Energiebenchmarks für IT-Systeme Bezug genommen, da diese von der LDEE zur Bewertung der IT genutzt

werden. Zudem wird auf die Modellierung der Leistungsaufnahme und der Performance von RZ-Komponenten eingegangen, da die LDEE sich aus entsprechenden Modellen zusammensetzt.

2.6.1 Energieeffizienzmetriken für Rechenzentren

In Tabelle 2.1 werden bekannte Energieeffizienzmetriken für Rechenzentren und ihre IT-Systeme aufgelistet. Die bekannteste Metrik ist dabei die PUE [SEZ13], deren verbreitete Nutzung sich unter anderem darin zeigt, dass sie in ISO/IEC 30134-2¹¹ standardisiert wird. Die PUE wurde ebenso wie die DCiE – der Kehrwert der PUE – von The Green Grid [BRPC07] entwickelt und hat zum Ziel, den energetischen Overhead durch die RZ-Infrastruktur zu quantifizieren. Entsprechend werden die Energiebedarfe der IT dem Gesamtenergiebedarf des RZ gegenübergestellt. Während in der ursprünglichen Definition [BRPC07] der PUE die Leistungsaufnahmen in einer Momentaufnahme genutzt wurden, ist die aktuelle Definition [AAF12] auf die jährlichen Energiebedarfe ausgelegt. Durch die Ermittlung eines Durchschnittswerts werden saisonale Einflüsse auf das Resultat verringert und die Vergleichbarkeit erhöht.

Auch wenn die PUE als Energieeffizienzmetrik verbreitet eingesetzt wird, kann sie diese nicht bewerten, da die hierfür zu betrachtende sinnvolle Arbeit, Produktivität oder Performance (vgl. Abschnitt 2.5) nicht berücksichtigt wird. Ein anschauliches Beispiel hierfür ist die Nutzung von Virtualisierung gegenüber einem dedizierten Betrieb bei einer identischen Infrastruktur. Durch die Virtualisierung wird die gleiche Arbeit unter Einsatz geringerer Hardwarekapazitäten erledigt. Obwohl insgesamt weniger Energie benötigt wird und die Energieeffizienz somit steigt, verschlechtert sich die PUE durch den sinkenden Anteil der IT-Energie an der Gesamtenergieaufnahme. Die PUE bewertet den nun größeren Overhead der Infrastruktur, der durch die gesunkene IT-Energie ebenso optimiert werden könnte.

Mit der FVER, der ITEU, der ITUE sowie der TUE fokussieren sich neben der PUE weitere Metriken auf die Leistungs- bzw. Energieaufnahme des RZ. Die FVER [NLLS12] repräsentiert die Proportionalität zwischen dem statischen und dem dynamischen Anteil der Leistungsaufnahme des Rechenzentrums, wobei kleinere Resultate besser sind. Die ITEU und ITUE sind systembezogene Metriken, die das Nutzungsverhältnis des IT-Systems in Form der Leistungs- bzw. Energieaufnahme widerspiegeln. Die ITEU [Gre12] setzt dazu die aktuell genutzte zur maximalen Leistungsaufnahme ins Verhältnis und repräsentiert damit im Prinzip die Auslastung. Sie wird in den übergeordneten Metriken CADE und DP-PE in Kombination mit weiteren Kennzahlen für ein gesamtes Rechenzentrum genutzt. Die ITUE [PPH⁺13] hingegen betrachtet die Energie, die das IT-System insgesamt benötigt, im Verhältnis zur Energie, die die Rechen- und Speicher-subkomponenten (CPU, RAM, HDD, ...) benötigen ($\text{energy}_{\text{com}}$). Die Messung

¹¹http://www.iso.org/iso/catalogue_detail.htm?csnumber=63451
(letzter Zugriff: 26.11.2015)

Name	Metrik	Ref.
CADE – Corporate Average DC Efficiency	$DCiE \cdot ITEU \cdot util_{IT} \cdot EE_{IT}$	[KFK08]
CPE – Compute Power Efficiency	$util_{IT} \cdot \frac{power_{IT}}{power_{DC}}$	[AMC+08]
DC-EEP – Data Center Energy Efficiency and Productivity	$PUE \cdot IT-PEW^{12}$	[Bri07]
DCeP – Data Center Energy Productivity	$\frac{useful\ work_{DC}}{energy_{DC}}$	[HMP+09]
DCiE – Data Center Infrastructure Efficiency	$\frac{1}{PUE}$	[BRPC07]
DCPpW – Data Center Performance per Watt	$\frac{useful\ work_{DC}}{energy_{DC}}$	[Pf08]
DPPE – DC Performance per Energy	$DCiE \cdot ITEU \cdot ITEE \cdot \frac{1}{1-GEC}^{13}$	[Gre12]
FVER – Fixed to Variable Energy Ratio	$1 + \frac{power_{min}}{power_{max} - power_{min}}$	[NLLS12]
IT-PEW – IT Productivity per Embedded Watt	$\frac{productivity_{sys}}{power_{sys}}$	[Bri07]
ITEE – IT Equipment Efficiency	$\frac{productivity_{sys}}{power_{sys}}$	[Gre12]
ITEU – IT Equipment Utilization	$\frac{power_{sys}}{power_{sysmax}}$	[Gre12]
ITUE – IT-Power Usage Effectiveness	$\frac{energy_{sys}}{energy_{com}}$	[PPH+13]
PUE – Power Usage Effectiveness	$\frac{energy_{DC}}{energy_{IT}}$	[AAF12]
SWaP – Space, Watts, and Performance	$\frac{performance_{sys}}{rack\ space_{sys} \cdot power_{sys}}$	[RSR+07]
TUE – Total-Power Usage Effectiveness	$ITUE \cdot PUE$	[PPH+13]

Tabelle 2.1: Übersicht bekannter Energieeffizienzmetriken

bzw. Ermittlung von Letzterem gestaltet sich allerdings schwierig [PPH+13]. Die ITUE fließt dann zusammen mit der PUE in die übergeordnete TUE ein. Insgesamt haben die genannten Metriken bzgl. der Energieeffizienzbewertung allerdings das gleiche Defizit wie die PUE: Es fehlt jeweils die Betrachtung der Arbeit/Produktivität.

Verschiedene Organisationen und Institute haben alternative Energieeffizienzmetriken entwickelt, die das beschriebene Defizit umgehen. Ein häufig gewählter Ansatz ist die Betrachtung der IT-Auslastung in Ergänzung zum Energiebedarf. Bekannte Beispiele hierfür sind die CADE der McKinsey&Company [KFK08] und die CPE von The Green Grid [AMC⁺08]. Allerdings reicht die isolierte Betrachtung der Auslastung nicht aus, um Aussagen über die sinnvolle Arbeit oder Produktivität zu erhalten, vgl. Abschnitt 2.5.4.

Daher hat The Green Grid die DCeP [HMP⁺09] und Dell die DCPpW [Pf08] entwickelt, welche jeweils die sinnvolle, getätigte Arbeit dem Energiebedarf gegenüberstellen. Während die Ermittlung der sinnvollen Arbeit im Rahmen der DCPpW nicht genauer definiert wurde, ist das Vorgehen bzgl. der DCeP sehr komplex. Jeder einzelne Task (vgl. Abschnitt 2.5.4) muss manuell von den Betreibern definiert, gemessen und in Relation zueinander gesetzt werden. Da die Resultate durch die manuelle Definition subjektiv sind, können diese nicht für Vergleiche zwischen Rechenzentren genutzt werden. Aufgrund der hohen Komplexität der Metrik wurden durch The Green Grid alternative Ersatzkennzahlen für die sinnvolle Arbeit beschrieben [HMP⁺09]. Diese reduzieren die sinnvolle Arbeit je nach Kennzahl auf die Performance oder Auslastung der IT-Systeme.

Als Alternative zur sinnvollen Arbeit berücksichtigen einige Metriken die Produktivität der IT, indem Kennzahlen wie Datendurchsatzraten und Rechen- sowie Speicherkapazitäten anstelle der konkret getätigten Arbeit betrachtet werden. Bekannte Beispiele hierfür sind die IT-PEW [Bri07] und die ITEE [Gre12], welche beide auf einzelne IT-Systeme angewandt werden. Die jeweils genutzten Kennzahlen orientieren sich dabei an den Hardwaretypen. Um Aussagen zur Energieeffizienz eines ganzen Rechenzentrums zu erhalten, werden die IT-PEW bzw. die ITEE in die übergeordneten RZ-Metriken DC-EEP bzw. DPPE integriert. Die Resultate sind durch ein objektives Bewertungsverfahren zwar zwischen Rechenzentren vergleichbar, die Beschränkung auf ausgewählte Kennzahlen der IT richtet den Fokus allerdings auf bestimmte (und ggf. nicht repräsentative) Arbeitslasten (z.B. auf Fließkommaberechnungen bei Nutzung von FLOPS – Floating Point Operations per Second).

Einen anderen Betrachtungswinkel besitzt die von Sun Microsystems definierte Metrik SWaP [RSR⁺07], welche auf einzelne IT-Systeme angewandt werden kann. Im Gegensatz zu den anderen Metriken wird zur Abschätzung der getätigten Arbeit keine Kennzahl ausgelesen, sondern es können beliebige Performancebenchmarks genutzt werden. Die Performanceresultate werden dann in Relation zu den Höheneinheiten (rack space) und der Leistungsaufnahme des IT-Systems gesetzt. Für Vergleiche zwischen IT-Systemen müssen die gleichen Benchmarks angewandt worden sein. Eine Anwendung auf ein ganzes Rechenzentrum wurde allerdings nicht definiert.

¹²Die PUE wird hier als SI-EER bezeichnet und auf ein Minimum von 1,6 definiert.

¹³Der Green Energy Coefficient (GEC) repräsentiert den Anteil genutzter, erneuerbarer Energien an der Gesamtenergie des RZ.

Ein detaillierter Vergleich der vorgestellten Metriken bzgl. der Anforderungen an Energieeffizienzmetriken für Rechenzentren (vgl. Abschnitt 3.2) wird in Abschnitt 3.3 beschrieben.

2.6.2 Energiebenchmarks für IT-Systeme

Die vorgestellten, systembezogenen Metriken (IT-PEW, ITEE, ITEU und ITUE) betrachten im Produktivbetrieb messbare Kennzahlen wie Datendurchsatzraten oder Energiebedarfe. Diese hängen von den im Rechenzentrum betriebenen Anwendungen ab und liefern somit ein Abbild der IT-Nutzung. Eine explizite Messung der erforderlichen Kennzahlen ist damit eine Voraussetzung zur Anwendung der entsprechenden Metriken.

Einen anderen Ansatz verfolgen Performancebenchmarks, die ein IT-System mit einem oder mehreren vordefinierten Workloads belasten und über die Ausführungszeiten oder die Anzahl abgearbeiteter Tasks die Performance bestimmen. Wird die so ermittelte Performance der zur Ausführung benötigten elektrischen Energie gegenübergestellt, erhält man die Energieeffizienz des IT-Systems. Eine Ermittlung der Energieeffizienz ist durch die festgelegten Arbeitslasten zwar nicht während des Produktivbetriebs möglich, jedoch können diese vorab bestimmt, gespeichert und bei Bedarf abgefragt werden. Durch die Ausführung identischer Arbeitslasten sind die Ergebnisse zudem direkt vergleichbar.

Die Standard Performance Evaluation Corporation (SPEC)¹⁴ hat mehrere Energiebenchmarks mit unterschiedlichen Zielrichtungen entwickelt. Der SPEC-power_2008-Benchmark [Lan09a] besteht aus einer Java-Serverapplikation, die die Anzahl der Anfragen an Server in 10%-Schritten skaliert und so Resultate für insgesamt elf Laststufen ermittelt. Das Server Efficiency Rating Tool (SERT) [LT11] wird ebenso von SPEC betrieben und belastet einen Server mit einer hohen Bandbreite unterschiedlicher Workloads. Im Gegensatz zu SPEC-power werden Bewertungen unabhängig vom installierten Betriebssystem und der Java-Laufzeitumgebung vorgenommen. Mit dem SPECvirt_sc2013 wird der Betrieb von virtualisierten Workloads auf Server- und Storage-Systemen untersucht. Die VM werden dabei mit unterschiedlichen SPEC-Benchmarks betrieben. Alle drei Benchmarks nutzen das Werkzeug SPEC PTDaemon [HLB12], welches die synchrone Messung der Leistungsaufnahme und der Temperaturen erlaubt.

SAP bietet einen Benchmark¹⁵ für einzelne Serversysteme als auch für Cluster bestehend aus mehreren Server- und Storage-Systemen basierend auf der eigenen Sales and Distribution Software an. In einer Erweiterung [SAP11] wird auch die zusätzliche Messung der Leistungsaufnahme unterstützt, so dass die Energieeffizienz über mehrere Laststufen bestimmt werden kann. Auch das Transaction Processing Performance Council (TPC)¹⁶ hat mit der TPC-Energy [YCN11] eine Spezifikation eingeführt, mit der zu jedem TPC-Benchmark zusätzliche Messun-

¹⁴<https://www.spec.org/> (letzter Zugriff: 26.11.2015)

¹⁵<http://global.sap.com/campaigns/benchmark/index.epx> (letzter Zugriff: 26.11.2015)

¹⁶<http://www.tpc.org/> (letzter Zugriff: 26.11.2015)

gen des Energiebedarfs definiert werden. Für jeden der TPC-Benchmarks (TPC-C, TPC-E, TPC-H) wird eine zusätzliche Metrik zur Darstellung der Energieeffizienz in Form der Leistungsaufnahme bzw. Power (Watt) pro Performance (Anzahl Transaktionen) definiert. Im Gegensatz zu den SPEC-Energiebenchmarks wird bei TPC allerdings nur die maximale Performance und somit auch nur die Energieeffizienz bei maximaler Auslastung ermittelt. Analog zum TPC bietet auch das Storage Performance Council (SPC)¹⁷ für seine Storagebenchmarks (SPC1, SPC2) eine optionale Erweiterung um eine Energiekomponente an. Die ermittelten Datendurchsatzraten werden dann der Leistungsaufnahme gegenübergestellt [Sto13]. Die Arbeitslasten der SPC-Benchmarks werden zudem über mehrere Laststufen skaliert, so dass die resultierende Energieeffizienz eines Systems in Abhängigkeit zur Auslastung betrachtet werden kann.

Im Bereich des High-Performance-Computings gibt es die Green 500 Liste¹⁸, bei der ganze HPC-Cluster mit tausenden CPU-Kernen und Leistungsaufnahmen von $\geq 50\text{kW}$ bewertet werden. Die Performanceresultate des LINPACK-Benchmarks, bei dem bei der Lösung von linearen Gleichungen die Anzahl der Gleitkommaoperationen (FLOPS) ermittelt wird, werden der Top 500 Liste¹⁹ entnommen. Die Resultate werden dann mit der durchschnittlichen Leistungsaufnahme während der Benchmarkausführung in Relation gesetzt.

Neben den Energiebenchmarks, die durch gemeinnützige Interessengruppen wie SPEC oder TPC entwickelt, verwaltet und angeboten werden, gibt es auch Forschungsgruppen, die sich mit dem Thema beschäftigen. So haben Rivoire et al. [RSRK07] den Benchmark JouleSort entwickelt, der einen Sortieralgorithmus auf unterschiedlichen Typen von IT-Systemen anwendet und neben der Geschwindigkeit auch die aufgenommene Energie berücksichtigt. Im Vergleich zu den genannten industriellen Benchmarks ist die angewandte Arbeitslast allerdings weniger repräsentativ für typische Anwendungen im Rechenzentrum. Komplett auf synthetische Workloads setzen Du Boias et al. [DBSP⁺11] mit ihrem Benchmarkframework SWEEP, bei dem die Arbeitslasten, mit denen die IT-Systeme belastet werden sollen, frei konfiguriert werden können. Für direkte Vergleiche müssen die Konfigurationen entsprechend identisch sein.

Im Rahmen dieser Arbeit werden bestehende Energiebenchmarks dazu genutzt, IT-Hardware bzgl. ihrer Performance und ihrer Leistungsaufnahme zu modellieren. Die konkrete, begründete Auswahl der geeigneten Benchmarks wird dabei in den Abschnitten 4.4 bzw. 4.5 beschrieben.

2.6.3 Powermodelle

Es gibt eine große Vielfalt von Forschungsarbeiten, die sich mit der energetischen Modellierung von Rechenzentren und den darin vorhandenen Komponenten beschäftigen. Nachfolgend werden die für diese Arbeit relevanten Ansätze

¹⁷<http://www.storageperformance.org/home/> (letzter Zugriff: 26.11.2015)

¹⁸<http://www.green500.org/> (letzter Zugriff: 14.12.2015)

¹⁹<http://www.top500.org/> (letzter Zugriff: 26.11.2015)

2 Grundlagen

vorgestellt, da die LDEE teilweise bereits vorhandene Modelle integriert oder auf diesen aufbaut.

Pelley et al. [PMWV09] kombinieren verfügbare Powermodelle zu einem kompletten RZ-Powermodell und abstrahieren die Interaktionen zwischen den Modellen über ein parametrierbares, abstraktes Modell. Das Modell erstreckt sich von der USV und den PDU über die Server bis zur Kühlinfrastruktur bestehend aus CRAH, Pumpe, Chiller und Kühlturm. Das Gesamtmodell schätzt die Leistungsaufnahme für gegebene Eingabeparameter der IT-Auslastung und der Außentemperatur ab. Eine optionale Simulationskomponente kann zudem die Auswirkungen konkreter Lastverteilungen der Anwendungen auf die Server ermitteln. Die einzelnen Powermodelle sind jeweils als lineare Modelle aufgebaut, deren Parameter mit realen Messdaten trainiert werden müssen. Eine Anwendung ist demnach nur möglich, wenn die Komponenten vermessen werden. Das RZ-Modell betrachtet zudem nur eine konkrete Kühlkette als Klimälösung und es fehlen Storage- und Netzwerkkomponenten. Breen et al. beschreiben in [BWP⁺10] ein kombiniertes Powermodell für eine ähnliche Klimakette (CRAH, Chiller, Kühlturm) eines Rechenzentrums. Die Modelle nutzen hierbei allerdings aus der IT-Last abgeleitete Wärme- und Volumenströme zur Abschätzung und keine linearen Modelle, die zuvor trainiert werden müssen. Im Gegensatz zum Ansatz von Pelley et al. werden jedoch mehr kontinuierliche Messdaten an den untersuchten Komponenten benötigt. Den gleichen Ansatz wie Breen et al. verfolgen auch Meisner et al. mit dem in [MW11] vorgestellten RZ-Powermodell. Die einzelnen Klimamodelle basieren auf Betriebsparametern wie Wärmestrom, Lüfterdrehzahlen und Temperaturen. Die Parameter werden dann aus dem laufenden Betrieb ermittelt und teils von der IT-Last abgeleitet.

Moore et al. definieren in [MCRS05] ein häufig verwendetes Powermodell für Chiller auf Basis des Coefficient of Performance (COP). Der temperaturabhängige COP wird dabei über ein lineares Modell, welches mit gemessenen Werten trainiert wird, berechnet. In [MCRS05] nutzen sie das Modell dazu, Entscheidungen für Lastoptimierungen zu treffen. Patnaik et al. [PMSR09] nutzen zwar auch den COP zur Abschätzung der Leistungsaufnahme, berücksichtigen allerdings nicht den Temperatureinfluss sondern nur einen Durchschnittswert. Dieser wird in Relation zur Kühllast des Chillers gesetzt, welche sich über die Auslastung und die Kühlkapazität ergibt.

Fan et al. beschreiben in [FWB07] zwei auslastungsabhängige Powermodelle für Server. Eines basierend auf einer linearen Funktion und ein nicht-lineares Modell, welche jeweils nur die CPU-Auslastung als Eingabe erfordern. Heath et al. [HDC⁺05] verwenden einen ähnlichen Ansatz, nutzen zusätzlich allerdings noch die Auslastung der Speicherkomponenten in Form des Datendurchsatzes als Eingangsparemeter. Economou et al. [ERKR06] lesen mehrere Performancecounter des untersuchten Servers aus. Neben der CPU-Auslastung sind dies u.a. Speicherzugriffsraten, Datendurchsatzraten des Netzwerks und der Festplatten oder Aktivitäten in den Caches und Gleitkommaeinheiten. Rivoire et al. haben in [RRK08] analoge Powermodelle miteinander verglichen. Die höchste Genauig-

keit besitzt dabei das Modell basierend auf Performancecountern, welches allerdings auch die meisten Messdaten benötigt. Prinzipiell liefern jedoch alle Modelle eine hinreichende Genauigkeit (Abweichungen $< 10\%$). Da alle Modelle vor dem Einsatz mit Messdaten, die bei der Anwendung von definierten Benchmarks ermittelt wurden, trainiert werden müssen, gestaltet sich ein Einsatz der Modelle im Produktivbetrieb als schwierig.

Hinsichtlich der Modellierung der Storagepower gibt es nur wenige Arbeiten, die ein komplettes Storagesystem umfassen. Hierzu gehört beispielsweise das Modellierungsverfahren von Allalouf et al. [AAF⁺09], welches die Leistungsaufnahme unter Berücksichtigung der anliegenden Arbeitslasten inkl. aktiver und Bereitschaftszustände abschätzt. Dazu werden Performancecounter für Zugriffsraten und Datendurchsätze ausgelesen und in Operationen auf Festspeichern übersetzt. Über Look-Up-Tables wird dann aus den Operationen die Leistungsaufnahme für den konkreten Workload interpoliert. Andere Ansätze konzentrieren sich hingegen auf die energetische Modellierung der Speicherkomponenten (HDD, SSD). So schätzen beispielsweise Basmadjian et al. [BAN⁺11] die Leistungsaufnahme in Abhängigkeit der Betriebszustände (idle, accessing und start-up) ab. Die Häufigkeiten der Zustände werden dann aus den Lese- und Schreibraten abgeleitet. Stoess et al. [SLB07] im Gegenzug definieren zwei energetische Zustände (active, idle) mit einer linearen, auslastungsabhängigen Interpolation zwischen diesen. Die Speicherauslastung wird dabei von den Datentransferraten und Antwortzeiten abgeleitet.

Mahadevan et al. [MSBR09] haben ein Framework zur energetischen Bewertung von Serverswitches entwickelt. Dazu werden Switches in unterschiedlichen Konfigurationen (z.B. Portnutzung, Bandbreite) mit diversen Trafficprofilen belastet, während gleichzeitig der Energiebedarf gemessen wird. Auf den Messdaten wird der eigens entwickelte Network Energy Proportionality Index, der den Grad der Übereinstimmung der vorliegenden Proportionalität zur idealen angibt, angewandt. Chabarek et al. [CSB⁺08] haben auf Basis empirischer Daten ein Powermodell für Netzwerkswitches entwickelt. Der darüber abgeschätzte Energiebedarf ist abhängig vom Chassis, der Linecard- und Portkonfiguration sowie der Auslastung der aktiven Ports. Vor einer Nutzung des Modells müssen die Abhängigkeiten allerdings anhand von Messdaten charakterisiert werden.

Powermodelle für Rechenzentren sowie für RZ-Komponenten sind ein wesentlicher Bestandteil der LDEE und dienen dort zur Abschätzung des Energiebedarfs. Einige der oben beschriebenen Ansätze werden im Rahmen dieser Arbeit in das Gesamtkonzept eines RZ-Powermodells integriert. Die konkret weitergenutzten Modelle werden in den jeweiligen Unterabschnitten des Abschnitts 4.5 behandelt.

2.6.4 Performancemodelle

Performancemodelle dienen dazu die Leistungsfähigkeit eines Rechenzentrums oder eines Systems, also wie viel Arbeit ein Rechenzentrum oder ein System in einer gewissen Zeit verrichten kann, abzuschätzen. Weit verbreitet ist hierzu

die Anwendung von Performancebenchmarks, die ein untersuchtes Testsystem bestehend aus einer oder mehreren Einzelkomponenten mit einer definierten Arbeitslast belegt und die Bearbeitungsdauer oder die Anzahl der abgearbeiteten Tasks misst. Performancebenchmarks, die zusätzlich die Energieaufnahme messen und somit die Energieeffizienz bewerten, wurden bereits in Abschnitt 2.6.2 beschrieben. Die gleichen Organisationen (SPEC, TPC, SPC) haben ebenso eine große Auswahl reiner Performancebenchmarks mit unterschiedlichen Arbeitslasten als auch Zielplattformen.

Es gibt zwar ein großes Spektrum unterschiedlicher Performancebenchmarks für Teilsysteme eines Rechenzentrums, eine Kombination von Teilergebnissen zu einer allgemeinen Gesamtbewertung eines ganzen Rechenzentrums wurde jedoch noch nicht definiert. Bislang gibt es nur situative – auf bestimmte Gegebenheiten angepasste – Verfahren. Ein beispielhafter Ansatz ist der von Google, die nicht mehr die Performance einzelner Server, sondern ihrer kompletten Rechenzentren untersuchen. So führte Google beispielsweise Sortieroperationen von Milliarden Datensätzen an einem ihrer Standorte durch und untersuchte damit Softwareplattformen wie MapReduce-Algorithmen [Pat08]. Das Verfahren verlangt allerdings nach einer spezifischen Infrastruktur und ist speziell auf die Bedürfnisse Googles ausgelegt. Eine allgemeine Anwendbarkeit ist somit nicht realistisch.

Ein Bewertungsverfahren zur Analyse der Performance von Cloud-Rechenzentren schlägt Bruneo in [Bru14] vor. Der Fokus des angewandten, stochastischen Modells liegt auf der Verteilung von Anfragen und Arbeitslasten in Form virtueller Maschinen auf die Hardwareressourcen. Die Bewertung wird dann anhand von übergreifenden Metriken, die die Auslastung, Verfügbarkeit, Wartezeiten, Servicezeiten und Ansprechbarkeit jeweils im Kontext des definierten, stochastischen Modells betrachten, vorgenommen. Zusammengenommen bilden diese Parameter die QoS-Anforderungen von Anwendern ab – werden diese nicht eingehalten, ist dies ein Anzeichen für einen Performanceengpass. Da die konkret zugrundeliegende Hardware bislang nicht berücksichtigt wird, ist eine genaue Quantifizierung der Performance über das beschriebene Modell nicht möglich.

Im US-Patent [BKR⁺13] wird ein Verfahren zur Arbeitslastzuordnung in einem Cloud-Rechenzentrum beschrieben, bei dem eine Teilkomponente eine Abschätzung der notwendigen Performance für den Betrieb gegebener Anwendungen auf der zugrundeliegenden Hardware ermittelt. Würde diese Abschätzung für alle Anwendungen des Rechenzentrums durchgeführt werden, erhielte man die Gesamtperformance des RZ. Als potentielle Ansätze zur Performanceabschätzung werden unter anderem analytische Modelle zur Abbildung von Applikationen auf die Performance, statische Analysen des Sourcecodes oder statistische Auswertungen vorhergehender Ausführungen der Anwendungen vorgeschlagen. Im Patent selbst wird ein MapReduce-Ansatz beschrieben, der eine Teilausführung der Applikationen auf der Hardware verlangt. Die Performance ist damit allerdings anwendungsspezifisch und lässt sich somit schlecht verallgemeinern.

Ein vollständiges Konzept zur allgemeinen, anwendungsunabhängigen Ermittlung der Performance einer kompletten Rechenzentrums-IT ist somit nicht ge-

geben. Im Rahmen dieser Arbeit werden daher vorhandene Teillösungen wie Performancebenchmarks über einen modellbasierten Ansatz miteinander kombiniert um letztlich eine Kennzahl für das gesamte RZ zu erhalten. Die dabei zu den IT-Komponenten entstandenen Performancemodelle sowie die Kombination zur Bewertung eines kompletten Rechenzentrums werden in Abschnitt 4.4 beschrieben.

3 Definition der LDEE Metrik

Die Idee hinter der Load Dependent Energy Efficiency (LDEE) ist die vollständige, objektive und korrekte Bewertung der Energieeffizienz von Rechenzentren. Sie soll die üblichen Anwendungsfälle von RZ-Betreibern, die in Abschnitt 3.1 vorgestellt werden, abdecken. Dazu muss die LDEE einige Anforderungen erfüllen. Diese werden in Abschnitt 3.2 erläutert und die Erfüllung dieser durch bekannte Energieeffizienzmetriken in Abschnitt 3.3 untersucht. Die genaue Definition und der Aufbau der LDEE wird dann anschließend in Abschnitt 3.4 behandelt.

3.1 Anwendungsfälle für Energieeffizienzmetriken

Energieeffizienzmetriken für Rechenzentren finden in den letzten Jahren immer mehr Anwendung. Wie bereits in Kapitel 1 motiviert wurde, sorgt vor allem ein erhöhter Kostendruck in Unternehmen dafür, dass vermehrt der Fokus auf einen energie- und damit kosteneffizienten Betrieb gerückt wird. Dies wirkt sich auch auf die Rechenzentren aus, indem die Verantwortlichen Auflagen zur Kosteneinsparung von ihren Vorgesetzten erhalten. Und der erste Schritt in diese Richtung ist die Kenntnis über den aktuellen Stand der Effizienz – abgebildet durch EE-Metriken.

Der typische Anwender von Energieeffizienzmetriken für Rechenzentren ist der Organisator der IT-Infrastrukturen des Unternehmens – zumeist der Chief Information Officer (CIO) bzw. der IT-Leiter, manchmal auch der Chief Technology Officer (CTO) bzw. der technische Leiter. Metriken dienen dem CIO/CTO dazu, sich einen Überblick über die Infrastrukturen zu verschaffen. Sie sind aber auch ein Mittel um Ineffizienzen oder Ungereimtheiten im Betrieb aufzudecken und können ebenso dazu genutzt werden, größere Veränderungen oder Optimierungen am IT-Betrieb bei den Vorgesetzten zu motivieren. Metriken sollten dementsprechend schnell und einfach verständlich und ihre Ergebnisse leicht vermittelbar sein – dennoch sollen sie eine gewisse Aussagekraft besitzen.

Die gängigsten Fragestellungen, zu deren Lösung Energieeffizienzmetriken unterstützend beitragen sollen, orientieren sich an den Aufgabenfeldern von CIO und CTO und werden nachfolgend aufgeführt:

1. Wie (energie)effizient ist mein Rechenzentrum? Wie ist das aktuelle Kosten-Nutzen-Verhältnis?
2. Wie (energie)effizient ist mein Rechenzentrum im Vergleich zu anderen Rechenzentren?

3 Definition der LDEE Metrik

3. Wo liegen die Ursachen für etwaige Ineffizienzen? Welche Kosten verursachen diese?
4. Welche Auswirkungen haben bzw. hatten Änderungen an der IT-Infrastruktur auf die Effizienz? Wie viel kann durch potentielle Maßnahmen eingespart werden?

Aus den Fragestellungen können die nachfolgend beschriebenen Anwendungsfälle für Energieeffizienzmetriken im Rechenzentrum extrahiert werden.

3.1.1 Bestimmung der Energieeffizienz

Für Energieeffizienzmetriken ist der obligatorische Anwendungsfall die Wiedergabe der Energieeffizienz als übersichtliche und leicht verständliche Kennzahl. Dies ist der Kernanwendungsfall für die Fragestellungen 1 und 2. Die Energieeffizienz kann je nach verwandter Metrik unterschiedlich definiert sein (z.B. Produktivität, Performance oder Auslastung) oder für unterschiedliche Teilbereiche gelten (z.B. Server oder Infrastruktur) – vgl. Abschnitt 2.5. Sie alle geben jedoch letztlich einen Überblick über die Effizienz der Energienutzung.

3.1.2 Vergleich der Energieeffizienz

Da ein isoliert betrachtetes Resultat einer Energieeffizienzkennzahl in der Regel nicht sonderlich aussagekräftig ist, ist der Vergleich mit Resultaten anderer Rechenzentren, wie er in Fragestellung 2 thematisiert wird, von großem Interesse. Durch die unterschiedlichen Architekturen sowie Ausrichtungen von Rechenzentren (Abschnitt 2.1) führt ein generischer Vergleich jedoch nicht unbedingt zum Ziel. Häufig ist ein Vergleich innerhalb von RZ-Klassen, welche die bezüglich Energieeffizienz relevanten Charakteristika abbilden, sinnvoller. Zudem wird die Vergleichbarkeit von den Umgebungsparametern der Rechenzentren beeinflusst. Wird von diesen Parametern (z.B. Arbeitslast, klimatische Bedingungen) abstrahiert, ist auch ein allgemeiner, fairer Vergleich möglich.

3.1.3 Identifizierung von Ineffizienzen

Mit der allgemeinen Energieeffizienz kann erkannt werden, ob es Ineffizienzen im Rechenzentrum gibt. Der nächste Schritt ist dann die Identifizierung der Quelle, welche in Fragestellung 3 zum Ausdruck kommt. Dies kann z.B. über feingranulare Metriken, welche auf einzelne Systeme anwendbar sind, realisiert werden. Im Idealfall werden dadurch Ansatzpunkte für zukünftige Optimierungsmaßnahmen erkannt.

3.1.4 Bewertung von Optimierungsmaßnahmen

Wurden Optimierungsmaßnahmen in einem Rechenzentrum durchgeführt, ist für CIO/CTO weiterhin von Interesse, was die durchgeführten Maßnahmen letztlich

gebracht haben. Um dies zu beurteilen kann ein Vorher-Nachher-Vergleich der Resultate einer Energieeffizienzmetrik unternommen werden. Wurden keine weiteren Betriebsparameter verändert, ist die Auswirkung der durchgeführten Optimierungen direkt ableitbar. Dieser Anwendungsfall leitet sich aus Fragestellung 4 her.

3.1.5 Exploration von potentiellen Modifikationen

Ähnlich zum vorhergehenden Anwendungsfall sollen Metriken dabei helfen, Modifikationen am Rechenzentrum zu beurteilen. In diesem Fall sind es allerdings geplante Anpassungen, die vor der Umsetzung untersucht werden sollen. Zur Abschätzung der Auswirkungen auf die Energieeffizienz werden dabei Annahmen getroffen und die Effizienz auf Basis dieser Annahmen berechnet. Die Genauigkeit hängt damit vollständig von der Güte der Annahmen ab. Indem Untersuchungen für unterschiedliche Modifikationen durchgeführt werden, kann eine komplette Exploration der Möglichkeiten erreicht werden. Auch dieser Anwendungsfall leitet sich aus Fragestellung 4 ab.

3.1.6 Auswirkung auf Kosten

Für die CIO/CTO ist bei der Ermittlung der Energieeffizienz auch die Auswirkung auf die Kosten von Interesse. Da die Energieeffizienzmetriken eine Relation zu der benötigten Leistungsaufnahme oder Energie herstellen, können die Kosten direkt aus den Parametern der Metriken berechnet werden. Je nach Fokus der Metrik gelten diese für das gesamte Rechenzentrum oder für Teilsysteme des Rechenzentrums. So könnte einerseits in Kombination mit der Produktivität/Performance das Kosten-Nutzen-Verhältnis des Gesamtbetriebs bestimmt werden (Fragestellung 1). Andererseits ist auch eine Kostenanalyse von Subsystemen, sofern passende Metriken vorhanden sind, möglich (Fragestellungen 3 und 4).

3.2 Anforderungen

Aus den beschriebenen typischen Anwendungsfällen von Energieeffizienzmetriken für Rechenzentren lassen sich die Anforderungen an eine zweckdienliche Metrik ableiten.

- R 3.1 **Repräsentation der Energieeffizienz:** Die Metrik muss die in Abschnitt 2.5.2 definierte Energieeffizienz, also die geleistete sinnvolle Arbeit in Relation zur Energie, die zur Vollbringung der Arbeit benötigt wird, bestimmen. Alternativ kann anstelle der sinnvollen Arbeit auch eine der in Abschnitt 2.5.4 beschriebenen Kennzahlen verwandt werden. Da das betrachtete Zielsystem der Metrik ein komplettes Rechenzentrum ist, muss die Metrik ebenso in der Lage sein, sowohl die sinnvolle Arbeit

3 Definition der LDEE Metrik

bzw. stellvertretende Kennzahlen als auch die Energie geeignet zusammenzufassen.

- R 3.2 **Auslastungsabhängigkeit:** Die Leistungsaufnahme der meisten Gewerke im Rechenzentrum hängt direkt von der anliegenden Last ab. Ebenso korreliert die Auslastung der IT mit der getätigten Arbeit bzw. Produktivität, vgl. Abschnitt 2.5.4. Da diese beiden Parameter die Energieeffizienz definieren, muss die Metrik so ausgelegt sein, dass die genannten Abhängigkeiten nachvollziehbare Auswirkungen auf das Resultat besitzen.
- R 3.3 **Vollständigkeit:** Zur Bewertung der Energieeffizienz eines Rechenzentrums müssen alle relevanten Systeme betrachtet werden. Es reicht beispielsweise nicht aus, nur die IT-Systeme, welche die Arbeit vollbringen, isoliert zu betrachten, sondern es muss ebenso die zum Betrieb der IT notwendige Infrastruktur in die Bewertung einfließen. Diese Anforderung ist zur Realisierung von fairen, korrekten Vergleichen zwischen Rechenzentren essentiell.
- R 3.4 **Objektivität:** Die Bewertung der Energieeffizienz muss objektiv und damit einhergehend deterministisch sein. Diese Anforderung schränkt vor allem die Betrachtung der geleisteten, sinnvollen Arbeit als Ausprägung des sinnvollen Outputs ein, da keine einheitliche Definition für sinnvolle Arbeit existiert und diese somit subjektiv ist, vgl. Abschnitte 2.5.2 und 2.5.4. Das bedeutet aber auch, dass die Energie über konsistente Methoden ermittelt werden muss. Die Objektivität ist die wichtigste Anforderung hinsichtlich der Vergleichbarkeit der Ergebnisse.
- R 3.5 **Vergleichbarkeit des Resultats:** Ergänzend zu den beiden vorhergehenden Anforderungen R 3.3 und R 3.4 müssen einzelne Resultate der Metrik vergleichbar sein. Das Format muss entsprechend einen relativen Wert darstellen, so dass Rechenzentren beliebiger Größe vergleichbar werden.
- R 3.6 **Unterscheidung von RZ-Typen:** Basierend auf dem Anwendungsfall des Vergleichs zwischen Rechenzentren in Abschnitt 3.1.2 sollte die Metrik eine Differenzierung der Ergebnisse bzgl. relevanter RZ-Eigenschaften erlauben. Die Relevanz der Eigenschaften bestimmt sich dabei durch ihren Einfluss auf die Resultate der Energieeffizienzmetrik sowie ihre Praktikabilität zur Klassifizierung.
- R 3.7 **Diverse Betrachtungsebenen:** Damit die Metrik zur Identifizierung von Ineffizienzen (vgl. Abschnitt 3.1.3) genutzt werden kann, muss sie mehrere Betrachtungsebenen besitzen. Neben der Aggregation der Parameter für ein zusammenfassendes Resultat muss die Metrik ebenso Ergebnisse zu Teilsystemen des Rechenzentrums liefern können. Anderen-

falls könnten keine Aussagen über die Effizienz einzelner Komponenten getroffen werden.

- R 3.8 **Unabhängigkeit von Messungen:** Der Anwendungsfall der Exploration potentieller Modifikationen (vgl. Abschnitt 3.1.5) erfordert die Möglichkeit einer Abschätzung der Energieeffizienz durch die Metrik, ohne dass gemessene Daten vorliegen. Diese Anforderung in Kombination mit der Auslastungsabhängigkeit R 3.2 hat auch Auswirkungen auf den Vergleich zwischen Rechenzentren. Durch die Ermittlung der Energieeffizienz für beliebige Arbeitslasten ist ein Vergleich mit identischen Umgebungsparametern möglich. Der Einfluss lokaler Arbeitslasten sowie klimatischer Bedingungen auf die Energieeffizienz kann so eliminiert und ein allgemeiner, fairer Vergleich realisiert werden.
- R 3.9 **Flexibilität:** Die Metrik sollte flexibel bezüglich potentieller Änderungen am Rechenzentrum sein. Eine proaktive Bewertung des Austauschs von Komponenten ist beispielsweise nur möglich, wenn die Betriebsparameter der neuen Komponente von der Metrik berücksichtigt werden können. Weitere Beispiele sind Konfigurationsänderungen an Hard- oder Software oder der Betrieb anderer Arbeitslasten. Diese Anforderung ist für den Anwendungsfall der Exploration essentiell.
- R 3.10 **Anwendbarkeit:** Die Anwendbarkeit der Metrik ist eine wichtige Anforderung, da sie direkten Einfluss auf ihre Nützlichkeit hat. Müssen zu viele Messdaten manuell erhoben und verarbeitet werden, ist die Integration der Metrik in die Managementwerkzeuge ggf. abschreckend und es würden einfachere Lösungen bevorzugt, auch wenn diese weniger aussagekräftig sind. Ebenso steigt mit der Komplexität die Fehleranfälligkeit und die Wahrscheinlichkeit von fehlenden Messdaten, so dass es zu einem Ausfall kommt.
- R 3.11 **Verständlichkeit:** Die Verständlichkeit des Resultats hat einen maßgeblichen Effekt auf die Anwendbarkeit und damit auch auf die Nützlichkeit der Metrik. Die typischen Anwender der Metrik (CIO/CTO) werden nicht viel Zeit aufwenden um sich in die Interpretation der Ergebnisse hineinzuarbeiten. Bekannte Einheiten (z.B. kW) und/oder Vergleichswerte könnten beispielsweise zu einem schnelleren und einfacheren Verständnis führen. Ebenso sollte aus dem Resultat der Metrik ein Rückschluss auf die relativen/absoluten Kosten möglich sein, welche einen weiteren Anwendungsfall der Metrik darstellen.

3.2.1 Anmerkung zu den Anforderungen

Jede der beschriebenen Anforderungen für sich betrachtet ist relativ weit gefasst und bietet somit etwas Implementierungsspielraum für die Entwicklung einer

3 Definition der LDEE Metrik

zweckdienlichen Energieeffizienzmetrik. Die Anwendungsbreite der resultierenden Metrik ist allerdings an die Menge der erfüllten Anforderungen gekoppelt. Und das Ziel ist die Entwicklung einer Metrik, welche alle Anforderungen bestmöglich erfüllt, um alle in Abschnitt 3.1 genannten Anwendungsfälle abzudecken.

Durch die Berücksichtigung aller Anforderungen ergeben sich jedoch Konflikte, die nach aktuellem Stand der Technik nicht gelöst werden können. So setzt beispielsweise die Definition für Energieeffizienz eine Ermittlung der getätigten sinnvollen Arbeit voraus, die bislang allerdings nur manuell und damit subjektiv für jedes Rechenzentrum separat definiert werden muss, vgl. Abschnitt 2.6.1. Damit stehen die Anforderungen R 3.1 und R 3.4 in einem direkten Konflikt zueinander und für die zu entwickelnde Metrik muss entschieden werden, welche Anforderung abgeschwächt werden muss.

Die konkreten Auswirkungen der Anforderungen auf die Konzeption der LDEE werden in Abschnitt 3.4 beschrieben.

3.3 Vergleich von Energieeffizienzmetriken

Es gibt unterschiedliche Ansätze für Metriken die Energieeffizienz eines Rechenzentrums oder einer ihrer Subkomponenten zu bewerten. Je nach Ansatz erfüllt die Metrik dabei nur eine Teilmenge der beschriebenen Anforderungen und ist damit auch nur für die entsprechenden Anwendungsfälle geeignet. In Tabelle 3.1 werden bekannte Metriken hinsichtlich ihrer Eigenschaften miteinander verglichen. Die Vergleichseigenschaften orientieren sich dabei an den zuvor beschriebenen Anforderungen bzw. Anwendungsfällen. Weitere Erläuterungen zu den jeweiligen Metriken inkl. der Zusammensetzungen finden sich in der Diskussion der verwandten Arbeiten in Abschnitt 2.6.1.

EE nach Definition Diese Eigenschaft gibt an, ob eine Metrik die Energieeffizienz nach der Definition aus Abschnitt 2.5.2 ermittelt oder zumindest eine der Ersatzkennzahlen aus Abschnitt 2.5.4 (z.B. Produktivität oder Performance kombiniert mit Auslastung) nutzt. Sie entspricht damit weitestgehend der Anforderung R 3.1 und ist die Grundvoraussetzung zur Lieferung eines repräsentativen Ergebnisses in den entsprechenden Anwendungsfällen.

Der Vergleich zeigt, dass einige der bekannten Metriken nicht die definierte Energieeffizienz widerspiegeln, sondern nur Teilaspekte betrachten. So berücksichtigen die Metriken PUE/DCiE, ITUE, TUE, FVER und ITEU nur das Verhältnis zwischen Leistungsaufnahme bzw. Energiebedarf unterschiedlicher Teilsysteme des Rechenzentrums. Diese Metriken beziehen sich also auf die effektive anstelle der effizienten Nutzung der Energie (vgl. Abschnitt 2.5.3), wie z.B. der Anteil an Energie für Infrastrukturen im Verhältnis zur Energieaufnahme der IT bei der PUE. Viele dieser Metriken sind daher entsprechend mit „Effectiveness“ betitelt. Zur Repräsentation der Energieeffizienz fehlt den Metriken ein Faktor, der die Leistungsfähigkeit der IT wiedergibt. Die CPE berücksichtigt zwar zu-

Metrik	EE nach De- finition	Vergleich- barkeit	Messunab- hängig	Betrachtungs- ebenen	Praktika- bilität
CADE	(ja)	ja	nein	Rechenzentrum	sehr gering
CPE	nein	nein	nein	Rechenzentrum	hoch
DC-EEP	ja	ja	nein	Rechenzentrum	durchschnittlich
DCeP	ja	nein	nein	Rechenzentrum	gering
DCPpW	ja	nein	nein	Rechenzentrum	gering
DPPE	ja	ja	nein	Rechenzentrum	sehr gering
FVER	nein	nein	nein	Rechenzentrum	hoch
IT-PEW/ITEE	ja	nein	nein	IT-System	durchschnittlich
ITEU	nein	nein	nein	IT-System	durchschnittlich
ITUE	nein	nein	nein	IT-System	durchschnittlich
PUE/DCiE	nein	nein	nein	Rechenzentrum	sehr hoch
SPECpower	ja	nein	nein	IT-System	durchschnittlich
SWaP	ja	nein	nein	IT-System	durchschnittlich
TPC-Energy	ja	nein	nein	IT-System	gering
TUE	nein	nein	nein	Rechenzentrum	durchschnittlich

Tabelle 3.1: Vergleich bekannter Energieeffizienzmetriken und -benchmarks bzgl. der Anforderungen. Details zu den jeweiligen Funktionsweisen finden sich in Abschnitt 2.6.1

3 Definition der LDEE Metrik

sätzlich zum Verhältnis der Energiebedarfe auch die Auslastung der IT, diese alleine spiegelt jedoch nicht die Leistungsfähigkeit wider, so dass auch hiermit keine Energieeffizienz nach Definition bestimmt werden kann.

Die weiteren Metriken verwenden entweder direkt die sinnvolle Arbeit oder eine der in Abschnitt 2.5.4 beschriebenen Ersatzkennzahlen Produktivität, gesamte getätigte Arbeit oder Performance. Die CADE bildet in der Theorie auch die Energieeffizienz ab, die konkrete Ermittlung dieser ist in [KFK08] allerdings nicht definiert.

Vergleichbarkeit Die Eigenschaft der Vergleichbarkeit signalisiert, ob ein Vergleich der Ergebnisse unterschiedlicher Rechenzentren mit dieser Metrik grundsätzlich sowohl möglich als auch sinnvoll ist. Sie fasst die Anforderungen R 3.3 für Vollständigkeit, R 3.4 für Objektivität und R 3.5 für die Vergleichbarkeit des Resultats zusammen. Ebenso fließt die erste Eigenschaft – die genutzte Definition der Energieeffizienz – und damit die Anforderung R 3.1 mit ein.

Alle Metriken, die nicht die definierte Energieeffizienz oder andere geeignete Kennzahlen wie die Produktivität oder Performance nutzen, sind nicht für Vergleiche zwischen Rechenzentren sinnvoll nutzbar. Das bedeutet nicht, dass sie nicht trotzdem zu diesem Zweck eingesetzt werden, wie im Falle der PUE. Allerdings leidet die Aussagekraft der Vergleiche darunter, dass die Energie-/Leistungsverhältnisse nicht in Relation zur Leistungsfähigkeit des Rechenzentrums gesetzt werden. Beispielsweise ist ein Rechenzentrum mit höherer Leistungsfähigkeit bei identischer Leistungsaufnahme energieeffizienter, auch wenn die Relation zwischen IT- und Infrastrukturleistungsaufnahme (bzw. die PUE) identisch ist.

Die Metriken und Benchmarks, welche auf einzelne IT-Systeme bezogen sind, können ebenfalls nicht für sich betrachtet für Vergleiche der Energieeffizienz ganzer Rechenzentren genutzt werden. Durch den Fokus auf ein IT-System erfüllen sie nicht die Anforderung R 3.3 für Vollständigkeit, die danach verlangt, dass ebenso die zum Betrieb notwendige Infrastruktur in die Energieeffizienz mit einfließt. Sie können zwar in der Regel für Vergleiche von IT-Systemen in unterschiedlichen Rechenzentren genutzt werden, die Eigenschaft der Vergleichbarkeit bezieht sich an dieser Stelle im Kontext der EE-Metriken allerdings nur auf den Vergleich ganzer Rechenzentren. Metriken/Benchmarks für IT-Systeme können jedoch als Bestandteil in übergeordnete Metriken integriert werden und so Vergleiche zwischen Rechenzentren ermöglichen, wie es beispielsweise mit der IT-PEW, welche eine Submetrik der DC-EEP darstellt, der Fall ist.

Des Weiteren ist mit den Metriken DCeP und DCPpW kein fairer Vergleich zwischen Rechenzentren möglich, da ihre jeweiligen Definitionen nicht objektiv sind und sie somit nicht die Anforderung R 3.4 erfüllen. Alle drei Metriken verlangen nach einer Messung der sinnvollen Arbeit, die allerdings für jedes Rechenzentrum manuell definiert werden muss. Aufgrund unterschiedlicher RZ-Ausrichtungen ist die Sinnhaftigkeit der Arbeitslasten immer vom Kontext

3.3 Vergleich von Energieeffizienzmetriken

abhängig und eine objektive Betrachtung daher nicht möglich.

Die restlichen Metriken verwenden zur Bestimmung der Energieeffizienz Kennzahlen, die objektiv und deterministisch bestimmt bzw. gemessen werden können und die IT als auch die Infrastruktur abdecken.

Messunabhängig Eine Metrik ist unabhängig von gemessenen Daten, wenn sie für beliebige Eingabedaten die Energieeffizienz des Rechenzentrums liefern kann (Anforderung R 3.8). Das bedeutet gleichzeitig, dass die Abhängigkeiten zwischen Lasten, Produktivität und Leistungs-/Energieaufnahme des Rechenzentrums von der Metrik abgebildet sein müssen (Anforderung R 3.2). Diese zwei Anforderungen sind eine Grundvoraussetzung für den Anwendungsfall der Exploration potentieller Modifikationen des RZ. Zudem ermöglichen sie einen fairen Vergleich von Rechenzentren, indem von den tatsächlichen Arbeitslasten abstrahiert wird um eine allgemeine Energieeffizienz zu ermitteln.

Der Vergleich offenbart, dass alle untersuchten Metriken basierend auf ihren Definitionen Messdaten benötigen aus denen die Resultate direkt berechnet werden. Zwar können für die Eingabedaten der Metriken Annahmen getroffen werden, aufgrund der Vielzahl der zu betrachtenden Komponenten sind Fehler hierbei jedoch praktisch unvermeidbar. Bezüglich des Energiebedarfs werden beispielsweise häufig die maximalen Leistungsaufnahmen aus den Datenblättern potentiell neu anzuschaffender Geräte verwandt, welche die typische Leistungsaufnahme im Teillastbereich deutlich überschätzen. Erschwerend kommt noch hinzu, dass zur Erreichung realistischer Resultate die Annahmen für unterschiedliche Betriebsmodi des Rechenzentrums (z.B. durch variierende Auslastungen oder klimatische Bedingungen) getroffen werden müssten, so dass dies auch einen beträchtlichen Aufwand darstellt.

Eine andere Möglichkeit ist die Ersetzung der konkreten Messdaten durch Modelle, Tabellen oder Funktionen, welche die Annahmen/Abschätzungen für die Eingabedaten liefern und die Abhängigkeiten dadurch implizit abbilden. Allerdings werden entsprechende Alternativverfahren in den Definitionen der vorgestellten Metriken nicht beschrieben, womit eine Ableitung dieser aus konkreten Messwerten notwendig wäre und somit wiederum keine Messunabhängigkeit erreicht würde.

Betrachtungsebenen Die Betrachtungsebene einer Metrik kann High- oder Low-Level sein oder sie kann mehrere Ebenen umfassen. High-Level steht dabei für einen abstrakten Blick auf das Ganze, in diesem Fall für ein gesamtes Rechenzentrum oder einen kompletten Teilbereich (z.B. Infrastruktur). Low-Level hingegen bezieht sich auf die Betrachtung einzelner Systeme oder Gruppen von Systemen. Dieser Vergleichsaspekt betrachtet die Anforderung R 3.7, welcher essentiell für den Anwendungsfall der Identifikation von Ineffizienzen ist.

Die untersuchten Metriken unterteilen sich in die beiden Gruppen „Rechenzentrum“ und „IT-System“. Die Metriken der ersten Gruppe betrachten dabei

3 Definition der LDEE Metrik

entsprechend ihrer Bezeichnung das komplette Rechenzentrum und repräsentieren somit die High-Level-Betrachtung. Mit Hilfe dieser Metriken kann zwar erkannt werden, ob ein Rechenzentrum ineffizient betrieben wird, allerdings lässt sich der Ursprung der Ineffizienz mangels detaillierter Daten eines niedrigeren Levels nicht identifizieren. Zu diesem Zweck ist die zweite Gruppe der Metriken nutzbar, die entsprechend der Low-Level-Betrachtung einzelne IT-Systeme bzgl. ihrer Energieeffizienz bewerten. Allerdings müssten diese Metriken bzw. Benchmarks für alle in Frage kommenden Systeme angewandt werden.

Da keine der Metriken sowohl eine Bewertung eines gesamten Rechenzentrums als auch von Einzelkomponenten liefert, wird der beschriebene Anwendungsfall bislang nicht abgedeckt. Dazu ist nach aktuellem Stand der Technik eine Kombination mehrerer Metriken und Benchmarks zur Berücksichtigung aller RZ-Komponenten notwendig.

Praktikabilität Die Praktikabilität einer Metrik fasst die beiden Anforderungen R 3.10 für Anwendbarkeit und R 3.11 für Verständlichkeit zusammen. Sie wird in mehreren abstrakten Stufen von sehr gering über durchschnittlich bis sehr hoch angegeben.

Die einzige Metrik mit sehr hoher Praktikabilität ist die PUE, deren Resultate bislang auch am häufigsten referenziert werden. Da mit der IT- und RZ-Leistungsaufnahme/Energie nur zwei aggregierte Messwerte benötigt werden, ist sie sehr schnell und einfach anwendbar. Zudem ist das Resultat der PUE nach kurzer Erläuterung zu verstehen. Die CPE und FVER setzen auf ähnlich einfach zu ermittelnde Messdaten, ihre Ergebnisse erfordern jedoch etwas mehr Zeit zum Verständnis und zur korrekten Einordnung. Ihre Praktikabilität ist somit hoch.

Die Metriken mit mittlerer Praktikabilität sind entweder gut verständlich aber nicht einfach anwendbar oder genau umgekehrt. Zur ersten Kategorie gehören die DC-EEP, IT-PEW/ITEE, SPECpower und SWaP. Alle diese Metriken ermitteln die Performance bzw. die getätigte Arbeit und setzen diese mit der dafür benötigten Energie in Relation. Daraus resultieren gut verständliche, brauchbare, relative Effizienzwerte – die Ermittlung der Performance/Arbeit ist allerdings mit einigem Aufwand verbunden. In die zweite Kategorie fallen die ITEU, ITUE und TUE. Sie benötigen jeweils wenige aggregierte Messwerte, allerdings gestaltet sich eine Einordnung der Resultate schwierig.

Die DCeP und DCPpW besitzen eine geringe Praktikabilität aufgrund der Schwierigkeiten die sinnvolle Arbeit zu messen. Sie bedürfen einer manuellen Bewertung aller Prozesse, die im Rechenzentrum laufen. Einzig die gute Verständlichkeit erhöht die Praktikabilität etwas. Die gleiche Einordnung erhält auch der Benchmark TPC-Energy, welcher aufgrund sehr komplexer Testumgebungen bestehend aus mehreren miteinander verbundenen Server- und Stagesystemen schwieriger anzuwenden ist als beispielsweise der SPECpower-Benchmark, der auf einem einzelnen Server ausgeführt wird. Die Metriken CADE und DPPE besitzen die geringste Praktikabilität, da sie beide auf vielen Teilmetriken auf-

setzen und die Messung dadurch mit hohem Aufwand verbunden ist. Zudem ist eine Einordnung der Resultate schwierig, da der Sinn der Kombination der Teilmetriken in beiden Fällen nicht offensichtlich ist.

3.3.1 Zusammenfassung

Keine der untersuchten Metriken kann alle Anforderungen erfüllen und damit für jeden der in Abschnitt 3.1 beschriebenen Anwendungsfälle genutzt werden. Zur einfachen Betrachtung und einer groben Einordnung der Energieeffizienz eines RZ-Standorts sind noch viele der vorgestellten Metriken anwendbar. Auch ein Vorher-Nachher-Vergleich zur Bewertung von Optimierungsmaßnahmen ist mit diesen Metriken möglich, so lange nur ein Standort betrachtet wird. Liegt jedoch ein Vergleich mit anderen Rechenzentren im Fokus des Anwenders, reduziert sich die Anzahl der dafür nutzbaren Metriken beträchtlich. Soll dieser Vergleich zudem unabhängig von den jeweils vorhandenen Prozessen bzw. Arbeitslasten sein, d.h. allgemein, kann keine der Metriken angewandt werden, da hierfür eine Unabhängigkeit von Messdaten inkl. einer Berücksichtigung des Einflusses der Auslastung notwendig ist. Auch eine Unterscheidung nach RZ-Merkmalen wie Ausrichtung oder Redundanzgrad ist bei den untersuchten Metriken nicht definiert.

Eine Identifizierung des Ursprungs von Ineffizienzen ist mit keiner einzelnen Metrik möglich. Hierzu bedarf es einer Kombination von High- und Low-Level-Kennzahlen und eine Abdeckung aller Komponenten. Der dafür notwendige manuelle Aufwand und die resultierende hohe Komplexität dürften für potentielle Anwender allerdings eher abschreckend wirken. Das gleiche gilt für die Exploration von möglichen Modifikationen. Die untersuchten Metriken können mit den korrekten Annahmen zwar als brauchbarer Indikator dienen, allerdings ist das Finden passender Annahmen nicht trivial und auch für diesen Fall ist eine High- und Low-Level-Betrachtung nötig.

Zur Realisierung dieser Anwendungsfälle ist eine direkt darauf ausgelegte Lösung sinnvoll. Wie bereits in Kapitel 1 motiviert wurde, war dies der Auslöser der Entwicklung der Load Dependent Energy Efficiency (LDEE). Die LDEE setzt auf auslastungsabhängigen Komponentenmodellen auf, um eine allgemeine, vergleichbare Energieeffizienz sowohl des gesamten Rechenzentrums als auch im Detail für die RZ-Komponenten bestimmen zu können. Die Definition und die Zusammensetzung der Metrik werden im folgenden Abschnitt beschrieben.

3.4 Definition der LDEE

Die Definition der Load Dependent Energy Efficiency (LDEE) orientiert sich an den in Abschnitt 3.2 beschriebenen Anforderungen – mit dem Ziel die typischen Anwendungsfälle abzudecken. Maßgeblichen Einfluss auf die Definition haben dabei die beiden Anforderungen R 3.8 (Unabhängigkeit von Messdaten) sowie R 3.2 (Berücksichtigung der Auslastungsabhängigkeit), da diese die Ausführung

3 Definition der LDEE Metrik

der Metrik für beliebige Eingabedaten ermöglichen. Um diese beiden Anforderungen zu erfüllen wird ein Ansatz gewählt, der die Metrik direkt mit Modellen der RZ-Komponenten koppelt. Diese Modelle integrieren dabei alle Abhängigkeiten der bzgl. Energieeffizienz (R 3.1) relevanten Parameter zu den möglichen Betriebszuständen des Rechenzentrums bzw. der jeweiligen Komponenten.

Nach der Definition der Energieeffizienz aus Abschnitt 2.5 sollen diese Modelle die sinnvolle Arbeit sowie die Leistungs- oder Energieaufnahme abbilden. Wie bereits in Abschnitt 3.2.1 angemerkt wurde, steht diese Definition der Energieeffizienz allerdings im Konflikt mit der Anforderung R 3.4 für Objektivität, da die sinnvolle Arbeit für jedes Rechenzentrum manuell definiert werden muss. Anstelle der sinnvollen Arbeit betrachtet die LDEE daher als Repräsentant für die Produktivität eine Kombination aus Auslastung und Performance, welche im Prinzip die tatsächlich genutzte IT-Kapazität darstellt. Dies hat mehrere Vorteile:

- Die Auslastung und Performance kann jeweils objektiv ermittelt werden (R 3.4).
- Viele Performancedaten zu diversen IT-Komponenten werden zu Vergleichszwecken veröffentlicht und können als Ersatz für eigene Messungen dienen (R 3.8, R 3.10).
- Die Auslastung der IT lässt sich relativ leicht messen ohne dabei den Produktivbetrieb negativ zu beeinflussen (R 3.10)
- Durch die im Modell abgebildete Auslastungsabhängigkeit (R 3.2) lassen sich Resultate für beliebige gegebene Auslastungen ermitteln.

Durch die Nutzung von Modellen anstelle direkter Messdaten lassen sich also viele Anforderungen abdecken. Bei der Definition der Energieeffizienz (R 3.1) wird zwar ein Kompromiss eingegangen, dafür erfüllt dieses Vorgehen die Auslastungsabhängigkeit (R 3.2), die Objektivität (R 3.4), die Messunabhängigkeit (R 3.8), die Flexibilität (R 3.9) und die Anwendbarkeit (R 3.10). Die Berücksichtigung von Modellen einzelner Komponenten ermöglicht zudem die Betrachtung der Energieeffizienz auf unterschiedlichen Ebenen (R 3.7).

Theoretisch könnten diese Modelle auch mit bekannten Metriken kombiniert werden, um Resultate für beliebige Eingaben zu erhalten. Allerdings entsprechen bei den meisten Metriken die Eingangsparameter nicht denen der Modelle. So ist beispielsweise durch die Anforderung der Auslastungsabhängigkeit die IT-Auslastung ein Parameter der Modelle, diese findet allerdings nur in wenigen Metriken Berücksichtigung. Aus diesem Grund werden die LDEE und die zugrundeliegenden Modelle in Kombination entwickelt. So berücksichtigt die Metrik die notwendigen Eingangsparameter der Modelle und die Modelle liefern die notwendigen Abschätzungen zur Ermittlung der für die LDEE definierte Energieeffizienz.

3.4.1 Definition der Hauptfunktion

Die LDEE ist als Funktion über die IT-Auslastung $\vec{l} \in \mathbb{R}^2$ des Rechenzentrums mit l_1 als Server- bzw. l_2 als Storageauslastung mit $l_1, l_2 \in [0, 1]$ und die Außentemperatur T [°C] am Freikühler definiert:

$$LDEE(\vec{l}, T) = \frac{perf_{DC}(\vec{l})}{power_{DC}(\vec{l}, T)} \quad (3.1)$$

Die LDEE ist demnach die Relation der Performance $perf_{DC}$ zur entsprechenden Leistungsaufnahme $power_{DC}$ des gesamten untersuchten Rechenzentrums DC für die gegebenen Parameter \vec{l} und T . Das Rechenzentrum ist dabei, wie in Abschnitt 2.1 beschrieben, ab der Haupteinspeisung definiert. Die Performance als auch die Leistungsaufnahme werden wie die Hauptmetrik über Funktionen, die mit den übergebenen Parametern aufgerufen werden, berechnet. Beide Funktionen fassen die Einzelwerte der RZ-Komponenten zu einem aggregierten Wert für das gesamte Rechenzentrum zusammen. Der jeweilige Aufbau der Funktionen ist in den Abbildungen 4.1 und 4.2 schematisch dargestellt und wird in den Abschnitten 4.4 und 4.5 im Detail beschrieben.

3.4.2 Funktionen für Performance und Power

Wie bereits andiskutiert wurde, wird für den Zähler der LDEE-Funktion nicht die der Standarddefinition für Energieeffizienz entsprechende (vgl. Abschnitt 2.5.2) sinnvolle Arbeit genutzt, da diese nicht objektiv ermittelt werden kann. Anstelle dessen wird eine Auswahl alternativer Parameter, die in Abschnitt 2.5.4 beschrieben werden, verwandt. Die Wahl hängt dabei direkt mit der Erfüllung der Anforderungen an die LDEE zusammen.

Ein häufig in Energieeffizienzmetriken verwandter Parameter ist die insgesamt getätigte Arbeit. Im Gegensatz zur sinnvollen Arbeit werden dabei auch die Arbeiten berücksichtigt, die nicht zu den Geschäftsprozessen beitragen (z.B. Altlasten). Dadurch entfällt die subjektive manuelle Einordnung der Arbeit und die Resultate sind objektiv (R 3.4) und vergleichbar (R 3.5). Die geeignete Wahl des Maßes (z.B. Operationen pro Sekunde) ist jedoch schwierig. Durch Arbeitslasten mit verschiedenen Schwerpunkten, die die Systemressourcen der IT-Systeme unterschiedlich belasten, ist die Repräsentativität der verrichteten Arbeit mit einer einzelnen Kennzahl nicht gewährleistet. Des Weiteren ist es sehr aufwändig, die Systemparameter aller IT-Systeme kontinuierlich zu messen und zu speichern (R 3.10).

Ein anderer in EE-Metriken betrachteter Parameter ist die Performance der IT-Systeme. Die Performance repräsentiert dabei die maximale Arbeit bzgl. einer vordefinierten Arbeitslast, die ein System verrichten kann. Da diese Arbeitslasten bereits mehrere Systemressourcen abdecken, sind sie repräsentativer als einzelne Systemparameter. Es werden zudem alle Systeme mit der gleichen Arbeitslast untersucht, so dass die Vergleichbarkeit (R 3.5) gegeben ist. Die Anwendbarkeit

3 Definition der LDEE Metrik

(R 3.10) kann für diesen Ansatz im Vergleich zu einer direkten Messung der Systemparameter erhöht werden, indem bereits verfügbare Performanceresultate verwendet werden. Wird allerdings nur die maximale Performance betrachtet, wird die Energieeffizienz (R 3.1) nicht korrekt abgebildet, da der Nutzungsfaktor fehlt.

Für die LDEE wird zur Abschätzung der verrichteten Arbeit die Performance der Einzelsysteme (möglichst vieler Benchmarks) mit der jeweiligen Auslastung kombiniert. Die $perf_{DC}$ -Funktion liefert dann die zusammengefasste auslastungsabhängige (und damit genutzte) Gesamtperformance des Rechenzentrums in Referenzeinheiten ru mit $perf_{DC} = \{p \in \mathbb{R} | p \geq 0\}$. Die Referenzeinheiten definieren sich aus den durchschnittlichen Server- bzw. Stagespezifikationen anhand derer die Performancemodelle charakterisiert wurden, vgl. Abschnitt 4.4. Die Skalierung mit der Auslastung repräsentiert dann den Anteil der vorhandenen Einheiten, welcher tatsächlich genutzt wird. Die Funktion $power_{DC}$ wiederum liefert die elektrische Leistung in Kilowatt, welche bei der ermittelten genutzten Performance durch die IT und die Infrastruktur aufgenommen wird, vgl. Abschnitt 4.5. Insgesamt gibt die LDEE damit die Energieeffizienz in Anzahl der genutzten IT-Referenzeinheiten pro Kilowatt (ru/kW) an.

3.4.3 Auslastungsabhängigkeit

Das zentrale Element der Definition der LDEE ist die Berücksichtigung der Auslastungsabhängigkeit der Performance und Power der RZ-Komponenten auf allen Ebenen, indem diese in den Einzelmodellen abgebildet und bis auf die höchste Ebene propagiert wird. Durch die Nutzung solcher lastabhängiger Modelle ist es möglich, die Energieeffizienz für beliebige gegebene Auslastungen zu ermitteln. Dies ermöglicht wiederum eine Beurteilung der Energieeffizienz von Rechenzentren unabhängig ihrer tatsächlich betriebenen Arbeitslasten. Die LDEE kann dadurch einerseits die konkret vorliegende Energieeffizienz basierend auf aktuellen Messdaten liefern, andererseits kann die LDEE auch die Energieeffizienz für alternative Betriebszustände abschätzen. Neben der für Energieeffizienzmetriken typischen Anwendung zur Ermittlung des Ist-Zustandes kann die LDEE somit auch zur Planung genutzt werden.

3.4.4 Eingabeparameter

Mit der Auslastungsabhängigkeit als zentrale Anforderung an die LDEE ist die Auslastung der IT der wichtigste Eingabeparameter der Metrik. Wie bereits erwähnt wurde, ist die Auslastung $\vec{l} \in \mathbb{R}^2$ als Spaltenvektor mit den beiden Komponenten l_1 für die Server- und l_2 für die Storageauslastung definiert. Die beiden Auslastungen stellen dabei jeweils die Relation der aggregierten genutzten Kapazitäten gegenüber der insgesamt verfügbaren Kapazität für alle entsprechenden Systeme des betrachteten Rechenzentrums dar. Der Wertebereich liegt damit für beide Komponenten zwischen 0 und 1. Die relative Auslastung wird

der absoluten Auslastung als Eingabe der Metrik vorgezogen, da sie die Größe der untersuchten Rechenzentren abstrahiert und die Resultate unterschiedlich großer Rechenzentren somit besser verglichen werden können bzw. solche Vergleiche nachvollziehbarer sind. Des Weiteren ist unter Kenntnis der konkreten IT-Systeme jederzeit eine nachträgliche Konvertierung zwischen relativen und absoluten Auslastungen möglich.

Die aggregierte als auch die maximal mögliche absolute Serverauslastung wird in Prozent bzgl. einer ausgewählten Referenzmaschine betrachtet. Das bedeutet, dass für jede 100% Auslastung eine Referenzmaschine komplett ausgelastet sein würde. Aufgrund unterschiedlicher Leistungsfähigkeiten der Servertypen ist die Auslastung allerdings nicht direkt übertragbar, vgl. Abschnitt 2.5.4. Die gleiche Arbeit wird von leistungsfähigeren Servern in einer kürzeren Zeitspanne erbracht, wodurch die durchschnittliche Auslastung über den gleichen Zeitraum bei diesen Servern geringer ist. Um diesem Aspekt Rechnung zu tragen, ist bei der Übertragung der Auslastung auf einzelne Serversysteme eine Angleichung notwendig. Demnach wird für jedes Serversystem die jeweilige Auslastung auf die hypothetische Auslastung dieser Arbeitslast auf der definierten Referenzmaschine umgerechnet. Zur Bestimmung der maximal möglichen absoluten Auslastung wird dies für jeweils 100% ausgelastete Serversysteme und für die aggregierte Auslastung mit den jeweils bekannten Auslastungswerten durchgeführt. Das genaue Vorgehen wird in Abschnitt 4.3.3 erläutert.

Die Auslastung l_2 der Stagesysteme definiert sich über die Summe der Speicheranforderungen in Relation zur insgesamt verfügbaren Speicherkapazität. Diese ist zwischen unterschiedlichen Systemen direkt übertragbar und bedarf damit im Gegensatz zur Serverauslastung keiner weiteren Anpassung. Da nur die tatsächlich zur Verfügung gestellte (Netto-)Speicherkapazität betrachtet wird, ist die Storageauslastung abstrahiert von den zugrundeliegenden Redundanzverfahren (z.B. RAID), so dass hier nicht weiter unterschieden werden muss.

Für Zusammenfassungen der Ergebnisse der LDEE wird anstelle des Auslastungsvektors \vec{l} übersichtshalber das arithmetische Mittel über alle Vektorkomponenten – also der Durchschnitt der relativen Auslastungen – genutzt.

3.4.5 RZ-Klassen

Zur Einordnung der Energieeffizienz eines Rechenzentrums ist ein Vergleich mit anderen Rechenzentren erstrebenswert. Wie allerdings im entsprechenden Anwendungsfall (vgl. Abschnitt 3.1.2) bereits erwähnt wurde, ist ein Vergleich zwischen unterschiedlichen Rechenzentrumstypen nicht unbedingt sinnvoll. Zur Erfüllung der Anforderung R 3.6 wurden daher zwei Eigenschaften von Rechenzentren identifiziert, die maßgeblichen Einfluss auf die Bewertung der Energieeffizienz besitzen.

Vergleich nach Ausrichtung

Die Ausrichtung eines Rechenzentrums (z.B. Unternehmens- oder Cloud-RZ) bestimmt die Zusammensetzung der IT-Systeme und der darauf ausgeführten Arbeitslasten, vgl. Abschnitt 2.1.1. Da man Serversysteme nicht direkt mit Storage- oder Netzwerkkomponenten vergleichen kann, sorgen unterschiedliche Verteilungen der IT auf diese drei Kategorien unter Umständen für deutlich abweichende Ergebnisse in der Performance und damit der Energieeffizienz. Unterschiedliche Arbeitslasten (z.B. HPC vs. Webservices) können diese Unterschiede noch weiter erhöhen. Ein Vergleich innerhalb ähnlicher RZ-Typen soll Diskrepanzen durch extreme Unterschiede in der IT unterbinden und versucht dadurch die Aussagekraft eines Vergleichs von Ergebnissen zu erhöhen.

Als erster Schritt wird dazu eine Klasseneinteilung von Rechenzentren nach ihrer Ausrichtung vorgenommen. Eine mögliche Einteilung wurde dazu bereits in Abschnitt 2.1.1 beschrieben. Basierend auf diesen RZ-Typen werden in einem zweiten Schritt Gewichte für unterschiedliche Relationen zwischen Server-, Storage- und Netzwerkkomponenten definiert. Wird nun ein Vergleich zweier Rechenzentren in der gleichen RZ-Klasse durchgeführt, werden die Performancewerte der IT-Komponenten mit den jeweiligen Gewichten der Klasse angepasst, um wichtige Aspekte hervorzuheben. So werden beispielsweise bei einem Vergleich der Energieeffizienz von speicherintensiven Rechenzentren (z.B. Cloud-RZ mit Fokus auf Storagediensten) die Storage-Systeme deutlich höher gewichtet als in den anderen Klassen, um den Speicheranforderungen dieser Rechenzentren Rechnung zu tragen. Vergleiche müssen dabei allerdings immer mit identischen Gewichten durchgeführt werden.

Für einen klassenübergreifenden Vergleich gibt es zwei weitere Möglichkeiten. Mit einer ausbalancierten Gewichtung fließen die IT-Komponenten zu gleichen Anteilen in die Bewertung ein. Standardmäßig wird hingegen gar keine Gewichtung vorgenommen, so dass die genutzte Performance aller IT-Komponenten ohne weitere Adaption aggregiert wird. Eine detaillierte Erläuterung der Klasseneinteilung und ihre jeweiligen Gewichte befindet sich in Abschnitt 4.4.8.

Vergleich nach Verfügbarkeit

Unabhängig von der Ausrichtung des Rechenzentrums beeinflusst die anvisierte Verfügbarkeitsklasse (s. Abschnitt 2.1.2) die Architektur und Redundanz – vor allem hinsichtlich der Infrastrukturkomponenten. Ein Rechenzentrum mit einer hohen Verfügbarkeit benötigt deutlich mehr Energie zur Bereitstellung der gleichen Leistungsfähigkeit als ein Rechenzentrum mit niedriger Verfügbarkeit und somit wenigen Redundanzen. Die Energieeffizienz verringert sich also maßgeblich mit der vorgehaltenen Redundanz. Um einen fairen Vergleich von Rechenzentren zu ermöglichen sollte die Verfügbarkeitsklasse und damit der Redundanzgrad berücksichtigt werden. An der Definition und der konkreten Bestimmung der Energieeffizienz ändert sich dabei im Gegensatz zum Vergleich nach Ausrich-

tung nichts. Das Rechenzentrum wird lediglich durch die Anwender der LDEE der korrekten Verfügbarkeitsklasse zugeordnet und diese dann für Vergleiche berücksichtigt.

3.4.6 Definition des Resultats

Die LDEE setzt sich aus dem Verhältnis der mit der Auslastung skalierten Performance und der zur Bereitstellung aufgenommenen elektrischen Leistung zusammen. Wie bereits in Abschnitt 3.4.2 beschrieben wurde, ist das Resultat entsprechend die pro benötigter Leistungseinheit tatsächlich genutzte Performance der IT in Anzahl der Referenzeinheiten – ru/kW (reference units per kilowatt). Eine Referenzeinheit bzw. Referenzmaschine ist das abstrakte, durchschnittliche Server- bzw. Storage-System, an denen die Performance in den jeweiligen Server- bzw. Storageperformancemodellen skaliert wird. Sie definiert sich durch die Menge aller Server- bzw. Storagekomponenten, die in die Charakterisierung der Performancemodelle einfließen. Die Menge dieser Systeme wird dabei einmal festgelegt, so dass die Referenzmaschinen für alle Anwendungen der Metrik gleich definiert sind um die Vergleichbarkeit (R 3.5) zu gewährleisten. Eine detailliertere Herleitung der Referenzmaschinen findet sich in Abschnitt 4.4.8.

Parameterabhängiger Effizienzverlauf

Das Resultat beschränkt sich gemäß der Auslastungsabhängigkeit der LDEE (R 3.2) nicht auf einen Einzelwert, sondern beschreibt einen Verlauf der Energieeffizienz in Abhängigkeit der Eingabeparameter der Auslastung und der Außentemperatur. Dieses Format zur Darstellung der Energieeffizienz wurde gewählt um den Einfluss zeitlich bedingter Nutzungsschwankungen und wechselnder Wetterbedingungen auf den Rechenzentrumsbetrieb direkt im Ergebnis zu berücksichtigen. Durch die Darstellung eines Einzelwerts, wie es für die gängigen Metriken gehandhabt wird, kann hingegen nur ein bestimmter Betriebszustand dargestellt werden. Ein Effizienzverlauf liefert somit deutlich mehr Informationen für den Anwender, die für weitere Planungen genutzt werden können.

Eine beispielhafte Darstellung des Energieeffizienzverlaufs findet sich in Abbildung 3.1. Die Eingabeparameter, die jeweils zur Berechnung der LDEE genutzt wurden, sind die durchschnittliche RZ-Auslastung in einem Intervall von $[0, 1]$ und die Außentemperatur in $^{\circ}C$. Die RZ-Auslastung wurde dabei für die Server- und Storage-Systeme per arithmetischem Mittel zusammengefasst, vgl. Abschnitt 3.4.4. Trägt man diese beiden Parameter gegen die LDEE auf, erhält man den dargestellten 3-dimensionalen Verlauf der LDEE in Referenzeinheiten pro Kilowatt Leistungsaufnahme.

Im Beispiel aus Abbildung 3.1 kann man sehr gut erkennen, dass die Energieeffizienz des Rechenzentrums proportional vom Nutzungsgrad der IT abhängt. In diesem Fall ist die Ursache die mangelnde energetische Skalierbarkeit der IT-Hardware, die unabhängig der anliegenden Arbeitslasten durchgehend betrieben

3 Definition der LDEE Metrik

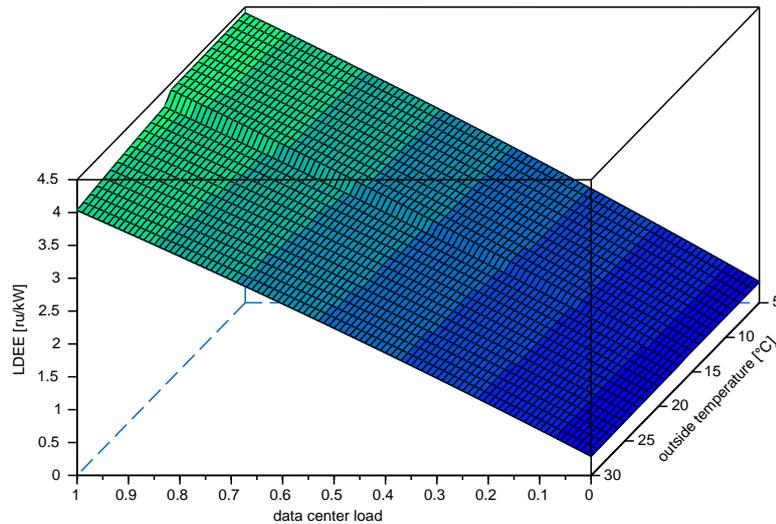


Abbildung 3.1: Verlauf der LDEE eines nicht optimierten Betriebs

wird und daher eine nahezu konstante Leistungsaufnahme besitzt. Dadurch können ebenso die Stromversorgung und die Klimatisierung nicht mit der Auslastung skalieren. Die gesamte RZ-Leistungsaufnahme bleibt letztlich konstant, während die genutzte Performance direkt mit der Auslastung skaliert. Als Vergleich wurde im Beispiel aus Abbildung 3.2 für das identische Rechenzentrum ein Last- und Powermanagement für Server angewandt, wie es in Abschnitt 2.4.3 beschrieben wurde. Die Leistungsaufnahme der IT-Hardware skaliert hierbei durch das dynamische Abschalten nicht benötigter Systeme deutlich besser mit der Auslastung, so dass sich neben der genutzten Performance auch die Leistungsaufnahme des Rechenzentrums mit sinkender Auslastung reduziert. Die LDEE bleibt damit länger auf einem höheren Niveau als im ersten Beispiel. Nur die nicht mitskalierende Infrastruktur sorgt in diesem Beispiel für einen starken Abfall der LDEE bei niedriger Auslastung. In den Beispieldiagrammen ist zudem zwischen 15°C und 20°C ein Versatz in der LDEE erkennbar. Dieser resultiert aus dem Übergang zwischen reiner Freiluftkühlung und dem Betrieb der Kältemaschinen und dem damit einhergehenden Anstieg in der Leistungsaufnahme.

Die dreidimensionale Darstellung der Energieeffizienz gibt den Anwendern der Metrik Hilfestellungen bei der Identifikation von ineffizienten Betriebsmodi. Starke Abfälle in der LDEE für bestimmte Betriebsparameter sind direkt ersichtlich und ermöglichen so die Eingrenzung der Ursachen. Eine optimale Skalierbarkeit eines Rechenzentrums ist erreicht, wenn sich die Ebene im dreidimensionalen Diagramm über einen konstanten (und möglichst hohen) LDEE-Wert erstreckt. In Bezug zur Auslastung sollte dies das Ziel der Rechenzentrumsbetreiber sein

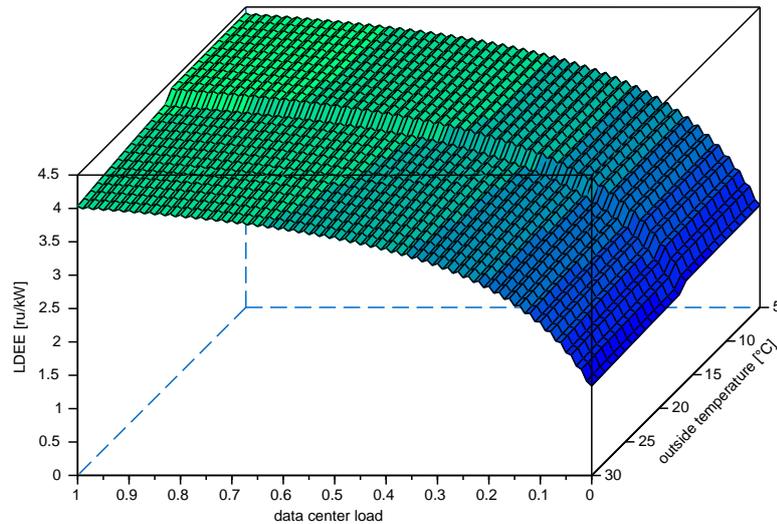


Abbildung 3.2: Verlauf der LDEE bei Nutzung von Last- und Powermanagement für Server

und ist durch sinnvolles Management der Arbeitslasten und der IT-Systeme annäherungsweise erreichbar. Bezüglich der Außentemperatur sind der optimalen Skalierbarkeit der temperaturabhängigen Effizienz der Klimatisierung aufgrund der thermodynamischen Gesetze Grenzen gesetzt. Kältere Außentemperaturen ermöglichen immer eine effizientere Kühlung, so dass für Rechenzentren mit sinkender Temperatur eine stetig steigende LDEE erkennbar sein sollte. Ist dies nicht der Fall, sind Optimierungen der Klimatisierung möglich.

Repräsentativer Einzelwert

Ein direkter, quantitativer Vergleich der beschriebenen Energieeffizienzkurven ist komplex, da viele Einzelwerte miteinander verglichen werden müssen. Dies kann zwar automatisiert in kurzer Zeit durchgeführt werden, für eine gute Anwendbarkeit der Metrik (R 3.10) ist es allerdings wichtig, dass die LDEE zusätzlich zur Kurve in einem einzelnen Wert zusammengefasst wird. Dieser sollte dabei so realisiert sein, dass ein Vergleich unterschiedlicher Rechenzentren (R 3.5) damit möglich ist. Die Ermittlung eines Einzelwerts bzw. Repräsentanten ist daher klar definiert. Für jeden der beiden Eingabeparameter wurde ein Vektor mit jeweils vier konkreten Werten festgelegt, für die in Kombination 16 Energieeffizienzwerte bestimmt werden, die einen Querschnitt der realistischen Betriebszustände darstellen.

Die RZ-Auslastung setzt sich aus den Werten $\{0.05, 0.3, 0.5, 0.8\}$ zusammen.

3 Definition der LDEE Metrik

Die Extremfälle des Leerlaufbetriebs (0.0) und der maximalen Auslastung (1.0) wurden bewusst nicht berücksichtigt, da diese Fälle in der Praxis nicht auftreten. Als Minimal- und Maximalauslastung werden die realistischeren Werte von 5% bzw. 80% betrachtet. Zudem wird neben der mittleren Auslastung (0.5) eine niedrigere RZ-Auslastung von 0.3 integriert, um den Fokus der Einzelwertberechnung auf die typischen, niedrigen RZ-Auslastungen zu lenken. Diese reichen je nach Untersuchung von 7% nach Liu [Liu11] bis zu ca. 30% nach Barroso et al. [BCH13, BH07]. Die RZ-Auslastung wird in diesem Betrachtungsfall jeweils gleichverteilt auf die Server- und Stagesysteme eines Rechenzentrums.

Für die Außentemperatur wurde ein Vektor mit den Werten $\{2, 8, 14, 20\}$ definiert. Die ausgewählten Temperaturen orientieren sich an den typischen Jahresverläufen in warmgemäßigten (nemoralen) Klimazonen. Der Großteil der Internetknotenpunkte weltweit und damit auch viele Rechenzentren, die sich in der Nähe dieser Verkehrsknoten ansiedeln, liegen in dieser Klimazone. Ein gutes Beispiel hierfür ist der Standort Frankfurt am Main, an dem mit dem DE-CIX einer der weltweit größten Internetknotenpunkte¹ beheimatet ist, weshalb dessen typische Außentemperaturen² als Basis für den Temperaturvektor dienen. Extrem hohe oder niedrige Temperaturen wurden dabei nicht betrachtet, da diese nur einen geringen Anteil der Zeit ausmachen.

Die für die beiden beschriebenen Vektoren der Eingabeparameter ermittelten Resultate der LDEE werden per Mittelwertbildung zu einem Repräsentanten zusammengefasst. Dabei wurde das arithmetische Mittel verwandt um den Einfluss von potentiellen Ausreißern in den berechneten Energieeffizienzwerten im finalen Repräsentanten abzubilden. Da die Eingabevektoren bereits so zusammengestellt wurden, dass unrealistische oder extrem selten auftretende Ausprägungen nicht enthalten sind, spielen erkannte starke Abweichungen im Guten wie im Schlechten eine relevante Rolle für eine repräsentative allgemeine Energieeffizienz.

Der Repräsentant der LDEE eines Rechenzentrums wird in erster Linie für direkte und schnelle Vergleiche mit anderen Rechenzentren benötigt. Entsprechend sollte das Resultat auch den RZ-Klassen der Ausrichtung und der Verfügbarkeit zugeordnet werden. Dazu werden die jeweiligen RZ-Klassen in eckigen Klammern dem Einzelresultat in der Form $LDEE[C_i, T_j] = x \text{ ru/kW}$ hinzugefügt, wobei die RZ-Ausrichtung C_i sich an den in Abschnitt 4.4.8 beschriebenen Einteilungen orientiert und die Verfügbarkeitsklassen T_j auf der Definition in Abschnitt 2.1.2 basieren.

Kennzeichnung der Eingabedaten

Um Anwendern der Metrik direkt mit den Resultaten weitere Informationen zu den verwandten Eingabedaten zu liefern, wurde eine zusätzliche Kennzeichnung definiert. So wird ein auf realen Messungen basierendes Einzelresultat der LDEE

¹https://prefix.pch.net/applications/ixpdir/?show_active_only=0&sort=traffic&order=desc (letzter Zugriff: 26.11.2015)

²<http://de.climate-data.org/location/447/> (letzter Zugriff: 26.11.2015)

als $LDEE_m$ gekennzeichnet, so dass bei der Betrachtung und weiteren Nutzung des Resultats direkt bekannt ist, dass dieses aus dem laufenden Betrieb ermittelt wurde. Die Nutzung der zuvor definierten Eingabevektoren zur Ermittlung des Repräsentanten wird hingegen als $LDEE_r$ dargestellt. Auch die beschriebene optionale Zuordnung zu RZ-Klassen kann inklusive der zusätzlichen Kennzeichnung vorgenommen werden, indem die Ausrichtung C_i und die Verfügbarkeit T_j des Rechenzentrums in der Form $LDEE_x[C_i, T_j]$ angehängt werden, wobei für x mit m oder r die Art der Eingabedaten gekennzeichnet wird. Durch die konsequente Kennzeichnung der Ergebnisse können unzulässige Vergleiche von Einzelresultaten mit unterschiedlichen Rahmenbedingungen verhindert werden.

Im Gegensatz zu den Einzelresultaten wird der parameterabhängige Effizienzverlauf nicht weiter gekennzeichnet, da dieser immer als Diagramm dargestellt und nicht für Vergleichszwecke zwischen Rechenzentren genutzt wird.

4 Modellbildung

Dieses Kapitel beschreibt die Modelle, auf deren Basis die LDEE-Metrik die Energieeffizienz eines Rechenzentrums bestimmt. Einleitend wird dazu in Abschnitt 4.1 ein Beispielszenario aufgeführt, welches den Vorteil der Nutzung von Modellen gegenüber realen Messungen anschaulich darstellt. Grundsätzlich besitzt jede Komponente des Rechenzentrums ein eigenes Powermodell (Abschnitt 4.5) und die IT-Hardware zusätzlich noch jeweils ein Performancemodell (Abschnitt 4.4). Die Modelle sind jeweils lastadaptiv, wobei die konkret vorliegende Last über ein IT-Lastmodell (Abschnitt 4.3) abgeschätzt wird. Zusammengesetzt kann ein komplettes Rechenzentrum hinsichtlich genutzter Performance und benötigter elektrischer Leistung abgeschätzt werden. Eine entsprechende Evaluation der Modelle wird in Abschnitt 4.6 beschrieben.

4.1 Beispielszenario

An einem Beispielrechenzentrum soll die erweiterte Anwendbarkeit und Nützlichkeit der LDEE gegenüber bislang verbreiteten Energieeffizienzmetriken veranschaulicht werden. Angenommen, ein mittelständisches Unternehmen besitzt einen Rechnerraum mit insgesamt 120 Server- und drei Storage-Systemen, auf denen gängige Unternehmens-SW wie Emailserver, CRM- und ERP-Systeme, Dateiserver und ähnliches betrieben und den Mitarbeitern zur Verfügung gestellt werden. Der Rechnerraum ist in der Verfügbarkeitsklasse (Tier) 1 einzuordnen. Die IT- und Infrastrukturkomponenten sind ca. sechs Jahre alt und die Verteilung der Leistungsaufnahme ist 70kW für die IT und 60kW für die Infrastruktur. Im Zuge eines unternehmensweiten Energieaudits möchte das Unternehmen den Rechnerraum nun gerne durch eine Modernisierung energieeffizienter gestalten. Der erste Schritt dazu ist die Analyse des Ist-Zustands. Zu diesem Zweck werden Messungen durchgeführt und die Daten über bekannte Metriken zusammengefasst.

Die Power Usage Effectiveness lässt sich schnell über wenige Messungen ermitteln und beträgt 1,86. Aus Erfahrung ist bekannt, dass dies kein schlechter Wert ist, dieser allerdings auch Spielraum für Optimierungen vermuten lässt¹. Diese sind auf Seiten der Infrastruktur anzusiedeln, da die PUE eine Metrik des Infrastrukturoverheads ist. Weiter eingrenzen lässt es sich allerdings nicht und über die IT-Seite kann mit der PUE überhaupt keine Aussage getroffen werden. Als nächstes wird also die Produktivitätsmetrik Data Center Energy Productivity

¹<http://www.vertatique.com/no-one-can-agree-typical-pue> (letzter Zugriff: 26.11.2015)

4 Modellbildung

(DCeP) angewandt. Nach einer manuellen Definition der sinnvollen Arbeit wurden ca. 6 Millionen abgeschlossene Tasks für einen Messzeitraum von einem Tag ermittelt. Im Verhältnis zur dafür benötigten Energie von 3120kWh erhält man eine DCeP von 1923 Tasks/kWh. Eine Aussagekraft für sich besitzt das Resultat nicht und ein Vergleich mit anderen Rechenzentren ist durch die manuelle Definition nicht möglich. Auch ein Herunterbrechen auf einzelne IT-Komponenten ist nicht sinnvoll, da viele Tasks über mehrere Systeme bearbeitet werden und der Einfluss einzelner Systeme somit schwierig zu bewerten ist.

Für eine Identifikation von Ansatzpunkten zur Optimierung sind die üblichen Metriken also nicht brauchbar, da hierzu neben der allgemeinen Bewertung auch eine Aufschlüsselung auf einzelne Komponenten notwendig ist. Diesen Ansatz verfolgt die LDEE mit der Nutzung von Performance- und Powermodellen für jede RZ-Komponente. Insgesamt liefert die LDEE als Ergebnis die zu einem Zeitpunkt genutzte IT-Performance im Verhältnis zur dafür benötigten elektrischen Leistung. Die Nutzung ergibt sich dabei über die Ressourcenauslastung der konkreten IT-Systeme. Diese wird mit der Performance kombiniert und in Einheiten von durchschnittlichen Referenzsystemen (ru) angegeben. Angenommen die einzelnen Server- und Storage-Systeme besitzen Performancekennzahlen zwischen 0,3 und 0,7 ru, so ergibt sich für das Beispielszenario bei einer Serverauslastung von 25% und einer Storageauslastung von 60% eine aggregierte, genutzte Performance von 15,2 und damit eine LDEE von $\frac{15,2}{130} \approx 0,12 \frac{\text{ru}}{\text{kW}}$. Dieser Wert ist direkt vergleichbar mit anderen Rechenzentren, optional auch bzgl. der Ausrichtung und der Verfügbarkeitsklasse. Über diese Vergleichswerte lässt sich eine grundsätzliche Tendenz der Energieeffizienz erkennen. Die Energieeffizienz in Referenzeinheiten pro kW lässt sich zudem für einzelne IT-Systeme ermitteln und miteinander vergleichen, so dass ineffiziente IT-Systeme identifiziert werden können. Ebenso lässt sich die Leistungsaufnahme der Powermodelle einzelner Infrastrukturkomponenten betrachten.

Nachdem das Unternehmen im ersten Schritt Ansatzpunkte zur Optimierung identifiziert hat, folgt nun eine Potentialanalyse der möglichen Maßnahmen. Energieeffizienzmetriken sind dabei ein nützliches Werkzeug, da sie einen Vorher-Nachher-Vergleich ermöglichen. Das Problem der gängigen Metriken ist allerdings, dass sie reale Messungen voraussetzen, so dass eine vollständige Bewertung erst nach der Umsetzung der Maßnahme vollzogen werden kann. Die LDEE hingegen kann auch mit Performance- und Powermodellen von Komponenten, die noch gar nicht im Rechenzentrum installiert sind, ermittelt werden. So können beispielsweise für die ineffizientesten Serversysteme Modelle moderner Systeme eingesetzt werden. Die genutzte Performance und die Leistungsaufnahme kann dann für beliebige Servertypen als auch Systemauslastungen sondiert werden, um die am besten geeignete Maßnahme zu ermitteln.

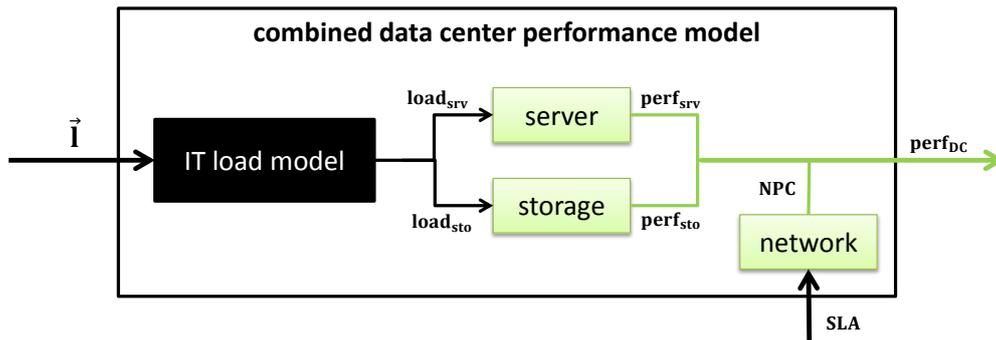


Abbildung 4.1: Kombiniertes RZ-Performancemodell

4.2 Aufbau der RZ-Modelle

Die LDEE ist definiert durch das Verhältnis der genutzten Performance zu der dafür benötigten elektrischen Leistung. Durch Messung der Parameter und einsetzen in die definierte Funktion in Gleichung 3.1 kann die LDEE für den Messzeitpunkt direkt ermittelt werden. Das zuvor beschriebene Beispielszenario hat jedoch aufgezeigt, dass die Nützlichkeit einer Metrik steigt, wenn sie auch ohne reale Messungen realistische Resultate für das gesamte RZ als auch für Teilkomponenten liefern kann. Um dies zu erreichen, werden für alle relevanten RZ-Komponenten jeweils eigene auslastungsabhängige Einzelmodelle erstellt, die dann zu einem RZ-Modell kombiniert werden. Dies wird zum einen für die genutzte Performance und zum anderen für die Leistungsaufnahme durchgeführt. Die Ausgaben der beiden resultierenden kombinierten RZ-Modelle bilden dann den Zähler sowie den Nenner der LDEE-Metrik.

Das RZ-Modell der genutzten Performance ist in Abbildung 4.1 dargestellt. Für alle Server- und Stagesysteme werden auf Basis öffentlich verfügbarer Benchmarkergebnisse lastadaptive Performancemodelle gebildet. Um die tatsächliche Nutzung des Rechenzentrums abzubilden, wird die resultierende Performance der Einzelmodelle der Komponenten mit einer abgeschätzten Auslastung entsprechend skaliert. Die Abschätzung der Auslastung bestimmt dabei ein IT-Lastmodell, welches etwaige Lastverteilungsverfahren im Rechenzentrum berücksichtigt, um eine realistische Verteilung der gesamten RZ-Arbeitslast auf die konkreten Komponenten ohne Notwendigkeit von Messungen zu erhalten. Die skalierte Performance der Server- und Stagesysteme fließt letztlich mit einem Koeffizienten der Netzwerkperformance (NPC) zusammen, der die Einhaltung der Netzwerk-SLA repräsentiert. Das Resultat des kombinierten RZ-Performancemodells ist dann im Prinzip eine mit der Auslastung und der Netzwerkperformance skalierte Summe der Server- und Storageperformance. Das Modell wird in Abschnitt 4.4 im Detail beschrieben.

Das RZ-Powermodell ist ähnlich wie das RZ-Performancemodell aufgebaut und wird in Abbildung 4.2 schematisch dargestellt. Es besteht aus lastadap-

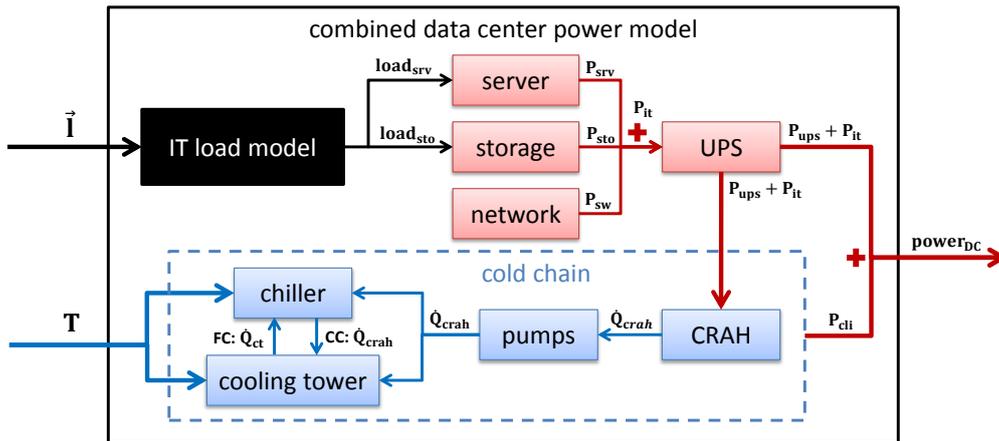


Abbildung 4.2: Kombiniertes RZ-Powermodell mit Unterscheidung zwischen Freikühlung (FC) und Kompressionskühlung (CC)

tiven Einzelmodellen aller relevanten RZ-Komponenten, die dann in Summe die Gesamtleistungsaufnahme des Rechenzentrums liefern. Die Verteilung der Auslastung auf die Server- und Storagekomponenten wird über das gleiche IT-Lastmodell abgeschätzt, so dass die elektrische Leistung bzgl. der vorliegenden genutzten Performance ermittelt wird. Da das IT-Lastmodell keine Auslastung zur Netzwerkhardware liefert, wird für die Netzwerkkomponenten nur eine statische Leistungsaufnahme abgeschätzt. Allerdings liegt der dynamische Anteil der Leistungsaufnahme nach Mahadevan et al. [MSBR09] bei unter fünf Prozent, so dass sich der Einfluss auf das RZ-Performancemodell in Grenzen hält.

Neben den IT-Komponenten werden auch die für den reibungslosen Betrieb der IT notwendigen Infrastrukturkomponenten betrachtet. Dies sind die Stromversorgung und die Klimatisierung. Die Leistungsaufnahme der IT-Komponenten läuft bei der unterbrechungsfreien Stromversorgung (USV bzw. UPS) zusammen, da die Verlustleistung der USV von der bereitgestellten elektrischen Leistung abhängt. Die Gesamtleistungsaufnahme der IT inkl. USV, welche der Abwärme entspricht, wird dann als Eingabe für die Klimamodelle benötigt. Die Klimamodelle besitzen untereinander weitere Abhängigkeiten, so dass es beispielsweise von der Außenlufttemperatur abhängt, ob Freikühlung genutzt wird oder nicht. In Abschnitt 4.5 wird die Modellierung der Leistungsaufnahme ausführlich erläutert.

4.3 IT-Lastmodell

Die Auslastungsabhängigkeit der LDEE-Metrik wird über Performance- und Powermodelle realisiert, die die Auslastung der Komponenten zur Abschätzung der Performance bzw. Leistungsaufnahme berücksichtigen und somit entsprechen-

de Eingabewerte benötigen. Die Auslastung der Komponenten ist dabei kein konstanter Wert, sondern hängt von diversen Faktoren wie z.B. der Gesamtnutzerzahl oder von ggf. verwendeten Last- und Powermanagementverfahren ab. Prinzipiell können die konkreten Lastverteilungen im Rechenzentrum gemessen und auf die Performance- und Powermodelle angewandt werden. Allerdings lässt sich damit nicht die Anforderung R 3.8 (Messunabhängigkeit) der LDEE-Metrik erfüllen. Um eine Messunabhängigkeit zu erreichen ist ein IT-Lastmodell notwendig, welches die Einflussfaktoren auf die Lastverteilung abstrahiert und somit die Verteilung der Last für beliebige Eingaben abschätzt.

4.3.1 Anforderungen

Die LDEE-Metrik liefert die Energieeffizienz für eine gegebene, relative, aggregierte Auslastung des gesamten Rechenzentrums. Das IT-Lastmodell erhält diese als Eingabe und muss die Verteilung der Last auf die IT-Hardware abschätzen. Falls im Rechenzentrum ein Last- und Powermanagement aktiv ist, muss das Modell zudem abschätzen, welche der IT-Hardwarekomponenten genutzt werden und welche aus Energiespargründen inaktiv sind. Entsprechend ergeben sich die folgenden Anforderungen für das IT-Lastmodell.

R 4.1 Unabhängigkeit von Messungen: Das IT-Lastmodell sollte Abschätzungen für Lastverteilungen liefern können, ohne auf historische Auslastungsdaten der IT-Systeme zurückgreifen zu müssen. Nur so kann die Anforderung R 3.8 der LDEE eingehalten werden.

R 4.2 Berücksichtigung der Betriebsart: Das IT-Lastmodell muss auf unterschiedliche Betriebsarten der IT eines Rechenzentrums anwendbar sein. Es muss die Lastverteilung in einem statischen ebenso wie in einem dynamischen Betrieb bestimmen können. Das heißt, neben der Auslastung der einzelnen Systeme muss das Modell auch die Wahl, welche Systeme überhaupt aktiv sind, abschätzen.

4.3.2 Modellansatz

Der Modellansatz zur Modellierung der IT-Last ist in Abbildung 4.3 dargestellt und unterteilt sich prinzipiell in zwei Schritte. Zunächst müssen in einem ersten Schritt die aktiven IT-Komponenten ermittelt werden, bevor für diese in einem zweiten Schritt die konkrete Auslastung abgeschätzt wird. Der erste Schritt berücksichtigt die Anwendung von dynamischen Powermanagementalgorithmen und damit die Anforderung R 4.2, durch die ein Teil der Serversysteme inaktiv sein kann, da sie temporär nicht benötigt werden. Da solche Algorithmen bislang nur für Serversysteme existieren, wird dieser Schritt zur Modellierung der Storgelast übersprungen. Dies gilt ebenso für Rechenzentren, die kein Powermanagement einsetzen. Der zweite Schritt – die Abschätzung einer konkreten

4 Modellbildung

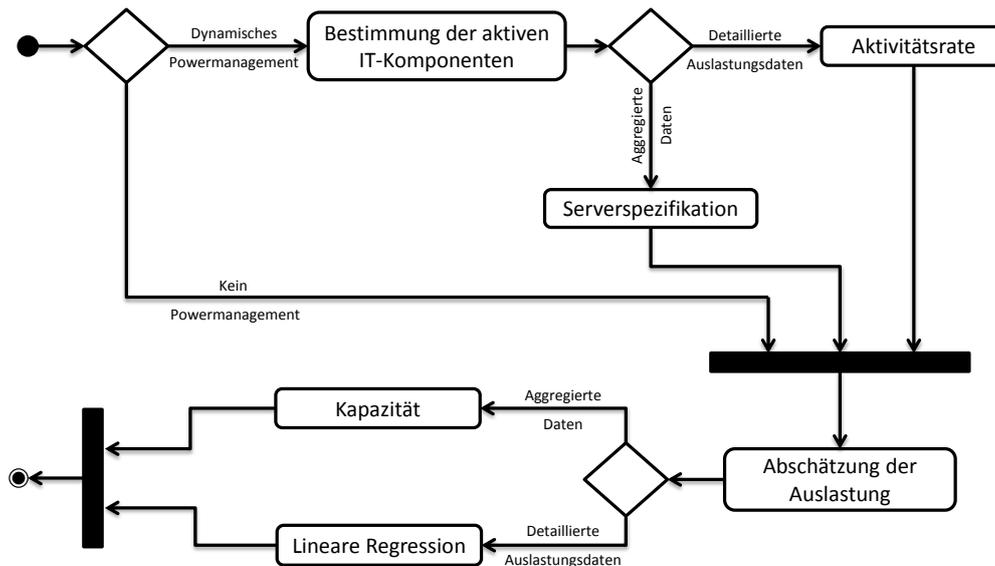


Abbildung 4.3: Aufbau des IT-Lastmodells, wobei die Wahl der ausgeführten Teilmodelle von der Konfiguration des Rechenzentrums abhängt

Auslastungsverteilung – wird dann für Server- sowie Storesysteme angewandt und ist unabhängig davon, ob diese energetisch gemanagt werden.

Die Wahl des genutzten Modells für jeden Schritt hängt dann von den vorhandenen Eingabedaten ab. Sind detaillierte historische Auslastungsdaten der einzelnen IT-Komponenten bekannt, projiziert das Modell diese in die Zukunft und erreicht damit relativ präzise Ergebnisse. Sind jedoch nur die aggregierte Rechenzentrumsauslastung sowie die Typen der IT-Komponenten bekannt, muss das Modell die Auslastung komplett unabhängig vergangener Verteilungen abschätzen, was in einem größeren Fehler resultiert. Der zweite Ansatz ist essentiell zur Erfüllung der Anforderung R 4.1.

Modellparameter

Die Art der Abschätzung der IT-Last und die Genauigkeit hängt von diversen Modellparametern ab, die nachfolgend vorgestellt werden.

- Aggregierte Rechenzentrumsauslastung: Die Summe der Auslastungen der IT-Systeme. Dies ist der Eingangsparameter des IT-Lastmodells, von dem die zu bestimmende Lastverteilung maßgeblich abhängt. Sie wird aus dem Eingangsvektor der LDEE – der relativen Auslastungen \vec{l} – berechnet. Die Serverauslastung wird dabei an die konkreten Servertypen angeglichen. Die Storageauslastung in Form der Speicherbelegung wird hingegen direkt anhand der vorhandenen Systemspezifikationen skaliert. Das jeweilige Vorgehen wird im folgenden Unterabschnitt 4.3.3 beschrieben.

- Betriebsart: Art des statischen oder dynamischen Betriebs. Wird ein dynamisches Powermanagement betrieben, sind zudem einige der Konfigurationsparameter erforderlich. Dies sind die Anzahl der in jedem Cluster gemagten VM und der Grad der parallelen Migrationen, vgl. Abschnitt 2.4.3.
- IT-Systeme: Liste aller im Rechenzentrum vorhandener IT-Systeme inkl. Spezifikation. Diese Informationen werden in der Regel durch die Anwendung der LDEE-Metrik (automatisch) erhoben und müssen dem IT-Lastmodell nur noch bereitgestellt werden.
- (optional) Auslastung der IT-Systeme: Analog zur aggregierten RZ-Auslastung, nur für jedes IT-System. Diese Messdaten ermöglichen genauere Resultate, beschränken aber den Wertebereich des Modells, welches dann nicht mehr messunabhängig (R 4.1) ist.

4.3.3 Konvertierung der Auslastung

Die LDEE besitzt als Eingangsparameter einen Vektor \vec{l} mit der relativen Auslastung der Server l_1 sowie der Stagesysteme l_2 . Das IT-Lastmodell hingegen benötigt als Eingabe die aggregierte, absolute Auslastung der IT-Systeme, deren Verteilung dann abgeschätzt wird. Vor der Anwendung des Modells müssen die relativen Auslastungen also in absolute konvertiert werden.

Für die Stagesysteme gestaltet sich dieses unkompliziert, da als Kriterium für Auslastung die absolute Speicherbelegung gegenüber der maximalen Speicherkapazität genutzt wird und dieses zwischen unterschiedlichen Stagetypen direkt übertragbar ist. Das bedeutet 500 GB belegter Speicher auf Stagesystem A entsprechen 500 GB belegtem Speicher auf System B. Da nur die tatsächlich verfügbaren (Netto-)Speicherbereiche berücksichtigt werden, ist dieses Kriterium unabhängig angewandter Redundanzverfahren (z.B. RAID) übertragbar. Somit wird zur Ermittlung der absoluten Auslastung der von allen Stagesystemen zur Verfügung gestellte Speicher (z.B. in GB) aggregiert und mit der relativen Auslastung l_2 multipliziert.

Die Auslastungen zwischen unterschiedlichen Serversystemen sind, wie bereits in Abschnitt 3.4.4 diskutiert wurde, nicht direkt übertragbar, wenn die Auslastung relativ angegeben wird. Eine Auslastung von 0,5 auf Server C entspricht keiner 50-prozentigen Auslastung auf Server D, wenn die Performance beider Server sich unterscheidet. Der leistungsfähigere Server arbeitet die gleichen Arbeitslasten grundsätzlich schneller ab, so dass die durchschnittliche Auslastung für den gleichen betrachteten Zeitraum in diesem Fall niedriger ist. Die prozentualen Auslastungen müssen daher für eine Aggregation an einer Referenzmaschine skaliert werden. Zudem müssen für Übertragungen prozentualer Auslastungen zwischen zwei Servern die entsprechenden Relationen bzgl. der Leistungsfähigkeit bekannt sein.

Zur Repräsentation der Leistungsfähigkeit gibt es unterschiedliche Ansätze. Ein häufig genutzter und einfach realisierbarer Ansatz ist die Relation der Per-

formance über die Taktfrequenz der installierten CPU. Die Taktfrequenz aller Kerne jeder CPU wird dabei zu einer Summe aggregiert, die dann die maximale Kapazität des Servers darstellt. Dieses Vorgehen wird unter anderem von Hankendi et al. in [HRC13] und Tighe et al. in [TKBL13] genutzt. Der große Vorteil dieses Ansatzes ist, dass die Taktfrequenzen genau spezifiziert und damit immer bekannt sind, so dass er sich jederzeit anwenden lässt. Das Verfahren besitzt allerdings auch einige Nachteile. So wird nur die CPU als einzige Systemressource berücksichtigt, welche zumindest den größten Einfluss auf die Performance bei Servern besitzt. Des Weiteren ist im Vergleich diverser CPU-Architekturen die Taktfrequenz je nach Arbeitslast nicht proportional zur Leistungsfähigkeit².

Ein anderer Ansatz ist die direkte Nutzung von Performancebenchmarks. Diese untersuchen die Leistungsfähigkeit eines Servers, indem sie auf diesen eine oder mehrere konkrete Arbeitslasten ausführen. Je nach Workload werden dabei neben der CPU auch weitere Systemressourcen eingebunden und bewertet. Durch die konkrete Messung der Performance fließen im Gegensatz zur alleinigen Nutzung der Taktrate unterschiedliche Architekturen mit in die Bewertung ein. Der größte Nachteil dieses Ansatzes ist allerdings, dass die Resultate verschiedener Servertypen verfügbar sein müssen. Beispiele für mögliche Benchmarks sind der PassMark CPU Mark³, LINPACK⁴ oder SPEC CPU2006⁵. Im Kontext der LDEE wird letztlich zur Konvertierung der Auslastung der Benchmark SPECpower_ssj2008⁶ verwandt, da dieser zum einen den Server als Ganzes mit einer repräsentativen Serverworkload belastet und zum anderen eine umfangreiche Datenbank mit Resultaten besitzt.

Die Konvertierung der relativen Auslastung l_1 vom Server s_1 zu Server s_2 wird über die maximal ermittelten Operationen p_1^{max} des SPECpower_ssj2008-Benchmarks (ssj_ops @ 100%) durchgeführt. Durch Multiplikation der Auslastung mit der maximalen Performance können die entsprechenden ssj_ops $p_1(l_1)$ berechnet werden: $p_1(l_1) = l_1 \cdot p_1^{max}$. Die der Auslastung l_1 entsprechenden ssj_ops können dann auf s_2 mit Hilfe der maximalen Performance p_2^{max} in die Auslastung l_2 , die diese erzeugen würden, umgerechnet werden: $l_2 = \frac{p_1(l_1)}{p_2^{max}}$. Zur Aggregation einer Gesamtlast werden die jeweiligen Serverauslastungen auf die Basis einer zuvor definierten Referenzmaschine s_r konvertiert. Das Vorgehen entspricht dabei dem beschriebenen, wobei für $s_2 = s_r$ gilt. Die aggregierte Gesamtlast ist dann immer unter Berücksichtigung der Referenzmaschine zu betrachten und stellt im Prinzip die Anzahl notwendiger Referenzmaschinen zur Handhabung der Last dar.

Prinzipiell könnte analog zum Storage auch direkt die absolute Auslastung in ssj_ops genutzt werden, so dass keine Umrechnung notwendig wäre. Allerdings

²http://www.tomshardware.de/x86-Prozessoren-Pro-Takt-Performance-Effizienz_tesberichte-240831.html (letzter Zugriff: 26.11.2015)

³<https://www.cpubenchmark.net/> (letzter Zugriff: 26.11.2015)

⁴<http://www.top500.org/project/linpack/> (letzter Zugriff: 26.11.2015)

⁵<https://www.spec.org/cpu2006/> (letzter Zugriff: 26.11.2015)

⁶https://www.spec.org/power_ssj2008/ (letzter Zugriff: 26.11.2015)

steht diese absolute Auslastung immer in Bezug zu der konkreten Arbeitslast des SPEC_power-Benchmarks. Die LDEE und das IT-Lastmodell sollen jedoch für Auslastungen, die durch beliebige Arbeitslasten erzeugt wurden, anwendbar sein, weshalb die von konkreten Arbeitslasten abstrahierende relative Auslastung genutzt wird.

4.3.4 Bestimmung der aktiven Server

Für den Fall, dass in einem Rechenzentrum dynamisches Powermanagement betrieben wird, verteilt sich die Last nicht auf alle Serversysteme sondern nur auf eine Teilmenge. Das IT-Lastmodell muss daher zunächst die wahrscheinliche Anzahl und Auswahl der Server für die gegebenen Eingangsdaten bestimmen. Diese Auswahl hängt stark vom eingesetzten Last- und Powermanagementalgorithmus ab. Das Modell muss daher dieses Verhalten abstrahieren um gute Ergebnisse zu erzielen.

Die wichtigste Anforderung eines dynamischen Powermanagements ist es, jederzeit ausreichende Rechenkapazitäten zur Verfügung zu stellen. Die bereitgestellten Kapazitäten werden von der aggregierten Rechenzentrumsauslastung sowie den reservierten Pufferressourcen beeinflusst. Zudem gibt es zeitliche Anforderungen, die in erster Linie davon abhängen, ob das Powermanagement proaktiv oder reaktiv arbeitet, vgl. Abschnitt 2.4.3. Unter Berücksichtigung dieser Parameter ermittelt das Modell in einem ersten Schritt die erforderlichen Systemressourcen mit der nachfolgenden Formel 4.1, die für das proaktive Verfahren aus dem AC4DC-Projekt [SJS⁺14] angepasst ist.

$$l_{max_i} = \max(l_j), \forall j \in \mathbb{N} : i \leq j \leq i + t_{pl} \quad (4.1)$$

$$t_{pl} = \frac{f_{sch} \cdot |V| \cdot t_m}{res} \quad (4.2)$$

Für jeden Zeitpunkt i der aggregierten RZ-Auslastung l wird das Maximum l_{max_i} eines Zeitfensters bestimmt. Dieses Zeitfenster repräsentiert den Zeitraum, für den eine konkrete Server-VM-Verteilung gültig ist. Bei proaktiven Verfahren werden diese Verteilungen bereits vor Auftreten der lokal maximalen Last l_{max_i} eingenommen, so dass das Zeitfenster entsprechend den Blick in die Zukunft darstellt. Dieser Zeitraum t_{pl} hängt von Parametern des Powermanagements ab. Im Falle des Verfahrens aus AC4DC sind dies die Größe der Menge V der gemanagten VM, ein Schedulingfaktor f_{sch} , der den Anteil der im dynamischen Betrieb migrierten VM mit Werten zwischen 0 und 1 widerspiegelt und der von der Zielauslastung abhängt, die Migrationszeit t_m einer VM in Minuten und die Auflösung res der Zeitreihe der aggregierten RZ-Last l , also die Anzahl der Minuten für die ein Zeitpunkt i in l steht. Der Schedulingfaktor multipliziert mit der VM-Anzahl schätzt damit die absolute Anzahl der vom LPM verschobenen VM ab, die multipliziert mit der Migrationszeit für VM die Länge des Planungs-

4 Modellbildung

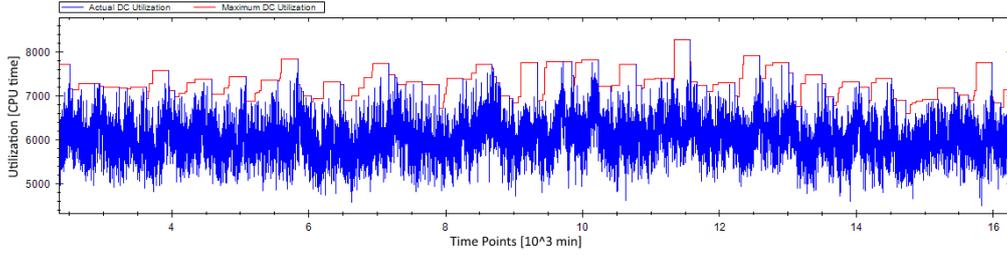


Abbildung 4.4: Maxima l_{max_i} (rot) des aggregierten Ressourcenbedarfs (blau) entsprechend der abgeschätzten Planungszeiträume des AC4DC-LPM

zeitraums t_{pl} des LPM ergibt. Durch die Division mit der Auflösung res wird dieser Zeitraum an die Auflösung der aggregierten Last angepasst.

Mit diesem Vorgehen kann für jede Systemressource, die das Powermanagement berücksichtigt, die erforderliche Kapazität abgeschätzt werden. Die Abbildung 4.4 zeigt die beispielhafte Anwendung auf der aggregierten CPU-Auslastung eines Rechenzentrums. Die rote Kurve gibt dabei die voraussichtlichen Kapazitäten an, mit denen der LPM-Algorithmus die Verteilung der VM plant.

Die nächste Aufgabe des IT-Lastmodells ist es nun, die konkreten Server zu ermitteln, die die zuvor abgeschätzten erforderlichen Kapazitäten – repräsentiert durch die Maxima l_{max_i} – bereitstellen müssen. Dazu wird für jeden Zeitpunkt i und jede Systemressource r (CPU, RAM, I/O) eine theoretische Verteilung der Last l auf die vorhandenen Server s bestimmt. Die primitiv-rekursive Funktion $f(l, n)$ in Gleichung 4.3 liefert dabei die Anzahl der benötigten Server. Die konkreten Server sind dann durch die Sortierung der geordneten Servermenge (S, \leq) gegeben, da die Sortierung die Priorität der Nutzung repräsentiert.

$$f(l, n) = \begin{cases} n & \text{falls } l \leq 0 \\ f(l - (c_{s_{n+1}}^r \cdot f_{ul}), n + 1) & \text{falls } l > 0 \end{cases} \quad (4.3)$$

Zu Beginn wird die Funktion mit $f(l_{max_i}^r, 0)$ aufgerufen. Dann wird für jeden Funktionsaufruf die Kapazität c_s^r der Ressource r des Servers s mit dem Index $n + 1$ der geordneten Menge (S, \leq) von der Last l subtrahiert, wobei zuvor die Zielauslastung über den zwischen 0 und 1 liegenden Faktor f_{ul} mit einfließt. Im Prinzip werden also der Reihe nach Server hinzugefügt, bis die Gesamtkapazität der betrachteten Ressource hinreichend groß ist, so dass die Zielauslastung nicht überschritten wird. Die erforderliche Kapazität ist erreicht, wenn die Last $l \leq 0$ ist. Die Anzahl der benötigten Server eines Zeitpunkts ist dann $\max(f(l_{max_i}^r, 0)) \forall r \in R$, wobei R die Menge der vom Powermanagement betrachteten Systemressourcen ist.

Da die Reihenfolge der Serverwahl immer gleich bleibt, muss für jeden Zeitpunkt nur die Anzahl der notwendigen und damit der angenommenen aktiven Server gespeichert werden. Die konkreten Server ergeben sich dann durch die

sortierte Serverliste. Die Sortierung der Server beeinflusst damit entscheidend die Abschätzung der Serverwahl. Hierbei gibt es, wie bereits in Abbildung 4.3 dargestellt, zwei Ansätze mit unterschiedlicher Genauigkeit: ein Verfahren basierend auf der Aktivitätsrate und eins basierend auf den Spezifikationen der Server.

Die Aktivitätsrate von Servern kann ermittelt werden, wenn historische Auslastungs- bzw. Aktivitätsdaten für alle gemanagten Server vorhanden sind. Für jeden Server wird dann der Anteil der Zeit berechnet, in der dieser aktiv war. Auf Basis dieser Daten wird dann eine Rangfolge der Server erstellt. Es wird angenommen, dass die Server mit der höchsten Aktivitätsrate auch zukünftig mit hoher Wahrscheinlichkeit genutzt werden.

Alternativ wird eine Rangfolge der Server anhand ihrer Spezifikationen erstellt, wenn keine historischen Nutzungsdaten vorliegen. Diese Methode versucht die Entscheidungen der Serverwahl eines Powermanagements anhand der bekannten Kriterien zu abstrahieren. Diese Kriterien umfassen einen Energieeffizienzfaktor sowie die Ressourcenkapazitäten für CPU und RAM, welche alle an den Servertypen gekoppelt sind. Analog zur Aktivitätsrate werden Rangfolgen für jedes Kriterium erstellt, welche anschließend in eine einzelne Rangfolge überführt werden. Dabei wird das gewichtete arithmetische Mittel der Indizes für jeden Server berechnet und diese untereinander aufsteigend sortiert.

Ein Vergleich der beiden Ansätze wird im Rahmen der Evaluation des IT-Lastmodells in Abschnitt 4.6.1 beschrieben.

4.3.5 Bestimmung der Auslastung

Die Performancemodelle als auch die Powermodelle der IT-Komponenten sind auslastungsabhängig und benötigen für gute Abschätzungen entsprechend detaillierte Eingabedaten. Das Wissen bzw. Abschätzungen über die Anzahl und die Auswahl der Komponenten alleine reicht also nicht aus. Das IT-Lastmodell schätzt daher neben den aktiven Komponenten auch die konkrete Auslastung dieser ab. Analog zur Serverwahl gibt es hier ebenso zwei Varianten, die sich nach den verfügbaren Daten richten. Das Modell ist allgemein gehalten, so dass es auf Server- als auch Stagesysteme anwendbar ist.

Die erste Methode nutzt die historischen Auslastungsverläufe der IT-Komponenten in Kombination mit der aggregierten RZ-Auslastung. Die Abhängigkeit zwischen diesen Datensätzen wird abstrahiert, um Auslastungen der IT-Komponenten für beliebige gegebene RZ-Lasten abschätzen zu können. Dazu wird die aggregierte RZ-Last zunächst in Auslastungsklassen eingeteilt. Man erhält damit eine Häufigkeitsverteilung der Einzelwerte über die Klassen wie in Abbildung 4.5 links dargestellt. In diesem Beispiel wurde die aggregierte RZ-Last in 20 Klassen von jeweils 5% Auslastung eingeteilt. Die Häufigkeit ist dabei die Anzahl der Zeitpunkte, an denen die Auslastung einer Klasse entspricht.

Als nächstes wird für jede Auslastungsklasse cl_y mit den korrespondierenden Zeitpunkten i aus den vorliegenden Auslastungsdaten l_{x_i} eines IT-Systems s_x

4 Modellbildung

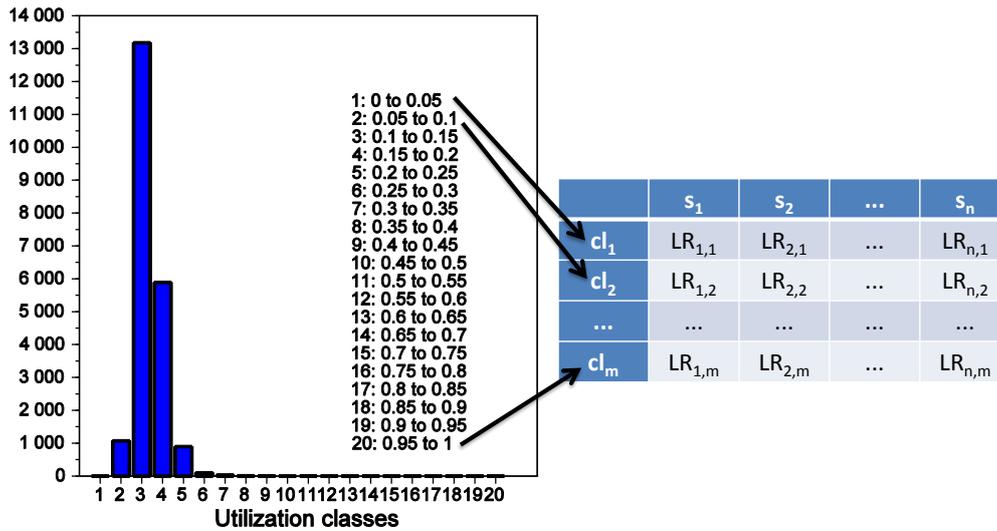


Abbildung 4.5: Abbildung der Auslastung eines IT-Systems s über eine lineare Regression LR für jede Auslastungsklasse cl .

– also den Auslastungen an allen Zeitpunkten, die der Klasse cl_y zugeordnet sind – und der aggregierten RZ-Last l_i eine lineare Regression entsprechend der Gleichung 4.4 berechnet. Diese stellt die Abhängigkeit der aggregierten RZ-Last in der definierten Klasse zu der Auslastung eines einzelnen IT-Systems dar. Prinzipiell kann diese Abhängigkeit auch für die gesamte Zeitreihe ermittelt werden, die Einteilung in Klassen erreicht jedoch eine höhere Genauigkeit, wie die Evaluation in Abschnitt 4.6.1 zeigt.

$$l_{x_i} = a + b \cdot l_i + e_i \quad (4.4)$$

Die Regressionsparameter $\begin{pmatrix} a \\ b \end{pmatrix} = LR_{x,y}$ werden in einer Lookup-Table für jedes IT-System s_x und jede Auslastungsklasse cl_y gespeichert, wie in Abbildung 4.5 rechts dargestellt. Die lineare Regression wird jeweils nach der Methode der kleinsten Quadrate ermittelt, bei der die summierten Quadrate der Residuen $\sum e_i^2$, also die Abweichungen der Regressionsgeraden von den Messwerten, minimiert wird. In Abbildung 4.6 wird die abgeschätzte CPU-Auslastung eines Servers beispielhaft dargestellt. Gibt es für eine Auslastungsklasse keine Referenzdaten, liefert die Abschätzung keinen Wert zurück, da die Abhängigkeit nicht modelliert werden konnte. In diesem Fall wird für die betroffenen Klassen das Verfahren basierend auf IT-Kapazitäten (vgl. Abbildung 4.3) genutzt, welches nachfolgend beschrieben wird.

Sind keine detaillierten Auslastungsdaten für jede IT-Komponente sondern nur die aggregierte RZ-Auslastung verfügbar, können die Abhängigkeiten zwischen IT- und RZ-Auslastung nicht abgebildet werden. Für die Modellierung der Auslastung wird daher eine Verteilung der aggregierten RZ-Last entsprechend der

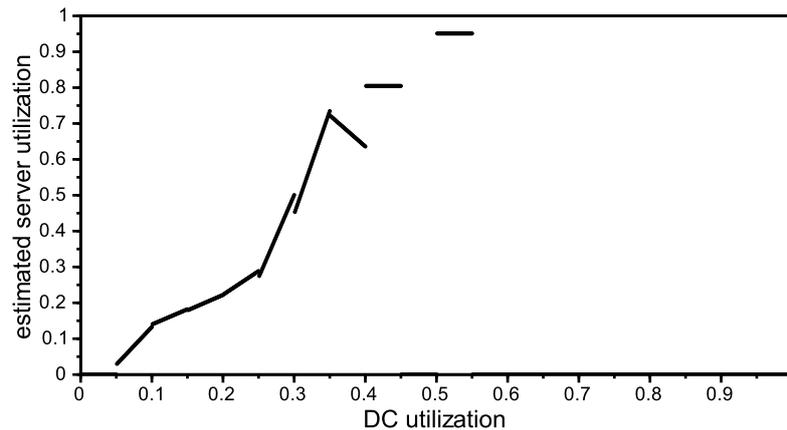


Abbildung 4.6: Beispiel für Abschätzung der CPU-Auslastung eines Servers auf Basis linearer Regression pro Auslastungsklasse.

verfügbaren Ressourcenkapazitäten der aktiven IT-Komponenten angenommen. Dargestellt ist dies in Gleichung 4.5. Die aggregierte RZ-Last l_i wird zunächst in Relation zur aggregierten maximalen Ressourcenkapazität der aktiven Komponenten gesetzt. Dadurch erhält man einen Auslastungsfaktor, der die Auslastung jedes aktiven IT-Systems x bestimmt. Dieser Wert wird daraufhin zur Abschätzung der konkreten relativen Auslastung l_{x_i} mit der maximalen Kapazität $l_{x_{max}}$ jedes IT-Systems multipliziert.

$$l_{x_i} = \frac{l_i}{\sum_x l_{x_{max}}} \cdot l_{x_{max}} \quad (4.5)$$

Durch Variationen in der aggregierten RZ-Last sowie in der Auswahl der aktiven IT-Systeme ändert sich der Auslastungsfaktor und die maximalen Ressourcen, so dass die Gleichung 4.5 für jeden Zeitpunkt i angewandt werden muss.

Die beiden Ansätze zur Abschätzung der IT-Auslastung – lineare Regression für Auslastungsklassen und Verteilung entsprechend aktiver Kapazitäten – werden in Abschnitt 4.6.1 miteinander und mit weiteren Ansätzen verglichen.

4.4 Performancemodelle

Neben der Auslastung der IT-Systeme ist die Kenntnis der jeweiligen Performance erforderlich. Erst eine Kombination der Auslastung mit der Performance kann die Arbeit, die ein IT-System theoretisch vollbringt, abbilden. Für alle IT-Systeme eines Rechenzentrums müssen entsprechend Performancemodelle verfügbar sein, um die theoretische Produktivität des gesamten RZ abschätzen zu können.

4.4.1 Anforderungen

Performancemodelle, die im Rahmen des kombinierten RZ-Performancemodells und der LDEE-Metrik Anwendung finden, müssen die nachfolgend beschriebenen Anforderungen erfüllen.

- R 4.3 **Vergleichbarkeit:** Die abgeschätzte Performance gleicher Hardwaretypen muss untereinander vergleichbar sein. Dies gilt in gleichem Maße für die Kombination der Performance mehrerer Hardwaresysteme. Um einen korrekten Vergleich zu ermöglichen, sollte die Abschätzung der Performance auf einem Mix aus unterschiedlichen, realistischen Arbeitslasten basieren, die möglichst einen Querschnitt typischer Anwendungen im RZ repräsentieren.
- R 4.4 **Vollständigkeit:** Da sich IT-Systeme eines Typs mitunter in der Auslegung der Subsysteme unterscheiden (z.B. CPU und RAM), sollten möglichst alle relevanten Subsysteme eines IT-Systems durch die Performancemodelle berücksichtigt werden. Anderenfalls ist keine vollständige Vergleichbarkeit gegeben.
- R 4.5 **Anwendbarkeit:** Die Performancemodelle müssen auf IT-Hardware im Produktivbetrieb anwendbar sein. Da in der Regel keine Performancemessungen an im Betrieb befindlicher Hardware durchgeführt werden können, bedeutet dies, dass die Modelle bereits vorab in einer charakterisierten Form für die IT-Systeme vorliegen müssen.
- R 4.6 **Auslastungsabhängigkeit:** Das Modell muss die Performance in Abhängigkeit einer gegebenen Auslastung des IT-Systems liefern können. Wird ein IT-System nur anteilig genutzt, trägt es auch nur in Höhe dieses Anteils zur Produktivität des RZ bei. Die Auslastungsabhängigkeit der Performance sowie der Power ist essentiell für eine korrekte lastabhängige Bewertung der Energieeffizienz.
- R 4.7 **Verständlichkeit:** Die durch die Modelle abgeschätzte Performance muss über eine nachvollziehbare Kennzahl abgebildet werden. Die Gesamtpformance eines Rechenzentrums fließt direkt in die LDEE, welche für Anwender leicht begreifbar sein muss, ein.

4.4.2 Bestimmung der Performance

Die Wahl des Verfahrens zur Abschätzung der IT-Performance richtet sich nach den beschriebenen Anforderungen. Ein Überblick der möglichen Ansätze mit Bezug zu den Anforderungen findet sich in Tabelle 4.1.

Ein typischer Ansatz betrachtet die Spezifikationen der IT-Systeme und vergleicht diese direkt miteinander, z.B. CPU und Arbeitsspeicher über Anzahl der Kerne, CPU-Takt und Speichergröße. Durch diverse Hardwarearchitekturen und

Methode	R 4.3	R 4.4	R 4.5	R 4.6	R 4.7
Systemspezifikationen	nein	ja	ja	nein	ja
Produktivanwendung	nein	nein	ja	ja	ja
Benchmarks direkt	ja	ja	nein	teils	teils
Benchmarks indirekt	ja	teils	ja	teils	teils

Tabelle 4.1: Vergleich von Methoden zur Bestimmung der Performance bzgl. der Anforderungen aus Abschnitt 4.4.1

-technologien, wie z.B. Hyperthreading oder Turbo Boost von Intel-Prozessoren, ist ein vollständiger Vergleich zwischen unterschiedlichen Modellen allerdings nicht möglich. Die genauen Auswirkungen der Technologien lassen sich in der Regel nicht den HW-Spezifikationen entnehmen, sondern müssen z.B. mittels Benchmarks vermessen werden. Da die Vergleichbarkeit der Ergebnisse – vor allem in Bezug zur LDEE – die wichtigste Anforderung (R 4.3) ist, eignet sich dieses Verfahren nicht zur Performancemodellierung.

Ähnlich ist es bei der Vermessung von Produktivanwendungen, die auf den zu bewertenden IT-Systemen laufen. Je nach Anwendung wird die Performance über unterschiedliche Kennzahlen ermittelt: Zum Beispiel kann man einen Server aufgrund seiner Bearbeitungsgeschwindigkeit von Datenbankbefehlen, der Antwortzeit beim Zugriff auf einen Webservice oder der Geschwindigkeit bei der Berechnung von Gleitkommaoperationen beurteilen. Gemeinsam haben diese Kennzahlen nur, dass sie jeweils den zeitlichen Faktor berücksichtigen. Durch die anwendungsspezifische Definition der Performance ist ein Vergleich also nicht möglich und dieses Verfahren für die Performancemodellierung nicht nutzbar.

Um eine faire Vergleichbarkeit zu erzielen, bieten sich standardisierte Performancebenchmarks, die direkt auf den zu bewertenden IT-Systemen angewandt werden, an. Nur durch den Einsatz von identischen Arbeitslasten auf den gleichen Hardwaretypen kann eine konsistente Bewertung der Performance von IT-Systemen erfolgen. Damit die Anforderung der Vollständigkeit (R 4.4) ebenso erfüllt wird, reicht es jedoch nicht die Bewertung von IT-Systemen nur anhand eines Benchmarks durchzuführen, da Benchmarks häufig nur einzelne Anwendungsfälle darstellen. Die Performancemodelle sollten demnach auf möglichst vielen unterschiedlichen Benchmarks und damit Anwendungen basieren, die auch die verschiedenen Ressourcen eines Systems belasten.

Eine Einschränkung bei der direkten Nutzung von Benchmarks zur Abschätzung der IT-Performance stellt allerdings die Anforderung R 4.5 – die Anwendbarkeit – dar. Um die konkrete Performance eines IT-Systems zu erhalten, müssten die entsprechenden Benchmarks auf diesem System ausgeführt werden. Im Produktivbetrieb eines Rechenzentrums ist dies praktisch nicht umzusetzen. Theoretisch ließen sich zwar einzelne Systeme vorübergehend isolieren und bewerten, allerdings sind der zeitliche Aufwand und damit die Kosten zu hoch, so

dass dieses Vorgehen in der Praxis nicht umgesetzt wird. Die direkte Anwendung von standardisierten Performancebenchmarks ist damit ausgeschlossen.

Die Resultate von Performancebenchmarks lassen sich aber indirekt zur Modellierung nutzen, wenn die ermittelte Performance von IT-Systemen anderweitig bereitgestellt wird, ohne dass Messungen auf den im Betrieb befindlichen Systemen notwendig wären. Ein verbreitetes Beispiel hierfür ist die Erfassung der Ergebnisse in öffentlich zugreifbaren Datenbanken. In diesem Fall wird auch die Anforderung R 4.5 erfüllt, da der Produktiveinsatz der IT-Systeme nicht unterbrochen wird. Da Performancewerte für viele unterschiedliche Systeme extern zur Verfügung stehen müssen, schränkt dieser Ansatz allerdings die Wahl der Benchmarks stark ein.

Die Auslastungsabhängigkeit (R 4.6) von Benchmarks ist für gewöhnlich nicht gegeben. Es gibt nur ein paar Ausnahmen, in denen bereits unterschiedliche Lastlevel berücksichtigt werden, wobei die unterschiedlichen Lastlevel in der Regel prozentual reduzierte Arbeitslasten darstellen. Diese direkte Skalierung der Arbeitslast kann auch für einige Benchmarks angewandt werden, die nur die maximale Performance liefern. Unter Annahme einer vollen Systemauslastung bei der ermittelten maximalen Performance und eines proportionalen Zusammenhangs zwischen Arbeitslast und Auslastung, kann eine direkte Skalierung der Performance mit der Auslastung abgeschätzt werden.

Die letzte Anforderung (R 4.7), die Verständlichkeit der Performancemodelle, ist durch die meisten Benchmarks zunächst nicht gegeben, da die Performance häufig nur durch anwendungsspezifische Operationen oder komplett abstrakte Zahlen dargestellt wird. Werden dann noch für die Abdeckung aller Subsysteme mehrere Benchmarks mit unterschiedlichen Kennzahlen kombiniert, sinkt die Nachvollziehbarkeit weiter. Um diese Anforderung bei Nutzung von Benchmarks zu erfüllen, müssen die Einzelergebnisse zu einer sinnvollen Kennzahl vereinigt werden.

Die Analyse der möglichen Ansätze zur Bestimmung der Performance bezüglich der Anforderungen zeigt, dass für die Performancemodellierung die Nutzung von Performancebenchmarks am sinnvollsten erscheint, da nur diese einen korrekten Vergleich von IT-Systemen eines Typs ermöglichen (R 4.3). Die notwendige Anwendbarkeit in Produktivumgebungen (R 4.5) verhindert allerdings die direkte Anwendung von Benchmarks auf der Hardware, so dass zuvor ermittelte Benchmarkresultate in die Performancemodelle einfließen müssen.

4.4.3 Aufbau der Performancemodelle

Das Diagramm in Abbildung 4.7 liefert einen Überblick über den prinzipiellen Ablauf der Performancemodellierung von Server- und Stagesystemen. Der erste Schritt ist dabei die Auswahl der Benchmarks, welche von den in Abschnitt 4.4.1 beschriebenen Anforderungen abhängt und die jeweils einmalig für das Server- sowie das Storageperformancemodell durchgeführt wird. Vor allem die Anforderungen R 4.4 (Vollständigkeit) und R 4.5 (Anwendbarkeit) haben einen großen

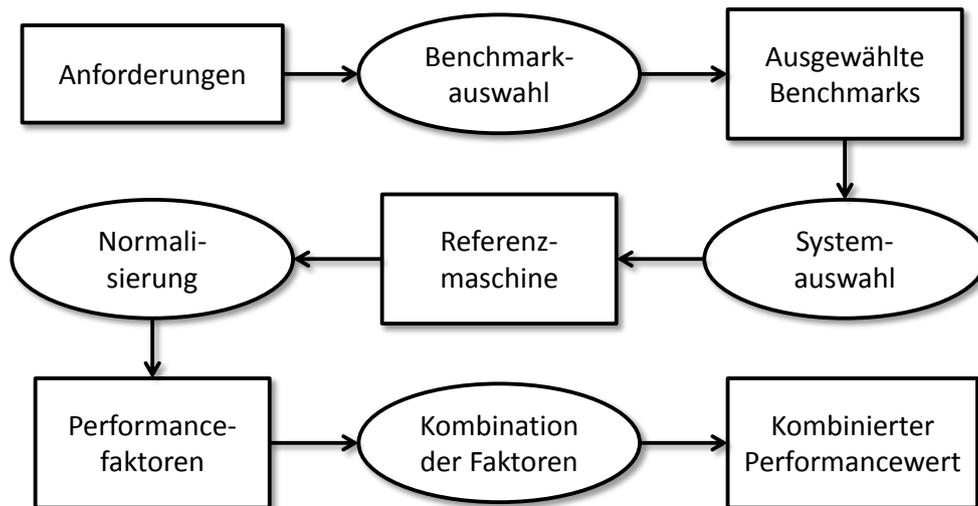


Abbildung 4.7: Prinzipieller Ablauf der Performancemodellierung von Server- und Storage-Systemen

Einfluss auf die Wahl – entsprechendes Gewicht haben demnach die Art der Arbeitslasten und die Verfügbarkeit von Benchmarkresultaten. In einem zweiten Schritt wird über alle IT-Systeme, die für alle ausgewählten Benchmarks mindestens ein Performanceresultat besitzen, eine Referenzmaschine definiert. Diese besitzt für jeden der Benchmarks einen dedizierten Performancewert, der sich aus dem Mittel aller ausgewählten Systeme berechnet. Über die Referenzmaschine können dann alle Benchmarkresultate normalisiert werden, so dass für jedes IT-System für jeden Benchmark ein Performancefaktor, der die Geschwindigkeit in Relation zur Referenzmaschine wiedergibt, resultiert. Für jedes IT-System werden letztlich die korrespondierenden Performancefaktoren aller Benchmarks zu einem Einzelwert kombiniert.

Im Folgenden werden die einzelnen Schritte im Detail beschrieben, wobei auf die Systemauswahl und Normalisierung aufgrund ihres direkten Zusammenhangs in einem gemeinsamen Abschnitt eingegangen wird.

Benchmarkauswahl

Die konkrete Auswahl der Benchmarks ist jeweils spezifisch für Server- sowie Storage-Systeme und wird daher in den nachgelagerten Abschnitten 4.4.4 und 4.4.5 diskutiert. Für die Beschreibung des allgemeinen Aufbaus des Performancemodells wird eine bereits vorgenommene Auswahl von geeigneten Benchmarks (im Weiteren als Menge A bezeichnet) angenommen.

Systemauswahl und Normalisierung

Idealerweise liefert jeder der ausgewählten Benchmarks einen Performancewert (also ein Benchmarkresultat) für jedes zu untersuchende IT-System. Da aus den separaten Performancewerten die Gesamtpformance des IT-Systems bestimmt wird, müssen die Einzelergebnisse geeignet kombiniert werden. Unterschiedliche Benchmarks bewerten die Performance jedoch in der Regel anhand unterschiedlicher Kennzahlen. Beispielsweise misst der LINPACK-Benchmark die Performance in GigaFLOPS (Floating Point Operations Per Second) während SPEC CPU2006 den SPEC_Speed als Faktor der durchschnittlichen Laufzeit gegenüber einer Referenzmaschine angibt.

Um eine Vergleichbarkeit (R 4.3) zwischen Arbeitslasten unterschiedlicher Benchmarks zu erreichen, muss vor der Kombination der Einzelresultate eine Normalisierung durchgeführt werden. Ansonsten hätten manche Anwendungen mehr Einfluss auf die Performancebewertung als andere und entsprechend spezialisierte IT-Systeme würden bevorzugt werden. Generell soll dieser Einfluss jedoch gleich verteilt sein oder – im Falle der Bewertung von bestimmten Rechenzentrumsklassen – bewusst gesteuert werden.

Die typische Vorgehensweise einer solchen Normalisierung ist die Division aller Benchmarkresultate durch das Resultat einer zuvor bestimmten Referenzmaschine, für die jeder Benchmark ein Resultat besitzt [FW86]. Man erhält damit einen Faktor, der angibt, wie schnell eine gewählte Maschine in Relation zur Referenzmaschine ist. Ein Nachteil dieses Ansatzes ist, dass nicht sichergestellt ist, dass die Referenzmaschine über alle Benchmarks ähnlich in Relation zu den anderen Maschinen einzuordnen ist. Es ist beispielsweise möglich, dass die Referenzmaschine bei Benchmark *A* zu den langsamsten 20% gehört, während sie bei allen anderen Benchmarks im Durchschnitt (also bei 50%) liegt. Durch die Division mit Resultaten der Referenzmaschine fallen normierte Ergebnisse der untersuchten Maschinen zu Benchmark *A* deutlich höher aus als bei den anderen Benchmarks. Somit erhält Benchmark *A* ungewollt eine höhere Gewichtung, weshalb eine gerechte Vergleichbarkeit nicht gegeben ist.

Ein alternativer Ansatz ist die Transformation aller Benchmarkresultate auf eine gemeinsame Skala, die z-Skala. Diese Methode nennt sich Standardisierung bzw. bei unvollständigen Messreihen Studentisierung [SSJ13]. Die transformierten Resultate jedes Benchmarks besitzen dann den arithmetischen Mittelwert 0 und geben jeweils die Abweichung vom Mittelwert in Vielfachen der Standardabweichung an. Ein Beispiel einer solchen Transformation wird in Abbildung 4.8 dargestellt. Auf der linken Seite werden für die Benchmarks SPECpower (oben) und SPEC CPU2006 (unten) die Verteilungen der Benchmarkresultate aller untersuchten Server als Histogramme dargestellt. Die Höhe der Balken gibt dabei jeweils an, wie viele Resultate in eine Klasse fallen. SPEC CPU2006 unterteilt sich zudem in die beiden Subbenchmarks CINT (rot) und CFP (blau). An der Klasseneinteilung (Abszisse) ist gut erkennbar, dass die Resultate des SPECpower-Benchmarks einige Größenordnungen über denen der

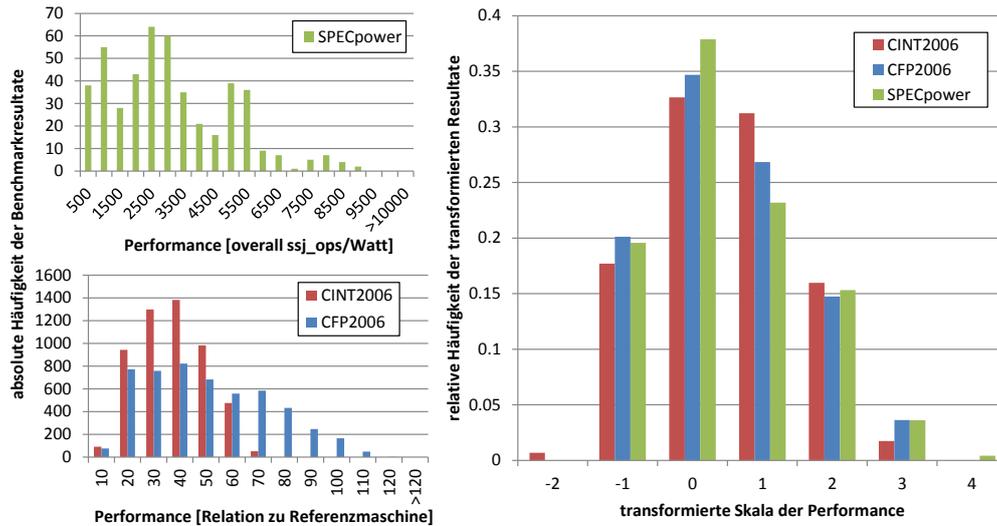


Abbildung 4.8: Beispiel der Skalentransformation der Performancebenchmarks SPECpower (oben links) und SPEC CPU2006 (unten links) auf die z-Skala (rechts)

CPU2006-Benchmarks liegen, was an den unterschiedlichen verwandten Metriken liegt. Während in SPECpower die Anzahl der abgearbeiteten Jobs ermittelt wird, werden für CPU2006 die Laufzeiten mit einer Referenzmaschine verglichen. Ein direkter Vergleich isolierter Resultate zwischen SPECpower und CPU2006 ist somit nicht möglich. Im rechten Histogramm der Abbildung 4.8 wurden die Resultate der Benchmarks z-transformiert, so dass ihre Wertebereiche nun alle auf der z-Skala liegen. Die transformierten Performancewerte können dann direkt benchmarkübergreifend miteinander verglichen werden.

Der Ansatz der Studentisierung ist für eine direkte Vergleichbarkeit der Benchmarks untereinander sehr gut geeignet. Allerdings lassen sich die transformierten Werte nicht zur Aggregation zu einer Gesamtpformance des Rechenzentrums verrechnen, da unterdurchschnittliche IT-Systeme negative Werte besitzen. Verschiebt man hingegen die Skala in den positiven Bereich, stimmen die Relationen unter den Einzelwerten nicht mehr und sehr performante Systeme werden gegenüber durchschnittlichen abgewertet.

Für die Performancemodellierung wird daher eine Mischung aus beiden Ansätzen gewählt. Die Benchmarkresultate werden mit Hilfe des Resultats einer Referenzmaschine PB_R normalisiert. Diese wird allerdings nicht ausgewählt, sondern analog zur Studentisierung aus der Menge aller IT-Systeme, die Resultate für jeden Benchmark PB^i der Auswahl A des Performancemodells besitzen, berechnet. Etwaige Stärken oder Schwächen der IT-Systeme für gewisse Benchmarks werden damit herausgemittelt. Man erhält letztlich für jedes IT-System s einen Performancefaktor PF_s^i , der die Performance des Systems gegenüber der Performance der Referenzmaschine bezüglich Benchmark PB^i angibt. Die Berechnung

4 Modellbildung

ist nachfolgend dargestellt.

$$R = \{s | \exists PB_s, \forall PB \in A\} \quad (4.6)$$

$$PB_R^i = \frac{\sum_{s \in R} PB_s^i}{|R|} \quad (4.7)$$

$$PF_s^i = \frac{PB_s^i}{PB_R^i} \quad (4.8)$$

R entspricht dabei der Menge aller IT-Systeme s , für die ein Benchmarkresultat PB_s für jeden der ausgewählten Benchmarks $PB \in A$ existiert (vgl. Gleichung 4.6). Über die IT-Systeme der Menge R kann für jeden Benchmark i die zugeordnete Referenzmaschine PB_R^i berechnet werden. Diese entspricht den durchschnittlichen Resultaten aller IT-Systeme $s \in R$ für Benchmark i (vgl. Gleichung 4.7). Anzumerken sei, dass die Referenzmaschine PB_R sich zwar immer aus den gleichen Systemen zusammensetzt, ihr Performancewert sich allerdings für jeden Benchmark i aufgrund der unterschiedlichen Benchmarkresultate unterscheidet. Letztlich kann für jedes IT-System s für jeden Benchmark i ein Performancefaktor PF_s^i errechnet werden, indem das Benchmarkresultat PB_s^i des Systems s in Relation zum durchschnittlichen Performancewert der Referenzmaschine PB_R^i für diesen Benchmark gesetzt wird (vgl. Gleichung 4.8). Entspricht der Performancefaktor $PF_s^i = 1$, ist das System s bzgl. Benchmark i genauso schnell wie die durchschnittliche Referenzmaschine. Bei einem $PF_s^i < 1$ ist sie hingegen langsamer und bei einem $PF_s^i > 1$ schneller.

Kombination der Faktoren

Nach der Angleichung der Benchmarks folgt die Kombination der Einzelergebnisse zu einem repräsentativen Einzelresultat $perf_s$ für ein IT-System s . Dies geschieht über eine gewichtete geometrische Mittelwertbildung über alle Performancefaktoren des IT-Systems mit den Gewichten w_i für die ausgewählten Performancebenchmarks $PB_i \in A$.

$$perf_s = \sqrt[w]{\prod_i (PF_s^i)^{w_i}} \quad \text{mit} \quad w = \sum_i w_i \quad (4.9)$$

Um eine möglichst allgemeine Bewertung des IT-Systems zu erreichen fließen die einzelnen Benchmarks generell zu gleichen Anteilen in das zusammengefasste Resultat mit ein, d.h. $\forall i w_i = 1$. Bei der Betrachtung von unterschiedlichen Rechenzentrumstypen kann es jedoch Sinn machen, eine Gewichtung entsprechend der Schwerpunkte der Benchmarks vorzunehmen, um den Fokus auf bestimmte Subsysteme wie beispielsweise die I/O-Verarbeitung zu legen. Dieser Ansatz wird in der beschriebenen Version der LDEE-Metrik allerdings nicht verfolgt, da

zusammen mit der unterschiedlichen Gewichtung zwischen Server- und Storage-Systemen zu viele Kombinationsmöglichkeiten entstünden. Letztlich leidet die Vergleichbarkeit der Ergebnisse durch eine hohe Vielfalt unterschiedlich einstellbarer Parameter. In Abschnitt 4.4.8 wird dieses Thema noch einmal aufgegriffen.

Zur Zusammenfassung der normalisierten Einzelergebnisse wird das geometrische dem arithmetischen Mittel vorgezogen, da nur das geometrische Mittel die Wertebereiche der normalisierten Benchmarkergebnisse angleicht. Das geometrische Mittel bleibt damit im Gegensatz zum arithmetischen Mittel konsistent, auch wenn andere Referenzen zur Normalisierung verwendet werden. Fleming und Wallace zeigen dies anhand von anschaulichen Beispielen und einem Beweis in [FW86].

4.4.4 Server

Die Auswahl der Benchmarks zur Modellierung der Serverperformance wird maßgeblich durch die Anforderung R 4.5, der Anwendbarkeit des Modells, beeinflusst. Wie bereits zuvor diskutiert wurde, lassen sich Benchmarks in der Regel nicht auf Serversystemen im Produktivbetrieb anwenden. Voraussetzung für die Benchmarks ist daher eine öffentlich zugreifbare Datenbank mit den Resultaten zu einer Vielzahl der Systeme.

Die Anforderung R 4.4 verlangt, dass möglichst viele Benchmarks mit unterschiedlichen Ausrichtungen in das Modell einfließen, damit die Server umfassend bewertet werden. Nun steht die Vollständigkeit der Anwendbarkeit allerdings gegenüber, so dass hier ein Zielkonflikt entsteht. Da nur wenige Benchmarks eine umfangreiche Ergebnisdatenbank besitzen, muss der Fokus nun entweder auf die Anwendbarkeit oder die Vollständigkeit gelegt werden.

Bei der Modellierung der Serverperformance wird ein Kompromiss zwischen diesen beiden Anforderungen durch die Auswahl folgender drei Benchmarks gebildet: SPEC CPU2006⁷, SPECpower_ssj2008⁸ und SAP SD Standard Application Benchmark⁹. Alle weiteren Serverperformancebenchmarks werden hauptsächlich aus zwei Gründen nicht für die Modellierung berücksichtigt. Zum einen besitzen die meisten Benchmarks gar keine Ergebnisdatenbank oder nur eine sehr kleine Menge an Resultaten, so dass eine Ausführung der Benchmarks auf den Servern nötig wäre, was jedoch aufgrund der Anforderungen nicht gewünscht ist. Und zum anderen analysieren einige Benchmarks eine komplette Systemumgebung bestehend aus mehreren Komponenten wie Server, Storage und dem verbindenden Netzwerk. Da somit keine Aussagen über die Einzelkomponenten sondern nur über das Gesamtsystem getroffen werden, sind die Ergebnisse für die modulare Modellierung der Serverperformance nicht nutzbar. Beispiele für solche Benchmarks sind SPECvirt_sc2010, VMmark 2.X oder TPC-C/E/H.

⁷<http://www.spec.org/cpu2006/> (letzter Zugriff: 26.11.2015)

⁸http://www.spec.org/power_ssj2008/ (letzter Zugriff: 26.11.2015)

⁹http://www.sap.com/campaigns/benchmark/appbm_sd.epx (letzter Zugriff: 26.11.2015)

4 Modellbildung

Nachfolgend werden die zur Performancemodellierung von Servern verwendeten Benchmarks näher beschrieben:

SPEC CPU2006 SPEC CPU 2006 ist ein CPU-intensiver Benchmark, der vornehmlich den Prozessor, den Arbeitsspeicher sowie den Compiler beansprucht. Die Workloads basieren auf realen Nutzeranwendungen in unterschiedlichen Anwendungsgebieten wie künstlicher Intelligenz, Optimierungs- und Suchproblemen, Kompression oder der Simulation der Dynamik von Flüssigkeiten/Molekülen. Die Workloads unterteilen sich in die zwei Kategorien 'Integer' (CINT) und 'Fließkomma' (CFP). Der Benchmark misst die benötigte Zeit zur Absolvierung jedes einzelnen Integer- und Fließkommaworkloads und setzt diese in Relation zu den Resultaten einer Referenzmaschine [Sta11]. Je nach Konfiguration der Benchmarkdurchläufe liefert SPEC CPU2006 das Resultat als SPECspeed (benötigte Zeit) oder als SPECrate (Datendurchsatz) jeweils zusammengefasst für die Integer- und Fließkommaworkloads. Diese Resultate werden anschließend in einen Basiswert (base) und einen Spitzenwert (peak) unterteilt, wobei der Basiswert strikteren Regeln unterliegt als der Spitzenwert, welcher aggressivere Optimierungen der getesteten Plattform erlaubt. Aus diesem Grund wird das Serverperformancemodell nur die ermittelten Basiswerte nutzen. Mit der SPECspeed und der SPECrate jeweils für CINT sowie CFP sind dies vier Performancewerte pro Server.

SPECpower_ssj2008 SPECpower_ssj2008 ist ein Performance/Powerbenchmark, der ein Serversystem auf elf unterschiedlichen Laststufen vom Leerlauf bis zur Volllast in 10%-Schritten belastet und dabei neben der Performance auch die entsprechende Leistungsaufnahme ermittelt. Der Workload besteht aus einer serverseitigen Java-Applikation (server side java - ssj) mit einer großen Anzahl simulierter Nutzer, die Anfragen in zufälligen Zeitintervallen an die Java-Anwendung absenden [Sta12]. Der Workload beansprucht auf der Hardwareseite in erster Linie die CPU, die Caches, den Arbeitsspeicher und die Skalierbarkeit des Shared Memories des Multiprozessorsystems. Zusätzlich hängt die Performance zum Teil auch von der Implementierung der Java Virtual Machine (JVM) sowie weiterer Eigenschaften der Java-Laufzeitumgebung und des Betriebssystems ab. Zu Beginn leitet SPECpower innerhalb einer Kalibrierungsphase die maximale Performance vom erzielten Datendurchsatz ab. Anschließend wird die Arbeitslast in 10%-Schritten bis zum Leerlaufbetrieb verringert und jeweils die Leistungsaufnahme gemessen. Letztlich liefert SPECpower_ssj2008 die Performance/Powerwerte für jede Laststufe sowie einen Durchschnittswert für schnelle Vergleiche. Zur Performancemodellierung wird der maximale, ermittelte Datendurchsatz in ssj_ops (server side java operations per second) verwendet.

SAP SD Standard Application Benchmark Der SAP Sales and Distribution (SD) Standard Application Benchmark deckt ein Lagerverkaufsszenario in

SAP ab, in dem Nutzeraufträge inklusive Verarbeitung und Auslieferung simuliert werden. Die Anzahl der simulierten Nutzer wird dabei stetig erhöht bis die Antwortzeit des untersuchten Systems eine Sekunde (zwei Sekunden vor 2009) erreicht. Der erzielte Datendurchsatz (bestellte Objekte pro Stunde) bei maximaler Arbeitslast wird dann in den SAPS-Wert (SAP Application Performance Standard) umgewandelt, welcher das Endresultat des Benchmarkdurchlaufs darstellt [MG08]. Der Workload belastet hauptsächlich die CPU und Speicherhierarchie eines Servers. Das Performancemodell für Server nutzt die Resultate zu den Zwei-Schichten-Architekturen (two-tier), bei denen die Datenbank und die SAP-Applikation auf einer einzelnen physikalischen Maschine installiert sind.

SERT Neben den drei verwendeten Benchmarks wird es zukünftig mit dem Server Efficiency Rating Tool (SERT), welches von SPEC entwickelt wird, noch ein weiteres Tool, das sich zur Performancemodellierung eignet, geben. SERT nutzt eine Menge synthetischer Workloads um die unterschiedlichen Systemressourcen eines Servers (CPU, Arbeitsspeicher, Netzwerk, Festspeicher) zu testen und detaillierte Energiebedarfe zu unterschiedlichen Laststufen zurückzuliefern [LT11]. Im Gegensatz zu den oben beschriebenen Benchmarks hat SERT das Ziel, die Serversysteme im Auslieferungszustand ohne besondere Konfigurationen oder Feintuning zu bewerten. Durch diese objektive Bewertung wäre SERT die ideale Lösung zur Nutzung in einer Energieeffizienzmetrik. Allerdings ist SERT noch immer in einer reinen Berichtsphase, in der die Ergebnisse zunächst gesammelt und analysiert werden ohne finale Resultate zu veröffentlichen [Sta13]. Auch das Format des letztlichen Resultats ist noch nicht entschieden. Erst wenn die Performance bzw. Effizienz der untersuchten Server veröffentlicht und hinreichend Daten vorhanden sind, wird SERT in das Serverperformancemodell integriert.

Wie in Abschnitt 4.4.3 beschrieben, werden die Resultate eines Serversystems für alle ausgewählten Benchmarks zunächst normalisiert. Zur Normalisierung wird eine einheitliche Menge von Systemen für alle Benchmarks ausgewählt, d.h. alle Server, die für jeden Benchmark mindestens ein Resultat besitzen. Gibt es für einen Server mehrere Resultate, wird zur Zusammenfassung das arithmetische Mittel angewandt. Aus allen zutreffenden Serversystemen wird die durchschnittliche Referenzmaschine PB_R berechnet, zu der die Relation jedes einzelnen Benchmarkresultats in Form des Performancefaktors PF_s^i (vgl. Gleichung 4.8) bestimmt wird. Als nächstes werden die einzelnen Performancefaktoren eines Servers mit dem gewichteten geometrischen Mittel zu einem Einzelwert $perf_s$ zusammengefasst (vgl. Gleichung 4.9). Für die allgemeine Bewertung sollen die Benchmarks zu gleichen Anteilen einfließen. Daher ergeben sich für SPECpower und dem SAP SD Standard Application Benchmark eine Gewichtung von 1, für die vier Einzelwerte des SPEC CPU2006 Benchmarks hingegen jeweils ein Gewicht von 0,25. Somit bleibt das Verhältnis der Benchmarks untereinander identisch.

4.4.5 Storage

Für das Storageperformancemodell gelten die gleichen Randbedingungen wie für das in Abschnitt 4.4.4 beschriebene Modell für Server. Die Auswahl der geeigneten Benchmarks wird beeinflusst durch die Anzahl der Benchmarks und den Umfang der veröffentlichten Ergebnisse der Benchmarks. Es gibt somit ebenso den Zielkonflikt zwischen den beiden Anforderungen R 4.4 und R 4.5, der Vollständigkeit und der Anwendbarkeit. Analog zu den Servern wurde bei der Auswahl der Benchmarks ein Kompromiss getroffen. In das Modell integriert wurden SPECsfs¹⁰ und SPC Benchmark 1¹¹. Alle anderen Benchmarks besitzen gar keine oder nur wenige öffentlich verfügbare Ergebnisse, so dass diese nicht zur Modellierung berücksichtigt werden. Die beiden zur Modellierung genutzten Benchmarks werden nachfolgend vorgestellt:

SPECsfs2008 SPECsfs2008 ist ein Benchmark für Stagesysteme, bei dem der Datendurchsatz und die Antwortzeit des Dateiservers gemessen werden. SPECsfs2008 beinhaltet jeweils einen Workload für die beiden verbreiteten Netzwerkspeicherprotokolle NFS (Network File System) und CIFS (Common Internet File System) bzw. SMB (Server Message Block), vgl. Abschnitt 2.3.2. Beide Workloads basieren dabei überwiegend auf Daten, die von zehntausenden Dateiservern der in SPEC kooperierenden Unternehmen stammen [Sta08]. Mehrere Clients belasten das Testsystem durch Anfragen, wobei zehn oder mehr gleichverteilte Laststufen erzeugt werden. Eine Laststufe entspricht damit einer bestimmten Anzahl an Operationen. Nach jedem Durchlauf wird der Datendurchsatz (in Operationen pro Sekunde) und die mittlere Antwortzeit pro Operation zurückgeliefert. Die Resultate werden zwischen den beiden Protokollen NFS und CIFS unterschieden und sind nicht direkt miteinander vergleichbar. Für die Performancemodellierung des Storage werden ausschließlich die Ergebnisse bzgl. des NFS berücksichtigt, da für das CIFS nur sehr wenige Resultate verfügbar sind.

SPC Benchmark 1 SPC-1 belastet Stagesysteme mit einem Workload, welcher typische Funktionen unternehmenskritischer Anwendungen abbildet. Dazu führt der Benchmark zufällige Input/Output-Operationen (I/O) aus und verwendet 'query'- sowie 'update'-Operationen exemplarisch für Online-Transaction-Processing-, Email- oder Datenbankapplikationen [Sto13]. SPC-1 kann ein Stagesystem anhand zwei unterschiedlicher Performancekennzahlen bewerten. Für hoch-parallele Systeme wird der maximale Durchsatz verarbeiteter I/O-Anfragen in SPC-1 IOPS (I/O operations per second) bestimmt und eine Gegenüberstellung der Antwortzeit mit dem Datendurchsatz für unterschiedliche Laststufen

¹⁰<http://www.spec.org/sfs2008/> (letzter Zugriff: 26.11.2015)

¹¹http://www.storageperformance.org/results/#spc1_overview
(letzter Zugriff: 26.11.2015)

geliefert. Bei Storage-Systemen für unternehmenskritische Anwendungen bewertet SPC-1 hingegen die niedrigste Antwortzeit der I/O-Anfragen in SPC-1 LRT (least response time).

Die beiden gewählten Benchmarks besitzen unterschiedliche Indikatoren zur Bewertung der Performance von Storage-Systemen. Zur Modellierung werden jeweils die Hauptmetriken mit der Darstellung der maximalen Durchsatzrate in SPECsfs2008-Operationen bzw. in SPC-1-I/O-Anfragen pro Sekunde sowie die Repräsentationen der durchschnittlichen Antwortzeit in Millisekunden (SPECsfs2008 overall response time bzw. SPC-1 LRT) genutzt. Dabei wird für die Antwortzeiten der Kehrwert des normalisierten Performancefaktors – also $1/PF_s^i$ – verwandt, da eine höhere Antwortzeit gleichbedeutend mit einer schlechteren Performance ist und der Performancefaktor bei steigender Antwortzeit entsprechend kleiner werden muss. Der prinzipielle Aufbau des Modells ist analog zum Serverperformancemodell. Nach der Normalisierung der Einzelwerte auf Basis der Referenzmaschine PB_R^i , die dem Durchschnitt aller Storage-Systeme mit Ergebnissen für beide Benchmarks entspricht, und der Bildung der Kehrwerte für die Antwortzeiten werden die vier Kennzahlen gleich gewichtet über das geometrische Mittel zu einem einzelnen Performancewert $perf_{sto}$ (vgl. Gleichung 4.9) zusammengefasst.

4.4.6 Netzwerk

Die Performanceanalyse von Netzwerken unterscheidet sich deutlich von den oben beschriebenen Server- und Storagebenchmarks. Es ist zwar möglich die einzelnen Netzwerkkomponenten wie Router und Switches zu bewerten, entscheidend für die Gesamtleistung eines Rechenzentrums ist jedoch ebenso die Netzwerkarchitektur, welche alle Systeme miteinander verbindet, vgl. Abschnitt 2.3.3. Da sich diese in jedem Rechenzentrum individuell zusammensetzt (u.a. aufgrund der Anzahl der Knoten und der konkret verbauten Komponenten) und somit nicht zwischen Rechenzentren vergleichbar ist, ist es sinnlos eine Ergebnisdatenbank für Netzwerkarchitekturen anzubieten. Das bedeutet, dass im Gegensatz zu den Server- und Storage-Systemen in jedem Rechenzentrum, welches vollständig bewertet werden soll, umfangreiche Performanceuntersuchungen des Netzwerks durchgeführt werden müssten.

Die meisten Netzwerkbenchmarks untersuchen die Geschwindigkeit zwischen zwei konkreten Knotenpunkten. Zur Beurteilung der gesamten Rechenzentrumsnetzwerke müssten allerdings deutlich mehr Knoten betrachtet werden, idealerweise alle (vgl. Abschnitt 2.3.3). Eine manuelle Installation und Ausführung von dedizierten Benchmarks ist aber nicht realisierbar – zumal diese Analysen einen spürbaren Einfluss auf den Produktivbetrieb hätten. Voraussetzung zur Bewertung ist also eine automatische Ermittlung der Netzwerkparameter ohne den Produktivbetrieb zu stören. Hierzu eignen sich die in Abschnitt 2.4.3 beschriebenen Managementwerkzeuge wie z.B. DCIM, welche bereits in vielen

Rechenzentren Anwendung finden.

Für alle Netzwerkkomponenten wird kontinuierlich der Datendurchsatz (Mbit/s), die Latenz (ms) und der relative Paketverlust (%) gemessen. Im Gegensatz zu den Server- und Storagesystemen wird aus diesen Parametern jedoch kein allgemeiner, vergleichbarer Performancewert berechnet. Die gemessenen Werte dienen dazu, die Auslegung der Netzwerkkapazitäten in Relation zu den verbauten IT-Systemen zu ermitteln. Der resultierende Netzwerkperformancekoeffizient NPC spiegelt den Einfluss der RZ-Netzwerke auf die Gesamtperformance des Rechenzentrums wider und wird wie nachfolgend dargestellt berechnet.

$$NPC = \frac{\sum_{i=0}^n \min \left(\min \left(1, \left\lfloor \frac{thr_{sla}}{thr_i} \right\rfloor \right), \min \left(1, \left\lfloor \frac{lat_{sla}}{lat_i} \right\rfloor \right), \min \left(1, \left\lfloor \frac{pl_{sla}}{pl_i} \right\rfloor \right) \right)}{n} \quad (4.10)$$

Für jeden Zeitpunkt i des Betrachtungszeitraums der Länge n wird für jeden der drei Parameter Datendurchsatz thr , Latenz lat und Paketverlust pl die Einhaltung der Service Level Agreements (SLA) des Rechenzentrumsnetzwerks untersucht. Übersteigt mindestens einer der Parameter die Vorgaben der SLA (thr_{sla} , lat_{sla} , pl_{sla}), wird der gesamte obere Term innerhalb der Aufsummierung für den entsprechenden Zeitpunkt i auf Null gesetzt, wodurch die Verletzung der SLA an diesem Zeitpunkt repräsentiert wird. Die Summenformel aggregiert dann alle Zeitpunkte, an denen die SLA eingehalten werden, und setzt diese in Relation zu der Anzahl aller Zeitpunkte n . Man erhält dadurch den zeitlichen Anteil, an dem die Performance des Netzwerks nicht negativ beeinflusst wird. Ein NPC von 1 entspricht demnach einer jederzeit hinreichenden Verfügbarkeit von Netzwerkkapazitäten.

Mit diesem Ansatz eines Koeffizienten fließt die Netzwerkperformance nicht direkt in die RZ-Performance ein, sondern nur indirekt als Einschränkung aufgrund des Netzwerks. Anstelle einer konkreten Performancekennzahl wird zunächst angenommen, dass das Netzwerk auf einen reibungslosen Betrieb der vorhandenen IT-Systeme ausgelegt ist. Der NPC gibt nun an, wie häufig diese Annahme zutrifft. Als Kriterium für nicht ausreichende Performance werden die SLA der RZ-Netzwerke verwendet. Werden diese verletzt, stellt dies laut Definition eine unzureichende Netzwerkperformance dar (vgl. Abschnitt 2.3.3) und der NPC ist entsprechend niedriger. Die Nutzung von SLA zur Abschätzung der Performance hat den Vorteil, dass diese für jedes Rechenzentrum speziell auf die Bedürfnisse der installierten Dienste und an die vorhandenen Netzwerktopologien angepasst sind. Die Netzwerkperformance müsste ansonsten für jedes Rechenzentrum separat modelliert werden.

Den umgekehrten Fall eines überdimensionierten Netzwerks deckt der NPC hingegen nicht ab. Eine potentiell höhere Netzwerkperformance als die tatsächlich genutzte fließt somit auch nicht in das RZ-Performancemodell ein. Da dieses die mit der Auslastung skalierte – und somit die genutzte – Performance beschreiben soll, ist dies jedoch kein Nachteil. Ein überdimensioniertes Netzwerk

wirkt sich allerdings negativ auf die Energieeffizienz aus, da mehr Kapazitäten bereitgestellt werden als tatsächlich nötig sind. Die in diesem Fall höhere Leistungsaufnahme der Netzwerkkomponenten wird jedoch vom RZ-Powermodell berücksichtigt (vgl. Abschnitt 4.5), so dass negative Auswirkungen auf die Energieeffizienz entsprechend in die LDEE einfließen.

Theoretisch ist die Einbindung der aggregierten RZ-Auslastung in die Berechnung des NPC denkbar, um die Auslastungsabhängigkeit auch auf der Netzwerkkomponenten umzusetzen. In praktischer Hinsicht stehen dem allerdings zwei Herausforderungen gegenüber. Zum einen müsste die aggregierte Auslastung von den RZ-Betreibern gemessen werden, um die Abhängigkeiten zwischen der Auslastung und SLA-Verletzungen abbilden zu können. Da diese allerdings in der Regel kein Bestandteil der Überprüfung der Netzwerk-SLA sind (vgl. Abschnitt 2.3.3), kann dies nicht vorausgesetzt werden. Zum anderen muss das Ausmaß der Performanceeinbußen quantifiziert werden, so dass der NPC den Grad der Einschränkung in Abhängigkeit der Auslastung wiedergeben kann. Um eine möglichst breite Anwendbarkeit zu erreichen, wurde der NPC daher von den Messwerten abstrahiert und als Koeffizient für den Anteil der Zeit eines reibungslosen Betriebs konzipiert.

Zur Bestimmung der LDEE unabhängig von konkreten Messdaten wird ein NPC von eins angenommen, da die Server- und Storageperformancemodelle die Netzwerkbelastung nicht berücksichtigen und die Einhaltung der Netzwerk-SLA somit nur für konkret gemessene Werte ermittelt werden kann.

4.4.7 Fehlende Performancewerte

Wenn für ein IT-System nur von einer Teilmenge der Benchmarks der Auswahl des jeweiligen Performancemodells Resultate vorliegen, werden auch nur die vorliegenden Ergebnisse genutzt. Das bedeutet, dass die Gewichtung der fehlenden Benchmarks damit 0 ist. Somit wird für das betroffene IT-System angenommen, dass die Relation in der Performance zu den anderen IT-Systemen bei den anderen Benchmarks gleich ist.

Schwieriger ist es, wenn es zu einem IT-System gar keine Performancewerte der ausgewählten Benchmarks gibt. Damit dennoch die Gesamtperformance des Rechenzentrums bestimmt werden kann, muss die Performance der fehlenden IT-Systeme abgeschätzt werden. Dies geschieht auf Basis der bekannten Hardwarespezifikationen der IT-Systeme, da diese in der Regel direkt vorliegen und somit ohne Eingriff in den Produktivbetrieb ermittelt werden können. Anhand der vorliegenden Spezifikationen können Performancewerte ähnlicher Systeme identifiziert werden.

Server: Für Serversysteme ist in erster Linie die Betrachtung der verbauten CPU-Typen wichtig, da von diesen viele Technologien wie Intels TurboBoost oder Hyperthreading, die maßgeblichen Einfluss auf die Performance besitzen,

abhängen. So können zwei CPU unterschiedlicher Typen trotz gleicher Kernanzahl und Taktrate deutlich voneinander abweichende Rechenleistungen erbringen, was beispielsweise mit der Diskrepanz zwischen Intel Xeon und Intel Core Prozessoren der Fall ist¹². Ausgehend von der CPU-Performance müssen außerdem noch die weiteren numerischen Spezifikationen wie die Anzahl der CPU und die Größe des Arbeitsspeichers berücksichtigt werden, um die Performanceskalierung abschätzen zu können.

Storage Performanceergebnisse von Storage-Systemen hängen hauptsächlich von der Art und Anzahl der Massenspeicherkomponenten ab. Solid State Disks bieten eine höhere Zugriffsgeschwindigkeit als Hard Disk Drives und eine größere Anzahl an Speicherkomponenten erhöht die parallele Bearbeitung. Hinzu kommt die Ausstattung der Rechenkomponenten (CPU, RAM) sowie die Netzwerkanbindung des Storage.

Netzwerk Für die Beurteilung der Netzwerkperformance werden Messreihen aus dem Rechenzentrum direkt genutzt. Die genauen Parameter richten sich jeweils nach den Netzwerk-SLA, die zur Überprüfung auch durchgehend gemessen werden müssen. Sollten in einem Rechenzentrum keine SLA definiert sein, wird angenommen, dass die Netzwerkkapazitäten so ausgelegt sind, dass kein negativer Einfluss auf die Gesamtperformance des Rechenzentrums vorliegt.

Es gibt unterschiedliche Ansätze die Performance eines Systems anhand der genannten Parameter abzuschätzen. Eine Möglichkeit besteht zum Beispiel in der Nutzung eines multiplen linearen Modells, welches die Systemspezifikationen als unabhängige Variablen verwendet und geeignet kombiniert. Problematisch sind hierbei nicht-numerische Faktoren wie einige der CPU-Technologien, die nur schwierig in ein lineares Modell integriert werden können. Als Alternative bietet sich daher eher eine Klassifizierung der Systeme an. Systeme ohne Performancedaten können dann über geeignete Verfahren (z.B. neuronale Netze) den Klassen und damit entsprechenden Performanewerten zugeordnet werden. Die Modellierung fehlender Daten wird in dieser Arbeit nicht näher betrachtet. Für die Evaluation der Metrik wird eine Klassenzuordnung der Systeme vorausgesetzt.

4.4.8 Kombiniertes RZ-Performancemodell

Die Gesamtperformance eines Rechenzentrums wird durch die Summe aller vorhandener IT-Komponenten und den Netzwerkverbindungen dazwischen bestimmt. Die vorgestellten Performancemodelle für Server- und Storage-Systeme liefern die allgemeine Performance einzelner Systeme im Vergleich zu den jeweiligen durchschnittlichen Referenzeinheiten. Der Netzwerkperformancekoeffizient hingegen

¹²http://www.cpubenchmark.net/high_end_cpus.html (letzter Zugriff: 26.11.2015)

gibt an, wie gut das Netzwerk für die übrigen IT-Komponenten skaliert ist bzw. wie hoch die Einschränkungen, die durch das Netzwerk verursacht werden, sind.

Aggregation der Performance

Die simple Aggregation aller Einzelresultate liefert einen allgemeinen Überblick über die potentielle Performance des Rechenzentrums. Zur Bestimmung der Energieeffizienz ist jedoch nicht die maximale Performance entscheidend, sondern die tatsächlich genutzte, was eine Annäherung der Produktivität des Rechenzentrums darstellt, vgl. Abschnitt 2.5.3. Noch vor der Aggregation zu einem Gesamtwert werden daher die einzelnen Performancewerte der IT-Systeme mit den durch das IT-Lastmodell abgeschätzten Auslastungswerten – wie in Abbildung 4.1 dargestellt – verrechnet. Der resultierende Wert entspricht dann der tatsächlich genutzten Performance und die Anforderung R 3.2 der Auslastungsabhängigkeit wird erfüllt. Durch dieses Vorgehen wird eine Unterauslastung der IT-Ressourcen bestraft und RZ-Betreiber letztlich dazu motiviert, die vorhandenen IT-Ressourcen – z.B. durch ein intelligentes Management – besser auszunutzen.

Die mit der Auslastung skalierten Performancewerte werden jeweils für alle vorhandenen Server- sowie Storage-Systeme aggregiert. Die beiden Resultate repräsentieren damit die genutzte Server- bzw. Storageperformance des gesamten Rechenzentrums im Verhältnis zu einem definierten durchschnittlichen Server- bzw. Storage-System. Beide Teilergebnisse werden anschließend gewichtet miteinander verrechnet. Standardmäßig sind die beiden Gewichte für Server- und Storageperformance mit $w_{srv} = w_{sto} = 1$ gleich und damit neutral. Bei Vergleichen innerhalb einer bestimmten RZ-Klasse werden die Gewichte hingegen adaptiert, um die Merkmale der Klasse hervorzuheben. Die konkreten Gewichtungen werden im nachfolgenden Abschnitt beschrieben.

Abschließend wird die genutzte Performance mit dem Netzwerkperformancekoeffizienten skaliert, welcher die Performanceeinschränkungen aufgrund der Netzwerkbeschaffenheiten darstellt. Der NPC ist somit – im Gegensatz zu den Performancewerten der Server- und Storage-Systeme – nicht von der Auslastung abhängig, sondern als statische Kennzahl realisiert, vgl. Abschnitt 4.4.6. Die gesamte genutzte RZ-Performance $perf_{DC}$ berechnet sich damit per Gleichung 4.11,

$$perf_{DC} = NPC \left(w_{srv} \sum_{i \in srv} load_i perf_i + w_{sto} \sum_{j \in sto} load_j perf_j \right) \quad (4.11)$$

mit dem Netzwerkperformancekoeffizienten NPC , den Gewichten w_{srv} und w_{sto} für Server bzw. Storage sowie der Auslastung $load$ und der Performance $perf$ jedes einzelnen Servers i oder Storage-Systems j .

Gewichtung

Ein Performance- bzw. Energieeffizienzvergleich sollte für RZ-Klassen durchgeführt werden, um den unterschiedlichen Ausrichtungen und Redundanzen Rechnung zu tragen, vgl. Abschnitt 3.4.5.

Die Ausrichtung (Geschäftsmodell) des Rechenzentrums hat dabei einen maßgeblichen Einfluss auf die Gewichtung der IT-Komponenten. Die Gewichte für die Server- w_{srv} und Stagesysteme w_{sto} in Gleichung 4.11 orientieren sich daher an dem RZ-Typ, welcher verglichen wird. Für die Netzwerkkomponenten gibt es hingegen kein Gewicht, da die Performance des Netzwerks nicht direkt betrachtet wird, sondern diese als Skalierungsfaktor NPC in die RZ-Performance einfließt. Die möglichen Gewichtungen zur Berechnung der RZ-Performance werden nachfolgend aufgelistet:

- Standard (S): $w_{srv}^S = 1$; $w_{sto}^S = 1$
- Balanced (B): $w_{srv}^B = \frac{|STO|}{|SRV|+|STO|}$; $w_{sto}^B = \frac{|SRV|}{|SRV|+|STO|}$
- Compute-Intensive (C): $w_{srv}^C = 0,8$; $w_{sto}^C = 0,2$
- Data-Intensive (D): $w_{srv}^D = 0,2$; $w_{sto}^D = 0,8$
- Free Choice (F): $w_{srv}^F = x$; $w_{sto}^F = 1 - x$; $\{x \in \mathbb{R} | 0 \leq x \leq 1\}$

Die Standardgewichtung (S) repräsentiert den Regelfall, der für allgemeine Betrachtungen der RZ-Performance ohne weitere Anforderungen genutzt wird. Die Gewichte haben beide den Wert 1 und verhalten sich neutral in der Multiplikation der Server- sowie Storageperformance. Die RZ-Performance ist damit die ungewichtete Aggregation beider Einzelwerte.

Möchte man die Server- und Storageperformance zu gleichen Teilen bewerten, können die ausbalancierten Gewichte (B) genutzt werden. Auch diese Gewichte können wie die Standardwerte für allgemeine Bewertungen oder für klassenübergreifende Vergleiche verwandt werden. Der Unterschied zum Standardfall liegt darin, dass das Verhältnis zwischen Server- und Stagesystemen des Rechenzentrums in die Bewertung mit einfließt. Die Gewichte berechnen sich dabei aus den auf ein Gesamtgewicht von 1 normierten Kehrwerten der Verhältnisse zwischen der Anzahl der Server $|SRV|$ und der Anzahl der Stagesysteme $|STO|$. Dieses Vorgehen beruht auf der Tatsache, dass sowohl die Server- als auch die Stagesysteme jeweils auf ihren eigenen Referenzeinheiten normiert werden und sich die jeweiligen Wertebereiche somit decken. Durch die typischerweise höhere Anzahl der Server im Vergleich zum Storage würde die Serverperformance bei einer gleichen Gewichtung bevorzugt werden. Der normierte Kehrwert gleicht dieses gekippte Verhältnis wieder aus.

Unterschiedliche Ausrichtungen von Rechenzentren können über die Gewichte der beiden Kategorien für rechenintensive (C/Compute-Intensive) und speicherintensive (D/Data-Intensive) Arbeitslasten, welche unter anderem in [FE10]

definiert wurden, abgebildet werden. Eine Gewichtung, die direkt auf den in Abschnitt 3.4.5 beschriebenen RZ-Klassen basiert, ist aufgrund fehlender Zahlen zur Hardwareverteilung der einzelnen Typen nicht umsetzbar. Zudem ist nicht sichergestellt, dass alle Rechenzentren eines Typs ein ähnliches Verhältnis zwischen rechen- und speicherintensiven Anwendungen besitzen. Durch die Einteilung der Gewichtungen in die drei Kategorien Balanced sowie Compute- und Data-Intensive können die Anwender der Metrik das betroffene Rechenzentrum unabhängig der Ausrichtung oder des Geschäftsmodells einordnen. Allerdings liegen auch bzgl. der typischen Verhältnisse zwischen rechen- und speicherintensiven Anwendungen in Rechenzentren keine Daten vor. Daher werden für die Gewichtungen der Kategorien C und D vorerst nur Empfehlungen angegeben, die bei Veröffentlichung entsprechender Daten aktualisiert werden. Für den rechenintensiven Fall werden dabei Gewichte der Serverperformance von 0,8 und der Storageperformance von 0,2 empfohlen, für die Kategorie der speicherintensiven Anwendungen ist die empfohlene Gewichtung genau umgekehrt. Die starke Gewichtung in die jeweilige Richtung soll die für rechen- bzw. speicherintensive Anwendungen relevanten Komponenten in der Bewertung deutlich hervorheben.

Prinzipiell können als Gewichtung der Hardware beliebige Relationen (F) verwandt werden, so lange die Gewichte nicht negativ sind und in Summe 1 ergeben. Für Vergleichszwecke muss dann allerdings die Performance der weiteren Rechenzentren mit den gleichen Gewichten bestimmt werden. Damit es ausreichend Vergleichswerte gibt, wurde die Anzahl der empfohlenen Gewichtungen bewusst begrenzt. Die gewählte Gewichtung wird dem Resultat in eckigen Klammern in der Form $LDEE[C_i, T_j] = x \text{ ru/kW}$ angefügt. Wird keine Gewichtung angegeben, wurde die Standardgewichtung (S) angewandt. Bei der freien Gewichtung (F) müssen ebenso die konkret genutzten Gewichte angegeben werden.

Beim Vergleich bzgl. der Verfügbarkeitsklassen kommen keine weiteren Anforderungen an die Ermittlung der Performance hinzu. Es werden die dem RZ-Typ entsprechenden oder die allgemeinen Gewichte verwandt und ansonsten ein Vergleich auf Rechenzentren mit dem gleichen Redundanzgrad beschränkt.

Format des Resultats

Die zusammengefasste RZ-Performance $perf_{DC}$ ist ein prinzipiell einheitenloser Faktor für die auslastungsabhängige und damit tatsächlich genutzte Performance gegenüber der definierten durchschnittlichen Server- und Storageeinheit, der in Abschnitt 4.4.3 eingeführten Referenzmaschine. Die Referenzmaschine PB_R definiert sich über die Menge R der IT-Systeme (Server bzw. Storage), die für alle Benchmarks $PB \in A$ der Menge der für das Performancemodell ausgewählten Benchmarks A mindestens ein Performanceresultat besitzen, vgl. Gleichung 4.6. Diese Menge R wird beispielhaft in Abbildung 4.9 dargestellt. Jeder schwarze Punkt stellt dabei ein konkretes IT-System aus der Menge aller Systeme S dar. Die blau hinterlegten Systeme besitzen jeweils Performanceresultate für die im Modell ausgewählten Benchmarks PB_1 bis PB_3 . Diese Systeme definieren die

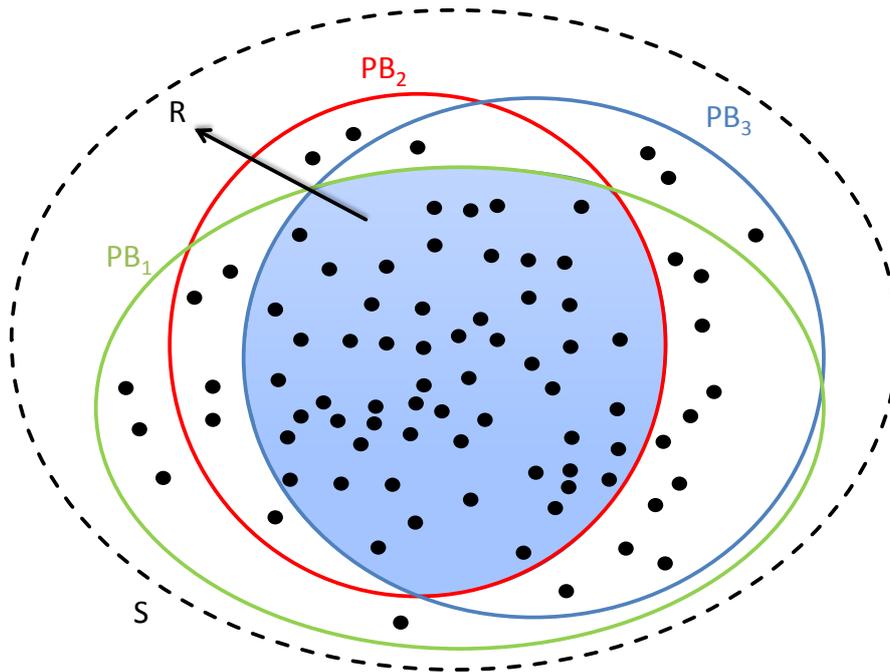


Abbildung 4.9: Beispiel für die Menge R aller IT-Systeme mit Resultaten für alle betrachteten Benchmarks PB_1 bis PB_3

Menge R und repräsentieren damit gemeinsam die Referenzmaschine. Die Menge R wird dabei einmalig festgelegt und ändert sich nicht über die Zeit, so dass die Performance immer gegenüber der gleichen durchschnittlichen Referenzmaschine bewertet wird und somit vergleichbar bleibt, auch wenn weitere Benchmarkresultate hinzukommen.

Die genutzte RZ-Performance $perf_{DC}$ beschreibt also den Faktor der Performance in Relation zur definierten Referenzmaschine. Durch diese Relation repräsentiert das Ergebnis die Anzahl der prinzipiell genutzten Referenzmaschinen, auch wenn diese gar nicht im betrachteten Rechenzentrum vorhanden sind bzw. gar nicht in dieser Form existieren. Allerdings bildet die definierte Referenzmaschine ein durchschnittliches Server- bzw. Stagesystem bezüglich der ausgewählten Benchmarks ab, so dass der Performancewert $perf_{DC}$ eine Einordnung der genutzten Performance ermöglicht. Die Einheit der $perf_{DC}$ ist damit die Anzahl der Referenzmaschinen bzw. -einheiten (ru , reference units). Nach oben ist das Resultat nicht begrenzt, so dass – im Gegensatz zu einer Normalisierung der Performance auf Werte zwischen null und eins – bei zukünftig performanteren Systemen keine erneute Anpassung des Wertebereichs vorgenommen werden muss.

4.5 Powermodelle

Das beschriebene Performancemodell liefert in Kombination mit der Auslastung einen Repräsentanten für die Produktivität des Rechenzentrums. Zur vollständigen Betrachtung der Energieeffizienz muss diese in Relation zum notwendigen Aufwand gesetzt werden. Dieser entspricht der benötigten Energie bzw. bei einer zeitpunktgenauen Betrachtung der Leistungsaufnahme des Rechenzentrums zur Bereitstellung der IT-Ressourcen. Entsprechend setzt die Load Dependent Energy Efficiency die genutzte Performance in Relation zu der korrespondierenden, abgeschätzten Leistungsaufnahme. Wie bereits in Abschnitt 4.2 beschrieben wurde, werden dazu analog zur Performance Powermodelle der RZ-Komponenten definiert, die in Kombination die Gesamtleistungsaufnahme des Rechenzentrums wiedergeben. Neben der IT-Hardware muss dazu die Leistungsaufnahme der Infrastrukturkomponenten (USV, Klimatisierung) betrachtet werden, da diese für den Betrieb der IT notwendig sind.

4.5.1 Anforderungen

Damit Powermodelle der RZ-Hardware für die LDEE-Metrik anwendbar sind, müssen diese die nachfolgend beschriebenen Anforderungen erfüllen.

- R 4.8 **Auslastungsabhängigkeit:** Analog zu den Performancemodellen müssen auch die Powermodelle die Auslastung der Systeme berücksichtigen und erben diese Anforderung damit direkt von der LDEE (vgl. R 3.2). Für die Powermodelle bedeutet dies, dass sie den Einfluss der Last auf die Leistungsaufnahme abbilden müssen.
- R 4.9 **Anwendbarkeit:** Die Powermodelle müssen ohne Beeinflussung des Produktivbetriebs anwendbar sein. Für RZ-Komponenten, die nicht direkt vermessen werden, müssen die Modelle eine Möglichkeit der Vorcharakterisierung bieten. Dies ermöglicht auch die Untersuchung des potentiellen Austauschs von Komponenten.
- R 4.10 **Temperaturabhängigkeit:** Einige der Klimakomponenten hängen maßgeblich von der Außentemperatur ab, da diese dazu genutzt wird, ein Kühlmedium abzukühlen. Die Modelle dieser Komponenten müssen daher den Einfluss der Außentemperatur auf die Leistungsaufnahme betrachten. Dies gilt auch für die übergeordneten Modelle wie die komplette Kühlkette und das kombinierte RZ-Powermodell.

4.5.2 Bestimmung der Leistungsaufnahme

Die Leistungs- bzw. Energieaufnahme der RZ-Komponenten wird direkt an den Geräten gemessen. Viele aktuelle Komponenten besitzen dazu bereits integrierte

Sensoren und ermöglichen die Abfrage der Leistungsaufnahmen über das Unternehmensnetzwerk beispielsweise über das Simple Network Management Protocol (SNMP). Da einige dieser Komponenten allerdings nur Modelldaten liefern – Modelldaten werden nicht gemessen sondern von in den Geräten hinterlegten Modellen anhand der Betriebsparameter berechnet und weichen daher zum Teil deutlich von der Realität ab – muss vor der Verwendung eine Plausibilitätsprüfung durchgeführt werden.

Besitzen die Komponenten keine interne Energiemessung, sind externe Messgeräte notwendig. Häufig finden intelligente PDU, die für jeden separaten Anschluss die Energieaufnahme über das Netzwerk bereitstellen, für die IT-Hardware Anwendung. Durch redundante Auslegungen der Netzteile lassen sich diese auch im Produktivbetrieb nachrüsten. Eine andere Möglichkeit ist die Messung über Induktion, bei der der Messsensor außerhalb des Stromkabels installiert wird ohne die Verbindung zu unterbrechen. Allerdings ist davon auszugehen, dass in den wenigsten Rechenzentren jede RZ-Komponente einzeln vermessen wird. Bei der Anwendung der LDEE sind daher üblicherweise zuvor charakterisierte Powermodelle notwendig um detaillierte Resultate auf Komponentenebene zu erhalten.

Auch wenn für alle Komponenten reale Messwerte vorliegen, können diese für eine detaillierte Betrachtung des Rechenzentrums unzureichend sein. So erfordern die Powermodelle eine Betrachtung der Leistungsaufnahme in Abhängigkeit von der Ressourcenauslastung (R 4.8) und der Temperatur (R 4.10). Bei Messungen an Produktivsystemen wird jedoch nie das ganze Spektrum möglicher Lasten und Temperaturen auftreten, so dass ein Powermodell basierend auf diesen Daten nur Aussagen über die typischen Betriebszustände liefert. Sollen allerdings potentielle Konfigurations- oder Architekturänderungen bewertet werden, die von den Messungen abweichende Eingabeparameter erfordern, muss auf vordefinierte Modelle zurückgegriffen werden.

Ein Beispiel für solch einen Fall ist die Virtualisierung von Servern. Möchte man die potentiellen Auswirkungen auf die Energieeffizienz durch Virtualisierung untersuchen, wendet man die LDEE und damit die Performance- und Powermodelle auf die angestrebte zukünftige Packdichte an. Die Modelle müssen dementsprechend die Power und Performance für höhere Auslastungen zurückliefern. Die Resultate sind allerdings nur adäquat, wenn die Modelle zuvor mit Trainingsdaten der gleichen Größenordnung charakterisiert wurden. Um die Anforderung der Anwendbarkeit (R 4.9) zu erfüllen, sollten also neben realen Messwerten auch zuvor erhobene Daten zur Modellierung herbeigezogen werden. Dies ist beispielsweise über Energiebenchmarks, die neben der Performance auch die korrespondierende Leistungsaufnahme messen, möglich.

4.5.3 Server

Der Server ist wahrscheinlich die Komponente des Rechenzentrums, die energie-technisch in der Forschung am häufigsten untersucht wurde. Es gibt dutzende

publizierte Powermodelle mit unterschiedlichen Ansätzen der Modellierung. Allerdings haben alle Modelle miteinander gemein, dass sie für jeden Servertyp auf Basis von Messdaten charakterisiert werden müssen, vgl. Abschnitt 2.6.3.

Damit die LDEE auch außerhalb der Verfügbarkeit von Messdaten anwendbar ist, müssen die Powermodelle bereits vor der Anwendung charakterisiert sein. Dies ist jedoch nur möglich, wenn ein Datenbestand zu allen in Frage kommenden Servertypen vorhanden ist. Aus diesem Grund basiert das Serverpowermodell, ebenso wie die in Abschnitt 4.4 beschriebenen Performancemodelle, auf öffentlich verfügbaren Power-/Performancedaten von standardisierten Energiebenchmarks. Es gibt zwar auch in der Forschung Powermodelle, die auf bekannten Benchmarks aufsetzen, allerdings erfordern diese die Anwendung der Benchmarks gefolgt von einer Charakterisierung anhand der dabei gemessenen Daten. Das Problem der Anwendbarkeit wird dadurch nicht gelöst.

Als Basis der Modellierung wird der bereits für das Serverperformancemodell in Abschnitt 4.4.4 beschriebene Benchmark SPECpower_ssj2008 genutzt. Dieser umfasst neben der Ermittlung der Performance die Messung korrespondierender Leistungsaufnahmen auf elf Laststufen und besitzt eine umfangreiche Datenbank mit den gemessenen Performance- und Powerdaten. Mit der SAP SD Server Power, TPC-Energy, SPECvirt_2013 Server PPW und dem Server Efficiency Rating Tool sind noch weitere Benchmarks, die die Performance der Leistungs- bzw. Energieaufnahme gegenüberstellen, vorhanden. Allerdings können diese Benchmarks nicht die Anforderung der Anwendbarkeit (R 4.9) erfüllen, da jeweils nur sehr wenige Benchmarkresultate, die zusätzlich zur Performance auch die entsprechende Leistungsaufnahme angeben, verfügbar sind. Server müssten daher zur Vermessung aus ihrem Produktivbetrieb abgezogen werden, was aufgrund des dafür zu betreibenden Aufwands nicht akzeptabel ist.

Der SPECpower_ssj2008-Benchmark ermittelt zunächst innerhalb einer Kalibrierungsphase die maximal erreichbare Performance. Anschließend wird die Arbeitslast in 10%-Schritten reduziert. Dabei wird jeweils die benötigte Leistungsaufnahme mitgeschrieben, so dass pro Laststufe ein Performance-zu-Power-Wert berechnet werden kann. Während das Performancemodell jeweils die maximal erreichbare Performance eines Servers berücksichtigt, ist für das Powermodell hingegen die für die unterschiedlichen Laststufen benötigte Leistungsaufnahme relevant. Genau diese Werte werden zur Charakterisierung des Serverpowermodells genutzt, welches in Gleichung 4.12 definiert ist.

$$P_{srv} = a_{srv} \cdot load_{srv}^3 + b_{srv} \cdot load_{srv}^2 + c_{srv} \cdot load_{srv} + d_{srv} \quad (4.12)$$

Die Leistungsaufnahme P_{srv} eines Servers srv hängt demnach einzig von der darauf anliegenden Last $load_{srv}$ ab. Die Regressionsparameter a_{srv} bis d_{srv} müssen dabei für jeden Servertyp auf Basis des SPECpower-Benchmarks charakterisiert werden. Dabei wird die Moore-Penrose-Pseudoinverse zur Bestimmung der kleinsten Quadrate verwandt [Alb72]. Ein nichtlineares Modell bestehend aus einem Polynom dritten Grades liefert dabei für die vorliegenden Trainings-

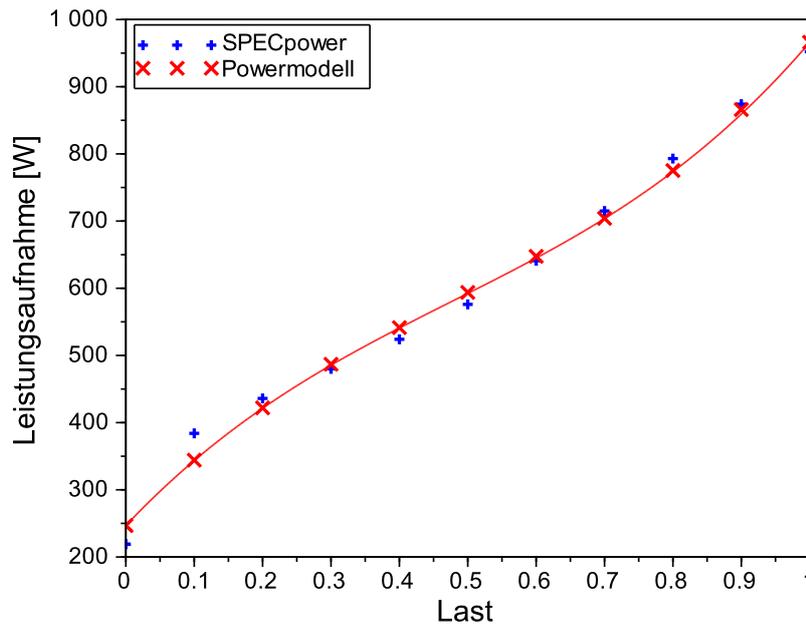


Abbildung 4.10: Vergleich der Abschätzung (rot) mit der bei SPECpower geführten Leistungsaufnahme (blau) für einen HP ProLiant DL170h G6

daten des SPECpower-Benchmarks die besten Ergebnisse. Der mittlere relative quadratische Fehler beträgt hierbei 2,9% (14,8% maximal). Im Vergleich dazu besitzt das Modell des Polynoms des zweiten Grades einen mittleren relativen Fehler von 5,4% (39,6%) und eine einfache lineare Regression 6,8% (44,2%). In Abbildung 4.10 ist ein Beispiel für eine solche Kurve der Leistungsaufnahme, die durch ein Polynom dritten Grades modelliert wird, dargestellt.

Sind die Servertypen identisch, lässt sich die Gleichung 4.12 direkt anwenden. Gleiche Servertypen besitzen aber gerade hinsichtlich des Arbeitsspeichers je nach Anwendungszweck große Unterschiede, so dass sich der relativ gesehen geringe Energiebedarf einzelner Module in der Masse deutlich auf den Energiebedarf des Servers auswirkt. Die Abschätzung der Leistungsaufnahme muss daher noch bzgl. des Arbeitsspeichers angepasst werden. Die in der Forschung vorhandenen RAM-Powermodelle wie z.B. [Vog10] und [BAN⁺11] können hierbei allerdings nicht angewandt werden, da in der Regel zu wenige Parameter zu den konkret verbauten RAM-Modulen und ihrer Nutzung vorliegen.

Die Anpassung wird daher aufgrund durchschnittlicher Powerwerte einzelner RAM-Module durchgeführt. Diese Werte reichen je nach RAM-Typ von 1,3W bis ca. 2,5W pro Modul, vgl. [BR10, Kin15, Kin12]. Der dynamische Anteil beläuft sich nach [Kin15] auf ca. 0,5W, wird jedoch mangels detaillierter Daten zur Auslastungsabhängigkeit der Leistungsaufnahme, welche zur Modellierung benötigt wären, vernachlässigt. Stattdessen wird pauschal für jedes Speichermodul

eine durchschnittliche Leistungsaufnahme von $P_{dimm} = 2W$ angenommen. Die Gleichung 4.12 wird somit folgendermaßen erweitert:

$$P_{srv} = a_{srv} \cdot load_{srv}^3 + b_{srv} \cdot load_{srv}^2 + c_{srv} \cdot load_{srv} + d_{srv} + P_{dimm}(\#DIMM_{srv} - \#DIMM_{ref}) \quad (4.13)$$

Man erhält damit eine Abschätzung der Leistungsaufnahme P_{srv} (in Watt) für einen Server srv , der einem in SPECpower enthaltenen Servertyp ref entspricht. Die Differenz der Anzahl der DIMM (Dual Inline Memory Module) zwischen dem Referenzserver ref und der untersuchten Maschine wird dabei mit der statischen Leistungsaufnahme P_{dimm} eines DIMM multipliziert um eine abweichende Arbeitsspeicherbelegung zu berücksichtigen.

Um den Einfluss des Arbeitsspeichers auf die Leistungsaufnahme zu bewerten, wurde das Serverpowermodell mit und ohne Berücksichtigung des RAM (also Gleichungen 4.12 und 4.13) mit Ergebnissen aus dem SAP SD Server Power Benchmark verglichen. Die dort untersuchten Serversysteme sind auch in SPECpower vorhanden, jedoch wurde für den SAP SD Benchmark die dreibis vierfache Menge an Arbeitsspeicher verbaut. Da die Powerwerte des SAP SD Benchmarks nur als prozentualer Anteil der maximalen Leistungsaufnahme verfügbar sind, wurden die Leistungsaufnahmen aus dem SPECpower-Benchmark und die Abschätzungen der Powermodelle entsprechend umgerechnet. Zur Bewertung wurde dann die durchschnittliche und maximale Abweichung beider Modelle ermittelt. In Abbildung 4.11 ist dies beispielhaft für einen der untersuchten Servertypen dargestellt. Das Serverpowermodell ohne RAM-Erweiterung (blau) weicht dabei deutlich stärker von den gemessenen Leistungsaufnahmen aus dem SAP SD Server Power Benchmark ab als das überarbeitete Powermodell (grün).

Dies wird auch in den berechneten Abweichungen für alle Beispielserver deutlich. Das angepasste Serverpowermodell besitzt im Durchschnitt eine Abweichung von 2,1% (maximal 5,6%) von den Ergebnissen des SAP SD Server Power Benchmarks. Das Powermodell, welches die Arbeitsspeicherbelegung nicht betrachtet, weicht hingegen um 5,3% (8,8%) ab.

4.5.4 Storage

Für Stagesysteme gibt es im Forschungsumfeld einige wenige Powermodelle. Diese benötigen jedoch zur Abschätzung des Energiebedarfs detaillierte Informationen über das Systemverhalten. Beispielsweise nutzt das von Allalouf et al. in [AAF⁺09] vorgestellte Powermodell die HW- und SW-Konfiguration des Stagesystems sowie die genaue Art der Workloads zur Abschätzung. Solche detaillierten Informationen sind bei der Anwendung der LDEE allerdings in der Regel nicht verfügbar, so dass analog zum Serverpowermodell ein eigenes Powermodell basierend auf öffentlich verfügbaren Benchmarkdaten definiert wird.

Das Modell greift dabei auf den in Abschnitt 4.4.5 beschriebenen SPC-1-Benchmark in der Erweiterung SPC-1/E zu, bei dem während des Benchmark-

4 Modellbildung

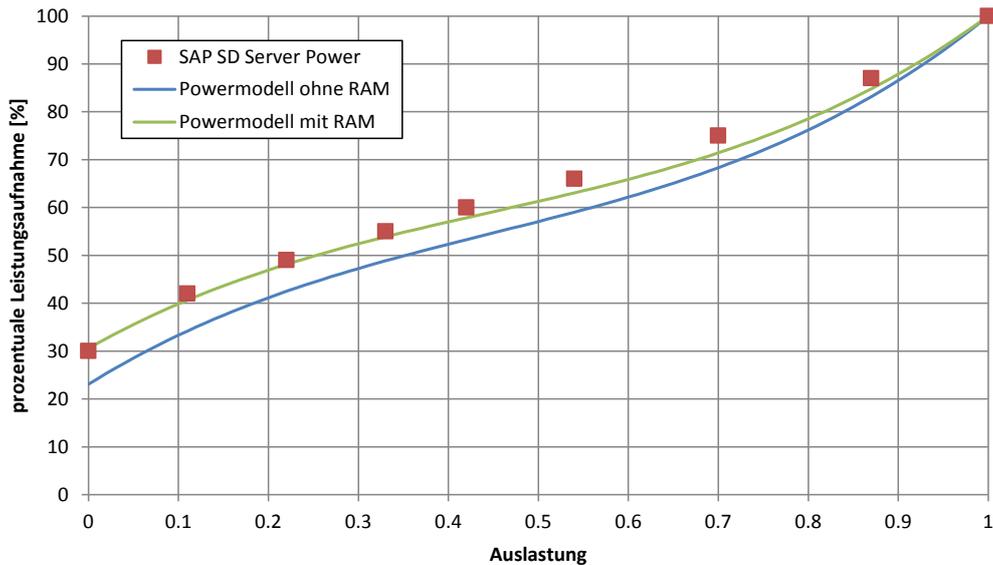


Abbildung 4.11: Anwendung der Serverpowermodelle ohne und mit Berücksichtigung des RAM auf einen Fujitsu PRIMERGY RX300 S8 im Vergleich mit Ergebnissen des SAP SD Server Power Benchmarks

durchlaufs zusätzlich die Leistungsaufnahme für verschiedene Laststufen gemessen wird. In Abbildung 4.12 wird ein Auszug der Daten des SPC-1/E für einige Storagesysteme dargestellt. Man kann gut erkennen, dass die Leistungsaufnahme linear von der vorliegenden Last abhängt. Der dynamische Anteil der Leistungsaufnahme ist mit 3% bis 16% allerdings sehr gering. Die in der Abbildung angegebenen Regressionsparameter zur Steigung fallen mit 0,1W bis 1,3W pro Prozent Auslastung entsprechend klein aus. Dies deckt sich auch mit den Untersuchungen von Allalouf et al. [AAF⁺09] und Lange [Lan09b].

Die Leistungsaufnahme der Storagesysteme wird also mittels linearer Regression über die Auslastungsdaten und Leistungsaufnahmen des SPC-1/E-Benchmarks modelliert. Bei Übereinstimmung der Storamodelle und der Speicherkonfiguration liegt der relative Fehler durch das Modell bei unter 1%. Aber gerade abweichende Speichermengen beeinflussen die Leistungsaufnahme des Storages deutlich, wie auch Liu et al. in [LWQ⁺10] und Lange in [Lan09b] gezeigt haben. Es wurde jeweils festgestellt, dass die Leistungsaufnahme des Storages proportional mit der Anzahl der Festplatten steigt, wobei eine Festplatte ca. 5% bis 10% der elektrischen Leistung des Storagecontrollers ausmacht. Die Speichermenge bzw. die Anzahl der Festplatten sollte daher in das Powermodell mit einfließen:

$$P_{sto} = a_{sto} \cdot load_{sto} + b_{sto} + P_{HD_{sto}}(\#HD_{sto} - \#HD_{ref}) \quad (4.14)$$

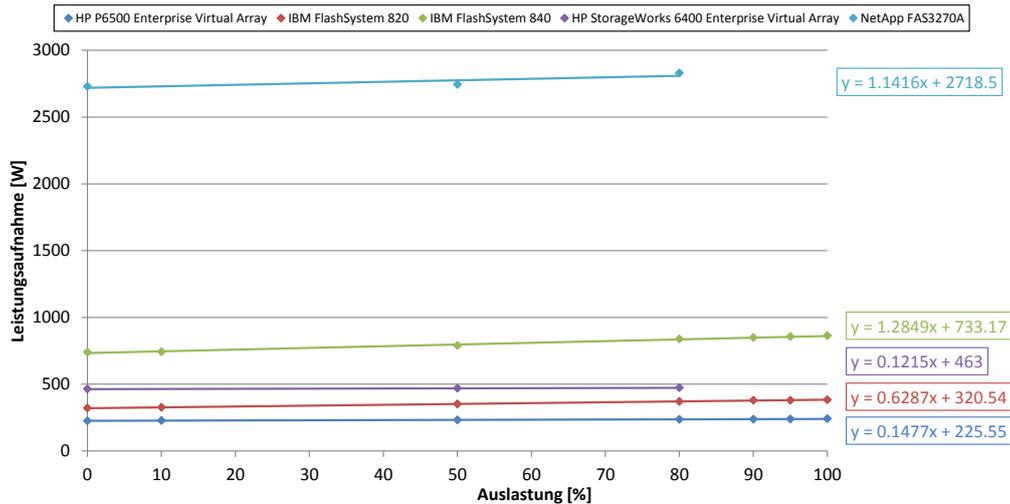


Abbildung 4.12: Leistungsaufnahme bei unterschiedlicher Systemauslastung durch den SPC-1-Benchmark und zugehörige lineare Regressionsparameter

Die Leistungsaufnahme P_{sto} des Storage systems sto setzt sich aus einem linearen auslastungsabhängigen Modell und einem Offset, der durch Anzahl und Typ der Festplatten (HD) bestimmt wird, zusammen. Die Parameter a_{sto} und b_{sto} werden dabei über die im SPC-1/E-Benchmark verfügbaren Daten mit der Methode der kleinsten Quadrate bestimmt. Der Offset ist die Differenz in der Leistungsaufnahme der verbauten Festplatten des untersuchten Systems sto zum Referenzsystem ref , welches in SPC-1/E vermessen wurde. Der grundsätzliche Aufbau des Modells ist somit analog zum Serverpowermodell. Allerdings wird für die Festplatten keine allgemeine durchschnittliche Leistungsaufnahme sondern für jedes Modell ein eigener Wert $P_{HD_{sto}}$ genutzt, da sich die Leistungsaufnahme aufgrund von Architektur (HDD vs. SSD) und Größe (128GB bis 6TB) deutlich unterscheiden kann. Für den Fall, dass im betrachteten Storage system diverse Festplattentypen verbaut sind, muss der Offset für jeden Festplattentyp separat berechnet werden. Dies umfasst die Anzahl ($\#HD$) sowie die Leistungsaufnahme ($P_{HD_{sto}}$) der Typen.

Aber auch innerhalb eines Speichermediums gibt es große Divergenzen. Allalouf et al. [AAF⁺09] und Lange [Lan09b] zeigen jeweils, dass die Leistungsaufnahme der Festplatten sehr stark von der Art der Arbeitslast abhängt. So liegt die Leistungsaufnahme bei 100% zufällig verteilten Schreibvorgängen um mehr als 50% höher als bei 100% sequentiellen Schreibvorgängen. Bei der allgemeinen Energieeffizienzabschätzung mit der LDEE können diese Auslastungstypen in der Regel jedoch nicht unterschieden werden, so dass zur Charakterisierung der Festplatten im Storagepowermodell durchschnittliche Leistungsaufnahmen basierend auf Datenblättern der Speichermedien oder Ranglisten bekannter HW-Webseiten

wie beispielsweise Tom's Hardware¹³ genutzt werden.

4.5.5 Netzwerk

Netzwerkkomponenten im Rechenzentrum sind ein viel untersuchter Forschungsgegenstand. Es gibt zu den diversen Typen an Komponenten (Switches, Hubs, Router, ...) viele unterschiedliche Powermodelle. Beispiele sind das Routerpowermodell von Chabarek et al. [CSB⁺08] und das Powermodell für Switches von Mahadevan et al. [MSBR09]. Allerdings besitzen diese Powermodelle die gleiche Problematik wie die verfügbaren Modelle für Server- und Stagesysteme: Sie verlangen gemessene Trainingsdaten um für die konkreten Geräte kalibriert zu werden. Dies steht jedoch im Konflikt zu den Anforderungen der LDEE.

Während für Server- und Stagesysteme Datenbanken mit Powerwerten von Energiebenchmarks verfügbar sind, gibt es keine solchen Benchmarks und damit Datenbestände für Netzwerkgeräte. Ein Powermodell analog zu den Server- und Stagesystemen ist damit für die Netzwerkkomponenten aktuell nicht anwendbar. Allerdings gibt es von Mahadevan et al. [MSBR09] sowie von Chabarek et al. [CB13] zwei Initiativen, in denen öffentliche Datenbanken mit Kennzahlen unterschiedlicher Netzwerkkomponenten angeboten werden sollen. Diese werden jeweils nach definierten Verfahren ermittelt und könnten als Ersatz für standardisierte Benchmarks genutzt werden.

Basierend auf [MSBR09] würde ein passendes Powermodell für die dann verfügbaren Eingabedaten folgendermaßen aussehen:

$$P_{sw} = P_{ch} + \sum_{lc} P_{lc} + \sum_{pt} Act_{pt} \cdot P_{cfg} \quad (4.15)$$

Die Leistungsaufnahme des Chassis P_{ch} sowie aller im Switch/Router installierten Linecards P_{lc} (mit deaktivierten Ports) bestimmt die statische Leistungsaufnahme der Komponente. Der dynamische Anteil wird über die Aktivität Act_{pt} (1/0) und die Leistungsaufnahme P_{cfg} bei gegebener Konfiguration cfg (10/100/1000 Mbps) der jeweiligen Ports pt definiert. Die Leistungsaufnahmen müssen dabei für die entsprechenden Netzwerkkomponenten in einer Datenbank hinterlegt sein. Die Typen der installierten Linecards und die Aktivität und Konfiguration der Ports kann direkt per SNMP vom Gerät über das Netzwerk ausgelesen werden.

Gegenüber der Definition von Mahadevan et al. [MSBR09] wurde die Skalierung des dynamischen Anteils mit einem Auslastungsfaktor weggelassen, da das IT-Lastmodell keine netzwerkbezogene Auslastung liefert. Der Einfluss der Auslastung der Ports liegt allerdings nach [MSBR09] bei unter 5%. Die dort ermittelte Abweichung des Modells von der Realität von unter 2% wird daher in der angepassten Form nur geringfügig größer sein.

¹³<http://www.tomshardware.com/> (letzter Zugriff: 26.11.2015)

Für den Fall, dass keine detaillierten Leistungsaufnahmen zu den Chassis, Linecards und Portkonfigurationen für Switches und Router zentral zur Verfügung stehen, kann ein simpleres, allgemeines Modell angewandt werden.

$$P_{sw} = P_{ch_{max}} + \sum_{lc} P_{lc_{max}} \quad (4.16)$$

Dieses verwendet die in Datenblättern veröffentlichten Leistungsaufnahmen $P_{ch_{max}}$ und $P_{lc_{max}}$ für die Chassis bzw. Linecards. Diese sind jedoch zumeist deutlich größer als die letztlich bezogene elektrische Leistung, da in der Regel die nominelle Maximalleistungsaufnahme angegeben wird. In [MSBR09] wurde diese Diskrepanz mit dem Ergebnis, dass die realen Maxima bis zu 70% (ca. 50% durchschnittlich) unter den angegebenen Werten liegen, untersucht. Dennoch kann dieses Powermodell als erste Abschätzung genutzt werden. Da der Einfluss des Netzwerks auf den Gesamtenergiebedarf eines Rechenzentrums nur ca. 5% beträgt [PMWV09, Koo11], können die Abweichungen des Netzwerkpowermodells an dieser Stelle in Kauf genommen werden.

4.5.6 USV

Die unterbrechungsfreie Stromversorgung (USV/UPS) speist die gesamte IT-Hardware mit Energie und sorgt für einen durchgängigen reibungslosen Betrieb. Dieser Vorgang ist jedoch mit Verlusten behaftet, die für eine Bewertung der Energieeffizienz berücksichtigt werden müssen. Entsprechend wird auch für die installierten USV-Systeme ein Powermodell benötigt. Dieses ist dabei im Gegensatz zu den IT-Systemen nicht direkt von der Auslastung abhängig, sondern nur indirekt über die Leistungsaufnahme der von der USV versorgten Systeme, welche entsprechend der Eingangsparameter des Modells ist. Das Modell aus Gleichung 4.17 orientiert sich dabei an dem Modell von Pelley et al. [PMWV09].

$$P_{ups} = (a_{ups} \cdot P_{it} + b_{ups}) - P_{it} \quad (4.17)$$

$$P_{it} = \sum_{srv} P_{srv} + \sum_{sto} P_{sto} + \sum_{sw} P_{sw} \quad (4.18)$$

Die Verlustleistung der USV P_{ups} berechnet sich demnach aus der Differenz zwischen der elektrischen Eingangs- und Ausgangsleistung. Die Ausgangsleistung entspricht der Summe der Leistungsaufnahmen aller angeschlossener IT-Komponenten (P_{it}) – Server (srv), Storage (sto) und Switches (sw). Die Eingangsleistung setzt sich hingegen aus einem statischen (b_{ups}) und einem dynamischen Anteil ($a_{ups} \cdot P_{it}$) zusammen. Der dynamische Anteil skaliert dabei proportional mit der bereitgestellten Ausgangsleistung. Die Modellparameter a_{ups} und b_{ups} werden per linearer Regression nach der Methode der kleinsten Quadrate ermittelt. Es gibt keine öffentlichen Datenbanken mit Leistungsdaten von USV-Systemen, d.h. die Trainingsdaten müssen entweder direkt gemessen oder aus

4 Modellbildung

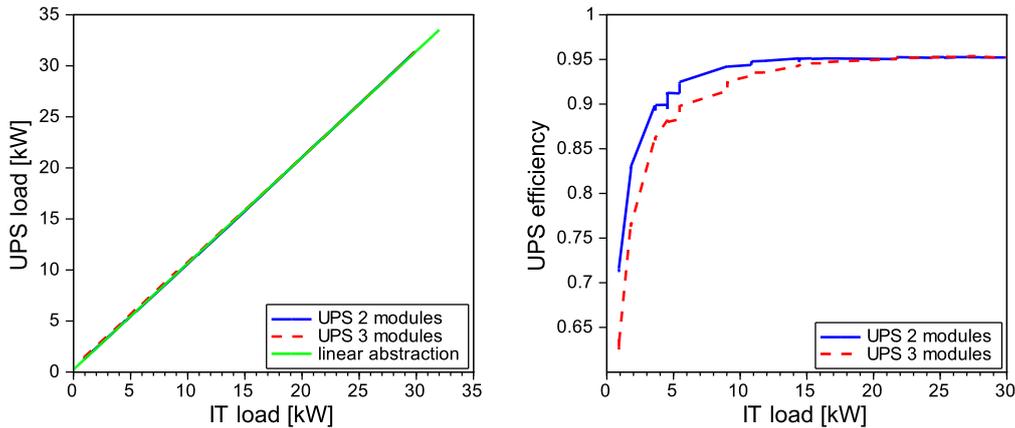


Abbildung 4.13: Elektrische Eingangsleistung (UPS load) und Wirkungsgrad (UPS efficiency) einer USV mit drei 20kW-Modulen gegenüber der Ausgangsleistung (IT load) [SJS⁺14]

öffentlich verfügbaren Datenblättern entnommen werden. Da in Datenblättern grundsätzlich der Wirkungsgrad in Prozent für unterschiedliche Auslastungen angegeben wird, können die lastabhängigen Verlustleistungen hieraus berechnet und anschließend zum Training des Modells genutzt werden.

Das Modell wurde mit realen Messdaten aus dem Projekt AC4DC für ein USV-System von Rittal mit drei 20kW-Modulen untersucht. Das Ergebnis für den Betrieb mit zwei bzw. drei Modulen ist in Abbildung 4.13 dargestellt. Die durchschnittliche Abweichung betrug dabei 1,4% (max. 5,1%). Ebenso wurde das Powermodell auf 243 anonymisierte Datensätze¹⁴ mit Effizienzkennzahlen, die über das Energy Star-Programm der US-amerikanischen Umweltschutzbehörde bereitgestellt werden, angewandt. Hier lag der durchschnittliche relative Fehler bei 0,5% (max. 3,0%).

4.5.7 Klimatisierung

Die RZ-Klimatisierung ist ein komplexes System bestehend aus mehreren untereinander zusammenhängenden Komponenten, deren Parameter voneinander abhängen und deren Betriebszustände sich somit gegenseitig beeinflussen. Grundsätzlich besteht die RZ-Klimatisierung aus der Kälteverteilung, dem Kälte-transport sowie der Kälteerzeugung, vgl. Abschnitt 2.2.2. Welche konkreten Komponenten dabei jeweils genutzt werden, unterscheidet sich von Rechenzentrum zu Rechenzentrum. Es ist sogar möglich und teilweise auch üblich, dass im gleichen Rechenzentrum unterschiedliche Verfahren parallel genutzt werden. So ist es z.B. denkbar Raumluftkühlung über Computer Room Air Handler (CRAH) für normale Enterpriseserver und parallel dazu Wasserkühlung für separate High-

¹⁴https://www.energystar.gov/index.cfm?c=new_specs.uninterruptible_power_supplies (letzter Zugriff: 26.11.2015)

Performance-Computing-Racks zu nutzen. Ein anderes Beispiel ist die parallele Nutzung von Freiluftkühlung neben der Kompressionskühlung (Chiller), welche bei bestimmten Temperaturen sogar in Kombination betrieben werden.

Letztlich umfasst ein Powermodell der Klimatisierung einzelne separate Modelle der verbauten Komponenten, die zu einem umfassenden Modell zusammengesetzt werden, wie es in Abbildung 4.2 dargestellt ist. Die Powermodelle der relevanten Klimakomponenten werden in den nachfolgenden Unterabschnitten beschrieben.

Lüfter

Lüfter werden im RZ dazu eingesetzt den Luftfluss in eine gewünschte Richtung zu gewährleisten. Sie sind zumeist Bestandteil größerer Systeme und sorgen dort in der Regel dafür, den Wärmetausch durch eine ideale Luftzufuhr zu optimieren. Die Leistungsaufnahme von Lüftern hängt dabei kubisch von ihrer Drehzahl und damit dem Volumenstrom ab, vgl. Lüftergesetze [KCR13]. Sie lässt sich entsprechend nach Meisner et al. [MW11] und Breen et al. [BWP⁺10] mit folgender Gleichung 4.19 abschätzen:

$$P_{fan} = P_{fan_{ref}} \left(\frac{\dot{V}}{\dot{V}_{ref}} \right)^3 \quad (4.19)$$

Die Leistungsaufnahme P_{fan} eines Lüfters wird also über das Verhältnis des Volumenstroms \dot{V} des untersuchten Lüfters zu einem Referenzlüfter ref bestimmt. Anstelle eines Referenzlüfters können auch die Maximalwerte (Leistungsaufnahme und Volumenstrom) des untersuchten Lüfters genutzt werden. Dennoch müssen zur Abschätzung der Leistungsaufnahme einige Parameter vorliegen, die in vielen Rechenzentren nicht gemessen werden.

Eine alternative Berechnung der Leistungsaufnahme von Lüftern präsentieren Liu et al. in [LCB⁺12] mit $f_a(d) = kd^3$. Aufgrund der Wärmeübertragungsgesetze lässt sich anstelle des Volumenstroms direkt die IT-Leistungsaufnahme d verwenden, da diese proportional voneinander abhängen. Diese wird mit einem Parameter k multipliziert, der die Temperaturdifferenz zwischen der Außenluft und der Luft an den Serverausgängen repräsentiert. Wie sich dieser Parameter allerdings genau zusammensetzt, wird in [LCB⁺12] nicht beschrieben.

Der Ansatz von Liu et al. lässt sich jedoch auf fehlende Parameter für das Modell in Gleichung 4.19 übertragen. So lässt sich der Volumenstrom von der abzuführenden Wärmelast ableiten. Diese bestimmt sich je nach vorgeschalteter Komponente – bei einem CRAH ist es beispielsweise die IT-Leistungsaufnahme bzw. der Wärmestrom. Aus dem Wärmestrom \dot{Q} wird nach [Inc11] unter Berücksichtigung der spezifischen Wärmekapazität c_p und der Dichte ρ des Übertragungsmediums (Luft) sowie der Temperaturdifferenz Δt am Lüfter der Volumenstrom \dot{V} berechnet.

$$\dot{V} = \frac{\dot{Q}}{c_p \cdot \rho \cdot \Delta t} \quad (4.20)$$

Sind die maximalen Spezifikationen für Leistungsaufnahme und Volumenstrom des Lüfters (z.B. anhand des Datenblatts) bekannt, lässt sich die Leistungsaufnahme durch Einsetzen des berechneten Volumenstroms in die Gleichung 4.19 abschätzen. Sind diese nicht bekannt, kann auch folgende Berechnungsformel für Lüfter verwendet werden [Inc11]:

$$P_{fan} = \frac{\Delta p \cdot \dot{V}}{\mu_{fan}} \quad (4.21)$$

In diesem Fall muss der Druckverlust bzw. die Druckdifferenz Δp am Lüfter bekannt sein. Die Effizienz des Lüfters μ_{fan} ist hingegen optional. Ist diese nicht bekannt, berechnet die Formel die ideale Leistungsaufnahme ohne Berücksichtigung der Verluste durch den Lüfter.

Für die LDEE werden Powermodelle bevorzugt, die die Lüfter umschließenden Systeme (z.B. CRAH oder Chiller) als komplettes System in Abhängigkeit zur Wärmelast abschätzen. Sind entsprechende Modelle allerdings nicht verfügbar, nutzt die LDEE die beschriebenen Lüfterpowermodelle aus den Gleichungen 4.19 und 4.21, je nach Verfügbarkeit von Daten. Die dafür notwendigen Parameter sollten dann – wie zuvor beschrieben – durch Managementsysteme im RZ zentral gesammelt, aus den bekannten Betriebsparametern berechnet und/oder aus öffentlich verfügbaren Quellen wie z.B. Datenblättern bezogen werden.

Pumpen

Die Pumpen transportieren das Kühlwasser zwischen den Klimakomponenten, z.B. vom Kühlturm zum Kaltwassersatz oder vom Kaltwassersatz zu den IT-Räumen und der dort installierten Kälteverteilung. Je nach Klimatisierung werden die Pumpen entweder mit einem konstanten Durchfluss betrieben oder sie werden adaptiv anhand des Kühlbedarfs geregelt. Nach Breen et al. [BWP⁺10] sowie Bean et al. [BD08] berechnet sich die Leistungsaufnahme analog zu Lüftern anhand des Druckverlusts Δp im Kaltwasserkreislauf, des Volumenstroms \dot{V} des Kühlwassers und der Pumpeneffizienz μ_{pmp} , wie in Gleichung 4.22 dargestellt.

$$P_{pmp} = \frac{\Delta p \cdot \dot{V}}{\mu_{pmp}} \quad (4.22)$$

Die Effizienz ist dabei als einziges vorab bekannt, z.B. aus Datenblättern. Die beiden anderen Parameter müssen hingegen entweder aus dem Betrieb ermittelt oder über weitere Parameter berechnet werden. Der Volumenstrom des Kühlwassers lässt sich dabei über Gleichung 4.20 ermitteln, wobei im Falle der Pumpen die spezifische Wärmekapazität und die Dichte des Kühlwasser anstelle der der Luft genutzt werden müssen. Sind nicht alle zur Berechnung notwendigen Parameter bekannt, kann stattdessen die maximale Leistungsaufnahme aus

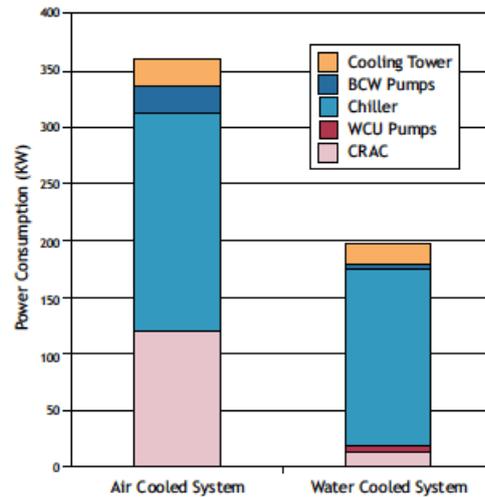


Abbildung 4.14: Vergleich der Leistungsaufnahmen luft- und wassergekühlter Klimaketten bei identischer Kühlleistung – Grafik unverändert aus [EI09] zitiert

den Produktspezifikationen verwandt werden. Der dadurch entstehende Fehler wirkt sich nur geringfügig auf das Gesamtergebnis aus, da die Kaltwasserpumpen einen geringen Anteil des Energiebedarfs des Rechenzentrums ausmachen, wie in Abbildung 4.14 zu sehen ist. In der dargestellten Untersuchung liegt der Anteil der Verlustleistung der Pumpen in luftgekühlten Systemen (BCW Pumps) im Vergleich zur gesamten Klimatisierung bei ca. 7% – bezogen auf das gesamte Rechenzentrum sind dies zwischen 2-3%.

Computer Room Air Handler

Ein Computer Room Air Handler (CRAH) – häufig auch Computer Room Air Conditioning (CRAC) – besteht im Prinzip aus einem Luft-Wasser-Wärmeübertrager und Lüftern. Den Durchfluss des Kühlwassers übernehmen dabei die Pumpen des Kühlkreislaufrs, die separat modelliert werden. Der CRAH stellt die Kaltlufttemperatur nur über die Öffnung des Ventils und die Drehzahlen der Lüfter ein. Die Leistungsaufnahme des CRAH wird somit maßgeblich durch die integrierten Lüfter bestimmt. Das bedeutet, die Leistungsaufnahme kann mit den in Gleichung 4.19 oder 4.21 beschriebenen Modellen abgeschätzt werden, sofern die erforderlichen Parameter verfügbar sind.

Auch typische Powermodelle aus der Forschung (z.B. [PMWV09], [BWP⁺10]) basieren auf diesem Ansatz und schätzen die Leistungsaufnahme P_{crah} des CRAH über die Drehzahl f der Lüfter ab. In Gleichung 4.23 wird ein solches Modell exemplarisch dargestellt. Neben einer drehzahlabhängigen Komponente besitzt das Modell mit b_{crah} eine statische Komponente, die die Controller, Sensoren und weitere Subsysteme des CRAH repräsentiert.

4 Modellbildung

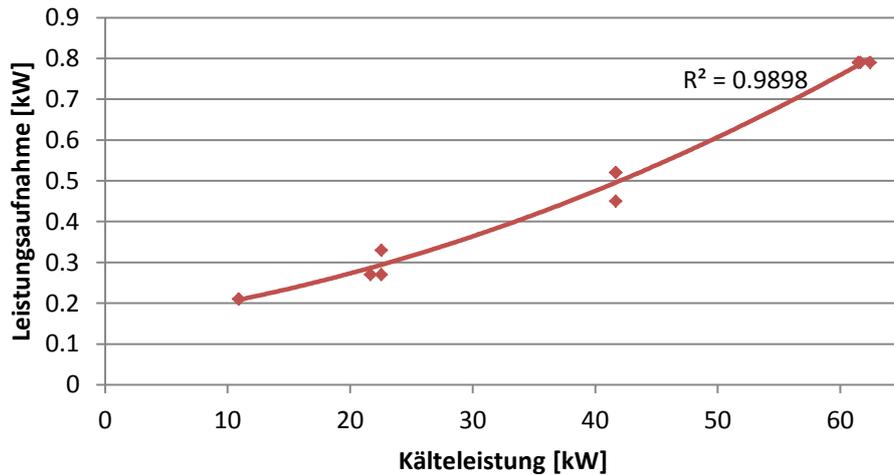


Abbildung 4.15: Modellierung der Leistungsaufnahme eines CRAH als Polynom zweiten Grades [SJS⁺14]

$$P_{crah} = a_{crah} \cdot f^3 + b_{crah} \quad (4.23)$$

Das Problem an diesem Modell liegt in der Anwendbarkeit für die LDEE. Die darin integrierten Powermodelle der IT-Systeme liefern als Ausgabe die Leistungsaufnahme und damit die Wärmelast der IT. Zur Anwendung dieses Klimamodells müsste die Wärmelast in die resultierende Lüfterdrehzahl konvertiert werden, wozu eine detaillierte Modellierung des Wärmeflusses notwendig ist. Da diese jedoch kein Bestandteil der LDEE ist, wurde ein Powermodell in Abhängigkeit zur Wärmelast des zu kühlenden IT-Equipments, welche dem Energiebedarf P_{it} (vgl. Gleichung 4.18) entspricht, definiert.

$$P_{crah} = a_{crah} \cdot P_{it}^2 + b_{crah} \cdot P_{it} + c_{crah} \quad (4.24)$$

Das Modell abstrahiert die Leistungsaufnahme durch ein Polynom zweiten Grades. Die Parameter werden per Regression und der Methode der kleinsten Quadrate berechnet. Pelly et al. [PMWV09] bestätigen die quadratische Abhängigkeit mit der Begründung, dass der kubische Anstieg der Leistungsaufnahme der Lüfter durch die erhöhte Effizienz des Wärmetransfers bei höherem Luftfluss etwas relativiert wird. Ein Beispiel für die Anwendung des Modells auf realen Messdaten aus dem Projekt AC4DC [SJS⁺14] ist in in Abbildung 4.15 gegeben. Es wurden Messungen zu drei unterschiedlichen Kaltlufttemperaturen (20°C, 24°C und 26°C) durchgeführt. Das Modell passt dabei mit einem Bestimmtheitsmaß R^2 von nahezu 1 sehr gut zu den Daten.

Durch eine Charakterisierung des Powermodells anhand von Messdaten werden implizit räumliche Gegebenheiten wie beispielsweise Kalt- oder Warmgangeinhausung sowie Betriebstemperaturen berücksichtigt, die ansonsten separat

über eine Modellierung des Wärmeflusses betrachtet werden müssten. Sind keine Messdaten verfügbar, müssen die Parameter aus Datenblättern abgeleitet werden, da es im Gegensatz zu Server- und Storage-Systemen für Klimageräte keine zentrale Anlaufstelle für entsprechende Daten gibt. Bei einer Charakterisierung anhand allgemeiner Daten kann allerdings nicht mehr die räumliche Umgebung durch das Modell abgebildet werden.

Chiller

Der Chiller bzw. Kaltwassersatz kühlt das erwärmte Wasser, welches von den Kühlgeräten in den Rechnerräumen zurückkommt, wieder herab. Dazu wird in einem internen Kreislauf ein Kältemittel abwechselnd komprimiert und expandiert sowie Wärmeenergie in den Zwischenschritten abgegeben bzw. aufgenommen. Diesem Kreislauf muss zur Aufrechterhaltung des Betriebs Energie, welche für den Kompressionsvorgang benötigt wird, zugeführt werden. Die Leistungsaufnahme P_{chi} eines Chillers lässt sich über den temperaturabhängigen Coefficient of Performance (COP) berechnen, welcher die Effizienz des Klimageräts abhängig zur eingestellten Kaltlufttemperatur (T_c) repräsentiert. Ein entsprechendes Powermodell wird unter anderem von Breen et al. in [BWP⁺10] beschrieben und ist in Gleichung 4.25 angegeben.

$$P_{chi} = \frac{Q}{COP(T_c)} \quad (4.25)$$

$$COP(T_c) = \frac{Q}{E} \quad (4.26)$$

Da der COP mit Q/E – also der abgeführten Wärmeenergie im Verhältnis zur dafür benötigten elektrischen Energie, vgl. Abschnitt 2.5.3 – definiert ist, kann durch eine Umstellung des Terms der Energiebedarf berechnet werden. Dies setzt die Kenntnis des COP (z.B. aus Datenblättern) sowie der abzuführenden Wärmeenergie Q , die sich aus dem Energiebedarf der IT ergibt, voraus. Ein anderer Ansatz von Pelley et al. [PMWV09] modelliert die Leistungsaufnahme des Chillers P_{chi} über ein lineares Modell mit einem Polynom zweiten Grades, da die Leistungsaufnahme quadratisch mit der Wärmelast ansteigt. Als unabhängige Variable wird die IT-Auslastung des gesamten Rechenzentrums U genutzt, welche nach [PMWV09] proportional zur IT-Wärmelast ist. Das Modell wird in der nachfolgenden Gleichung dargestellt.

$$P_{chi} = a_{chi} \cdot U^2 + b_{chi} \cdot U + c_{chi} \quad (4.27)$$

Obwohl Pelley et al. in ihrem Papier zeigen, dass auch die Außentemperatur einen Einfluss auf die Leistungsaufnahme besitzt – schließlich wird die Wärmelast vom Chiller an die Außenluft abgegeben –, wird diese im vorgestellten Modell nicht berücksichtigt. Dies führt mitunter zu hohen Abweichungen. Wendet man das Modell auf Messdaten eines Chillers aus dem AC4DC-Projekt [SJS⁺14] an,

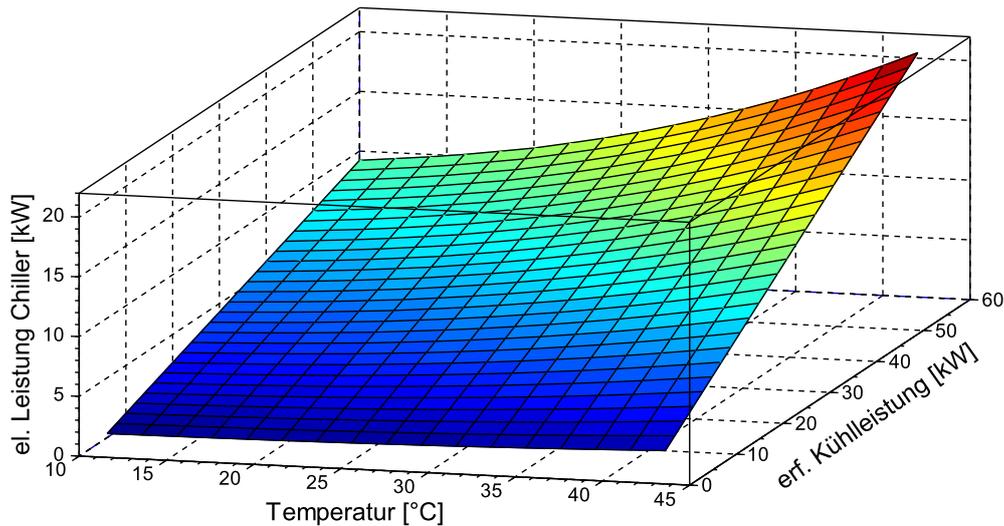


Abbildung 4.16: Powermodell angewandt auf einen Chiller aus dem AC4DC-Projekt [SJS⁺14]

so erhält man eine durchschnittliche Abweichung von 2,1kW bzw. eine relative Abweichung von 28,5%. Das für die LDEE entwickelte Powermodell lässt daher die Außentemperatur als Umgebungsparameter mit einfließen. Es basiert ebenso wie das Modell von Pelley et al. auf einem linearen Modell, allerdings mit multiplen unabhängigen Variablen – dem Kältebedarf cd und der Außentemperatur t_{oa} .

$$P_{chi} = a_{chi} \cdot cd \cdot t_{oa}^2 + b_{chi} \cdot cd \cdot t_{oa} + c_{chi} \cdot cd + d_{chi} \cdot t_{oa} + e_{chi} \quad (4.28)$$

Im Vergleich zum Modell von Pelley et al. liegt die Abweichung bei durchschnittlich 0,03kW bzw. bei knapp 1%. Auch das Bestimmtheitsmaß ist mit nahezu 1 im Vergleich zu 0,77 fast optimal. Dass kein Overfitting des Modells vorliegt, zeigte eine 3-fache Kreuzvalidierung, bei der der relative Fehler ebenso unter einem Prozent lag. In Abbildung 4.16 ist das angewandte Modell exemplarisch dargestellt. Die Einflüsse der unabhängigen Variablen sind dabei gut zu erkennen – die Leistungsaufnahme steigt linear mit der erforderlichen Kühlleistung und überproportional mit der Außentemperatur an.

Kühltürme

Kühltürme sorgen für einen Austausch von Wärme mit der Außenluft. In den meisten RZ-Klimainstallation werden sie an den Chiller angeschlossen, um im Kompressorbetrieb die anfallende Wärme des Verflüssigers abzuführen. Häufig ist aber auch ein reiner Freikühlbetrieb möglich, bei dem das Kühlmedium um den Chiller umgeleitet und direkt vom Kühlturm herabgekühlt wird. Dies ist

natürlich nur bei entsprechenden Außentemperaturen möglich. Außerdem unterscheiden sich zwischen den beiden Betriebsarten auch die Leistungsaufnahme des Kühlturms, die im Freikühlbetrieb durch die höheren Anforderungen steigt.

Um Kühltürme energetisch zu modellieren, wird bei bekannten Powermodellen vom vorliegenden Volumendurchfluss (Luft und Wasser) abstrahiert, da die Hauptkomponenten von Kühltürmen Lüfter und Pumpen sind. Ein entsprechendes Modell von Breen et al. [BWP⁺10] ist nachfolgend dargestellt.

$$P_{ct} = P_{fan} + P_{pmp} \quad (4.29)$$

Die Leistungsaufnahme der Lüfter und Pumpen berechnet sich anhand der Gleichungen 4.19 und 4.22. Die Parameter für den jeweiligen Volumendurchfluss sowie Druckverlust werden dabei durch den Aufbau der Kühlkette, dem aktiven Betriebsmodus und der abzuführenden Wärmeenergie bestimmt. Die Leistungsaufnahme wird also beispielsweise dadurch beeinflusst, ob der Kühlturm das erwärmte Kühlwasser vom Kaltwassersatz oder direkt von den CRAH erhält. Ebenso besitzt die Außentemperatur, die bei der Berechnung des Volumenflusses einfließt, Auswirkungen auf die Leistungsaufnahme. Diese Faktoren werden bei der nachfolgend beschriebenen Kombination der Powermodelle berücksichtigt.

4.5.8 Kombiniertes RZ-Powermodell

Die Leistungsaufnahme des gesamten Rechenzentrums ist die Summe der Leistungsaufnahmen aller ihrer Komponenten. Entsprechend werden die Ausgaben der Powermodelle zu allen vorhandenen Komponenten addiert, wie in Abbildung 4.2 dargestellt. Als Eingabe benötigen alle Powermodelle eine Form der Last. Für die Server- und Storage-Systeme wird die vom vorgeschalteten IT-Lastmodell abgeschätzte Verteilung der RZ-Last verwendet. Die Eingaben für die Infrastrukturkomponenten ergeben sich jeweils aus den vorgeschalteten Modellen. Die Abhängigkeiten zwischen den Einzelmodellen werden nachfolgend näher erläutert.

Abhängigkeiten zwischen Modellen

Für jede USV-Komponente sollte bekannt sein, welche IT-Systeme von dieser versorgt werden. Das USV-Powermodell (UPS in Abbildung 4.2) erhält dann als Eingabe jeweils die Summe der Leistungsaufnahmen der angeschlossenen IT-Komponenten – also die aggregierten Ausgaben der entsprechenden Modelle. Das USV-Modell berechnet dann für jede USV die elektrische Eingangsleistung, die sich aus der IT-Leistung und der Verlustleistung der USV zusammensetzt. Die Summe aller Eingangsleistungen sowie die IT-Leistungsaufnahme, die nicht über USV-Systeme bereitgestellt wird, entspricht dann der anfallenden Wärmelast, die über die Kühlkette abgeführt werden muss. Nach dem Energieerhaltungssatz

in der Thermodynamik [BK12] entspricht die elektrische Leistung dann dem Wärmestrom \dot{Q} .

Dieser Wärmestrom muss von den Kühlgeräten in den Rechnerräumen abgeführt werden und ist damit ein Eingangsparameter des CRAH-Powermodells. Sind mehrere Kühlgeräte vorhanden, muss die Abwärme bzw. elektrische Leistung auf die Geräte und ihre entsprechenden Powermodelle aufgeteilt werden. Dies geschieht entweder getrennt nach Räumlichkeit oder per Gleichverteilung für Geräte, die sich im gleichen Raum befinden. Die CRAH übertragen die Wärme dabei über einen Wärmetauscher an das Kühlwasser, welches in einem Kreislauf zu den Kaltwassersätzen und wieder zurück geführt wird. Zu der Abwärme der IT addiert sich nun noch die Verlustleistung des CRAH – also die Ausgabe des CRAH-Powermodells.

Der Kühlkreislauf wird über Pumpen angetrieben, deren Leistungsaufnahme über den Wärmestrom sowie die Druckdifferenz berechnet wird. Ist mehr als eine Pumpe im gleichen Kreislauf in Betrieb, hat dies einen entsprechenden Einfluss auf die Rechenparameter. Sind mehrere Pumpen parallel geschaltet, wird der Volumendurchfluss, der sich über den Wärmestrom ergibt, auf die Pumpen verteilt. Sind die Pumpen hingegen in Reihe geschaltet, verteilt sich das hydraulische Potential auf die Pumpen, wodurch die Druckdifferenz sinkt [GRU04]. Die Leistungsaufnahme der einzelnen Pumpe ist damit bei einer Aufteilung erwartungsgemäß geringer als wenn eine Pumpe die gesamte Arbeit alleine leisten müsste.

Weitere Abhängigkeiten zwischen den Kühlkomponenten – vornehmlich den Pumpen, dem Kaltwassersatz und den Kühltürmen – ergeben sich durch die mögliche Nutzung von Freiluftkühlung, vgl. Abschnitt 2.2.2. Ist diese Option vorhanden, hängen die Betriebsparameter der Einzelkomponenten vom aktuellen Betriebszustand ab, d.h. ob ein reiner Freikühlbetrieb, ein reiner Kompressorbetrieb oder ein Mischbetrieb aktiv ist [SPX13]. Der Betriebsmodus richtet sich dabei nach den Differenzen zwischen der Außentemperatur und der zu erreichenden Wasservorlauftemperatur bzw. der vorliegenden Wasserrücklauftemperatur. Unter Berücksichtigung dieser Parameter kann zunächst die erreichbare Kälteleistung (bzw. abgeführte Wärmeleistung) \dot{Q}_{ct} des Kühlturms berechnet werden [Alf04]. Dazu wird die nachfolgende Gleichung angewandt:

$$\dot{Q}_{ct} = \dot{V} \cdot c_p \cdot \rho \cdot \Delta t \quad (4.30)$$

Entscheidend für die Kälteleistung sind der Volumendurchfluss \dot{V} und die Differenz der Wasservor- und -rücklauftemperaturen Δt im Kühlturm – die spezifische Wärmekapazität c_p sowie die Dichte ρ sind hingegen konstant. Die Wasserrücklauftemperatur ist dabei durch die abzuführende Wärme des Rechenzentrums gesetzt, während sich die mögliche Wasservorlauftemperatur über die Außenluft, an die die Wärme übertragen wird, und die Effizienz des Kühlturms bestimmt. Bei bekannter Wasserrücklauftemperatur T_i , Kühlgrenztemperatur der Außenluft T_{wb} sowie Effizienz μ_{ct} des Kühlturms kann die potentielle Wasservor-

lauftemperatur T_o mit Gleichung 4.31 bestimmt werden [Nap05].

$$T_o = T_i - \mu_{ct}(T_i - T_{wb}) \quad (4.31)$$

Die durch Gleichung 4.30 berechnete Kälteleistung der Freiluftkühlung muss der Kaltwassersatz (Chiller), sofern ein Mischbetrieb unterstützt wird, nicht mehr erbringen. Das herabzukühlende Kühlwasser wird bereits vor dem Eintreffen beim Kaltwassersatz über einen Wärmetauscher durch das vom Kühlturm kommende kalte Wasser vorgekühlt. Das Chillerpowermodell wird daher mit einem entsprechend geringeren Kältebedarf ausgeführt. Sollte der notwendige Kältebedarf komplett von den Kühltürmen bzw. über die Freikühlung gedeckt werden, benötigt die Kompressionskühlung des Chillers keine Energie, da sie für diesen Zeitraum abgeschaltet wird.

Im Falle zu hoher Außentemperaturen/Kühlgrenztemperaturen oder wenn die Option zur Freikühlung aus anderen Gründen nicht genutzt wird, wird zunächst das Chillerpowermodell auf dem Wärmestrom des Kühlkreislaufs angewandt. Die Wärme wird dann vom Chiller über den Verflüssiger an die Kühltürme abgeführt, so dass die Leistungsaufnahme des Kühlturms über Gleichung 4.29 berechnet werden kann. Der Wärmestrom verändert sich zwar nicht, allerdings kommt dieser mit deutlich höheren Temperaturen von ca. 35°C bis 40°C [Tec08, FM05] bei den Kühltürmen an. Durch die höhere Temperaturdifferenz sinkt der Volumenstrom an der Pumpe und dem Lüfter (vgl. Gleichung 4.20), so dass letztlich auch der Energiebedarf dieser Komponenten sinkt.

In Abbildung 4.17 ist der beschriebene Zusammenhang der Powermodelle der Klimakomponenten grafisch dargestellt. Das Temperaturverhältnis bestimmt dabei den Betriebsmodus, sofern Freiluftkühlung unterstützt wird. Sollte dies nicht der Fall sein, so wird immer der rechte Pfad ausgeführt, bei dem der Kompressor im Chiller die Kühlung übernimmt und der Kühlturm nur die Abwärme des Verflüssigers abführt. Bei der reinen Freiluftkühlung (linker Pfad) kommt der Kompressor des Chillers überhaupt nicht zum Einsatz, da der Kühlturm die komplette Abwärme des Rechenzentrums handhaben kann.

Format des Resultats

Jedes einzelne Powermodell besitzt als Ausgabe die Leistungsaufnahme der korrespondierenden Komponente. Die Gesamtleistungsaufnahme des Rechenzentrums setzt sich entsprechend Abbildung 4.2 aus den Abschätzungen der Modelle aller IT-Komponenten, der USV-Systeme sowie der Klimakomponenten zusammen. Das Resultat ist ein einzelner, aggregierter Wert der elektrischen Leistung, der hinsichtlich der Größe typischer Rechenzentren auf kW skaliert wird.

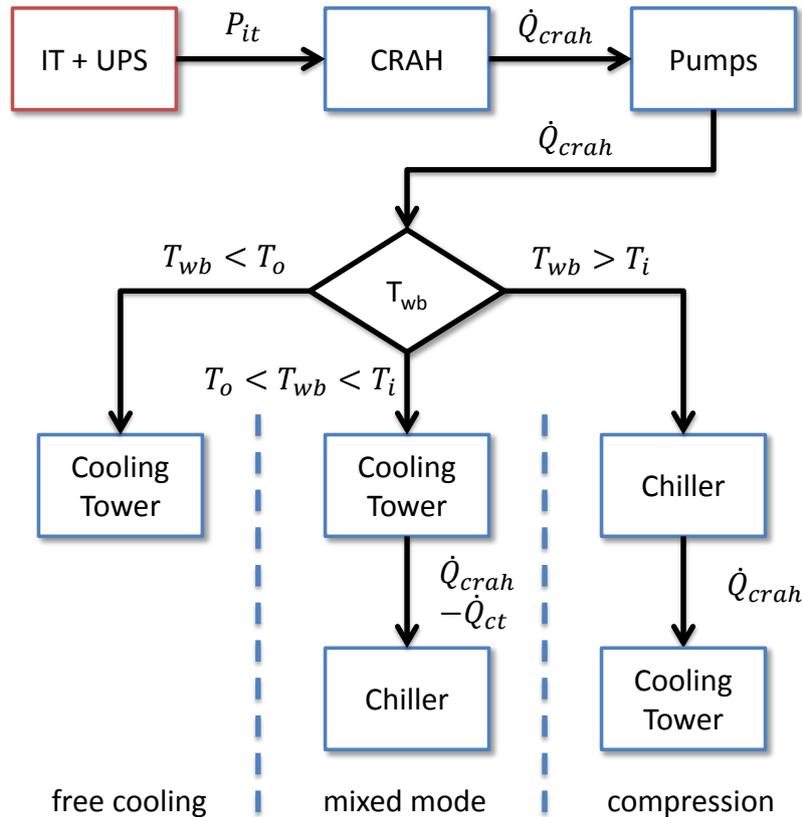


Abbildung 4.17: Der Betriebsmodus des Rechenzentrums hängt von der Kühlgrenztemperatur T_{wb} im Vergleich zu den Wasservor- und -rücklauftemperaturen T_o und T_i ab. Die einzelnen Powermodelle liefern als Ausgabe jeweils die eigene Verlustleistung, die in Summe die Leistungsaufnahme der Klimatisierung ergibt.

4.6 Evaluation

Nachfolgend werden Evaluationen der in diesem Kapitel vorgestellten Modelle beschrieben. In der Regel wurden dabei Teillösungen als auch die integrierten Modelle untersucht und mit alternativen Lösungen verglichen.

4.6.1 IT-Lastmodell

Das IT-Lastmodell wird in zwei Schritten ausgeführt. Für Rechenzentren, in denen ein Powermanagement der Server aktiv ist, bestimmt das Modell zunächst, welche Server wahrscheinlich genutzt werden. Für die aktiven Server- und Storagekomponenten wird anschließend eine konkrete Auslastung abgeschätzt. Die beiden Teilverfahren wurden jeweils im Vergleich zu alternativen Ansätzen evaluiert.

Bestimmung der aktiven Server

Zur Bestimmung der aktiven Server bei einer gegebenen, aggregierten Rechenzentrumsauslastung wurde in Abschnitt 4.3.4 ein Modell vorgestellt, welches je nach verfügbaren Eingabedaten eines von zwei Verfahren anwendet (vgl. Abbildung 4.3). In der nachfolgend beschriebenen Evaluation wird das Modell basierend auf der Aktivitätsrate der Server mit AR und das Modell, welches die Serverspezifikationen nutzt, mit Sp betitelt.

Die Basisdaten der Evaluation stammen aus der LPM-Simulation des AC4DC-Projekts [SJS⁺14]. Mit diesem Werkzeug wurden diverse Konfigurationen eines Rechenzentrums durchlaufen, wobei nur die Anzahl der virtuellen Maschinen festgelegt wurde – die simulierten Servertypen sowie die Arbeitslasten jeder VM wurden zufällig bestimmt. Pro vorgegebener VM-Anzahl wurden zehn Konfigurationen generiert und das LPM auf diesen jeweils simuliert. Daraus resultierten Zeitreihen mit der Serveraktivität und -auslastung sowie der aggregierten RZ-Auslastung für die Ressourcentypen CPU und RAM. Auf diesen Daten wurden die Modelle zur Bestimmung der aktiven Server angewandt, wobei die Daten in Trainingsdaten und Testdaten unterteilt wurden.

Die im IT-Lastmodell umgesetzten Verfahren AR und Sp wurden jeweils mit zwei alternativen Ansätzen verglichen. Die Methoden mit der Namens Erweiterung $w/o\ max$ entsprechen dabei den zuvor genannten Verfahren mit dem einzigen Unterschied, dass der erste Schritt der Bestimmung der Maximalwerte ausgelassen wurde. Die Varianten berücksichtigen also nicht die Planungszeiträume des LPM, sondern nutzen die unveränderten Eingabewerte. Zudem wurden die Verfahren Av , welches die Serverauswahl aufgrund der durchschnittlichen Auslastung der Server vornimmt, und Rdm , welches eine zufällige Auswahl der Server trifft, als Vergleich herbeigezogen. Av berücksichtigt dabei ebenso wie AR die detaillierten Eingabedaten, während Rdm genauso wie Sp nur die aggregierten Auslastungen nutzt.

Insgesamt wurden die Ergebnisse anhand von zwei Metriken untersucht, jeweils ein Maß für die Anzahl und die Auswahl der aktiven Server. Der relative quadratische Fehler repräsentiert die Differenz der abgeschätzten Serveranzahl zu der tatsächlichen für jeden Zeitpunkt der Testdaten. Die Trefferrate hingegen gibt den Anteil der Server an, die vom angewandten Modell ausgewählt wurden und mit der tatsächlichen Serverwahl übereinstimmen. Der relative Fehler der Methoden wird in Abbildung 4.18 und die Servertrefferrate in Abbildung 4.19 dargestellt.

Die Verwendung von detaillierten Auslastungsdaten pro Server (AR , $AR\ w/o\ max$, Av) ermöglicht erwartungsgemäß deutlich bessere Abschätzungen als wenn die Abschätzungen nur auf der aggregierten RZ-Last basieren (Sp , $Sp\ w/o\ max$, Rdm). Die durchschnittliche Abweichung der abgeschätzten Serveranzahl bei der Aktivitätsrate AR liegt bei ca. 4%, während Abschätzungen anhand der Serverspezifikationen Sp ca. 12% abweichen. Noch deutlicher ist der Unterschied bei der Serverwahl. Hier treffen die Abschätzungen über AR mit 90% recht gut

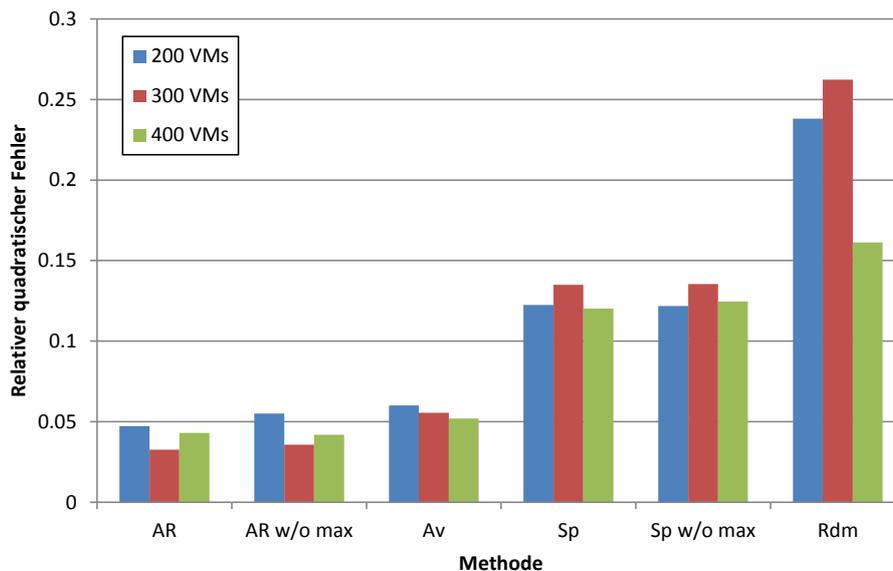


Abbildung 4.18: Durchschnittliche Differenz der abgeschätzten zur tatsächlichen Anzahl aktiver Server

zu, bei *Sp* hingegen entsprechen die Abschätzungen einer zufälligen Auswahl *Rdm*.

Wird die Bildung der Maxima vor der Anwendung weggelassen (*w/o max*), verschlechtern sich die Ergebnisse nur marginal (0,3% bei *AR* bzw. 0,1% bei *Sp*). Auf die letztliche Bewertung durch die LDEE-Metrik hat dies keinen signifikanten Einfluss, so dass dieser Schritt nicht unbedingt notwendig ist. Dies bedeutet gleichzeitig, dass das IT-Lastmodell flexibler auf andere LPM-Implementierungen angewandt werden kann, die andere oder gar keine Planungszeiträume nutzen.

Bestimmung der Auslastung

Die beiden in Abschnitt 4.3.5 beschriebenen Verfahren wurden anhand realer Auslastungszeitreihen, die im Rahmen des AC4DC-Projekts [SJS⁺14] im Produktivbetrieb erhoben wurden, mit alternativen Ansätzen verglichen. Die Referenzdaten repräsentieren die CPU-Auslastung von sieben ESX-Hosts mit je zwei Quad-Core-Prozessoren und liegen in einer zeitlichen Auflösung von einer Minute vor. Es wurde insgesamt ein Auszug von genau einem Monat für die Evaluation genutzt.

Jede untersuchte Methode wurde auf allen Servern angewandt, um für jeden Zeitpunkt der Evaluationsdaten eine konkrete Auslastung abzuschätzen. Diese wurde dann mit der tatsächlich aufgetretenen Auslastung verglichen und die absolute Abweichung zu dieser berechnet. Da das Ziel des IT-Lastmodells die Bestimmung der durchschnittlichen Auslastung für die wahrscheinlich genutzten Server ist, wird die durchschnittliche absolute Abweichung als Metrik für diese

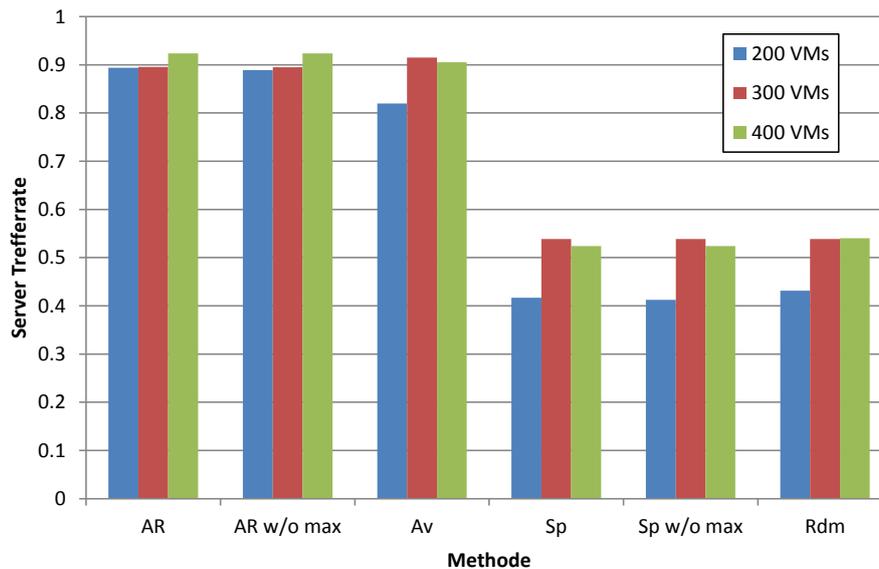


Abbildung 4.19: Anteil der durch die Abschätzungen korrekt bestimmten aktiven Server im Vergleich zu den tatsächlich genutzten

Evaluation angewandt.

Die Ergebnisse der Evaluation für alle untersuchten Methoden ist in Abbildung 4.20 dargestellt. Es wird jeweils die durchschnittliche absolute quadratische Abweichung der CPU-Auslastung, die einen Wertebereich von 0 bis 1 besitzt, dargestellt. Die blauen Balken sind dabei die direkten Ergebnisse der Verfahren wohingegen die roten Balken zusätzlich eine nachträgliche Interpolation repräsentieren, so dass die Summe aller Einzelauslastungen der vorgegebenen aggregierten Auslastung entspricht. Die im IT-Lastmodell integrierten Verfahren sind die lineare Regression für Auslastungsklassen (Cl+LR) und die Kapazitätsabhängigkeit (Cap). Zusätzlich wurden das harmonische (H) und das arithmetische (A) Mittel sowie der Modus für gruppierte Daten (Mode) auf Basis von Auslastungsklassen (Cl) untersucht. Des Weiteren wurden die lineare Regression (LR), das harmonische (H) und das arithmetische (A) Mittel auch auf die kompletten Serverauslastungen ohne vorherige Klassifizierung der Auslastungen angewandt.

Man kann gut erkennen, dass die Verfahren, die Abschätzungen für vordefinierte Auslastungsklassen vornehmen (Cl), die besten Ergebnisse erzielen. Eine vorherige Klassifikation der Auslastungsdaten macht also durchaus Sinn. Die unterschiedlichen Methoden zur konkreten Abschätzung der Auslastung unterscheiden sich allerdings nur geringfügig, so dass für das IT-Lastmodell die lineare Regression gewählt wurde, da diese keine weitere Interpolation benötigt. Das alternative Verfahren auf Basis der maximalen Kapazität (Cap), welches ins IT-Lastmodell integriert wurde, besitzt die größte Ungenauigkeit. Es ist jedoch auch das einzige Verfahren, welches keine detaillierten Auslastungsdaten

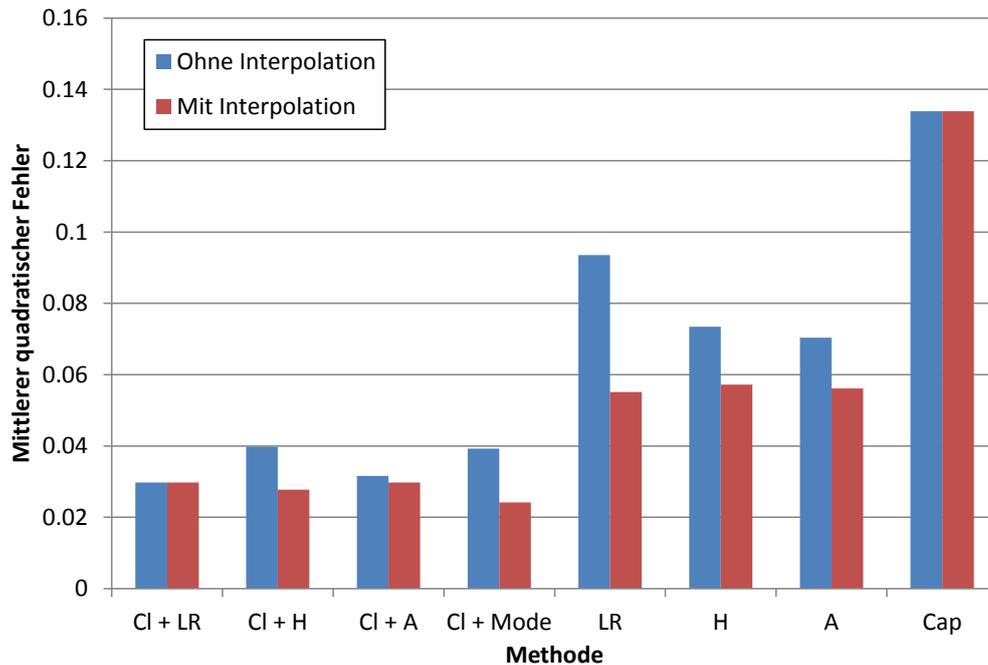


Abbildung 4.20: Evaluation unterschiedlicher Methoden zur Abschätzung der Auslastung eines Servers

pro IT-Komponente benötigt.

4.6.2 RZ-Performancemodell

Die in Abschnitt 4.4 vorgestellten Performancemodelle zu Server- und Storage-systemen basieren auf den meistgenutzten Performancebenchmarks. Ein direkter Vergleich dieser Modelle mit den meistgenutzten Verfahren zum Performancevergleich ist durch die somit vorhandenen Abhängigkeiten nicht aussagekräftig. Aber auch der Vergleich mit anderen Performancebenchmarks, die nicht für die Modelle berücksichtigt wurden, oder mit alternativen Methoden zur Bestimmung der Performance ist nicht sinnvoll. Da die Performance eines Systems nicht eindeutig ist, sondern je nach untersuchter Charakteristik interpretiert werden muss, kann nicht evaluiert werden, welches Verfahren besser geeignet ist.

Aus diesem Grund werden die Performancemodelle zu den Teilkomponenten Server, Storage und Netzwerk nicht für sich betrachtet evaluiert, sondern nur in Kombination als gesamtes RZ-Performancemodell betrachtet. Da bislang keine Konzepte zur Ermittlung der kompletten RZ-Performance verbreitet sind (vgl. Abschnitt 2.6.4), werden die Ergebnisse des Performancemodells mit bekannten Metriken, die die sinnvolle Arbeit bzw. Produktivität berücksichtigen, verglichen. Im Konkreten sind dies die Data Center Energy Productivity (DCeP) von The Green Grid und die Digital Service Efficiency (DSE), welche von Ebay für

ihre eigenen Systeme entwickelt wurde und angewandt wird. Die Evaluationen nehmen dabei Rückbezug auf die in Abschnitt 4.4.1 definierten Anforderungen.

Data Center Energy Productivity

Die DCeP von The Green Grid setzt sich nach [AMC⁺08] folgendermaßen zusammen:

$$\text{DCeP} = \frac{\text{Useful Work Produced}}{\text{Total Data Center Energy Consumed Producing this Work}} \quad (4.32)$$

Für die Evaluation des RZ-Performancemodells wird nur der Zähler, also die im Evaluationszeitraum vollbrachte sinnvolle Arbeit, herbeigezogen. Die Gegenüberstellung zur dafür aufgewandten Energie spielt an dieser Stelle keine Rolle, sondern wird erst für die Gesamtevaluation der LDEE in Abschnitt 6.2.4 betrachtet. Die sinnvolle Arbeit ist nach [AMC⁺08] wie folgt definiert:

$$\text{Useful Work Produced} = \sum_{i=1}^M V_i \cdot U_i(t, T) \cdot T_i \quad (4.33)$$

Die sinnvolle Arbeit ist demnach die Summe aller abgeschlossener Tasks T_i , deren Laufzeit jeweils durch die zeitabhängige Funktion $U_i(t, T)$ – mit den Parametern t für die zur Bearbeitung benötigte Zeit und T für den Zeitpunkt des Abschlusses – bestimmt wird. Der Normalisierungsfaktor V_i erlaubt eine numerische Summierung der Tasks und stellt gleichzeitig die Priorität und damit den Einfluss der Task auf das Gesamtergebnis dar.

Damit diese Metrik ein korrektes Ergebnis liefert, müssen alle im Rechenzentrum vorhandenen Typen von Tasks T_i vorab mit ihren jeweiligen Zeitfunktionen U_i und Normalisierungsfaktoren V_i definiert werden. Dies ist für jedes Rechenzentrum unterschiedlich und muss manuell vorgenommen werden, so dass die Bewertung letztlich subjektiv ist und die Ergebnisse nur innerhalb eines Rechenzentrums vergleichbar sind.

Da für die Evaluation nur Auslastungsdaten ohne Bezug zu konkreten Applikationen vorlagen, wurden mehrere Tasks manuell konstruiert. Die Summe der bearbeiteten Tasks wurde anschließend aus den simulierten Auslastungs- und Performancedaten berechnet. Als Ergebnis erhält man eine vereinfachte Darstellung der sinnvollen Arbeit als Anzahl der absolvierten und nach Priorität normierten Tasks pro Sekunde.

The Green Grid hat zudem mehrere Ersatzkennzahlen zur Ermittlung der sinnvollen Arbeit veröffentlicht, vgl. [HMP⁺09]. Eine dieser Metriken berechnet aus der durchschnittlichen Serverauslastung multipliziert mit der in SPECpower_ssj2008 maximal erreichten Performance eine Abschätzung für die Produktivität und verfolgt damit den gleichen Ansatz wie die LDEE. Diese Ersatzkennzahl wird ebenso im Rahmen der Evaluation mit dem vorgestellten RZ-Performancemodell verglichen.

Digital Service Efficiency

Die DSE-Metrik [Eba13] wurde von Ebay zur Kontrolle der Produktivität der eigenen Rechenzentren entwickelt und besitzt Analogien zur DCeP. Die DSE implementiert die Produktivität über zwei Schlüsselindikatoren der Plattform: Die Anzahl der zur Abwicklung von Käufen und Verkäufen generierten URL. Dies entspricht prinzipiell den manuell zu definierenden Tasks bei der DCeP. Aus $DSE = \frac{\text{Tasks Completed}}{\text{Energy Consumed for Task Completion}}$ wird dann im Konkreten:

$$DSE = \frac{\#Buy\ URL + \#Sell\ URL}{Energy_{DC}} \quad (4.34)$$

Die DSE liefert damit die Effizienz in der Einheit tasks/kWh. Für die Evaluation wurden für die beiden Indikatoren die Anzahl der erstellten URL auf Basis der durch Simulationen erzielten Auslastungs- und Performancedaten berechnet.

Ergebnisse

Mit Hilfe der Simulation des Last- und Powermanagements aus dem Forschungsprojekt AC4DC wurde ein Testszenario erstellt, auf dem unterschiedliche Workloads simuliert werden konnten. Es wurde für alle Durchläufe die gleiche Hardwareplattform angenommen. Diese bestand aus einem Rack mit 21 Servern gemischt aus vier Servertypen und einem Storage-System. Die darauf simulierten Workloads wurden je nach Durchlauf aus 100, 200 bzw. 300 VM-Profilen, die in einem Produktiv-Rechenzentrum aufgenommen wurden, generiert. Dabei wurde für jede VM-Anzahl zusätzlich zwischen zwei unterschiedlichen Volatilitäten (Schwankungen) der Auslastung unterschieden. Bei niedriger Volatilität mitteln sich die Einzelprofile in Summe gegenseitig aus, bei hoher Volatilität besitzen die VM ähnliche Profile und es kommt zu sehr hohen Schwankungen der Auslastung. In Abbildung 4.21 werden zwei entsprechende, aggregierte Auslastungsprofile dargestellt.

Insgesamt wurden somit sechs Simulationen jeweils über eine 14-tägige Simulationszeit durchgeführt. Für jede Simulation wurden die einzelnen Server- sowie Storageauslastungen gespeichert und ausgewertet. Neben der genutzten Performance des untersuchten RZ-Performancemodells – mit identischen Gewichten $w_{srv}^F = w_{sto}^F = 0,5$ für Server und Storage sowie einem NPC von 0,95 – wurden vergleichsweise die beiden Produktivitätsanteile der vorgestellten Metriken DCeP und DSE sowie die Ersatzkennzahl (DCeP-Proxy) basierend auf dem SPECpower-Benchmark berechnet. Zudem wurde jeweils die maximale Performance nach dem RZ-Performancemodell sowie die durchschnittliche Server- und Storageauslastung ermittelt. Die niedrige Auslastung der Serversysteme bei der Evaluation resultiert aus den Anforderungen des VM-Betriebs, dass die insgesamt verfügbare Ressourcenkapazität jederzeit ausreichen muss, auch wenn alle VM ihren maximalen Ressourcenbedarf zur gleichen Zeit benötigen. Die Ergebnisse der Simulationen werden in Tabelle 4.2 dargestellt.

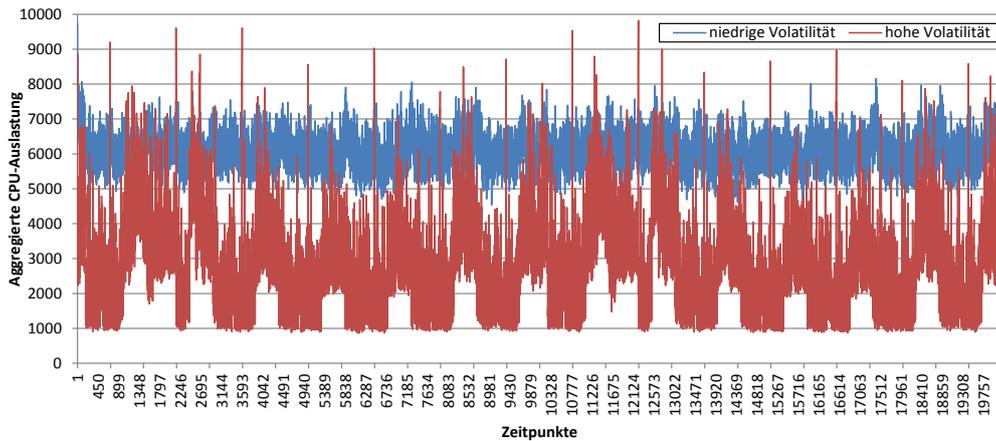


Abbildung 4.21: Vergleich der Schwankungen der simulierten CPU-Auslastungen für den Testfall mit 300 VM

VM-Profil	hohe Volatilität		
Anzahl VM	100	200	300
genutzte RZ-Perf. [ru]	0,419487955	0,836794363	1,252353494
DCeP [tasks]	13.651.200,00	26.222.400,00	40.003.200,00
DSE [tasks]	4.838.400,00	9.676.800,00	14.515.200,00
DCeP-Proxy [ssj_ops]	2,81772E+15	5,45166E+15	7,89689E+15
max. Serverperf.	32,18902688	32,18902688	32,18902688
max. Storageperf.	0,713350719	0,713350719	0,713350719
Serverauslastung	0,020318347	0,040492293	0,060668287
Storageauslastung	0,321166992	0,642411665	0,958407315

VM-Profil	niedrige Volatilität		
Anzahl VM	100	200	300
genutzte RZ-Perf. [ru]	0,72658576	1,785434542	2,482674977
DCeP [tasks]	23.198.400,00	57.888.000,00	80.481.600,00
DSE [tasks]	7.257.600,00	19.353.600,00	29.030.400,00
DCeP-Proxy [ssj_ops]	5,60458E+15	1,02115E+16	1,68264E+16
max. Serverperf.	32,18902688	32,18902688	32,18902688
max. Storageperf.	0,713350719	0,713350719	0,713350719
Serverauslastung	0,040403506	0,102536332	0,141135165
Storageauslastung	0,321166992	0,642411665	0,958407315

Tabelle 4.2: Performance und Produktivität für unterschiedliche Workloads

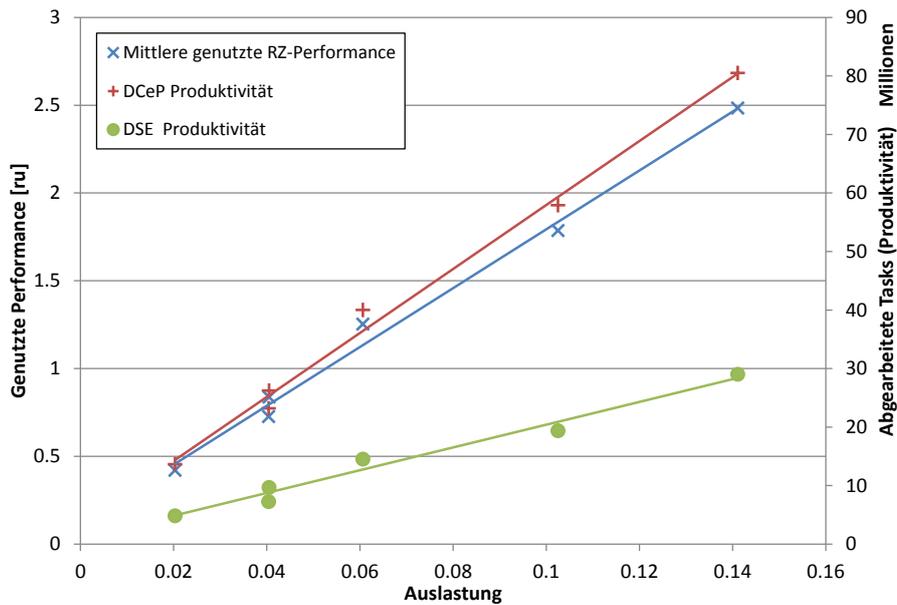


Abbildung 4.22: Abhängigkeit der genutzten Performance und der Produktivität von der Auslastung

Um die Produktivitätsmetriken anwenden zu können, wurde die simulativ ermittelte Auslastung auf der verwendeten IT-Hardware in abgeschlossene Tasks umgerechnet. Da die DSE speziell auf die Applikationsumgebungen von Ebay angepasst ist, wurde dieser Fall auch für die DCeP genutzt. Die abweichenden Werte ergeben sich dabei daraus, dass für die DCeP neben den generierten URL für Käufe und Verkäufe weitere Tasks definiert wurden, die in der DSE keine Berücksichtigung finden. Die definierten Tasks umfassen beispielhaft die Produktivität in den Bereichen Seitenaufrufe, Kleinanzeigen und Werbung.

Nachfolgend werden die Ergebnisse bezüglich der in Abschnitt 4.4.1 definierten Anforderungen diskutiert.

Auslastungsabhängigkeit (R 4.6) Anhand der Resultate kann man gut erkennen, dass die genutzte RZ-Performance der LDEE ebenso wie die Produktivitätskennzahlen DCeP und DSE von der Auslastung abhängen, so wie es die Anforderung R 4.6 für das Performancemodell verlangt. In Abbildung 4.22 wird dies nochmals verdeutlicht, indem jeweils lineare Approximationen berechnet und dargestellt wurden. Auch die Ersatzkennzahl der DCeP-Metrik (DCeP-Proxy) ist auslastungsabhängig, was bereits aus ihrer Definition hervorgeht.

Allerdings sind die Produktivitätsmetriken aufgrund der Erfassung abgeschlossener Tasks, die jeweils gewisse Laufzeiten besitzen, nur für einen ganzen zu betrachtenden Zeitraum definiert. Das RZ-Modell der genutzten Performance wird im Rahmen der LDEE hingegen für jeden einzelnen Zeitpunkt betrachtet.

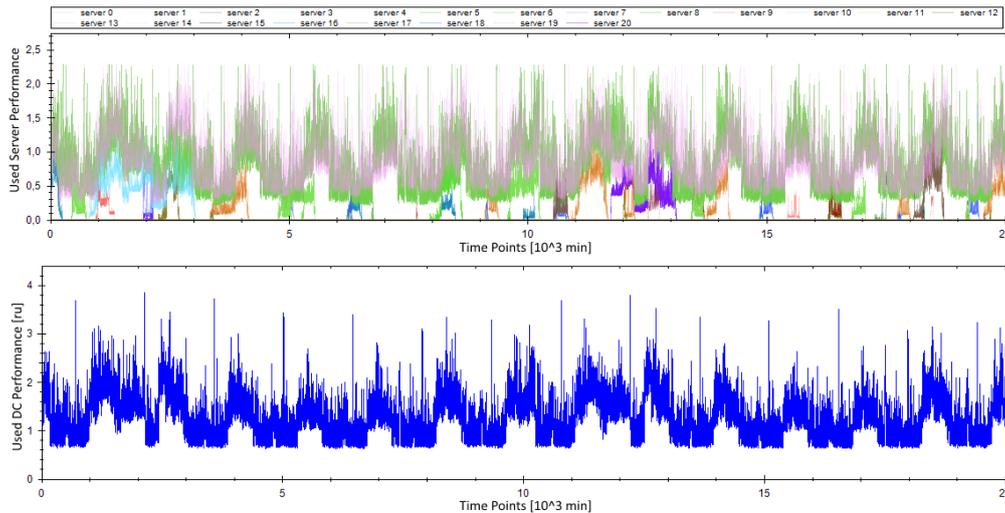


Abbildung 4.23: Genutzte Performance der Server (oben) und des gesamten Rechenzentrums (unten) inkl. Storage und Netzwerk jeweils für einen Simulationszeitraum von 14 Tagen

Entsprechend erhält man anstelle von Summen/Durchschnittswerten eine ganze Zeitreihe der genutzten Performance, wie sie beispielhaft in Abbildung 4.23 dargestellt wird. Für die Darstellung der Ergebnisse in Tabelle 4.2 wurde aus Gründen der Vergleichbarkeit der Durchschnitt dieser Zeitreihen verwendet.

Vergleichbarkeit (R 4.3) An den Resultaten in Tabelle 4.2 sieht man, dass die einzelnen untersuchten Metriken Werte in komplett anderen Größenordnungen liefern und damit nicht direkt vergleichbar sind. Auch die sehr ähnlich definierten Metriken DCeP und DSE liegen weit auseinander. Der Grund hierfür liegt in der manuellen und damit subjektiven Definition der Tasks eines Rechenzentrums. So wurde für die DSE beispielsweise nur eine Teilmenge der Tasks des betroffenen Rechenzentrums betrachtet, so dass die Anzahl abgearbeiteter Tasks hier geringer ist.

Neben dem ersten Testszenario (TS_1) des Ebay-Anwendungsfalls wurde auch ein zweites Szenario (TS_2) betrachtet, bei dem die Tasks aus Anfragen an ERP-, CRM- und Datenbanksysteme definiert wurden. Ansonsten wurden die gleichen Workloads auf der gleichen Hardware simuliert, es wurde lediglich angenommen, dass die Workloads in diesem Fall durch die definierten Anfragen erzeugt wurden. Der Vergleich zwischen beiden Testszenarien ist in Tabelle 4.3 dargestellt.

Die genutzte RZ-Performance ebenso wie der DCeP-Proxy betrachten nicht die Anwendung im Detail, sondern abstrahieren von der geleisteten Arbeit anhand der Auslastung und der Hardwareperformance. Die Resultate sind daher identisch für beide Testszenarien. Die DCeP hingegen fällt in TS_2 deutlich kleiner aus, was daran liegt, dass die definierten Tasks für sich betrachtet mehr Hardwa-

Testszenario Anzahl VM	TS_1		
	100	200	300
genutzte RZ-Perf. [ru]	0,419487955	0,836794363	1,252353494
DCeP [tasks]	13.651.200,00	26.222.400,00	40.003.200,00
DCeP-Proxy [ssj_ops]	2,81772E+15	5,45166E+15	7,89689E+15

Testszenario Anzahl VM	TS_2		
	100	200	300
genutzte RZ-Perf. [ru]	0,419487955	0,836794363	1,252353494
DCeP [tasks]	4.233.600,00	8.467.200,00	12.700.800,00
DCeP-Proxy [ssj_ops]	2,81772E+15	5,45166E+15	7,89689E+15

Tabelle 4.3: Performance und Produktivität für zwei Testszenarien TS_i

ressourcen benötigen und damit zu einer höheren Auslastung führen, so dass insgesamt weniger Tasks als in TS_1 ausgeführt werden. Damit die DCeP vergleichbar ist, müssten zu jeder Task die benötigten Hardwareressourcen betrachtet werden. Prinzipiell ist dies auch das Vorgehen des RZ-Performancemodells der LDEE-Metrik, mit dem Unterschied, dass von den Tasks abstrahiert und die gesamte Arbeit zusammengefasst wird. Der einzige Nachteil an diesem Vorgehen ist, dass nicht zwischen sinnvoller und unnötiger getätigter Arbeit bzw. den entsprechenden Tasks unterschieden werden kann.

Fazit Nur zwei der untersuchten Kennzahlen können die beiden Anforderungen für Vergleichbarkeit (R 4.3) und Auslastungsabhängigkeit (R 4.6) erfüllen. Diese sind das RZ-Performancemodell, das in der LDEE genutzt wird, und die Ersatzkennzahl der DCeP, welche den gleichen Ansatz der Kombination von Auslastung und Performance verfolgt. Bei den Produktivitätsmetriken verhindert die manuelle Definition der Tasks eine allgemeingültige Vergleichbarkeit – eine Abstraktion von den konkreten Anwendungen ist daher notwendig.

Diese Tatsache hat auch Einfluss auf die Anwendbarkeit (R 4.5) der Metriken. Für jeden definierten Task muss auch eine Messung implementiert werden, wodurch die Anwendbarkeit deutlich eingeschränkt wird. Das RZ-Performancemodell sowie die DCeP-Ersatzkennzahl nutzen hingegen vorab bekannte Benchmarkresultate und kombinieren diese mit den Auslastungen der Komponenten, die zumeist gemessen werden oder alternativ über das IT-Lastmodell ermittelt werden können. Eine Anwendung ohne in das Produktivsystem einzugreifen ist somit möglich.

In Anbetracht der Vollständigkeit (R 4.4) kann die DCeP-Ersatzkennzahl die Anforderungen nicht erfüllen. Sie betrachtet nur die Serverseite und dort mit dem Server-Side-Java-Workload des SPECpower-Benchmarks einen einzigen Workload. Das RZ-Performancemodell der LDEE hingegen berücksichtigt

auf Seiten der Server drei Benchmarks mit insgesamt mehr als zehn unterschiedlichen Workloads. Ebenso werden Storagesysteme mit zwei Benchmarks und die Netzwerkperformance über die Einhaltung der SLA bewertet. Die Produktivitätsmetriken DCeP und DSE erfüllen die Anforderung R 4.4 der Vollständigkeit, sofern alle relevanten Tasks berücksichtigt werden. Diese umfassen bereits alle wichtigen Workloads auf allen involvierten IT-Systemen.

Die Verständlichkeit (R 4.7) der Metriken ist sehr subjektiv. Die DCeP, ihre Ersatzkennzahl und die DSE zählen die Anzahl der abgeschlossenen Tasks. Dies ist für Anwender zunächst leicht begreiflich, die Größenordnung der Resultate im sieben- bis 15-stelligen Bereich allerdings auch unübersichtlich. Eine geeignete Normierung könnte hier die Lösung sein. Genau diese ist bereits in das Konzept der LDEE-Performancemodelle integriert. Ein Resultat in genutzten Referenzmaschinen (r_u) muss zwar zunächst erläutert werden, unter Angabe der durchschnittlichen Referenzmaschine ist sie jedoch schnell zu verstehen. Wichtig ist vor allem der Hinweis auf die genutzten Kapazitäten gegenüber den insgesamt vorhandenen.

Zusammengefasst kann nur das RZ-Performancemodell der LDEE jede der genannten Anforderungen erfüllen.

4.6.3 RZ-Powermodell

Zur Evaluation des RZ-Powermodells wurde ein Szenario basierend auf realen Messungen an einem Containerrechenzentrum aus dem AC4DC-Projekt [SJS⁺14] erstellt. Zur Kälteverteilung im Container wurden sogenannte Zero-U-Space-Cooling-Systems (ZUCS) – direkt im Doppelboden unter den Racks verbauten Wärmetauscher mit Lüftern – eingesetzt. Die Kälteerzeugung übernahm ein Chiller mit kombinierter Freiluftkühlung, wobei während der Messungen nur die Freikühleinheit aktiv war. Es wurden keine realen IT-Komponenten verbaut, sondern die Wärmelast über Heizlüfter erzeugt. Die Kühlkette wurde dann für simulierte IT-Auslastungen zwischen 10kW und 60kW und für unterschiedliche Wasservorlauftemperaturen vermessen.

Die sechs ZUCS-Geräte wurden mit dem in Gleichung 4.24 beschriebenen linearen Modell abgeschätzt. Die durch die Heizlüfter simulierte Wärmelast der IT war der Eingabeparameter des Modells, welches eine durchschnittliche relative Abweichung von 4,8% von den Messdaten besaß.

Für die Freikühlung wurde das Berechnungsmodell aus Gleichung 4.29 angewandt. Aus dem Wärmestrom, der den aggregierten Leistungsaufnahmen der IT und der Kühlgeräte im IT-Raum entspricht, wurde der vorliegende Volumenstrom berechnet. Anhand des Volumenstroms, den gemessenen Temperaturen an den Klimageräten und den Kennlinien der Lüfter bzw. Pumpen konnte die Leistungsaufnahme der Geräte berechnet werden. Da allerdings keine Daten zum vorliegenden Druck und zur Effizienz der Geräte verfügbar waren, mussten hierfür Annahmen getroffen werden. Der Fehler ist daher mit 18,7% vergleichsweise hoch.

Insgesamt betrug die Abweichung der Abschätzung von den Messwerten 11,7%.

Anmerkung zum Powermodell der Klimatisierung

Das Evaluationsszenario zum Container-RZ hat gezeigt, dass die Abschätzung der Leistungsaufnahme der Klimakette stark von realen Messungen im Rechenzentrum abhängt. Im Gegensatz zu Server- und Stagesystemen gibt es keine Datenbanken mit ausgiebigen Performance/Power-Werten und auch die Hersteller von Klimakomponenten veröffentlichen nur sehr allgemeine Daten zu ihren Produkten – zumeist nur unzureichende maximale Betriebsparameter. Werden die benötigten Parameter nicht gemessen, müssen im schlechtesten Fall Annahmen für diese getroffen werden. Um die Notwendigkeit von detaillierten Messungen zu umgehen sind zwei mögliche Vorgehensweisen denkbar. Zum einen könnten die Hersteller detaillierte Parameter zu den unterschiedlichen potentiellen Betriebszuständen zur Verfügung stellen. Auf deren Basis könnten korrespondierende Powermodelle charakterisiert werden, von denen sich konkrete Analysen und Aussagen während des Betriebs als auch bereits vor einer Installation der entsprechenden Geräte ableiten ließen. Zum anderen wäre ein Verfahren zur Abschätzung der Leistungsaufnahme, welches auf allgemeine, aggregierte Messdaten anwendbar ist und als Ersatzkennzahl für die in diesem Kapitel vorgestellten detaillierten Infrastrukturpowermodelle dienen kann, sinnvoll. Eine solche Ersatzmetrik wird im folgenden Kapitel beschrieben.

5 Definition der IPA Metrik

Die Leistungsaufnahme des Rechenzentrums ist ein maßgeblicher Faktor für die LDEE. Auch wenn diese nicht im Detail oder nur unvollständig gemessen wird, soll die LDEE eine realistische Abschätzung in Abhängigkeit zur Auslastung liefern. Die Powermodelle der IT-Komponenten liefern auch in einem solchen Fall Abschätzungen, da öffentlich verfügbare Benchmarkresultate genutzt werden. Für die Infrakstrukturkomponenten ist dies allerdings nicht möglich, so dass in diesem Fall auf eine Ersatzmetrik basierend auf PUE-Resultaten zurückgegriffen wird. Diese wird in Abschnitt 5.1 motiviert und ihre beiden Teilmetriken Power Variability sowie Infrastructure Power Adaptability in den Abschnitten 5.2 sowie 5.3 vorgestellt. Abschnitt 5.4 beschreibt die Anwendung als Ersatzkennzahl. Zum Abschluss des Kapitels werden in Abschnitt 5.5 die Evaluationen beschrieben. In [SSN14] finden sich weitere Details zu den entwickelten Metriken.

5.1 Motivation

Die in Abschnitt 4.5 vorgestellten Powermodelle für Klimageräte sind nur genau, wenn detaillierte Betriebsdaten, seien es direkte Messdaten oder Charakteristika aus Datenblättern, verfügbar sind. Für fehlende Parameter können zwar Annahmen getroffen werden, allerdings erfordert dies einerseits einen manuellen Eingriff, wodurch die Anforderung für die Objektivität (R 3.4) der LDEE nicht mehr eingehalten werden kann. Andererseits entsteht dadurch auch ein wesentlicher Modellfehler, der bei ungewissen Annahmen nur noch weiter steigt. Sollte also keine ausreichende Datenbasis für die Anwendung der Powermodelle vorhanden sein, ist es sinnvoller anstelle von Annahmen oder Inter-/Extrapolationen nur auf die tatsächlich gemessenen Daten zurückzugreifen, auch wenn diese stark aggregiert sein sollten.

Prinzipiell kann anstelle von detaillierten Komponentenmodellen auch die Power Usage Effectiveness eines Rechenzentrums für eine allgemeine Bewertung der Infrastruktur genutzt werden. Die PUE hat sich in den letzten Jahren als wichtigste Energieeffizienzkennzahl etabliert und wird daher in den meisten Rechenzentren angewandt. Die PUE alleine besitzt allerdings keine Aussagekraft über die Lastadaptivität der Infrastruktur. Eine Nutzung der PUE für die LDEE hilft zwar bei der durchschnittlichen Bewertung der LDEE für den Messzeitraum des PUE-Resultats, allerdings kann sie nicht dafür genutzt werden die Energieeffizienz für unterschiedliche RZ-Lasten zu bestimmen. Da dies eine maßgebliche Anforderung (R 3.2) der LDEE ist, wurde im Rahmen dieser Arbeit eine zu-

sätzliche Kennzahl geschaffen, die diese Anforderung trotz eingeschränkter bzw. aggregierter Messdaten erfüllt.

Die zu diesem Zweck entwickelte und als Erweiterung der PUE konzipierte Kennzahl ist die Infrastructure Power Adaptability (IPA). Während die PUE den relativen statischen Overhead der Infrastruktur angibt, repräsentiert die IPA die Anpassungsfähigkeit der Infrastruktur hinsichtlich sich verändernder IT-Lasten. Dazu nutzt die IPA mit der Power Variability eine Hilfsmetrik, welche die Veränderlichkeit eines Verlaufs der Leistungsaufnahme quantifiziert. Durch Kombination der PUE mit der IPA kann für eine gegebene RZ-Auslastung sowie Leistungsaufnahme der IT, welche sich durch die IT-Powermodelle ergibt, die auslastungsabhängige Leistungsaufnahme der Infrastruktur ermittelt werden. Während aus der PUE der durchschnittliche absolute Overhead der Infrastruktur berechnet wird, wird die IPA dazu genutzt, den durchschnittlichen Overhead an die gegebene Auslastung anzupassen. Damit können PUE und IPA als Ersatzkennzahlen im Falle fehlender oder nicht charakterisierbarer Powermodelle für Infrastrukturkomponenten im Rahmen der LDEE angewandt und die Anforderung für Auslastungsabhängigkeit (R 3.2) weiterhin erfüllt werden.

5.2 Power Variability

Die Power Variability (PVar) ist eine Metrik zur Veranschaulichung der Veränderlichkeit von Leistungsaufnahmen in einer Zeitreihe. Sie dient als Teilmetrik der IPA: Indem sie jeweils auf die Kurven der IT- und Infrastrukturleistungsaufnahmen angewandt wird, können die Variabilitäten direkt in Relation zueinander gesetzt werden um Aussagen über die Adaptivität der Infrastruktur zur IT treffen zu können. Die Variabilität wird für einen untersuchten Zeitraum zwischen 0 und 1 ohne Einheit angegeben, wobei 0 für keine Veränderungen und 1 für fortlaufende extreme Veränderungen in der Zeitreihe steht.

Zur Bestimmung der Variabilität wurde ein eigens definierter Ansatz gewählt, da bekannte statistische Verfahren nicht ideal für den beschriebenen Einsatzzweck sind. So beschreiben die Standardabweichung als auch der Variationskoeffizient zwar die Streuung der Einzelwerte um einen Erwartungs- bzw. Mittelwert, allerdings differenzieren diese Maße nicht zwischen Veränderungen zwischen direkt aufeinanderfolgenden und weit auseinanderliegenden Zeitpunkten. Zur Beurteilung der Adaptivität ist jedoch gerade die Veränderung an aufeinanderfolgenden Zeitpunkten oder Zeitfenstern relevant. Dieser Aspekt einer Zeitreihe kann mit Hilfe der Volatilität, welche die Schwankung innerhalb der Zeitreihe darstellt, untersucht werden. Die Volatilität ist die Standardabweichung der absoluten, relativen oder logarithmischen Differenzen zwischen gewählten Zeitpunkten. Es werden allerdings nur Differenzen zwischen vordefinierten Zeitpunkten betrachtet, d.h. Veränderungen, die zwischen diesen Zeitpunkten auftreten, werden nicht berücksichtigt. Je nach Abstand der Zeitpunkte sind die erkannten Differenzen recht zufällig. Für die PVar werden daher nicht nur einzelne Zeit-

punkte, sondern die maximale Differenz in ganzen Zeitfenstern untersucht. Des Weiteren ist das Ergebnis der Volatilität nicht auf einen Wertebereich (z.B. zwischen 0 und 1) begrenzt, sondern nach oben hin offen, was sich negativ auf die Vergleichbarkeit und Verständlichkeit auswirkt.

5.2.1 Konzeption

Bevor die PVar auf einer Zeitreihe angewandt wird, muss die Auflösung der Daten angepasst werden. Optimal ist eine Auflösung von ca. fünf Minuten, da kleinere Spitzen somit bereits herausgemittelt sind. Liegen die Daten in einer höheren (feingranulareren) Auflösung vor, sollten diese zuvor entsprechend durch Mittelwertbildung heruntergerechnet werden. Da es eine gewisse Zeit dauert, bis sich eine steigende oder fallende IT-Last spürbar auf die Abwärme auswirkt, macht es keinen Sinn Lastspitzen, die nur für wenige Minuten auftreten, für die Variabilität zu berücksichtigen.

Nach der Vorbereitung der Zeitreihe wird in einem zweiten Schritt die Zeitreihe in Abschnitte segmentiert. Jeder Abschnitt umfasst dabei $t + 1$ Zeitpunkte und überlappt sich mit jeweils genau einem Zeitpunkt mit den beiden benachbarten Segmenten. Für eine Zeitreihe mit n Zeitpunkten erhält man somit n/t Segmente. Ein Beispiel hierfür ist in Abbildung 5.1 dargestellt. Für jedes der Segmente i wird dann die relative Spannweite RR_i (relative range) berechnet. Diese ergibt sich aus der absoluten Spannweite $R_i = x_{max}^i - x_{min}^i$ (range) aller Werte des Zeitfensters i dividiert durch die größte Ausprägung x_{max}^i des Zeitfensters.

$$RR_i = \frac{R_i}{x_{max}^i} = \frac{x_{max}^i - x_{min}^i}{x_{max}^i} \quad (5.1)$$

Ein einzelner Repräsentant für die Power Variability $PVar$ berechnet sich durch eine Mittelwertbildung über die relativen Spannweiten RR_i aller Segmente i , wie in Gleichung 5.2 dargestellt. Durch die Ermittlung von relativen, normierten Werten für jedes Segment ist auch das letztliche Resultat der PVar auf den Wertebereich zwischen 0 und 1 normiert. Die PVar gibt somit den durchschnittlichen Anteil der Variation der Ausprägungen in Relation zur maximalen Ausprägung pro Zeitraum an.

$$PVar = \frac{1}{n} \sum_{i=1}^n RR_i \quad (5.2)$$

Einerseits werden durch die Mittelung der Zeitreihe auf eine Auflösung von ca. fünf Minuten kurze Lastspitzen, die sich nicht spürbar auf die klimatischen Bedingungen und damit auf den Betrieb der Kühlinfrastruktur auswirken, herausgefiltert. Andererseits sorgt die anschließende Segmentierung in Zeitfenster dafür, dass höhere und längerfristige Variationen, für welche die Adaptivität der Infrastruktur untersucht werden soll, entsprechend dargestellt werden. Würde die Variation direkt in fünf-Minuten-Schritten gemessen, wäre die relative

5 Definition der IPA Metrik

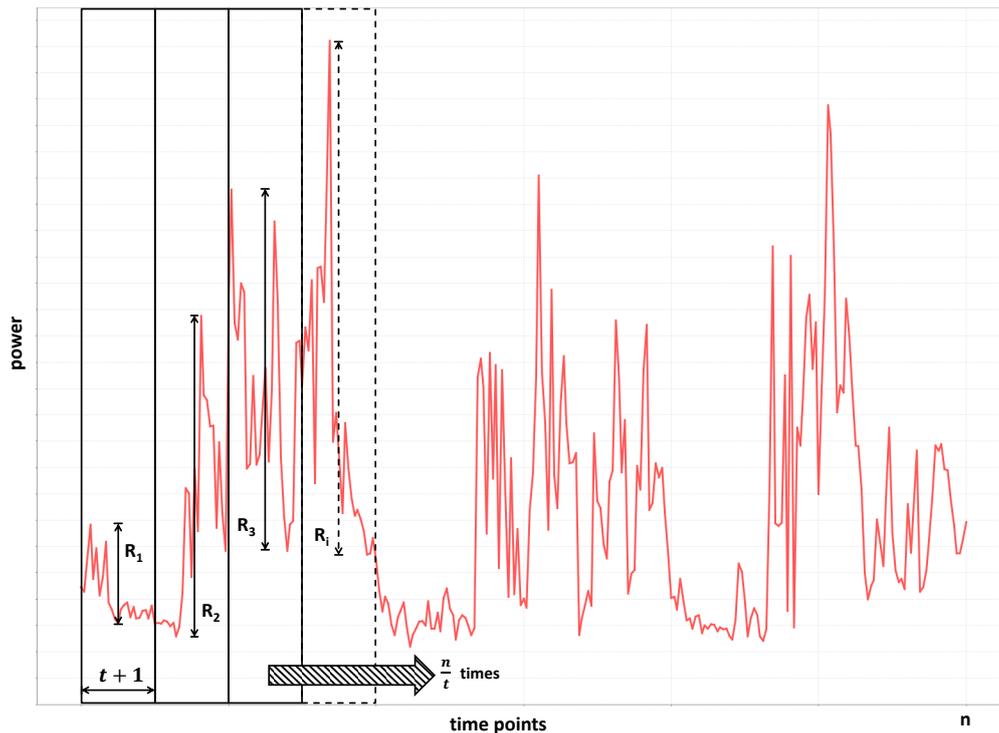


Abbildung 5.1: Unterteilung der Zeitreihe in n/t Segmente für die Anwendung der Power Variability.

Spannweite zumeist sehr gering, da größere Änderungen in der IT-Last (z.B. Arbeitsbeginn/Mittagszeit oder Wartungsfenster) in der Regel einen deutlich längeren Zeitraum benötigen. Als Größe des Zeitfensters zur Segmentierung sollten daher mindestens 30 Minuten gewählt werden.

5.2.2 Beispiel

Zur Veranschaulichung der Power Variability wurde die Metrik auf realen Messdaten eines Testrechenzentrums angewandt. Die Leistungsaufnahme der IT und Infrastruktur eines Tages sowie die resultierende PUE sind in Abbildung 5.2 dargestellt. Durch die indirekte Kühlung per Außenluft über einen Luft-Luft-Wärmetauscher ist der Overhead der Infrastruktur- gegenüber der IT-Leistungsaufnahme und damit auch die PUE sehr gering.

Die Power Variability der IT-Leistungsaufnahme ist mit 0,13 moderat während die der Infrastruktur mit 0,41 mehr als drei mal so hoch ist. Betrachtet man die Darstellung der Kurven in Abbildung 5.2, mögen diese Werte irreführend sein, da die Variation der IT-Leistungsaufnahme deutlich stärker ins Auge fällt. Die PVar betrachtet allerdings die relative Variation und nicht die absolute, so dass die Schwankungen der Infrastruktur deutlich höher einzuordnen sind

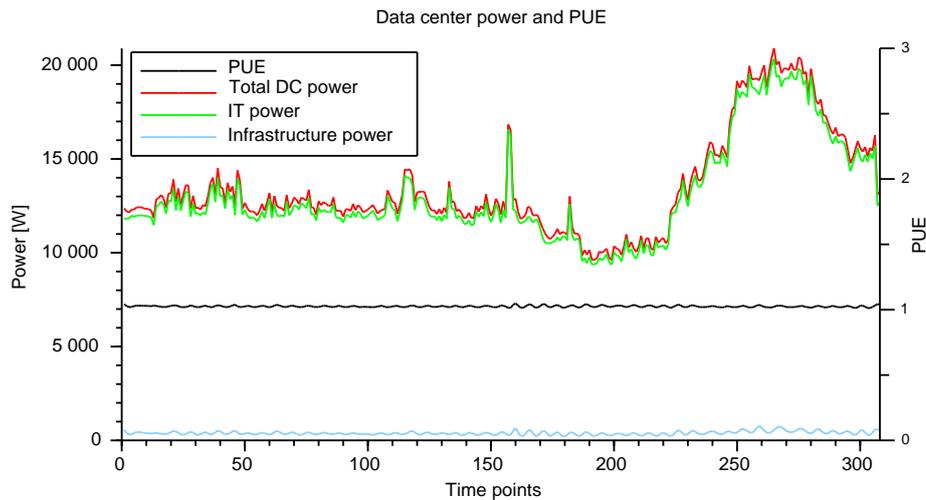


Abbildung 5.2: Auszug der Leistungsaufnahmen für IT und Infrastruktur eines Testrechenzentrums mit einem Luft-Luft-Wärmetauscher. Ein Zeitpunkt entspricht 5 Minuten.

Kennzahl	IT	Infrastruktur	Relation
Power Variability	0,13	0,41	0,32
Standardabweichung	2607,4W	95,5W	27,3
Variationskoeffizient	0,20	0,25	0,81
Volatilität abs. Differenzen	1433,9W	114,1W	12,6
Volatilität rel. Differenzen	0,11	0,29	0,37

Tabelle 5.1: Variabilität, Schwankung und Streuung der Leistungsaufnahmen aus Abbildung 5.2

als die der IT. Auch die Regelmäßigkeit der Schwankungen bei der Infrastruktur ist in der Abbildung gut zu erkennen und wird durch die Betrachtung von Zeitfenstern in der PVar berücksichtigt. Zum direkten Vergleich wurden neben der PVar auch die Standardabweichung, der Variationskoeffizient und die Volatilität angewandt. Der Variationskoeffizient ist als relatives Maß der Streuung über die Division der Standardabweichung durch den Erwartungswert definiert und stellt somit die relative Standardabweichung dar. Die Volatilität bildet hingegen die Schwankungen zwischen definierten Zeitabständen (z.B. 60 Minuten) ab und berechnet sich aus der Standardabweichung der Differenzen zwischen den entsprechenden Zeitpunkten. Hierbei wurde zwischen absoluten und relativen Differenzen unterschieden. Die jeweiligen Ergebnisse sind in Tabelle 5.1 dargestellt.

Anhand der Ergebnisse ist ersichtlich, dass die Kennzahlen Standardabwei-

chung und Volatilität der absoluten Differenzen nicht für den Einsatzzweck der IPA anwendbar sind, da bei großen Unterschieden im Wertebereich die ebenso in Tabelle 5.1 angegebene Relation der Variabilität verfälscht wird. Die PVar, der Variationskoeffizient sowie die Volatilität relativer Differenzen setzen die Abweichungen jeweils in Relation zur Höhe der Ausprägung, so dass unterschiedliche Zeitreihen miteinander vergleichbar werden. Aber auch zwischen den drei relativen Kennzahlen gibt es Unterschiede. Nach dem Variationskoeffizienten besitzen die IT- und die Infrastrukturleistungsaufnahmen eine ähnliche Variation. Dies liegt daran, dass der Koeffizient die Differenz jedes Einzelwerts zum Erwartungswert bzw. in diesem Fall zum Mittelwert berücksichtigt. Für die IPA ist hingegen die Differenz zwischen aufeinanderfolgenden Zeitpunkten oder -fenstern relevant. Diese wird von der PVar und der Volatilität relativer Differenzen berücksichtigt, weshalb sich bei diesen beiden Kennzahlen für die IT-Leistungsaufnahme eine geringere Variabilität ergibt.

Dem Beispiel folgend könnte auch die Volatilität relativer Differenzen anstelle der PVar für die IPA genutzt werden. Allerdings tritt bei der Volatilität eine gewisse Inkonsistenz dadurch auf, dass nur ausgewählte Zeitpunkte und keine Zeitfenster betrachtet werden. Dies lässt sich auch am obigen Beispiel zeigen. Führt man die gleiche Berechnung für ein Zeitfenster bzw. – im Falle der Volatilität – Abstand der Differenzbildung von 65 anstatt 60 Minuten durch (13 statt 12 Zeitpunkte), kommt für die Volatilität ein deutlich anderes Ergebnis heraus. Während die relative Volatilität der IT-Leistungsaufnahme geringfügig auf 0,15 steigt, fällt sie für die Infrastruktur deutlich von 0,29 auf 0,17. Der große Unterschied liegt daran, dass bei der zweiten Berechnung viele der Variationen genau zwischen zwei gewählten Zeitpunkten liegen und damit nicht berücksichtigt werden. Die PVar hingegen liefert für das veränderte Zeitfenster mit 0,15 für die IT und 0,41 für die Infrastruktur mit einer Relation von 0,36 fast die gleichen Ergebnisse. Der Ansatz der PVar durch die Betrachtung der maximalen Differenz ganzer Zeitfenster ist somit konsistenter.

5.3 Infrastructure Power Adaptability

Die Infrastructure Power Adaptability (IPA) bewertet die Anpassungsfähigkeit der Infrastruktur an Änderungen in der IT-Leistungsaufnahme. Variationen in der Leistungsaufnahme treten unter anderem durch Nutzungsschwankungen aufgrund von Tagesabläufen (Mittagspause, Wochenende, ...) oder durch geplante Auszeiten für Wartungszwecke auf. Sinkt die IT-Leistungsaufnahme, muss weniger Abwärme aus den IT-Räumen abgeführt werden, so dass die Leistungsaufnahme der Infrastruktur ebenso sinken sollte. Ist dies nicht der Fall, kühlt die Klimatisierung mehr als eigentlich notwendig, so dass elektrische Energie verschwendet wird. Die IPA stellt diesen Zusammenhang in einer repräsentativen Kennzahl dar, um RZ-Betreibern neben der PUE ein weiteres Hilfsmittel zur Erreichung und Kontrolle eines energieeffizienten Betriebs zur Verfügung zu stellen.

Dabei ist die IPA für jedes Rechenzentrum anwendbar, welches kontinuierliche PUE-Messungen durchführt.

Die IPA bildet die Adaptivität als Prozentzahl inkl. Verweis auf die Variabilität der IT-Leistungsaufnahme in der Form $IPA [\%] @ PVar(IT)$ ab. So bedeutet beispielsweise ein Resultat von $63\% @ 0,24$, dass die Infrastruktur des untersuchten Rechenzentrums zu 63% mit Schwankungen in der IT-Leistung skaliert, wobei die IT-Leistungsaufnahme selbst um 24% variiert. Die moderaten Schwankungen in der IT-Leistungsaufnahme werden also zu einem großen Anteil von der Infrastruktur nachgezogen. Eine Adaptivität von 100% ist dabei gleichbedeutend mit einer idealen Anpassung der Infrastruktur bei Änderungen in der IT-Leistungsaufnahme. 0% hingegen bedeutet, dass sich die Infrastruktur überhaupt nicht anpasst. Werte dazwischen geben den entsprechenden Prozentsatz an, wie gut die Infrastruktur mit der IT skaliert. Es sind auch Resultate über 100% möglich, wenn die Infrastrukturleistungsaufnahme relativ betrachtet mehr skaliert als die der IT. Ist die Adaptivität dabei deutlich über 100%, ist dies ebenso wie sehr niedrige Werte ein Anzeichen für Ineffizienzen in der Klimatisierung.

Da die IPA keine Informationen über die vorliegende Variabilität liefert, sollte die PVar der IT für eine bessere Einordnung des Resultats mit angegeben werden. Ist diese nicht bekannt, ist die resultierende IPA nicht eindeutig. Wenn beispielsweise die IT- und Infrastrukturleistungsaufnahmen überhaupt nicht variieren, wird die IPA trotzdem einen sehr guten Wert liefern, da sich die Infrastruktur schließlich entsprechend der IT verhalten hat. Wenn keine Varianz vorliegt, kann diese auch nicht beurteilt werden. Um diesem Sachverhalt Rechnung zu tragen, sollte die IPA immer in Kombination mit der PVar der IT angegeben werden. Somit ist auf den ersten Blick erkennbar, wie aussagekräftig das Ergebnis ist.

5.3.1 Konzeption

Zur Ermittlung der Adaptivität setzt die IPA die Leistungsaufnahmen der IT und der Infrastruktur in Relation. Es müssen dabei Zeitreihen mit kontinuierlichen Messungen vorliegen, damit ihre jeweiligen Eigenschaften und Verhaltensweisen berücksichtigt werden können. Diese Voraussetzung ist für viele Rechenzentren gegeben, da sie zum Beispiel bei einer kontinuierlichen PUE-Ermittlung notwendig sind.

Die IPA betrachtet die Variabilität beider Zeitreihen der Leistungsaufnahmen, indem jeweils die im vorherigen Abschnitt beschriebene PVar ermittelt wird. Zudem berücksichtigt die IPA die Ausprägung und den Umfang dieser aufgetretenen Variationen, um ein vollständiges Abbild der Verhaltensweisen der Zeitreihen zu erhalten. Zu diesem Zweck werden die Abweichungen für jede Zeitreihe ausgehend von einer Basislinie, welche dem am häufigsten aufgetretenen, energetischen Zustand entspricht, berechnet. Dieser Zustand lässt sich mit Hilfe des Modus berechnen, wobei in diesem Fall der Modus für gruppierte Daten genutzt wird. Da die Leistungsaufnahme in den Zeitreihen als reelle Zahl angegeben ist und die

5 Definition der IPA Metrik

Häufigkeit des Auftretens nur über genaue Übereinstimmungen der Leistungsaufnahmen bestimmt wird, ist eine direkte Betrachtung der Zeitreihe nicht sinnvoll. Anstelle dessen wird zunächst eine Diskretisierung der Leistungsaufnahmen in Klassen durchgeführt und die Häufigkeit der Klassen ermittelt. Als Klassengröße wird jeweils 5% der maximalen Leistungsaufnahme genutzt, so dass jede Zeitreihe in 20 Klassen eingeteilt wird. Anschließend wird über Gleichung 5.3 der Modus für gruppierte Daten *mode* bestimmt, welcher in der Modalklasse m , der Klasse mit der höchsten Häufigkeit, liegt. Die gewählte Klassengröße stellt dabei einen guten Kompromiss aus berücksichtigter Schwankung der Basislinie und Genauigkeit des Modus dar. Bei kleineren Klassengrößen ist die Wahrscheinlichkeit größer, dass sich die Basislinie aufgrund von Schwankungen auf mehrere Klassen aufteilt und die Modalklasse ggf. nicht mehr eindeutig ist. Bei größeren Klassen sinkt hingegen die Genauigkeit der Abbildung der Basislinie durch den Modus, da die Modalklasse einen größeren Wertebereich beschreibt.

$$mode = L_m + h_m \frac{f_m - f_{m-1}}{2f_m - f_{m-1} - f_{m+1}} \quad (5.3)$$

L_i ist die untere Grenze der Klasse i , d.h. die niedrigste in i liegende Leistungsaufnahme. f_i entspricht der Häufigkeit und h_i dem Intervall bzw. Wertebereich der Klasse i . Die Klassen $m - 1$ und $m + 1$ beziehen sich auf die Klasse vor bzw. nach der Modalklasse m . Der so berechnete Modus wird nun als Basislinie der Zeitreihe angenommen, von der die durchschnittliche relative Abweichung (RAD, relative average deviation) bestimmt wird. Die Berechnung der RAD wird in Gleichung 5.4 mit den konkreten Ausprägungen x_i und der Länge n der untersuchten Zeitreihe dargestellt.

$$RAD = \frac{\sum_{i=1}^n |x_i - mode|}{n \cdot mode} \quad (5.4)$$

Die IPA kombiniert nun die Variabilität (PVar) mit der durchschnittlichen Abweichung vom typischen Betrieb (RAD) für die IT- als auch die Infrastrukturleistungsaufnahme jeweils durch eine Multiplikation, so dass für jede Zeitreihe ein kombinierter Repräsentant als Verhaltensbeschreibung resultiert. Die beiden Repräsentanten werden dann – wie in Gleichung 5.5 dargestellt – in Relation zueinander gesetzt. Man erhält einen Faktor, der den Unterschied der Zeitreihen bzgl. Variabilität und Abweichungen beschreibt.

$$IPA = \sqrt{\frac{PVar_{inf} \cdot RAD_{inf}}{PVar_{it} \cdot RAD_{it}}} \cdot 100[\%] \quad (5.5)$$

Da mit der PVar und der RAD zwei relative Werte miteinander multipliziert wurden, wird die Quadratwurzel des Resultats gezogen um wieder einen normierten Wert zu erhalten, der dann letztlich in einen Prozentsatz umgerechnet wird.

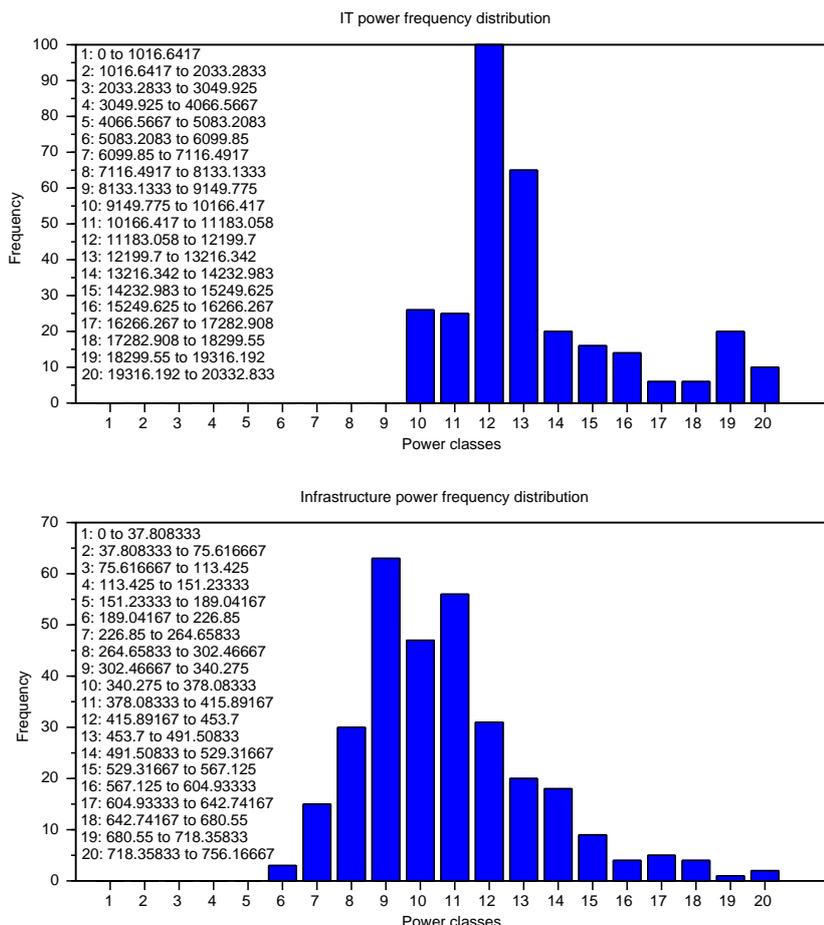


Abbildung 5.3: Histogramm der Häufigkeit jeder Klasse für die Leistungsaufnahmen der IT (oben) und der Infrastruktur (unten). Die Klassengrenzen werden jeweils auf der linken Seite dargestellt.

5.3.2 Beispiel

Wendet man die IPA auf das Anwendungsbeispiel für die PVar in Abschnitt 5.2.2 an, erhält man eine IPA von 223,7% bei einer $PVar(IT)$ von 0,13. Die Klasseneinteilung sowie die Häufigkeiten jeder Klasse für die Zeitreihen der IT- als auch Infrastrukturleistungsaufnahme werden in Abbildung 5.3 dargestellt.

Die größte Häufigkeit bei der IT besitzt die Klasse 12 – die Basislinie liegt damit zwischen 11,2 kW und 12,2 kW, vgl. Abbildung 5.2. Bei der Infrastruktur ist dies die Klasse 9 bzw. 302 W bis 340 W. Von diesen Basislinien wurde die relative durchschnittliche Abweichung RAD berechnet, welche bei der IT 0,16 und bei der Infrastruktur 0,25 beträgt. Bzgl. der Abweichungen ist die Diskrepanz zwischen IT und Infrastruktur demnach nicht so groß wie bei der PVar (Faktor

5 Definition der IPA Metrik

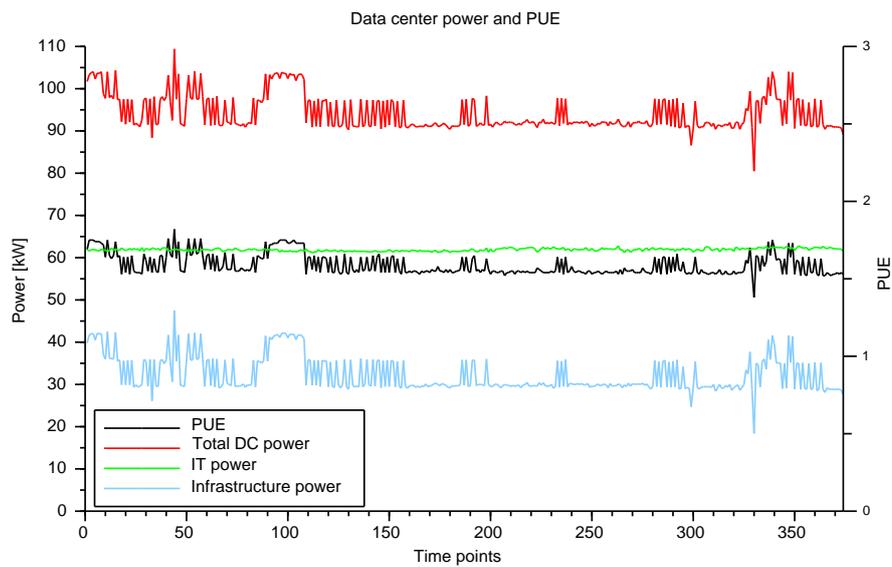


Abbildung 5.4: Auszug der Leistungsaufnahmen eines Produktivrechenzentrums

0,62 anstatt 0,32). Die resultierende IPA von 223,7% deutet darauf hin, dass bei der Infrastruktur durch die viel zu starke Skalierung Ineffizienzen vorliegen. Nach Rücksprache mit den Betreibern des Testrechenzentrums wurde bestätigt, dass der Regler für die dem Luft-Luft-Wärmetauscher nachgeschalteten Ventilatoren noch nicht optimal eingestellt ist und zu stark um die Sollwerte schwankt.

Ein weiteres Anwendungsbeispiel – in diesem Fall eines Produktivrechenzentrums – ist in Abbildung 5.4 dargestellt. Wie sich in der Abbildung bereits erkennen lässt und die PVar quantitativ darstellt, besitzt die Infrastrukturleistungsaufnahme im Vergleich zur IT eine relativ hohe Variabilität (0,07 gegenüber 0,004). Dies wirkt sich auch auf die IPA aus, welche mit einer Adaptivität von 1295% auf eine extreme Ineffizienz der Infrastruktur hindeutet. Nach eingehender Untersuchung der RZ-Komponenten wurde die Quelle dieser Ineffizienzen auch identifiziert: Dadurch, dass der Außentempersensor der Klimaregelung am Auslass des Kühlturms angebracht war, lieferte dieser im Freikühlbetrieb nach kurzer Zeit im Vergleich zur tatsächlichen Außentemperatur viel zu hohe Werte zurück, so dass automatisch zurück in den Chillerbetrieb geschaltet wurde. Nachdem der Sensor sich abgekühlt hatte, wurde wiederum in den Freikühlbetrieb gewechselt. Bei bestimmten Außenbedingungen wurde somit ständig zwischen den Kühlmodi hin- und hergeschaltet, was letztlich zu einer unerwartet hohen Variabilität und ebenso zu einem deutlichen Anstieg der Leistungsaufnahme der Infrastruktur führte.

Die beiden Anwendungsbeispiele haben gezeigt, dass die IPA als Kennzahl für einen schnellen Überblick über die Effizienz des Betriebs geeignet ist. Prinzipiell

lassen sich die gleichen Schlüsse auch durch manuelle Betrachtung der Leistungsaufnahmen im Vergleich ziehen, allerdings bevorzugen RZ-Betreiber Kennzahlen, die Informationen auf einen Blick liefern, vor Diagrammen.

5.4 Anwendung als Ersatzkennzahl

Neben der direkten Nutzung der IPA und PVar als Kennzahlen für die Effizienz des RZ-Betriebs ist die Verwendung der Metriken zur auslastungsabhängigen Abschätzung der Infrastrukturleistungsaufnahme ein wichtiger Aspekt, welcher bereits in Abschnitt 5.1 motiviert wurde. Dazu wird eine Kombination der Metriken PUE, IPA und PVar verwandt. Ein großer Vorteil ist dabei, dass alle drei Metriken als Datenbasis die aggregierten Verläufe der Leistungsaufnahmen der IT und der Infrastruktur voraussetzen, welche in allen Rechenzentren, in denen kontinuierliche PUE-Messungen durchgeführt werden, verfügbar sind.

Die PUE ist ein Maß für den relativen Overhead der Infrastrukturleistungsaufnahme gegenüber der IT-Leistungsaufnahme. Für eine gegebene IT-Leistungsaufnahme kann die PUE dazu genutzt werden die absolute Leistungsaufnahme der Infrastruktur zu bestimmen. Diese gilt allerdings nur für eine konkrete Auslastung, welche der mittleren RZ-Systemlast, für die die PUE ursprünglich berechnet wurde, entspricht. Für ein möglichst realistisches Resultat sollte demnach die mit der Auslastung korrespondierende IT-Leistungsaufnahme genutzt werden, welche durch eine Ausführung der IT-Powermodelle der LDEE mit der entsprechenden Auslastung berechnet wird. Neben den historischen Verläufen der Leistungsaufnahmen für IT und Infrastruktur ist die durchschnittliche Auslastung bei der PUE-Ermittlung der einzige weitere notwendige Parameter, um die auslastungsabhängige Infrastrukturleistungsaufnahme mit Hilfe der Ersatzkennzahlen abzuschätzen.

Da die PUE ein konstanter Wert ist und keine Information über die Auslastungsabhängigkeit liefert, wird die PVar als Metrik für Variabilität der Leistungsaufnahme hinzugezogen. Als Bestandteil der IPA fließt die PVar dabei indirekt in die Berechnung mit ein. Zudem zeigt die PVar der IT für welche IT-Lastbereiche die meisten Daten vorliegen bzw. in welchen Bereichen das Rechenzentrum typischerweise operiert. Für diese Lastbereiche ist die Abschätzung über die Ersatzkennzahlen am genauesten. Die IPA beschreibt dann letztlich den dynamischen Anteil der Infrastruktur und ermöglicht die Skalierung der statischen Leistungsaufnahme der Infrastruktur, die über den PUE berechnet wird, mit der Auslastung. Die Gleichungen 5.6 bis 5.9 zeigen, wie sich die Metriken als Ersatzkennzahl in der LDEE anstelle von Powermodellen nutzen lassen.

$$LDEE(l, l_{pue}) = \frac{perf_{dc}(l)}{power_{it}(l) + power_{inf}(l, l_{pue})} \quad (5.6)$$

$$power_{it}(l) = \sum_s P_s(l) \quad , s \in srv, sto, net \quad (5.7)$$

$$power_{inf}(l, l_{pue}) = power_{it}(l_{pue}) \cdot (PUE - 1) \cdot f(l, l_{pue}) \quad (5.8)$$

$$f(l, l_{pue}) = 1 - \frac{IPA}{100} \left(1 - \frac{power_{it}(l)}{power_{it}(l_{pue})} \right) \quad (5.9)$$

Anstelle des kombinierten RZ-Powermodells im Zähler der LDEE (vgl. Gleichung 3.1 und Abbildung 4.2) wird nur der IT-Anteil $power_{it}$ über die Powermodelle der IT-Komponenten berechnet, der Infrastrukturanteil $power_{inf}$ wird hingegen mit Hilfe der Ersatzkennzahlen abgeschätzt, wie in Gleichung 5.8 dargestellt. Als Ausgangsbasis wird der absolute Infrastrukturoverhead zur RZ-Last l_{pue} , der bei der PUE-Ermittlung vorliegenden durchschnittlichen Systemauslastung, berechnet. Dazu wird die IT-Leistungsaufnahme bei entsprechender Auslastung l_{pue} mit dem relativen Infrastrukturoverhead ($PUE - 1$) multipliziert. Dieser Ausgangswert wird mit einem Faktor, der mittels der Funktion f in Gleichung 5.9 berechnet wird, auf die zu untersuchende Auslastung l skaliert. Der rechte Teil der Gleichung 5.9 beschreibt dabei die relative Veränderung der IT-Leistungsaufnahme zwischen der Zielauslastung l und der PUE-Auslastung l_{pue} . Aus der Multiplikation mit der Adaptivität der Infrastruktur ($IPA/100$, vgl. Gleichung 5.5) resultiert dann ein Faktor, der die notwendige Anpassung des Ausgangswerts (Infrastrukturleistungsaufnahme bei l_{pue}) aufgrund von Adaptivität und zu untersuchender Auslastung darstellt.

Mit dem beschriebenen Verfahren kann die IPA also in Kombination mit der PUE als Ersatzkennzahl zur Abschätzung der Infrastrukturleistungsaufnahme anstelle entsprechender Powermodelle dienen, sofern kontinuierliche PUE-Messungen in einem operativen Rechenzentrum durchgeführt werden. Damit erweitert sich die Anwendbarkeit der LDEE auf viele Produktivrechenzentren, die zwar grundlegende Effizienzmessungen (PUE) durchführen, jedoch keine detaillierte Datenerfassung einzelner Infrastrukturkomponenten vornehmen. Durch Nutzung der IPA wird sogar die Auslastungsabhängigkeit der aggregierten Infrastruktur umgesetzt, so dass eine Anwendung der LDEE für beliebige Auslastungen ebenso möglich ist. Die Anwendung von Ersatzkennzahlen hilft allerdings nicht, wenn neben dem Fehlen von Infrastrukturmodellen auch keine Messdaten vorhanden sind. In einem solchen Fall, welcher zum Beispiel bei der Konzeption eines geplanten Rechenzentrums mit bislang nicht modellierten bzw. charakterisierten Klimakomponenten gegeben ist, kann die LDEE nicht angewandt werden. Es werden für eine Anwendung der LDEE immer entweder Modelle der Komponenten oder (kontinuierliche) Messdaten benötigt.

5.5 Evaluation

Die Evaluation der IPA umfasst einerseits einen ausgiebigen Vergleich mit der PUE Scalability von The Green Grid und andererseits eine Untersuchung der Genauigkeit bei der Nutzung zur Abschätzung der Infrastrukturleistungsaufnahme für die LDEE als Ersatz für Infrastrukturpowermodelle.

5.5.1 Vergleich IPA mit PUE Scalability

Zur Evaluation der IPA wird diese für Extremfälle ermittelt und das Verhalten gegenüber der PUE Scalability erläutert. Anschließend wird die Anwendung auf generierte Datensätze eines RZ-Containermodells demonstriert. Zunächst wird dazu kurz die PUE Scalability vorgestellt.

PUE Scalability

Die PUE Scalability wurde von The Green Grid als Erweiterung der PUE entwickelt. Sie soll Rechenzentrumsbetreibern die potentielle Skalierbarkeit der gesamten RZ-Leistungsaufnahme bei sich ändernder IT-Leistungsaufnahme in einer Kennzahl darstellen [AAF12]. Mit Kenntnis der PUE und der PUE Scalability könnten so Aussagen getroffen werden, welche Auswirkungen Änderungen in der IT-Leistungsaufnahme auf die gesamte RZ-Leistungsaufnahme und damit implizit auch auf die Infrastrukturleistungsaufnahme besitzen. Diese allgemeine Zielbeschreibung ist ähnlich zur IPA, wobei die PUE Scalability den Fokus auf die Gesamt-RZ-Leistungsaufnahme legt, während die IPA insbesondere die Infrastrukturleistungsaufnahme betrachtet.

Ähnlich der IPA nutzt auch die PUE Scalability Verläufe der Leistungsaufnahmen, die für die Ermittlung der PUE gemessen werden, um mittels statistischen Verfahren auf diesen eine Aussage zur Skalierbarkeit zu treffen. Nach Azevedo et al. [AAF12] ist die PUE Scalability dabei definiert durch das Verhältnis der Steigungen der tatsächlichen PUE Scalability (m_{Actual}) zu der proportionalen PUE Scalability (mean PUE, m_{PUE}), vgl. Gleichung 5.10. Die PUE Scalability repräsentiert die Skalierbarkeit in Prozent, wobei 100% für die optimale Skalierbarkeit, die es zu erreichen gilt, steht. Werte darunter stehen für eine entsprechend schlechtere Skalierbarkeit. Werte über 100% hingegen sind nach Azevedo et al. ungültig und sollen auf eine unzureichende Anzahl Messwerte hindeuten.

$$\text{PUE Scalability} = \frac{m_{\text{Actual}}}{m_{\text{PUE}}} 100\% \quad (5.10)$$

In Abbildung 5.5 ist ein Beispiel für die Auswertung der PUE Scalability dargestellt. Die tatsächliche PUE Scalability (Actual Scalability) wird per linearer Regression mit der Methode der kleinsten Quadrate aus den rohen Messdaten zur PUE-Ermittlung (blaue Punkte) berechnet. Sie spiegelt damit das Verhältnis zwischen IT- und RZ-Leistungsaufnahme wider. Die proportionale PUE Scalability (Proportional Scalability) nimmt den durchschnittlichen PUE

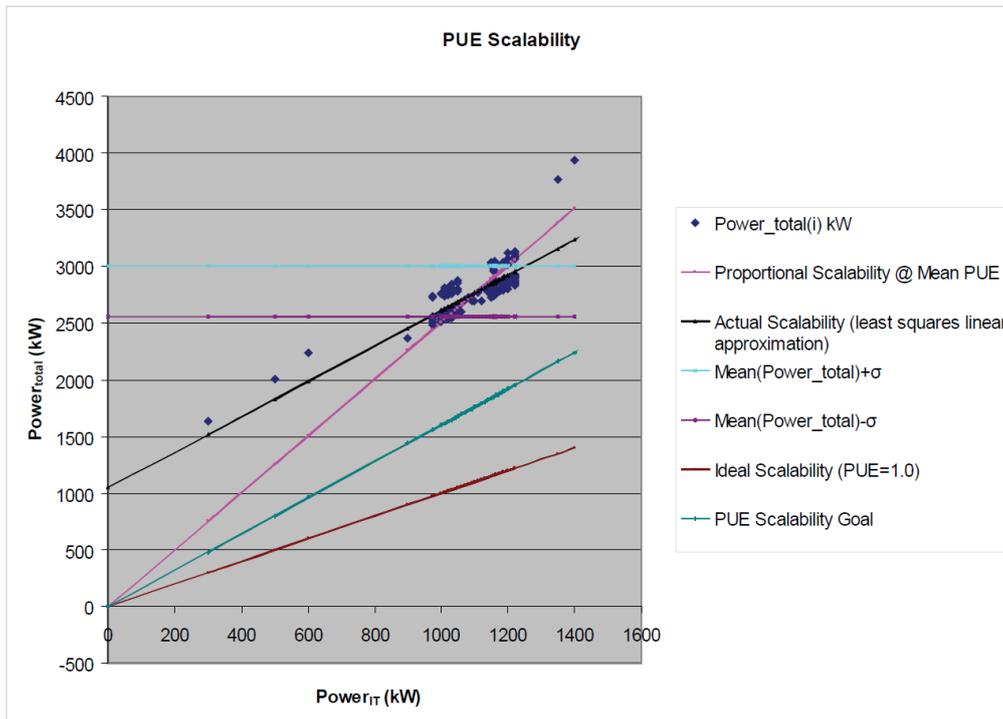


Abbildung 5.5: Beispielanwendung der PUE Scalability mit den gemessenen Leistungsaufnahmen (blau) und der proportionalen (rosa) sowie tatsächlichen (schwarz) PUE Scalability, die weiteren Kurven sind für diese Arbeit nicht relevant – Grafik unverändert aus [AAF12] zitiert

m_{PUE} und geht ansonsten von einer perfekten Proportionalität zwischen IT- und RZ-Leistungsaufnahme aus. Die Gerade verläuft entsprechend vom Koordinatenursprung aus durch m_{PUE} . Die Steigungen im Beispiel sind $m_{Actual} = 1,57$ und $m_{PUE} = 2,51$. Die resultierende PUE Scalability ist damit 62,5%.

Die PUE Scalability als auch die IPA repräsentieren die Skalierbarkeit/Adaptivität eines Rechenzentrums jeweils als prozentualen Wert. Der Unterschied in der Zielsetzung beider Metriken liegt dabei darin, dass die PUE Scalability das Verhältnis zwischen IT- und RZ-Leistungsaufnahme untersucht, während die IPA das Verhältnis zwischen IT- und Infrastrukturleistungsaufnahme betrachtet. Auch wenn dieser Unterschied auf den ersten Blick marginal erscheint, führt er zu maßgeblich unterschiedlichen Ergebnissen, wie die nachfolgende Evaluation aufzeigt.

Verhalten in konstruierten Extremfällen

Für die erste Evaluation wurden manuell Datensätze, die Extremfälle für Skalierbarkeit und Adaptivität darstellen, erzeugt. Diese Fälle kommen zwar in der

Regel nicht vor, dennoch sollten die untersuchten Metriken in der Lage sein, diese adäquat abzubilden. Die Datensätze unterteilen sich in eine vollständig adaptive RZ-Infrastruktur (DS_A), eine absolut konstante Infrastrukturleistungsaufnahme (DS_B) und ein überproportionales Verhalten der Infrastruktur (DS_C). Die drei Datensätze sind in Abbildung 5.6 dargestellt und die Ergebnisse der IPA @ PVar sowie der PUE Scalability werden in Tabelle 5.2 zusammengefasst.

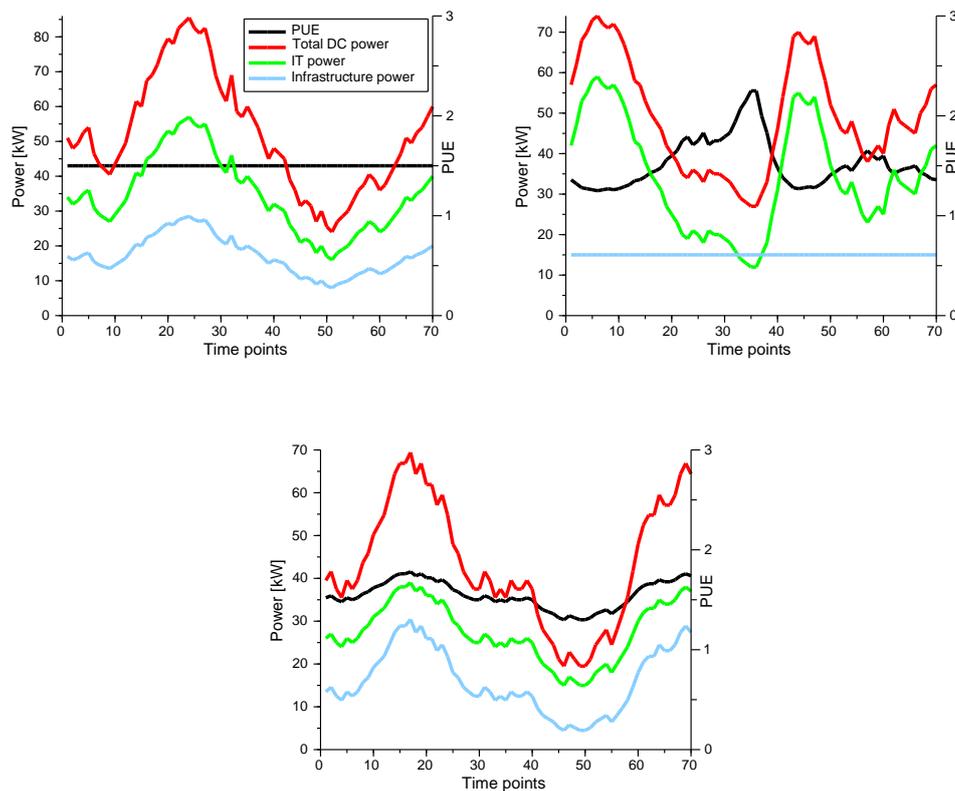


Abbildung 5.6: oben links: DS_A ; oben rechts: DS_B ; unten: DS_C

Im ersten Datensatz (DS_A) ist die Infrastrukturleistungsaufnahme durchgehend halb so hoch wie die IT-Leistungsaufnahme, es liegt also eine perfekte Adaptivität der Infrastruktur vor. Entsprechend ist die PUE konstant bei 1,5. Die untersuchten Metriken liefern hier beide erwartungsgemäß ein Ergebnis von 100%.

Im zweiten Datensatz DS_B ist die Infrastrukturleistungsaufnahme durchgehend konstant. Die Infrastruktur ist demnach überhaupt nicht adaptiv, was die IPA mit einem Resultat von 0% korrekt abbildet. Die PUE Scalability hingegen liefert ein Ergebnis von 69,5%. Wie bereits im vorherigen Abschnitt beschrieben, bewertet die PUE Scalability die Skalierbarkeit der RZ-Leistungsaufnahme und

5 Definition der IPA Metrik

Metric	DS _A	DS _B	DS _C	DS _D	DS _E	DS _F	DS _G
mean PUE	1.50	1.44	1.58	1.22	1.12	1.19	1.10
PUE Scalability [%]	100	69.5	100	86.4	94.5	88.6	98.6
IPA [%]	100	0	196.3	27.0	56.0	29.1	85.6
PVar _{IT}	0.17	0.24	0.14	0.34	0.35	0.13	0.19
PVar _{Inf}	0.17	0	0.26	0.11	0.22	0.03	0.17

Tabelle 5.2: Ergebnisse der Metriken für unterschiedliche Datensätze

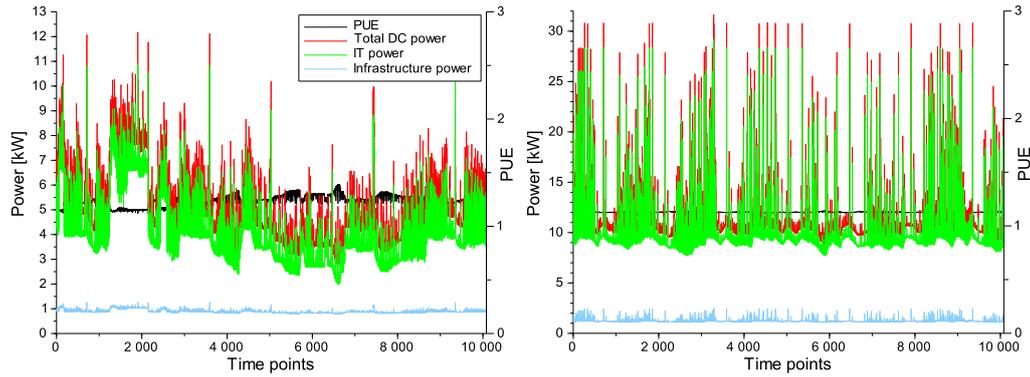
diese verändert sich immer proportional zur IT-Leistungsaufnahme. Das Ergebnis wird lediglich durch die relative Höhe der Infrastruktur- zur IT-Leistungsaufnahme beeinflusst – etwaige Variationen in der Infrastrukturleistungsaufnahme fließen somit nur indirekt ein.

Der dritte Datensatz DS_C repräsentiert den Fall, dass die Infrastrukturleistungsaufnahme im Vergleich zur IT-Leistungsaufnahme deutlich mehr skaliert. Dieser Fall kann darauf hindeuten, dass in der Infrastruktur Ineffizienzen vorhanden sind, wie sie zum Beispiel in Abschnitt 5.3 in den Anwendungsbeispielen beschrieben wurden. In den beschriebenen Beispielen konnte die IPA das Verhalten bereits abbilden, ebenso wie in diesem Datensatz. Die PUE Scalability ist nach Azevedo et al. [AAF12] nicht für Werte über 100% definiert. So kann sie diesen Fall nicht abdecken, obwohl eine Anwendung der Metrik auf das Beispiel mit 131,6% ein passendes Resultat liefern würde.

Obwohl die PUE Scalability im Rahmen ihrer Definition korrekte Ergebnisse liefert, ist die Aussagekraft dieser jedoch eingeschränkt. Die Skalierbarkeit der RZ-Leistungsaufnahme in Relation zu Änderungen der IT-Leistungsaufnahme ist kein geeignetes Maß – da die IT-Leistungsaufnahme Teil der RZ-Leistungsaufnahme ist, skaliert die RZ-Leistungsaufnahme automatisch mit Änderungen an der IT. Der Einfluss der Infrastruktur auf das Resultat wird damit nur über die absolute Höhe der Leistungsaufnahme im Verhältnis zur IT bestimmt. Ist die Infrastrukturleistungsaufnahme im Vergleich zur IT sehr gering, beeinflusst fast ausschließlich die IT die RZ-Leistungsaufnahme und das Ergebnis der PUE Scalability besitzt nur eine geringe Aussagekraft. Erst wenn die Infrastrukturleistungsaufnahme selbst einen maßgeblichen Anteil der RZ-Leistungsaufnahme einnimmt ($PUE \geq 2,0$), liefert die PUE Scalability brauchbare Ergebnisse bzgl. der Skalierung der Infrastruktur.

Anwendung auf RZ-Daten

In einer zweiten Evaluation wurden die IPA und die PUE Scalability auf Daten eines RZ-Containers angewandt. Diese wurden mit dem Simulationswerkzeug für Serverlastmanagement aus dem AC4DC-Projekt [SJS⁺14] generiert. Dazu wurden sechs 19“-Racks jeweils bestückt mit 21 Hewlett-Packard ProLiant DL380 G6 Enterpriseservern angenommen. Als unterbrechungsfreie Stromversorgung soll-

Abbildung 5.7: links: DS_D – 900 VM; rechts: DS_E – 1300 VM

te eine USV mit zwei 20kW-Modulen dienen und zur Klimatisierung wurden sechs Wärmetauscher mit EC-Lüftern (electronically commutated) im Doppelboden des Containers, Pumpen zum Kälte-transport sowie ein Freikühlmodul zur Kälteerzeugung modelliert. Diese Kühltette besitzt eine maximale Kälteleistung von 60kW. Der Betrieb des so modellierten Containerrechenzentrums wurde dann für diverse VM-Mengen mit dem Simulationswerkzeug simuliert. Die Profile der simulierten VM wurden zuvor in einem Produktivrechenzentrum aufgenommen. Unter Verwendung von Powermodellen zu allen RZ-Komponenten (vgl. Abschnitt 4.5) konnte dann die Leistungsaufnahme, auf deren Basis die Metriken angewandt werden, abgeschätzt werden.

Insgesamt wurden die vier Datensätze DS_D bis DS_G erzeugt. Diese sind in den Abbildungen 5.7 und 5.8 dargestellt. In den einzelnen Simulationen wurden unterschiedlich stark variierende VM-Profile genutzt, um unterschiedliche Variabilitäten der IT-Leistungsaufnahme zu erzeugen. In DS_G wurde zudem mit dem NEC Express5800/A1080a-E ein anderes Servermodell verwandt, um die Kühltette an die Grenze ihrer Leistungsfähigkeit zu bringen. Für diese Simulation besaß die USV drei 20kW-Module. Die Resultate der Metriken für die jeweiligen Datensätze sind in Tabelle 5.2 dargestellt.

Man kann gut erkennen, dass die PUE Scalability für das simulierte Containerrechenzentrum einen viel kleineren Wertebereich besitzt als die IPA. Dadurch, dass die Infrastrukturleistungsaufnahme nur ca. ein fünftel der IT-Leistungsaufnahme beträgt, kann auch nur ein entsprechend kleiner Anteil der RZ-Leistungsaufnahme potentiell nicht adaptiv zur IT-Leistungsaufnahme sein bzw. nicht mit ihr skalieren, da der restliche Anteil der RZ-Leistungsaufnahme genau der IT-Leistungsaufnahme entspricht. Je höher der Anteil der Infrastrukturleistungsaufnahme ist, desto höher ist dementsprechend der Wertebereich der PUE Scalability. Ein Vergleich der Resultate verschiedener Rechenzentren ist dadurch nicht sinnvoll. Selbst ein Vorher-Nachher-Vergleich des selben Rechenzentrums ist nicht praktikabel, wenn sich das Verhältnis zwischen IT- und Infrastrukturleistungsaufnahme zwischenzeitlich merklich ändert.

5 Definition der IPA Metrik

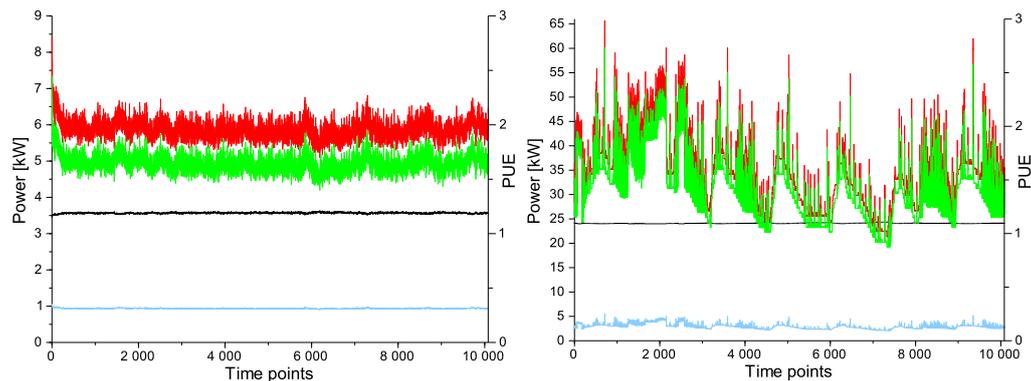


Abbildung 5.8: links: DS_F – 1000 VM; rechts: DS_G – 1100 VM; die Legende entspricht Abbildung 5.7

Eine weitere Erkenntnis der Evaluation ist, dass die Adaptivität bzw. Skalierbarkeit stark von den Arbeitslasten im Rechenzentrum abhängt. So reicht die IPA bei gleicher Infrastruktur (mit Ausnahme eines USV-Moduls in DS_G) von 27,0% bis 85,6%. Je höher dabei die durchschnittliche absolute Leistungsaufnahme der IT war, desto höher fiel auch die Adaptivität aus. Daraus lässt sich schließen, dass die Infrastrukturkomponenten des Container-RZ eine relativ hohe statische Verlustleistung im Vergleich zur dynamischen besitzen, so dass die Adaptivität erst bei höheren Lasten deutlich wird. An diesem Beispiel wird auch deutlich, warum für eine vollständige, allgemeine Bewertung der Energieeffizienz eines Rechenzentrums die Auslastungsabhängigkeit relevant ist. Mit den entsprechenden Modellen kann die LDEE dazu genutzt werden, die Energieeffizienz für beliebige Arbeitslasten zu bewerten.

5.5.2 Genauigkeit bei Nutzung als Ersatzkennzahl

Die zweite Evaluation der IPA betrachtet den Anwendungsfall, dass die IPA in Kombination mit der PUE als Ersatzkennzahl für fehlende oder nicht charakterisierbare Powermodelle der Infrastruktur genutzt wird. Zu diesem Zweck wurde für die Datensätze DS_D bis DS_G der vorherigen Evaluation mit dem in Abschnitt 5.4 vorgestellten Verfahren die elektrische Leistung der Infrastruktur abgeschätzt und mit den vorhandenen Daten auf Basis der Powermodelle verglichen.

Die zur Abschätzung notwendigen Resultate für die IPA sowie die mittlere PUE wurden der Tabelle 5.2 entnommen. Als weiterer Parameter war noch die durchschnittliche IT-Auslastung für den Zeitraum der PUE-Berechnung l_{pue} notwendig, welcher aus der IT-Auslastungskurve berechnet wurde. Das Verfahren wurde dann für jeden einzelnen Zeitpunkt der IT-Leistungsaufnahmekurve mit korrespondierender IT-Auslastung angewandt und die jeweilige Infrastrukturleistungsaufnahme abgeschätzt. Die Abweichungen dieses Ansatzes im Vergleich zu

Fehler	DS_D	DS_E	DS_F	DS_G
Mittl. quadr. abs. Fehler [kW]	0.007	0.016	0.000	0.029
Mittl. quadr. rel. Fehler [%]	0.8	1.1	0.04	1.1
Max. abs. Fehler [kW]	0.023	0.079	0.001	0.250
Max. rel. Fehler [%]	1.9	4.1	0.1	4.5

Tabelle 5.3: Abweichung bei Nutzung der IPA zur Berechnung der Infrastrukturleistungsaufnahme in der LDEE

den Resultaten der Infrastrukturpowermodelle werden in Tabelle 5.3 für die einzelnen Datensätze zusammengefasst. Mit einem mittleren relativen Fehler von 1,1% oder weniger über die untersuchten Datensätze ist die Abschätzung per Ersatzkennzahl in diesem Fall eine gute Alternative für Powermodelle der Infrastruktur.

Im Falle des simulierten Containerrechenzentrums ist die Lastabhängigkeit der Infrastruktur und damit der Powermodelle allerdings sehr hoch, so dass über die Auslastungsrelation aus Gleichung 5.9 zuverlässige Ergebnisse erzielt werden können. Wendet man die Abschätzung über die IPA allerdings auf Verläufen der Leistungsaufnahme mit stark disproportionalen Verhalten an, ist die Abweichung deutlich größer. Im Falle des Produktivrechenzentrums aus Abbildung 5.4 erhält man beispielsweise einen mittleren relativen Fehler von 15% gegenüber der real gemessenen Infrastrukturleistungsaufnahme. Powermodelle können bei Berücksichtigung der relevanten Parameter mit solchen Verhaltensstrukturen besser umgehen. Die Abschätzung der Infrastrukturleistungsaufnahme über die Ersatzkennzahl sollte also nur genutzt werden, falls keine Powermodelle der Infrastruktur verfügbar bzw. anwendbar sind.

6 Prozessintegration

Dieses Kapitel beschreibt die Integration der LDEE in die Unternehmensprozesse zur Überwachung und Kontrolle des Rechenzentrumsbetriebs. Dazu wird in Abschnitt 6.1 eine potentielle Umsetzung als Erweiterung typischer Managementsysteme in Rechenzentren beschrieben. Der zweite Teil des Kapitels beschäftigt sich dann in Abschnitt 6.2 mit einem Anwendungsbeispiel der LDEE und vergleicht die Ergebnisse mit verbreiteten Energieeffizienzmetriken hinsichtlich der definierten Anforderungen.

6.1 Umsetzung

Eine Anwendung der LDEE setzt die Ausführung von Modellen für alle Systeme des Rechenzentrums voraus, was bei großen Rechenzentren zu entsprechend hohem Aufwand führt. Nach der Identifizierung der konkreten Systeme müssen die zugeordneten Modelle anhand der Systemdaten trainiert und untereinander verknüpft werden. Anschließend werden die Modellinstanzen mit den zutreffenden Eingabedaten aufgerufen. Prinzipiell ist eine manuelle Auswahl und Verknüpfung der Modelle beispielsweise über Skripte möglich, die Anforderung der Anwendbarkeit der LDEE (R 3.10) erfordert allerdings einen automatisierten Ansatz. Anderenfalls sinkt die Akzeptanz bei zu hohem Aufwand einer Anwendung.

Eine automatisierte LDEE-Anwendung muss die folgenden Aspekte umsetzen:

1. Identifikation der von der LDEE betrachteten Systeme
 - Typen/Modelle und ihre Spezifikationen
 - Verknüpfung voneinander abhängiger Modelle
2. Erfassung der Trainings- sowie Betriebsdaten
 - Zutreffende Benchmarkresultate
 - Leistungsaufnahmen, Auslastungen
3. Verknüpfung der Systeme mit korrespondierenden Modellen
 - Performance und Power für IT-Systeme
 - Power für Infrastrukturkomponenten
 - Charakterisierung der verknüpften Modelle
4. Realisierung der Ausführbarkeit
 - Zusammenfassung der Teilergebnisse zur LDEE

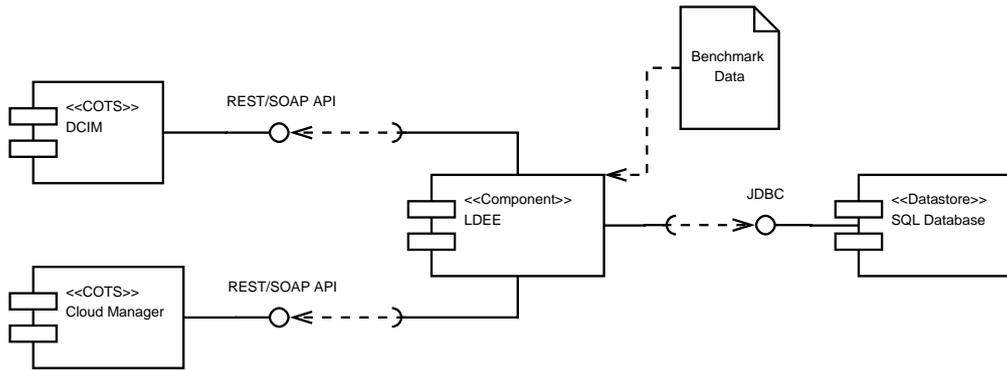


Abbildung 6.1: Komponentendiagramm zur Umsetzung einer automatisierten Anwendung der LDEE

Die Liste zeigt, dass für eine automatisierte Umsetzung einer Anwendung der LDEE neben den essentiellen, vorgestellten Verfahren und Modellen zusätzliche, unterstützende Maßnahmen wie die Inventarisierung der RZ-Systeme oder die Bereitstellung von Daten notwendig sind. Alle benötigten Informationen sollten dabei idealerweise bereits verfügbar sein, so dass die Umsetzung der LDEE sich auf die Nutzung dieser zur korrekten Zuordnung von Daten und Modellen beschränken kann. Vorteilhaft hierbei ist die immer stärkere Verbreitung von Managementsystemen in Rechenzentren, die automatisch Verwaltungsaufgaben wie Inventarisierung oder Aggregation von Systemdaten inkl. Berichterstattung übernehmen. Auf Seiten der Hardware und Infrastruktur ist das in Abschnitt 2.4.3 beschriebene Data Center Infrastructure Management (DCIM) nützlich. Für die Softwareseite der Dienste und Arbeitslasten hingegen wird auf übergreifende Cloud Computing Plattformen, welche in ebenso Abschnitt 2.4.3 beschrieben werden, zurückgegriffen.

6.1.1 Gesamtkonzept

Das Konzept der automatisierten Anwendung der LDEE sieht vor auf bestehende Managementsysteme eines Rechenzentrums aufzusetzen, um die darüber bereitgestellten Systeminformationen und Daten mit den LDEE-Modellen zu verknüpfen. Sind alle notwendigen Informationen vorhanden, kann die Energieeffizienz der Komponenten bestimmt und zu einer Gesamteffizienz kombiniert werden. Die Darstellung der Ergebnisse wird wiederum in das betreffende Managementwerkzeug eingebettet, um den Anwendern eine zentrale Anlaufstelle für das Controlling und Reporting zur Verfügung zu stellen.

Der grundsätzliche Aufbau der automatisierten LDEE-Ermittlung ist im Komponentendiagramm in Abbildung 6.1 dargestellt. Die zentrale Komponente ist die LDEE-Applikation, welche die Koordination zwischen den Modellen und den Trainings-/Betriebsdaten übernimmt.

Die LDEE-Komponente unterteilt sich dabei in mehrere Aufgabenpakete, wel-

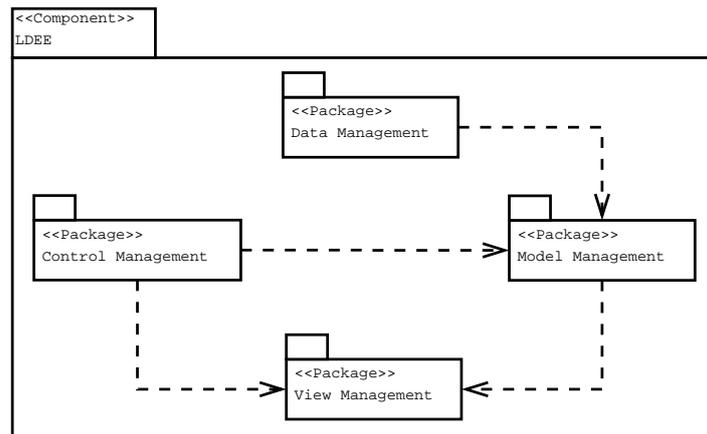


Abbildung 6.2: Pakete der LDEE-Komponente aus Abbildung 6.1

che in Abbildung 6.2 dargestellt werden. Über Schnittstellen zu den im Rechenzentrum vorhandenen Managementwerkzeugen (DCIM, Cloud Manager) liest die LDEE-Komponente notwendige Daten aus und speichert diese geeignet in einer relationalen Datenbank ab. Ebenso verfährt sie mit den Speicherausügen von öffentlich zugreifbaren Datenbanken mit Benchmarkergebnissen. Eine Aufgabe ist entsprechend das Datenmanagement. Ein zweites Paket arbeitet dann rein auf der Datenbank, indem es die RZ-Komponenten mit den in der Datenbank verfügbaren Modellen der LDEE verknüpft und die Modelle anschließend bzgl. der Komponenten charakterisiert (Model Management). Das dritte Aufgabenpaket ist dann die Ausführung der Metrik auf den Daten und Modellen (Control Management). Die Ergebnisse werden letztlich in einem weiteren Paket aufbereitet und dem DCIM über die Schnittstelle bereitgestellt (View Management).

6.1.2 Anbindung an DCIM

Der erste Schritt in Richtung automatisierter Anwendung der LDEE ist die Anbindung der Metrik und ihrer Modelle an ein bestehendes DCIM-System. Hersteller von DCIM bieten unterschiedliche Mechanismen zur Interaktion der sonstigen in Rechenzentren betriebenen (eigenen) Software mit den Funktionen und Daten des DCIM an. Die Bereitstellung solcher Schnittstellen wird dadurch motiviert, dass bereits vorhandene Werkzeuge für das Monitoring oder Controlling oder auch eigene implementierte Software in ein nachträglich installiertes DCIM integriert werden können, so dass das DCIM zur zentralen Anlaufstelle wird.

DCIM-Schnittstelle

Für Schnittstellen in DCIM-Systemen kommen unterschiedliche Technologien zum Einsatz. Ein typischer Ansatz ist das Simple Open Access Protocol (SOAP) zum Austausch strukturierter Informationen – zumeist im XML-Format – im-

plementiert als Webservice. Zur Übertragung der Nachrichten stützt sich SOAP auf weitere Protokolle der Anwendungs- und Transportschicht – z.B. HTTP und TCP. Da SOAP durch einen umfangreichen Funktionsumfang auch einen großen Overhead besitzt, kommen inzwischen vermehrt RESTful API (Application Programming Interface) in DCIM-Systemen zum Einsatz. REST steht für Representational State Transfer und bezeichnet ein Paradigma für Softwarearchitekturen basierend auf skalierbaren Webservices für verteilte Systeme, vgl. [Mas11]. Eine RESTful API ist demnach eine Schnittstelle, die sich an den architektonischen Bedingungen von REST, die die heutigen Technologien des World Wide Webs abbilden, orientiert. REST ist schlanker, schneller und effizienter als SOAP sowie leichter erlernbar und findet dadurch immer mehr Verwendung.

Das Konzept zur Anbindung der LDEE an ein DCIM wird beispielhaft für StruxureWare von Schneider Electric vorgestellt. Die Schnittstelle von StruxureWare basiert noch auf dem SOAP-Ansatz, wird zukünftig aber nach und nach auf eine RESTful API umgerüstet. Nach Analysen der IDC [Kop13] besaß StruxureWare im Jahr 2013 den zweithöchsten Marktanteil der DCIM-Systeme hinter Trellis von Emerson Network Power. Da für Trellis allerdings im Gegensatz zu StruxureWare¹ keine ausführliche Dokumentation der API-Spezifikationen verfügbar ist, wird das Konzept am Beispiel von StruxureWare gezeigt. Durch einen ähnlichen Aufbau der Schnittstellen unterschiedlicher DCIM-Systeme ist das Konzept zwischen diesen allerdings übertragbar.

Auslesen des Inventars

Ein Einsatzzweck von DCIM ist die automatisierte Inventur der an das Netzwerk angeschlossenen RZ-Komponenten sowie die Abfrage der Betriebsparameter dieser. In der GUI wird dann eine Liste der Komponenten inklusive der abgefragten Parameter zur Verfügung gestellt, die bei Planungszwecken hilfreich ist. Neben der Komponenten und ihrer Betriebsparameter werden auch topologische Abhängigkeiten hinterlegt, z.B. welcher Server in welchem Rack installiert ist. Da diese Liste alle bzgl. Performance und Leistungsaufnahme relevanten Komponenten enthält, ist sie für die Anwendung der LDEE äußerst nützlich.

Die DCIM-Schnittstellen bieten grundsätzlich Funktionen zum Auslesen solcher Inventarlisten an. Für StruxureWare werden diese Funktionen im *InventoryWebService* in einem Top-Down-Ansatz zusammengefasst. So kann man sich über Funktionen wie beispielsweise *getAllServerRooms*, *getRowsInRoom* und *getRacksInRoom/Row* gezielt auf bestimmte Bereiche – wie in diesem Fall Racks – einschränken, um konkrete Hardwarekomponenten über Funktionen wie *getServersInRack/Room* oder *getRackMountPdusInRack* zu selektieren. Alternativ ist auch die Rückgabe aller Geräte über *getAllDataCenterItems* möglich, wobei zusätzlich die Option einer Filterung z.B. nach Typ über *getDataCenterItemsByFilter* besteht.

¹<http://dcimsupport.apc.com/display/public/DCIMDEVELOPER/StruxureWare+Data+Center+Operation+Web+Service+API> (letzter Zugriff: 26.11.2015)

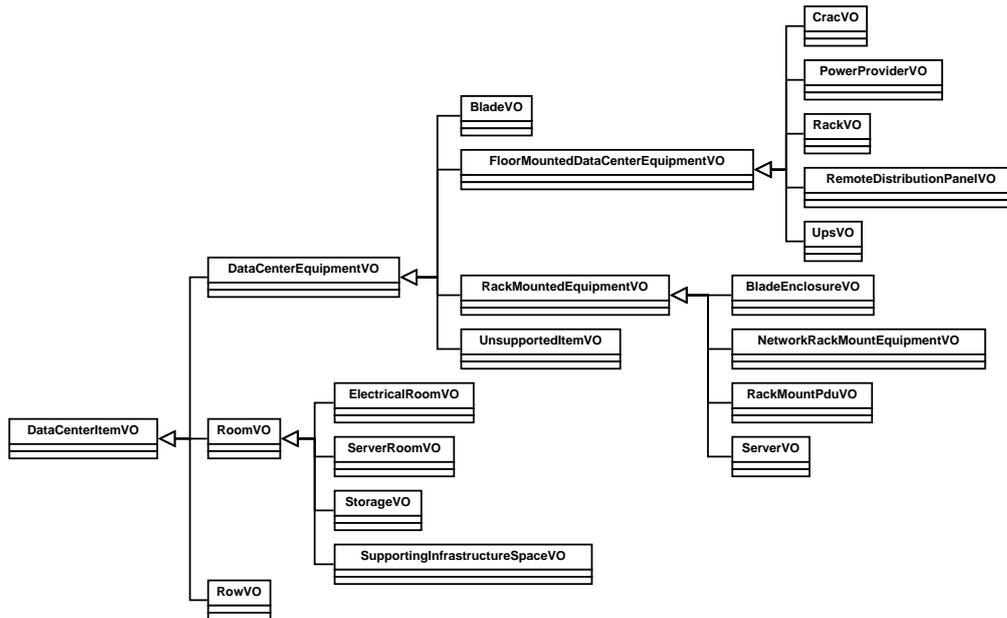


Abbildung 6.3: Klassenhierarchie für RZ-Equipment in der StruxureWare API

Das Auslesen des Inventars für die LDEE wird über die zweite Variante (*getAllDataCenterItems*) umgesetzt, da eine vollständige Liste der Komponenten benötigt wird. Diese Funktion liefert ein Array mit einem *DataCenterItemVO* pro Bestandteil des Rechenzentrums, wobei die konkrete Klasse der jeweiligen Instanz den Typ dieses Bestandteils definiert. Die zugehörige Klassenhierarchie ist in Abbildung 6.3 dargestellt. So werden Server beispielsweise durch die Klassen *ServerVO* und *BladeVO* beschrieben, während eine USV eine Instanz der *UpsVO* ist. Das genaue Modell ist wiederum durch Eigenschaften der *DataCenterEquipmentVO* definiert. Für die LDEE werden nur die von *DataCenterEquipmentVO* ererbenden Komponenten betrachtet.

Die extrahierten *DataCenterEquipmentVO* müssen zu einer weiteren Datenverarbeitung geeignet gespeichert werden. Dazu werden die gesammelten Informationen automatisiert in eine relationale Datenbank importiert, wobei alle RZ-Komponenten zunächst in einer gemeinsamen Tabelle abgelegt werden. Das entsprechende Entity-Relationship-Modell ist in Abbildung 6.4 dargestellt. Die Unterscheidung zwischen unterschiedlichen Komponenten wird einerseits durch die Klasse (z.B. *UpsVO*, im Diagramm 'Type') und andererseits durch die Produktinformationen (*getManufacturer* und *getModelName*) vorgenommen. Dies ist beispielsweise für Stagesysteme der Fall, da diese in StruxureWare zusammen mit Servern der Klasse *ServerVO* zugeordnet sind. Über einen Abgleich mit den Benchmarkdaten, die zur Charakterisierung der LDEE-Modelle genutzt werden, ist eine korrekte Zuordnung möglich.

Mittels der Methode *getCustomProperties* lassen sich die (eingetragenen) Spe-

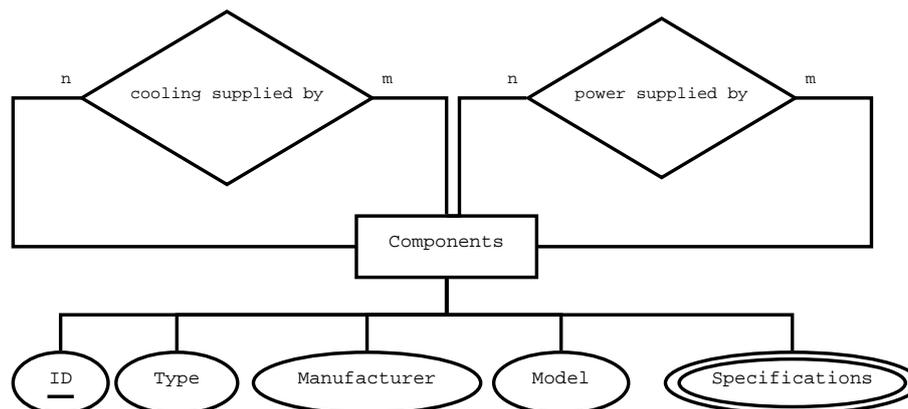


Abbildung 6.4: ER-Diagramm zur Speicherung der Komponentendaten

zifikationen der einzelnen Komponenten auslesen. Dazu gehören beispielsweise die CPU- und RAM-Ausstattung bei Servern oder die Speicherkapazität des Storages. Diese Daten werden der Komponententabelle der Datenbank hinzugefügt, welches in Abbildung 6.4 als 'Specifications' zusammenfassend dargestellt wird.

Neben der Liste der Komponenten sind auch die Abhängigkeiten bzw. Relationen untereinander wichtig, da diese die Verknüpfung der Modelle der LDEE bestimmen. Zu diesem Zweck werden für alle ermittelten *DataCenterEquipmentVO* über *getPowerConsumers* die angeschlossenen Verbraucher und über *getPowerProviders* die Komponenten, welche die elektrische Leistung bereitstellen (USV), abgefragt. Diese Daten werden als n-zu-m-Beziehung zwischen Lieferant und Verbrauchern in eine Relationstabelle der Datenbank eingefügt. Das Vorgehen bezüglich der Klimatisierung ist analog. Hier werden die Relationen (welches Gerät welche Komponenten klimatisiert) über die Verteilung der Komponenten auf die Räume bestimmt, indem die eingangs beschriebenen Methoden (z.B. *getServersInRoom*) genutzt werden. Die Zuordnung von IT-Komponenten zu den Klimageräten (*CracVO*) wird dann wiederum in einer eigenen Tabelle der Datenbank festgehalten.

Der erste Punkt der für die Umsetzung zu betrachtenden Aspekte (vgl. Abschnitt 6.1) ist damit berücksichtigt.

Auslesen verfügbarer Betriebsparameter

Werden Betriebsparameter der RZ-Komponenten im DCIM mitgeloggt, können diese ebenso über die API ausgelesen werden. Im Falle von StruxureWare sind dies die folgenden für die LDEE verwertbaren Parameter²:

- Ambient_Temperature: Umgebungstemperatur des IT-Raums abseits der Racks

²<http://dcimsupport.apc.com/ws-javadoc/com/apc/webservice/history/Type.html>
(letzter Zugriff: 26.11.2015)

- `Auxiliary_Load_Connected`: Abschätzung der Last (Leistungsaufnahme) von Infrastrukturkomponenten
- `Cooling_Capacity`: Die maximale Kühlleistung von Klimageräten
- `Cooling_Load`: Die beanspruchte Kühlleistung von Klimageräten
- `Inlet_Temperature`: Servereingangstemperatur am Rack
- `Measured_A/B_Feed_Average/Peak`: Durchschnittliche/Maximale elektrische Leistung von USV-Geräten auf Pfad A/B
- `PUE`: Power Usage Effectiveness des gesamten Rechenzentrums
- `Total_Estimated_Load`: Abschätzung der Last (Leistungsaufnahme) von IT-Komponenten
- `Total_Load_Connected`: Maximal mögliche Last (Leistungsaufnahme) von IT-Komponenten
- `Total_Network_Ports`: Anzahl der Netzwerkports von Netzwerkkomponenten
- `Used_Network_Ports`: Belegte Netzwerkports von Netzwerkkomponenten

Für die angegebenen Komponenten lassen sich diese Parameter jeweils als aktueller Einzelwert oder als ganze Zeitreihe für einen gegebenen Zeitraum auslesen. Die Betriebsdaten werden nicht permanent in der Datenbankkomponente gespeichert, sondern bei Bedarf über die API erhoben. Auffällig ist, dass in StruxureWare die Auslastung der IT-Systeme durch die Arbeitslasten nicht berücksichtigt wird. Diese müssen über andere Managementplattformen wie die nachfolgend betrachteten Cloud Computing Manager ausgelesen werden.

6.1.3 Anbindung an Cloud Computing Manager

In Managementsystemen für Cloud Computing fließen unter anderem die Ebenen der Virtualisierung, der betriebenen Dienste, des Storages und des Netzwerks zusammen, vgl. Abschnitt 2.4.3. Prinzipiell decken Cloud Manager damit alle Bereiche ab, die nicht von DCIM-Systemen erfasst werden. Über bereitgestellte Schnittstellen lassen sich somit die fehlenden Betriebsparameter auslesen. Die Umsetzung wird dabei am Beispiel des Cloud Managers OpenStack³ gezeigt, da dieser am weitesten fortgeschritten ist und am häufigsten eingesetzt wird⁴.

OpenStack setzt auf eine RESTful API für unkomplizierte Abfragen und Zugriffe über Webservices mit geringem Overhead. Die Auslastung der Server wird dabei über die API-Erweiterung Healthnmon (Health and Monitoring Service) abgefragt. Der Befehl `'GET /servers/id/metrics'` mit einem optional übergebenen Zeitraum liefert die Systemauslastungen aller Server als JavaScript Object

³<https://www.openstack.org/> (letzter Zugriff: 26.11.2015)

⁴<https://www.openstack.org/assets/presentation-media/72ForresterTrueStateofCloudAdoption.pptx> (letzter Zugriff: 26.11.2015)

Notation (JSON), welches automatisiert über Skripte weiterverarbeitet und mit den gespeicherten Informationen zu den IT-Komponenten verknüpft wird. Von Interesse ist dabei allen voran das Attribut *cpuSystemLoad* – es sind aber auch Daten zur Arbeitsspeichernutzung und zum I/O der Festplatte und des Netzwerks verfügbar. Für Storagekomponenten wird die OpenStack Block Storage API verwandt, über die per Befehle wie *'GET /v2/{tenant_id}/volumes/details'* die verfügbaren Speichergeräte und ihre Eigenschaften abgefragt werden können. Auch zur Überwachung des RZ-Netzwerks bietet OpenStack Funktionen an, die über die korrespondierende Networking API abgefragt werden können.

Ebenso wie die Betriebsparameter zur Infrastruktur werden die Daten zur IT nur bei Bedarf ausgelesen und nicht permanent gespeichert.

6.1.4 Zugriff auf Benchmarks

Die Power- und Performancemodelle der LDEE basieren zu einem großen Teil auf öffentlich verfügbaren Daten, die bei der Ausführung von Energie- und Performancebenchmarks erhoben wurden. Zur Charakterisierung der Modelle für die konkret vorhandenen Komponenten des Rechenzentrums werden daher die korrespondierenden Benchmarkdaten benötigt. Diese erhält man für die in Abschnitt 4.4 beschriebenen Benchmarks grundsätzlich als Speicherauszug der Datenbankeinträge z.B. im csv-Format (comma separated values). Diese Speicherauszüge können dann automatisiert über Skripte eingelesen, weiterverarbeitet und schließlich in der beschriebenen relationalen Datenbank abgelegt werden. Die Zuordnung der Benchmarkdaten zu den IT-Komponenten geschieht dann über den eindeutigen Bezeichner (Name des Herstellers und des Modells) der Komponente. Die prinzipielle Struktur zur Speicherung dieser Daten wird im ER-Diagramm in Abbildung 6.5 gezeigt, welches die Erweiterung des in Abbildung 6.4 dargestellten Diagramms repräsentiert. Jedes Benchmarkresultat ist jeweils einem Benchmark und einer Komponente zuzuordnen. Für eine Komponente werden mehrere Benchmarks angewandt und jeder Benchmark wiederum für mehrere Komponenten.

Mit der Umsetzung der Abfrage aller benötigten Betriebsdaten sowie der Integration der Benchmarkresultate wird der zweite Punkt der in Abschnitt 6.1 aufgelisteten Aspekte abgedeckt.

6.1.5 Verknüpfung von Komponenten und Modellen

Der dritte Aspekt der Umsetzung (vgl. Abschnitt 6.1) behandelt die Verknüpfung der Modelle der LDEE mit den konkreten Komponenten des Rechenzentrums. Diese Verknüpfung wird im ER-Modell in Abbildung 6.5 durch die n-zu-m-Relation 'represented by' dargestellt. Eine Komponente kann also durch unterschiedliche Modelle repräsentiert werden (z.B. Leistungsaufnahme und Performance). Ein einzelnes Modell ist aber auch für verschiedene Komponenten anwendbar (z.B. für unterschiedliche Servertypen). Die für die jeweiligen Kom-

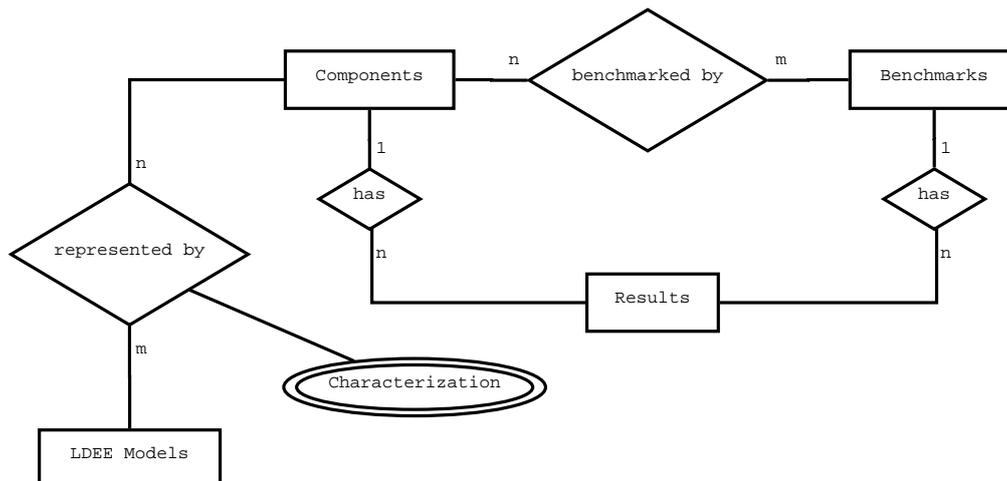


Abbildung 6.5: ER-Diagramm zur Integration der Benchmarkresultate und Verknüpfung von Komponenten mit den Modellen der LDEE

ponenten charakterisierten und somit anwendbaren Modelle werden als (mehrwertiges) Attribut der Relation gespeichert.

Für Modelle der LDEE, die nur Benchmarkdaten zur Charakterisierung bzw. zum Training benötigen, kann diese bereits durchgeführt und die resultierenden Modellparameter in der Datenbank gespeichert werden. Für alle anderen Modelle, die weitere Daten benötigen, sind weitere Voraussetzungen zur Charakterisierung zu erfüllen. Die konkrete Umsetzung richtet sich nach der Art des Modells und wird nachfolgend vorgestellt.

IT-Lastmodell: Das IT-Lastmodell wird nicht mit konkreten IT-Komponenten verknüpft, sondern gilt für die gesamte gemanagte IT. Vielmehr müssen daher die verfügbaren Daten mit dem Modell verknüpft werden, so dass dieses für das Rechenzentrum trainiert werden kann. Das Modell setzt dabei auf den Auslastungen und der Aktivität (an/aus) der IT-Komponenten auf. Wird ein Last- und Powermanagementverfahren im Rechenzentrum eingesetzt, kann dieses direkte Daten für das Modell liefern. Dies bedarf einer speziellen Anpassung der Schnittstellen, es sei denn, das Lastmanagementverfahren nutzt bereits bekannte Cloud Manager wie OpenStack. In diesem Fall kommt das in Abschnitt 6.1.3 beschriebene Verfahren zum Einsatz. Alternativ können die notwendigen Auslastungs- und Aktivitätsdaten auch unabhängig vom Lastmanagement gesammelt werden.

Performancemodelle: Da die Modelle der genutzten Server- und Storageperformance auf vorab bekannten Benchmarkresultaten aufsetzen, liegen die charakterisierten Modelle bereits vor. Die relative, maximale Performance $perf_i$ jedes IT-Systems i ist entsprechend in der Datenbank hinterlegt. Es müssen lediglich für alle Server- und Storagekomponenten die korrespondierenden Performan-

cewerte zugeordnet werden. Dies wird anhand der eindeutigen Bezeichner der Hardware (Name des Herstellers und Modells) bewerkstelligt. Der Netzwerkperformancekoeffizient richtet sich hingegen nach den Netzwerk-SLA, die häufig nicht in DCIM-Systemen kontrolliert werden. Diese werden allerdings vermehrt in Werkzeugen zum Cloud Computing Management berücksichtigt, so dass hierfür der in Abschnitt 6.1.3 beschriebene Ansatz genutzt werden kann.

Powermodelle der IT: Die lastabhängigen Powermodelle für Server und Storage setzen analog zu ihren Performancemodellen rein auf Messdaten von Energiebenchmarks. Entsprechend liegen charakterisierte Modelle bereits vor und es muss nur noch eine Verknüpfung über die Bezeichner der Komponenten hergestellt werden. Anschließend lassen sich Abschätzungen der Leistungsaufnahme für beliebige gegebene Auslastungen ermitteln. Dafür notwendige Modellvariablen wie die Anzahl der Festplatten im Storagepowermodell werden – wie zuvor beschrieben – über die Methode *getCustomProperties* der jeweiligen Komponente über die API abgefragt und sind ebenfalls in der Datenbank hinterlegt. Alternativ können diese Informationen auch über die API der Cloud Manager abgefragt werden.

Powermodelle der Infrastruktur: Zur Anwendung der Powermodelle für Infrastrukturkomponenten sind deren Betriebsparameter notwendig. Diese werden in StruxureWare über den *CustomPropertiesWebService* und den *HistoryReportService* für die entsprechenden Komponenten ausgelesen. Anschließend kann je nach Art des Modells entweder die Charakterisierung auf Basis der Betriebsdaten vorgenommen oder die Berechnungsformeln direkt auf den Betriebsdaten angewandt werden.

6.1.6 Ablauf der Anwendung

Nach der Schaffung der Voraussetzungen zur Ermittlung der Energieeffizienz (Schnittstellen, Datenquellen, Datenbankschema) ist der letzte Aspekt der Umsetzung die Realisierung der konkreten Anwendung. Diese wird über eine eigene Softwareapplikation, die Zugriff auf die beschriebene relationale Datenbank hat und die Schnittstellen zu den Managementwerkzeugen nutzt, realisiert. Die Applikation entspricht der in den Abbildungen 6.1 und 6.2 dargestellten LDEE-Komponente und orchestriert somit den Gesamtablauf von der Datensammlung, -verarbeitung und -speicherung über die Charakterisierung der Modelle bis zur Lieferung/Darstellung der resultierenden Energieeffizienz.

Die Applikation liefert für eine gegebene RZ-Auslastung und Außentemperatur das Ergebnis der LDEE. Hierzu werden die kombinierten RZ-Modelle zur Performance und zur Leistungsaufnahme entsprechend der Beschreibung in Abschnitt 4.2 angewandt. Zunächst wird dazu mit Hilfe des IT-Lastmodells aus den über die Schnittstellen ausgelesenen Daten die Verteilung der Arbeitslast abgeschätzt. Diese wiederum ist ein Eingabeparameter der den IT-Komponenten

zugeordneten Performance- und Powermodellen. Die Zuordnung wird dabei der Datenbank entnommen, ebenso wie die zuvor charakterisierten Modellparameter. Für die IT-Komponenten können somit die Performance und Leistungsaufnahme abgeschätzt und bereits als Teilresultate zurückgegeben werden.

Die Infrastrukturkomponenten hingegen benötigen als Eingabeparameter die Leistungsaufnahmen der abhängigen IT sowie Infrastruktur. Für USV-Geräte ist dies die Gesamtleistungsaufnahme der Geräte, die von der jeweiligen USV mit Strom versorgt wird. Für Klimageräte ist dies analog dazu die Leistungsaufnahme der gekühlten Komponenten. Diese werden ebenfalls aus der Datenbank den entsprechenden Relationen entnommen. Die abgeschätzten Leistungsaufnahmen der Infrastrukturkomponenten können dann als Teilresultate zurückgeliefert werden. Umgebungsdetails wie die Architektur der Raumluftklimatisierung (z.B. Einhausung) oder Kühltemperaturen fließen dabei implizit über die Modelle, die mit realen Messdaten charakterisiert wurden, in die Resultate ein. Aus den Einzelwerten werden dann die kombinierten Ergebnisse für das gesamte Rechenzentrum berechnet, die ins Verhältnis gesetzt die LDEE ergeben. Insgesamt liefert die Softwareapplikation neben der LDEE für das gesamte Rechenzentrum auch die Energieeffizienz jeder IT-Komponente sowie geeignete Effizienzdaten (Wirkungsgrad, EER, ...) jeder Infrastrukturkomponente. Mit diesen Informationen lassen sich auch für einzelne Systeme Aussagen ableiten.

Im beispielhaften Anwendungsfall der DCIM StruxureWare kann die LDEE-Komponente als Java-Bibliothek eingebettet und aus StruxureWare heraus per JavaScript aufgerufen werden. Die Resultate können dann wiederum ins StruxureWare Dashboard zur individuellen Darstellung integriert werden. Über angepasste Skripte lässt sich zudem die Berechnung der LDEE mehrfach für variable Eingangsparameter aufrufen. So lassen sich die Resultate in den in Abschnitt 3.4.6 beschriebenen Formaten darstellen – zum einen als dreidimensionale Darstellung eines Effizienzverlaufs und zum anderen als repräsentativen Querschnitt für einen direkten Vergleich zwischen Rechenzentren.

6.2 Anwendungsbeispiel

Bislang wurden im Rahmen der Modellbeschreibung (Kapitel 4) und der Konzeption der IPA (Kapitel 5) die folgenden Teilevaluationen von Subkomponenten der LDEE beschrieben:

- Anwendung einzelner Powermodelle (Server, USV, CRAH, Chiller) auf verfügbaren Messdaten in den jeweiligen Beschreibungen des Abschnitts 4.5
- Vergleich des IT-Lastmodells mit alternativen Ansätzen in Abschnitt 4.6.1
- Vergleich des kombinierten RZ-Performancemodells mit bekannten Produktivitätsmetriken in Abschnitt 4.6.2

- Anwendung des Klimapowermodells auf ein Containerrechenzentrum mit Freikühleinheit in Abschnitt 4.6.3
- Vergleich der PVar mit alternativen Ansätzen in Abschnitt 5.2.2
- Anwendung der IPA für zwei konkrete Rechenzentren in Abschnitt 5.3.2
- Vergleich der IPA mit der PUE Scalability in Abschnitt 5.5.1
- Anwendung der IPA als Ersatzkennzahl in Abschnitt 5.5.2

Die Teilevaluationen dienen dem Zweck, die Validität, Plausibilität und/oder Anwendbarkeit der Teilkomponenten der LDEE zu zeigen. In Ergänzung dazu wird in diesem Abschnitt die Anwendung der gesamten zusammengesetzten Metrik anhand eines Beispielrechenzentrums demonstriert. An dieser Beispielanwendung wird dann die Erfüllung der in Abschnitt 3.2 definierten Anforderungen im Vergleich zu ausgewählten, verbreiteten Metriken (vgl. Abschnitt 2.6.1) gezeigt. Zur Evaluation war kein vollständiger Datensatz eines einzelnen realen Rechenzentrums verfügbar. Aus diesem Grund wurde der Betrieb eines realistischen Rechenzentrums simuliert, dessen Modelle auf realen Messdaten aus dem AC4DC-Projekt [SJS⁺14] und aus öffentlichen Energiebenchmarks beruhen. Die Zusammensetzung wird dabei in den Abschnitten 6.2.1 bis 6.2.3 erläutert.

Die Nutzung von Simulationsdaten stellt keine Einschränkung der Aussagekraft der Analysen dar. Die Effekte und Auswirkungen der Metrik bzgl. der Anforderungen können auch anhand von Simulationsdaten gezeigt werden. Reale Messdaten wären hingegen bei der Untersuchung der Genauigkeit der Modelle der LDEE essentiell, welche allerdings bereits in den Evaluationen der Teilkomponenten untersucht wurde. Neben der Bereitstellung eines vollständigen Datensatzes ist ein weiterer Vorteil der Nutzung von Simulationsdaten die Möglichkeit unterschiedliche Rechenzentrumskonfigurationen zu vergleichen. Damit lassen sich die Auswirkungen auf die angewandten Metriken untersuchen und bestimmte (ggf. unerwünschte) Effekte erzeugen. Dies ist mit realen Messdaten in der Regel nicht möglich, da die Änderungen, die solche Effekte hervorrufen, sich über einen längeren Zeitraum erstrecken.

6.2.1 RZ-Hardware

Das aufgestellte Rechenzentrumsszenario ist dem Betrieb eines realen Rechenzentrums nachempfunden. Die Auswahl der Geräte wird entsprechend manuell vorgenommen und untereinander abgestimmt. Die Art der Kühlösung ist dabei der begrenzende Faktor, da nur für eine kombinierte Kühlkette ausreichend Daten, die eine ausgiebige Simulation mit unterschiedlichen Variablen (z.B. IT-Last, Temperaturen) ermöglichen, verfügbar sind. Das dabei genutzte Kühl- bzw. Infrastrukturkonzept ist die RiMatrix S⁵ von Rittal, welche im Projekt AC4DC

⁵http://www.rittal.com/de_de/rimatrixs/ (letzter Zugriff: 26.11.2015)

vermessen und modelliert wurde, vgl. [SJS⁺14]. Da dieses System modular skalierbar ist, wirkt sich diese Beschränkung nicht auf die Größe des Beispielrechenzentrums aus.

Die Auswahl der Größe des Rechenzentrums basiert auf der Studie zum Materialbestand der Rechenzentren in Deutschland von Hintemann et al. (vgl. [HFS10]). Während Serverschränke und Serverräume mit 33.000 bzw. 18.000 Lokationen den größten Anteil an den Rechenzentrumstypen ausmachen, ist eine Anwendung der LDEE vornehmlich für größere Rechenzentren ab 50kW Gesamtleistungsaufnahme interessant. Die hauptsächliche Zielgruppe der Metrik beginnt bei den größeren Serverräumen und beinhaltet überwiegend kleine und mittlere Rechenzentren mit 1800 bzw. 350 Lokationen. Für das Beispielszenario wird eine Rechenzentrumsgröße einer Gesamtleistungsaufnahme von ca. 550kW, welches nach Hintemann et al. dem Durchschnitt der mittleren Rechenzentren entspricht, gewählt.

In der gleichen Studie [HFS10] wird ebenso eine typische Zusammenstellung der IT-Komponenten für Rechenzentren der jeweiligen Größe genannt. Auf Basis dieser Angaben setzt sich auch das Beispielszenario zusammen. Dieses besteht aus 600 Servern, 600TB Speicherkapazität der Stagesysteme, 18 Serverswitche mit insgesamt 1800 Ports und 54 SAN-Switche mit 2400 SAN-Ports. Die konkreten Komponenten sind dabei nachfolgend aufgelistet:

- Serversysteme
 - Dell PowerEdge R820, 4x Intel Xeon E5-4610 v2
 - Dell PowerEdge R720, 2x Intel Xeon E5-2650
 - Dell PowerEdge R710, 2x Intel Xeon E5645
 - Fujitsu PRIMERGY RX300 S7, 2x Intel Xeon E5-2650
 - Fujitsu PRIMERGY RX300 S6, 2x Intel Xeon E5645
 - HP ProLiant DL385 G5, 2x AMD Opteron 2356
- Stagesysteme
 - Dell Compellent FS8600 v2
 - NetApp Data ONTAP 8.1 Cluster
 - NetApp FAS3160
 - Hitachi Unified Storage File
- Netzwerkkomponenten
 - Brocade VDX 8770-4 (192 10GbE-Ports)
 - Brocade VDX 6940-36Q (144 10GbE-Ports)
 - Brocade 6520 FC-Switch (96 16Gbps-Ports)

Insgesamt beträgt die maximale IT-Leistungsaufnahme mittlerer Rechenzentren damit im Durchschnitt ca. 250kW nach Hintemann et al. An dieser installierten, maximalen IT-Leistungsaufnahme wird das komplette Rechenzentrum skaliert. Auf der einen Seite muss die Energieversorgungsinfrastruktur die maximale elektrische Leistung der IT bereitstellen können, auf der anderen Seite muss die Klimatisierung die in Wärme umgewandelte Leistung abführen können. Die RiMatrix S ist ein modularisiertes Infrastruktursystem inkl. Reihenklimatisierung, wobei jedes Modul aus sechs IT-Racks und drei Netzwerk- und Infrastrukturracks besteht. Die enthaltene Infrastruktur ist unter anderem eine USV sowie die Unterstromverteilung. Die Klimatisierung ist im Unterboden der IT-Racks angeordnet und benötigt zum Betrieb einen Kaltwasseranschluss. Da ein einzelnes Modul auf eine IT-Leistungsaufnahme von 60kW ausgelegt ist, werden für die 250kW maximaler IT-Leistungsaufnahme des Beispielszenarios insgesamt fünf Module benötigt. Um eine n+1-Redundanz bzgl. der Infrastruktur zu erreichen (Tier 2, vgl. Abschnitt 2.1.2), wird zudem ein sechstes Modul hinzugefügt. Die Module werden dann so ausgerichtet, dass man eine Unterteilung in Kalt- und Warmgänge erhält.

Die RiMatrix S RZ-Module müssen für die volle Funktionalität an eine externe Quelle zur Kaltwasserlieferung angeschlossen werden. Auch hierzu wurde im Projekt AC4DC eine Komponente durch Rittal vermessen, vgl. [SJS⁺14]. Diese besteht aus einem Kaltwassersatz bzw. Chiller mit integriertem Kühlturm, der alternativ die Klimatisierung per Freiluftkühlung ermöglicht. Ein solches Modul ist wie die RiMatrix S selbst auf 60kW Kühlleistung ausgelegt, so dass hiervon ebenso sechs Module benötigt werden. Als Kaltwassertemperatur des Kühlkreislaufs werden 22°C angenommen, so dass eine Kaltgangtemperatur von < 24°C erreichbar ist. Das Kühlwasser kann dann bis 20°C Außentemperatur über einen reinen Freikühlbetrieb gekühlt werden. Zwischen 20°C bis hin zur Wasserrücklauftemperatur (ca. 27°C) kann die Freiluftkühlung in Kombination mit dem Kaltwassersatz genutzt werden, um diesen durch eine Vorkühlung zu entlasten. Ab 27°C wird nur noch der Kaltwassersatz genutzt.

6.2.2 RZ-Software

Auf der Softwareseite des Beispielrechenzentrums wurde eine vollständige Virtualisierung der Betriebssysteme und Anwendungen angenommen. In der Realität ist die Virtualisierungsrate in der Regel noch deutlich niedriger, aktuelle Analysen⁶ zeigen aber einen Trend stetig steigender Virtualisierungsraten. Das Anwendungsbeispiel repräsentiert damit ein Rechenzentrum, wie es zukünftig häufiger vertreten sein wird.

Durch die Nutzung einer vollständigen Virtualisierung werden auf jedem Server mehrere in virtuelle Maschinen gekapselte Dienste betrieben. Die statische Zuordnung der Dienste bzw. VM zu den Servern wird dabei anhand des poten-

⁶<http://www.idc.com/getdoc.jsp?containerId=prCZ24997714> (letzter Zugriff: 26.11.2015)

tiellen maximalen Ressourcenbedarfs der Dienste vorgenommen. Durch dieses Vorgehen kann die Auslastung der Server gegenüber nicht-virtualisierten Umgebungen deutlich erhöht werden. Das Beispielszenario repräsentiert damit bereits im statischen Betrieb ein State-of-the-Art-Rechenzentrum.

Neben dem statischen Verfahren unterstützt das Beispielszenario auch die dynamische Virtualisierung über ein Last- und Powermanagement. Dieses nutzt die Möglichkeit zur unterbrechungsfreien Verschiebung von virtuellen Maschinen per Live-Migration, um dadurch dynamische Zuordnungen von VM und Servern anhand des tatsächlichen anstelle des maximalen Ressourcenbedarfs vorzunehmen, vgl. Abschnitt 2.4.3. Aktuell nicht genutzte physikalische Server werden zur Energieeinsparung heruntergefahren und erst bei Bedarf wieder reaktiviert.

Durch den Fokus der RZ-Betreiber auf Verfügbarkeit kommen derartige Last- und Powermanagementverfahren allerdings kaum produktiv zum Einsatz. Durch eine immer breitere Nutzung von Cloud Managern wie OpenStack oder OpenNebula könnte sich dies in Zukunft ändern, da diese bereits mit simplen Lastmanagern ausgeliefert werden und komplexere Verfahren über Plugins einfach nachinstalliert werden können. Aktuell repräsentiert der dynamische Betrieb des Beispielrechenzentrums allerdings noch ein Zukunftsszenario.

6.2.3 Datenbasis

Nach der Definition des Aufbaus des Beispielrechenzentrums wird zur Realisierung der Simulation eines Rechenzentrumsbetriebs eine Datenbasis bzgl. der betriebenen Dienste sowie der Klimadaten benötigt. Die Anwendungsdaten basieren auf realen, gemessenen Auslastungsdaten, die im Rahmen des AC4DC-Projekts [SJS⁺14] im Produktivbetrieb erhoben wurden. Die Daten umfassen pro virtueller Maschine die CPU- sowie die Arbeitsspeicherauslastung und besitzen eine zeitliche Auflösung von einer Minute. Die gemessenen Auslastungen werden für die simulierten virtuellen Maschinen des Anwendungsbeispiels direkt übernommen, so dass die zeitliche Auflösung der Simulation ebenso auf eine Minute festgelegt wird. Für die Simulationen wird die Anzahl der betriebenen VM so gewählt, dass im statischen Betrieb alle Server genutzt werden. Dieser Fall tritt bei knapp 15 VM pro Server ein, so dass insgesamt ca. 9000 VM auf 600 Servern simuliert werden. Pro VM wird zudem eine permanente Speicherbelegung der Stagesysteme zwischen 40 und 80 Gigabyte angenommen. Insgesamt resultiert daraus eine Storagebelegung von 540 Terabyte.

Der Simulationszeitraum jedes Simulationsdurchlaufs beträgt eine Woche, so dass Tages-/Nachtzyklen ebenso wie Lastschwankungen durch Wochenendtage Berücksichtigung finden. Der Zeitraum ist von Ende Juni bis Anfang Juli angesiedelt. Dieser wirkt sich in erster Linie auf die klimatischen Bedingungen des RZ-Standorts aus. Die entsprechenden Außentemperaturen wurden dabei über die inzwischen eingestellte Google-Wetter-API⁷ bezogen und werden als Eingangs-

⁷<http://www.programmableweb.com/news/google-weather-api/2012/08/28>
(letzter Zugriff: 26.11.2015)

be für die Klimamodelle des Simulators genutzt. Als Standort wurde Frankfurt am Main ausgewählt. Die dabei vorherrschenden Außentemperaturen von 16°C bis 31°C sorgen dabei für einen steten Wechsel zwischen den Betriebsmodi in der Kälteerzeugung (Freiluftkühlung, Kaltwassersatz oder kombiniert), so dass direkte Auswirkungen auf die Resultate der Metriken bzgl. der Klimatisierung untersucht werden können.

6.2.4 Evaluation

Im Rahmen des Anwendungsbeispiels werden unterschiedliche Untersuchungen durchgeführt. In erster Linie wird die Energieeffizienz nach der LDEE mit den typischen, verbreiteten EE-Metriken für verschiedene Testszenarien verglichen. Daran soll demonstriert werden, wie die Metriken auf bestimmte RZ-Konstellationen oder -Konfigurationen reagieren und welchen Einfluss diese auf die Aussagekraft der Resultate besitzen. Des Weiteren werden die LDEE und die Vergleichsmetriken in Hinblick der in Abschnitt 3.2 definierten Anforderungen untersucht.

Vergleich mit etablierten Metriken

Zum Vergleich werden neben der LDEE weitere Energieeffizienzmetriken auf die identischen Testszenarien angewandt. Die Auswahl der Vergleichsmetriken richtet sich dabei allen voran an die verbreitete Nutzung in der Industrie. Eine Untersuchung diesbezüglich wurde von Schödwell et al. [SEZ13] durchgeführt. Die Power Usage Effectiveness ist demnach mit Abstand die bekannteste Metrik, gefolgt vom Energiebenchmark SPECpower. Da der Benchmark allerdings nur einzelne Systeme und nicht das ganze Rechenzentrum betrachtet, lässt er sich im Rahmen dieser Evaluation nicht für einen Vergleich nutzen. Mit der CADE und der DCeP folgen mit etwas Abstand zwei Metriken, die neben der Energie auch die Produktivität des gesamten Rechenzentrums betrachten. Der Vergleich der Metriken im Rahmen der Evaluation wird damit anhand der in Tabelle 6.1 aufgeführten Metriken durchgeführt. Der Übersicht halber wird der jeweilige grundsätzliche Aufbau der Metriken an dieser Stelle wiederholt. Eine detailliertere Beschreibung der Metriken findet sich in Abschnitt 2.6.1.

Die PUE wird mit Hilfe der im Simulationswerkzeug integrierten Powermodelle berechnet. Aus der abgeschätzten Leistungsaufnahme der RZ-Komponenten wird über die Simulationszeit die Energieaufnahme der IT sowie des gesamten Rechenzentrums berechnet, woraus sich die PUE ergibt. In der aktuellen Definition der PUE (vgl. Abschnitt 2.6.1) wird nur ein einzelnes Resultat für den gesamten Betrachtungszeitraum berechnet. Unter Nutzung der ursprünglichen Definition basierend auf den Leistungsaufnahmen kann allerdings auch ein Verlauf der PUE über den Simulationszeitraum ermittelt werden.

Die CADE erfordert etwas mehr Parameter als die PUE. Die DCiE entspricht dem Kehrwert der ursprünglichen Definition der PUE, also $\frac{\text{power}_{IT}}{\text{power}_{DC}}$. Diese wird

Name	Metrik
Load Dependent Energy Efficiency	$LDEE = \frac{\text{perf}_{\text{DC}}(\vec{I})}{\text{power}_{\text{DC}}(\vec{I}, T)}$
Power Usage Effectiveness	$PUE = \frac{\text{energy}_{\text{DC}}}{\text{energy}_{\text{IT}}}$
Corporate Average DC Efficiency	$CADE = \text{DCiE} \cdot \frac{\text{power}_{\text{IT}}}{\text{power}_{\text{ITmax}}} \cdot \text{util}_{\text{IT}} \cdot \text{EE}_{\text{IT}}$
Data Center Energy Productivity	$\text{DCeP} = \frac{\text{useful work}_{\text{DC}}}{\text{energy}_{\text{DC}}}$

Tabelle 6.1: In der Evaluation betrachtete Metriken

multipliziert mit der aktuellen IT-Last und damit in diesem Fall mit der Relation der momentan benötigten zur maximalen IT-Power der Komponenten, die derzeit installiert sind. Des Weiteren wird die durchschnittliche CPU-Auslastung der IT util_{IT} benötigt, die sich aus der Simulation direkt ergibt. Der vierte Parameter – die Energieeffizienz der IT-Komponenten EE_{IT} – ist nach Kaplan et al. [KFK08] noch nicht definiert und soll durch zukünftige Metriken dargestellt werden. Für diese Evaluation wird daher auf normierte SPECpower-Resultate als Ersatzkennzahl für den offenen Parameter zurückgegriffen. Analog zur ursprünglichen PUE-Definition basiert die CADE auf der Leistungsaufnahme der RZ-Komponenten, so dass die CADE für einen einzelnen Zeitpunkt bestimmt werden kann und man somit einen Verlauf erhält.

Die Berechnung der DCeP ist nochmals aufwändiger, da nach der Definition in [AMC⁺08] jede sinnvolle Arbeit quantitativ gemessen werden muss. Da die konkreten Anwendungen hinter den Auslastungsdaten der virtuellen Maschinen nicht bekannt sind, wird für diese Evaluation das in Abschnitt 4.6.2 beschriebene Vorgehen angewandt, um die von der IT bearbeiteten Tasks zu simulieren. Die Anzahl der absolvierten Tasks kann dabei nur über einen bestimmten Zeitraum und nicht für einzelne Zeitpunkte ermittelt werden. Im Gegensatz zu den anderen drei Metriken erhält man für die DCeP somit keinen Verlauf der Energieeffizienz sondern nur einen Einzelwert, der einen definierten Zeitraum repräsentiert. Die DCeP wurde daher im Rahmen dieser Evaluation für das in Abschnitt 4.6.2 beschriebene Testszenario 2 (TS_2) neben der Gesamtauswertung zusätzlich stundenweise bestimmt, um zumindest näherungsweise einen Verlauf zu erhalten.

Nach der Simulation eines statischen sowie eines dynamischen Rechenzentrumsbetriebs wurden die vier Metriken auf den ermittelten Zeitreihen der Auslastung sowie Leistungsaufnahmen angewandt. Die Verläufe der Metriken werden für den dynamischen Betrieb in Abbildung 6.6 und für den statischen Betrieb in Abbildung 6.7 dargestellt, wobei für die DCeP stündliche Werte ermittelt wurden. Die jeweiligen Resultate werden zudem in Tabelle 6.2 für den gesamten Simulationszeitraum zusammengefasst.

Man kann gut erkennen, dass die LDEE, die CADE und die DCeP das dynamische, bedarfsgerechte Hinzuschalten von Rechenkapazitäten durch das Last-

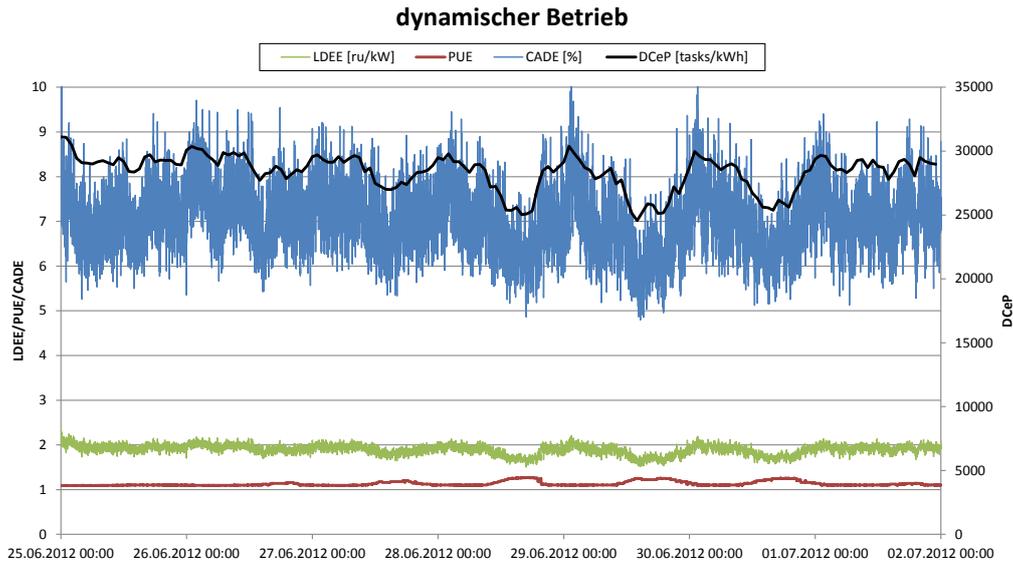


Abbildung 6.6: Vergleich der Metriken für dynamischen Betrieb

Metrik	dynamischer Betrieb	statischer Betrieb	Einheit
LDEE	1,892	1,129	ru/kW
PUE	1,132	1,126	
CADE	7,0	3,1	%
DCeP	28287,9	16874,7	tasks/kWh

Tabelle 6.2: Resultate der EE-Metriken im Vergleich

und Powermanagement positiv bewerten. Die PUE hingegen, bei der das Resultat den Overhead der Infrastruktur repräsentiert und größere Werte damit schlechter sind, ist im dynamischen Betrieb marginal schlechter. Dies demonstriert sehr gut, dass die PUE nicht sinnvoll zur Beurteilung der Energieeffizienz genutzt werden kann. Sie zeigt in diesem Anwendungsbeispiel jedoch, dass die Leistungsaufnahme der Infrastruktur sehr gut mit der IT-Leistungsaufnahme skaliert, da das Verhältnis fast gleich bleibt.

Auffällig ist das Ergebnis der CADE, wenn man dieses mit den in [KFK08] beschriebenen CADE Tiers, welche im Prinzip Rangeinstufungen der Energieeffizienz entsprechend der CADE-Resultate sind, vergleicht. Die CADE Tiers unterteilen sich in fünf Level mit der nachfolgenden Zuordnung:

- Level 1: 0% bis 5%
- Level 2: 5% bis 10%
- Level 3: 10% bis 20%
- Level 4: 20% bis 40%

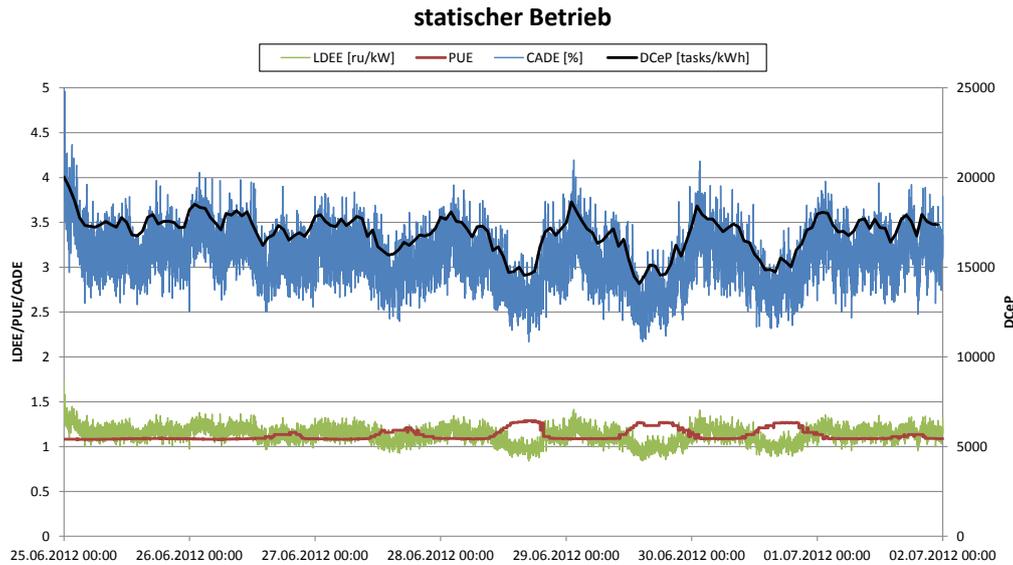


Abbildung 6.7: Vergleich der Metriken für statischen Betrieb

- Level 5: >40%

Dieser Einordnung zufolge wäre das Beispielerchenzentrum im statischen Betrieb im CADE Level 1, welches nach [KFK08] dem Stand der meisten Rechenzentren des Jahres 2008 entspräche. Da für die Evaluation allerdings modernere und damit energieeffizientere IT- als auch Infrastrukturkomponenten der Jahre 2010 bis 2014 angenommen wurden und die durchschnittliche Auslastung der aktiven IT-Systeme mit 19,1% im statischen sowie 73,3% im dynamischen Betrieb normal bzw. sehr hoch war, ist die Definition der CADE Tiers fragwürdig. Selbst wenn der Parameter EE_{IT} für die Energieeffizienz der Server, welcher in [KFK08] nicht näher definiert wurde, weggelassen wird, erhöht sich die CADE nur auf 9% bzw. 21% für den statischen bzw. dynamischen Betrieb. Ein höheres Resultat bzgl. der CADE kann nur erreicht werden, wenn die gesamte installierte IT-Kapazität $power_{IT_{max}}$ hoch ausgelastet wird. Werden Server bei Nichtnutzung abgeschaltet, wirkt sich dies entsprechend negativ auf die CADE aus. Die CADE ist letztlich so konstruiert, dass sich nur eine bestimmte Art von Optimierungen positiv auf das Resultat auswirkt. Damit verfehlt diese Metrik das Ziel eine allgemeine Energieeffizienz darzustellen.

Die LDEE und die DCeP haben zwischen beiden Betriebsmodi des Rechenzentrums das gleiche Verhältnis. Dies resultiert aus dem in Abschnitt 4.6.2 beschriebenen Verfahren, welches die sinnvolle Arbeit (Anzahl abgearbeiteter Tasks) aus der simulierten Hardwarenutzung, welche ein Hauptaspekt der LDEE ist, berechnet. Diese Relation ist realistisch, da die genutzte Performance bei gleichbleibender Anzahl abgearbeiteter Tasks trotz unterschiedlicher Hardwarenutzung gleich bleibt. So werden beispielsweise im dynamischen Betrieb für die gleiche Arbeit

weniger IT-Systeme genutzt, diese dafür allerdings höher ausgelastet, so dass letztlich die gleiche genutzte Performance resultiert.

Analysen bezüglich der Anforderungen

Im Folgenden werden die erzielten Resultate der untersuchten Metriken hinsichtlich der wichtigsten in Abschnitt 3.2 definierten Anforderungen für Energieeffizienzmetriken untersucht.

Repräsentation der Energieeffizienz (R 3.1) Wie bereits im vorherigen Abschnitt beschrieben wurde, fällt die PUE bzgl. dieser Anforderung direkt heraus, da sie nur die Leistungsaufnahme der IT bzw. des gesamten Rechenzentrums betrachtet. Die drei anderen Metriken hingegen besitzen unterschiedliche Auslegungen der Energieeffizienz. Die reinste Definition besitzt dabei die DCeP, bei der die tatsächlich getätigte, sinnvolle Arbeit gemessen und der dafür benötigten Energie gegenübergestellt wird. Die LDEE wiederum nutzt nicht die ursprüngliche Definition der Energieeffizienz, sondern abstrahiert diese durch die Kombination der Auslastung mit der Performance. Das Verhalten der beiden Metriken ist dabei sehr ähnlich, wie sich schon beim Verhältnis der Resultate des dynamischen zum statischen Betriebs gezeigt hat. Entsprechend haben die Metriken einen Korrelationskoeffizienten von 0,83.

Anders verhält es sich mit der CADE. Diese besitzt im dynamischen Betrieb nur eine Korrelation von 0,58 mit der DCeP, während diese im statischen Betrieb immerhin bei 0,71 liegt, was sich ebenso anhand der Abbildungen 6.6 und 6.7 erahnen lässt. Dieser große Unterschied liegt an der Form, in der die Energieeffizienz in die Berechnung der CADE einfließt. Im Gegensatz zu anderen Metriken liegt die Energieeffizienz nicht im Fokus sondern ist nur einer von vier Faktoren, vgl. Tabelle 6.1. Zudem wird die prozentuale Energieeffizienz aller IT-Geräte verwendet, auch wenn diese gar nicht aktiv sind. Die LDEE hingegen betrachtet nur die tatsächlich genutzte IT und die DCeP abstrahiert komplett von dieser Ebene, indem direkt die Arbeit, welche von der aktiven IT geschafft wird, gemessen wird. Dies erklärt die Differenz zwischen dem statischen Betrieb, in dem alle IT-Systeme aktiv sind, und dem dynamischen.

Auslastungsabhängigkeit (R 3.2) Die Auslastungsabhängigkeit der Metriken kann mit Hilfe des Korrelationskoeffizienten quantifiziert werden. Die errechneten Koeffizienten sind in Tabelle 6.3 angegeben.

Die PUE besitzt erwartungsgemäß keinerlei positive als auch negative Korrelation zur Auslastung. Dies lässt sich auch bereits aus der PUE-Formel (vgl. Tabelle 6.1) ableiten, in der nur die Energieaufnahmen der IT und des Rechenzentrums berücksichtigt werden. Da diese zudem ins Verhältnis zueinander gesetzt werden, wirken sich höhere Energieaufnahmen durch eine höhere Auslastung der IT nicht merklich auf die PUE aus, da sie sowohl die IT- als auch RZ-Energie beeinflussen.

Metrik	dynamischer Betrieb	statischer Betrieb
LDEE	0,596	0,784
PUE	-0,018	-0,088
CADE	0,897	0,864
DCeP	0,247	0,523

Tabelle 6.3: Korrelationskoeffizienten der EE-Metriken zur Auslastung

Bei der DCeP ließ sich nur eine geringe bis mittlere Abhängigkeit ermitteln. Dies resultiert aus dem Fokus der Metrik auf die getätigte Arbeit, welche zwar die Ursache für die vorhandene Auslastung darstellt, selbst allerdings vielmehr von der Performance der zugrundeliegenden IT-Systeme abhängt. Schließlich erzeugt die gleiche Arbeit auf einem performanteren System eine geringere Auslastung. Gleichzeitig bedeutet dies, dass die Auslastung alleine kein Ersatz für die sinnvolle Arbeit ist, sondern für eine Abschätzung dieser in Kombination mit der Performance betrachtet werden muss.

Diesen Ansatz verfolgt die LDEE, welche aufgrund der direkten Integration der Auslastung als Eingangsparameter der Funktion eine deutlich höhere Abhängigkeit zur Auslastung besitzt als die DCeP. Es liegt jedoch keine vollständige Abhängigkeit aus oben genanntem Grund vor. Die Auslastung wird mit der Performance kombiniert, um so die tatsächlich genutzten Kapazitäten abzuschätzen. Somit hat auch die Performance der IT-Systeme Einfluss auf das Ergebnis der LDEE. Dies kann man auch in Abbildung 6.8 erkennen, bei der die Ergebnisse der LDEE gegenüber der jeweils vorliegenden Auslastung der aktiven IT-Systeme aufgetragen wurden. Diese sind bei genauer Betrachtung linienförmig angeordnet, wobei eine „Linie“ eine nahezu lineare Abhängigkeit der LDEE zur Auslastung darstellt. Jede „Linie“ repräsentiert dabei eine bestimmte Konstellation genutzter IT-Systeme. Im dynamischen Betrieb sind die Ergebnisse der LDEE weiter auseinander gefächert, da nicht genutzte Server temporär heruntergefahren werden und unterschiedliche Konstellationen der Servernutzung sich dadurch stärker auf das Ergebnis der LDEE auswirken. Im statischen Betrieb ergeben sich die verschiedenen Konstellationen nur durch Auslastung unterschiedlich performanter sowie effizienter Server.

Die CADE besitzt im Vergleich zu den anderen Metriken eine sehr hohe Korrelation zur Auslastung. Der Grund hierfür liegt im Aufbau der Metrik, welche die Auslastung gleich durch zwei von vier Faktoren einfließen lässt (vgl. Tabelle 6.1) – zum einen direkt über die durchschnittliche IT-Auslastung $util_{IT}$ und zum anderen indirekt über die Relation der aktuellen IT-Leistungsaufnahme zum Maximum. Des Weiteren variiert die DCiE (bzw. die PUE) im Anwendungsbeispiel nur marginal und die Energieeffizienz der IT fließt als konstanter Faktor in die CADE ein, so dass das Ergebnis hauptsächlich durch die IT-Last sowie -Auslastung bestimmt wird.

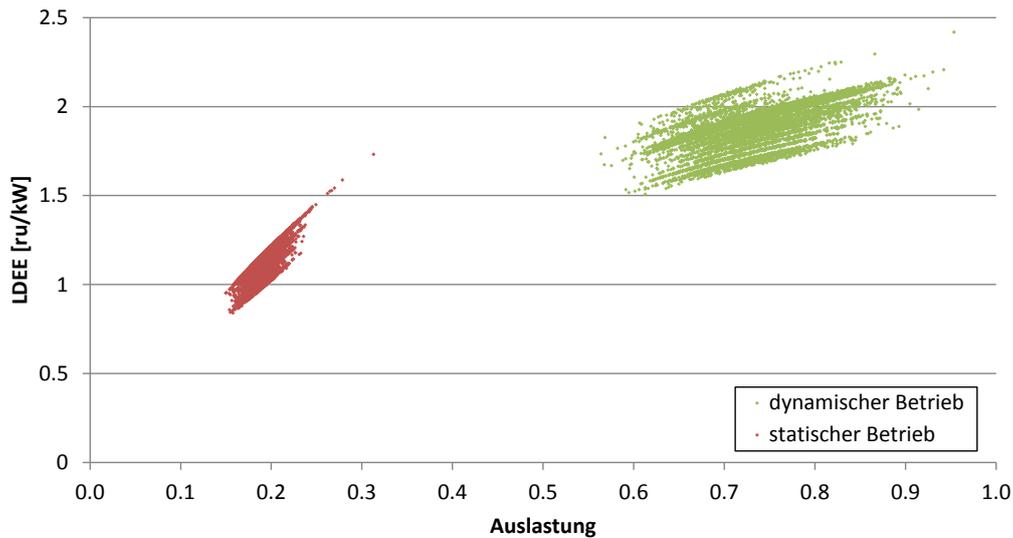


Abbildung 6.8: Relation der LDEE zur Auslastung der aktiven IT-Systeme

Vergleichbarkeit (R 3.3, R 3.4, R 3.5, R 3.6) Als Grundvoraussetzung zur Realisierung einer Vergleichbarkeit der Resultate müssen die Anforderungen R 3.3 für die vollständige Betrachtung der Systemumgebung und R 3.4 für die objektive Bewertung erfüllt sein. Hinzu kommt mit der Anforderung R 3.5, dass die Resultate an sich zwischen Rechenzentren unterschiedlicher Ausprägung vergleichbar sein müssen. Die letztgenannte Anforderung kann direkt an den definierten Formaten der Resultate untersucht werden.

So liefert die PUE einen einheitenlosen Faktor und die CADE eine prozentuale Angabe der Energieeffizienz, welche beide von der Größe des Rechenzentrums abstrahieren. Eine prozentuale Energieeffizienz hat allerdings das Problem, dass sie eine Relation zu einer vordefinierten Energieeffizienz (in der Regel das Maximum) darstellt. Mit zukünftig performanteren Systemen kann die Energieeffizienz dann entweder über 100% liegen oder der Referenzwert wird angepasst, wodurch jedoch Vergleiche mit älteren Ergebnissen aufgrund der unterschiedlichen Referenzwerte nicht mehr möglich sind. Die DCeP und die LDEE geben anstelle abstrakter Zahlen hingegen ein konkretes Verhältnis der Produktivität bzw. einer geeigneten Ersatzkennzahl zur Energie- bzw. Leistungsaufnahme an. Durch die Offenheit des Wertebereichs nach oben können auch zukünftige IT-Systeme mit dem identischen Verfahren bewertet und untereinander verglichen werden.

Neben dem Format des Resultats ist die konsistente vollständige und objektive Betrachtung der Energieeffizienz für einen Vergleich zwischen mehreren Rechenzentren essentiell. Eine konsistente Vollständigkeit wird durch das Abstecken der betrachteten Systemumgebung und die Definition der Messpunkte erzielt. Im Rahmen dieser Evaluation sind diese durch die vorhandenen Simulationsmodelle

gegeben. Im Produktiveinsatz der LDEE wird die konsistente Abfrage der Systeme des Rechenzentrums über die Realisierung als Plugin eines DCIM-Systems vorgenommen (vgl. Abschnitt 6.1.2).

Die Objektivität hängt ebenso von der Definition der Metriken ab. Es muss dabei klar herausgestellt sein, wie die notwendigen Eingangsparameter der Metriken ermittelt werden, ohne dass der Anwender eigene Interpretationen hinzufügt. Bei der PUE sind die Parameter in [AAF12] klar definiert. Die CADE hingegen lässt in [KFK08] den Parameter für die Energieeffizienz der IT-Hardware offen, so dass eine objektive Bewertung nur bei konsequentem Weglassen des Parameters möglich ist. Die DCeP erfordert nach [AMC⁺08] eine manuelle Definition der sinnvollen Arbeit bzw. der Tasks des Rechenzentrums, wodurch die Ergebnisse naturgemäß subjektiv sind. Dies wurde bereits in der Evaluation der Performancemodelle in Abschnitt 4.6.2 demonstriert.

Die LDEE ist aufgrund ihres modellbasierten Ansatzes durchgängig objektiv. Die Komponenten des Rechenzentrums bestimmen, aus welchen Modellen sich die Bewertung der LDEE zusammensetzt. Gestützt wird dies, wie bereits erwähnt, durch automatische Inventurverfahren in DCIM-Systemen. Die Teilmodelle sind dann für alle Rechenzentren identisch und werden anhand der vorliegenden Umgebungsparameter charakterisiert. Zwei Parameter, die vor allem bei einem RZ-übergreifenden Vergleich einer komplett objektiven Bewertung entgegenstehen, sind die Arbeitslasten bzw. die daraus resultierende Auslastung des Rechenzentrums sowie die bei der Bewertung vorliegenden klimatischen Bedingungen. Die LDEE ist daher so definiert, dass sie zusätzlich zur Bewertung der Energieeffizienz anhand der gemessenen Daten ($LDEE_m$) einen repräsentativen Einzelwert ($LDEE_r$) für vordefinierte Auslastungen und Außentemperaturen liefern kann, der dann für schnelle, unkomplizierte Vergleichszwecke nutzbar ist. Die genaue Definition dieses Repräsentanten findet sich in Abschnitt 3.4.6.

Die Repräsentanten $LDEE_r$ für das Anwendungsbeispiel sind für den dynamischen Betrieb 2,22 ru/kW und für den statischen Betrieb 1,79 ru/kW. Sie sind damit deutlich höher als die LDEE-Resultate basierend auf den gemessenen Daten – bzw. in diesem Fall Simulationsdaten – ($LDEE_m$), da das Mittel der untersuchten Auslastungen für den Repräsentanten mit 0,41 höher als die gemessene RZ-Gesamtauslastung von 0,19 ist. Die Resultate werden in Tabelle 6.4 zusammengefasst.

Die LDEE-Resultate für das eingangs beschriebene Beispielrechenzentrum (RZ A) können direkt ohne weitere Umrechnung mit Rechenzentren beliebiger Größe verglichen werden. Um dies beispielhaft zu demonstrieren wurden zwei weitere Rechenzentren definiert, simuliert und anschließend die LDEE angewandt. Die jeweiligen Ergebnisse sind ebenso in Tabelle 6.4 angegeben. Das RZ B ist dabei bzgl. der HW-Komponenten ähnlich dem RZ A aufgebaut, es besitzt allerdings nur zwei RiMatrix-Module im Vergleich zu den sechs Modulen des RZ A. Analog wurde auch die Anzahl der Server von 600 auf 200 und die elektrische Anschlussleistung von 550 kW auf ca. 180 kW skaliert. Allerdings werden in RZ B deutlich weniger VM pro Server betrieben als möglich, so dass die RZ-Auslastung sehr

	RZ A	RZ B	RZ C
el. Anschlussleistung [kW]	550	180	270
Server	600	200	200
VMs	9000	600	1400
RZ-Gesamtauslastung	0,19	0,04	0,19
LDEE _m [ru/kW] (dynamisch)	1,89	0,67	0,57
LDEE _m [ru/kW] (statisch)	1,13	0,55	0,29
LDEE _r [ru/kW] (dynamisch)	2,22	2,22	0,72
LDEE _r [ru/kW] (statisch)	1,79	1,80	0,49

Tabelle 6.4: Vergleich der gemessenen LDEE (LDEE_m) und des von der Arbeitslast abstrahierten Repräsentanten (LDEE_r) für verschiedene Rechenzentren

gering ausfällt. Diese geringe Auslastung wirkt sich entsprechend negativ auf die LDEE_m der Messdaten (bzw. Simulationsdaten) aus. Der Repräsentant LDEE_r abstrahiert hingegen von der tatsächlichen Arbeitslast im Rechenzentrum und bewertet alle Rechenzentren für die gleichen vordefinierten Auslastungen und Temperaturen. Da die Ausstattung der Rechenzentren A und B fast identisch ist und nur skaliert wurde, liefert die LDEE_r auch das gleiche Ergebnis.

Das Rechenzentrum C besitzt im Vergleich zu A und B deutlich ältere Serversysteme (im Mittel 3 bis 4 Jahre älter) und kann aufgrund dessen nicht so viele VM pro Server (7 anstatt 15) betreiben. Durch die niedrige Performance der IT ist die Auslastung dennoch so hoch wie in RZ A. Die älteren und ineffizienteren Server wirken sich letztlich auch negativ auf die LDEE aus, sowohl auf die LDEE_m als auch auf die LDEE_r. Die LDEE ermöglicht somit objektive Vergleiche wahlweise mit Berücksichtigung der konkreten Arbeitslasten der Rechenzentren (LDEE_m) als auch abstrahiert für vordefinierte Eingangsparameter (LDEE_r), falls nur die Hardware selbst von Interesse ist.

Als weitere Anforderung R 3.6 für den Vergleich von Rechenzentren müssen Energieeffizienzmetriken die Option für den Vergleich gleichgearteter Rechenzentren ermöglichen. Ist eine geeignete Klassifikationsmethode vorhanden, können allerdings alle Metriken diese Anforderung im Rahmen ihrer Vergleichbarkeit erfüllen, da die Vergleiche nur auf die Rechenzentren einer Klasse eingeschränkt werden. Die Metriken sollten jedoch zumindest eine Definition zur Unterscheidung der RZ-Klassen liefern. Diese ist bislang nur bei der LDEE mit der in den Abschnitten 3.4.5 und 4.4.8 beschriebenen Definition der Klassen vorhanden, wobei die LDEE wahlweise für die Ausrichtung und/oder die Verfügbarkeitsklasse des RZ angegeben werden kann. Die Ausrichtung gibt dabei den Fokus des Rechenzentrums an (ausbalanciert, rechen-/speicherintensiv) und die Verfügbarkeitsklasse den Grad der HW-Redundanzen. Der LDEE-Repräsentant eines Cloud-Rechenzentrums mit Fokus auf Big Data und der Verfügbarkeitsstufe 'Tier 3' wird beispielsweise durch LDEE_r[C_D, T₃] gekennzeichnet.

Server	LDEE _m	LDEE _m	LDEE _r
	dynamisch	statisch	
IBM System x3650 M2 (8 x 2 GHz)	7,57	4,80	4,16
IBM System x3650 M2 (8 x 2 GHz)	0,08	2,33	4,16
HP ProLiant DL380 G5 (8 x 3 GHz)	1,36	1,12	0,82
HP ProLiant DL385 G5 (8 x 2,3 GHz)	0	0,25	0,68
HP ProLiant DL380 G5 (4 x 1,6 GHz)	0	0,12	0,48

Tabelle 6.5: LDEE in ru/kW auf Komponentenebene für Beispiel-Server des RZ C, vgl. Tabelle 6.4

Diverse Betrachtungsebenen (R 3.7) Wie bereits in Abschnitt 3.3 dargestellt, besitzt von den im Rahmen der Evaluation untersuchten Metriken nur die LDEE mehrere Betrachtungsebenen. Die PUE, die CADE und die DCeP bewerten jeweils nur das gesamte Rechenzentrum basierend auf den gemessenen Daten. Zumindest die DCeP ließe sich theoretisch noch für einzelne Serversysteme anwenden, um diese bzgl. ihrer Effizienz miteinander zu vergleichen. Auf Storage-Systeme oder Infrastrukturkomponenten ist sie allerdings nicht isoliert anwendbar.

Die LDEE besitzt durch die Nutzung von Modellen für die einzelnen RZ-Komponenten neben der aggregierten Bewertung auch die Möglichkeit die Energieeffizienz auf Komponentenebene zu ermitteln und auszugeben. Dies ist sowohl für die tatsächlich vorliegenden Betriebsparameter als auch für beliebige übergebene Eingangsparameter möglich, also für die zuvor beschriebenen LDEE_m und LDEE_r. Da eine Auflistung von Einzeleffizienzen in Rechenzentren mit hunderten Servern schnell unübersichtlich wird, sollte die Ausgabe der Daten zur Erhöhung der Anwendbarkeit an die Anzeigemethodiken von DCIM-Systemen gekoppelt werden.

Die Einzelbewertung der RZ-Komponenten wurde beispielhaft für das im vorherigen Abschnitt simulierte Rechenzentrum C durchgeführt. In Tabelle 6.5 wird dabei ein Auszug der Ergebnisse für unterschiedliche Servertypen dargestellt. Die repräsentative LDEE_r gibt die von der tatsächlichen Arbeitslast abstrahierte Energieeffizienz an und liefert damit einen Überblick über die allgemeine Effizienz der Komponente. Mit der LDEE_m kann hingegen untersucht werden, welche Systeme zudem nur geringfügig genutzt werden. Zum Beispiel wurde einer der IBM System x3650 M2 Server sowohl im dynamischen als auch im statischen Betrieb viel genutzt, ein weiterer identischer Server jedoch aufgrund der Arbeitslastverteilung deutlich weniger.

Die Angabe von Effizienzen der Einzelkomponenten ist nicht auf Serversysteme beschränkt. Es lässt sich ebenso die Effizienz der Storage-Systeme als ru/kW ausgeben. Da für Netzwerkkomponenten keine Energieeffizienz bestimmt wird, wird hier der Energiebedarf der Anzahl der Ports einer vorgegeben Bandbreite gegenübergestellt. Für die USV-Geräte wird der Wirkungsgrad zwischen der

bereitzustellenden IT-Leistung und der Eingangsleistung geliefert. Die Effizienz der Klimageräte wird in Form der erzielten Kühlleistung in Relation zum Bedarf an elektrischer Leistung dargestellt. Dies entspricht prinzipiell einer für den tatsächlichen Betrieb definierten Energy Efficiency Ratio (vgl. Abschnitt 2.5.3).

Auf Basis dieser Informationen können Anwender der LDEE Entscheidungen für Optimierungsmaßnahmen treffen. Ein geplanter Austausch von Komponenten kann dabei bereits vor der Installation untersucht werden, wie in der nachfolgenden Diskussion zur Anforderung R 3.9 der Flexibilität am Beispiel des RZ C demonstriert wird.

Unabhängigkeit von Messungen (R 3.8) Die LDEE ist die einzige Metrik, die unabhängig von Messungen im Rechenzentrum die Energieeffizienz für beliebige Arbeitslasten bzw. Auslastungen liefern kann. Prinzipiell können zwar auch alle anderen Metriken auf beliebige Größen ihrer jeweiligen Parameter (wie z.B. Produktivität, Energiebedarf) angewandt werden, allerdings bestimmen sich die Abhängigkeiten jeweils direkt aus diesen. Somit müssen die Parameter bereits geeignet abgeschätzt worden sein um ein sinnvolles Ergebnis zu erzielen. Die LDEE hingegen benötigt die aggregierte RZ-Auslastung sowie die Außentemperatur und kann daraufhin auf Basis der vorliegenden Komponentenmodelle die Energieeffizienz bestimmen.

Wie bereits unter dem Aspekt der Vergleichbarkeit (R 3.3 ff.) im Rahmen der Evaluation dargelegt wurde, kann die LDEE eine von realen Arbeitslasten abstrahierte Energieeffizienz ($LDEE_r$) zu Vergleichszwecken ermitteln. Gleichzeitig lässt sich jedoch auch ein Verlauf der Energieeffizienz bezüglich der Eingangsparameter RZ-Auslastung und Außentemperatur erstellen, welcher die unterschiedlichen Betriebspunkte des Rechenzentrums in einem Diagramm darstellt. Die Ermittlung und Darstellung dieses Verlaufs wurde in Abschnitt 3.4.6 definiert. Im Rahmen der Evaluation wurden für das Anwendungsbeispiel (RZ A) für den dynamischen und den statischen Betrieb jeweils der Verlauf der LDEE gebildet und in den Abbildungen 6.9 bzw. 6.10 dargestellt.

Beide Verläufe deuten eine Abhängigkeit der LDEE von der Auslastung an, wobei die Energieeffizienz im dynamischen Betrieb bei sinkender Auslastung länger höher bleibt und erst ab einer Auslastung von 20% oder weniger stark abfällt. Dies wird in Abbildung 6.11 für eine Außentemperatur von 15°C nochmals verdeutlicht. Der Grund für den rapiden Abfall im dynamischen Betrieb ist, dass sich bei niedrigen Auslastungen statische Leistungsaufnahmen schlecht skalierender RZ-Komponenten durch das Herunterfahren ungenutzter Server erheblich stärker auswirken als bei hohen Auslastungen, wenn mehr Serversysteme aktiv sind. Gleichmaßen lässt sich in beiden Betriebsarten erkennen, dass die Außentemperatur während des Freikühlbetriebs nur geringen Einfluss auf die Energieeffizienz besitzt. Erst wenn der Kaltwassersatz des Beispielrechenzentrums ab 20°C zugeschaltet wird, sinkt die Energieeffizienz durch den erhöhten elektrischen Leistungsbedarf merklich.

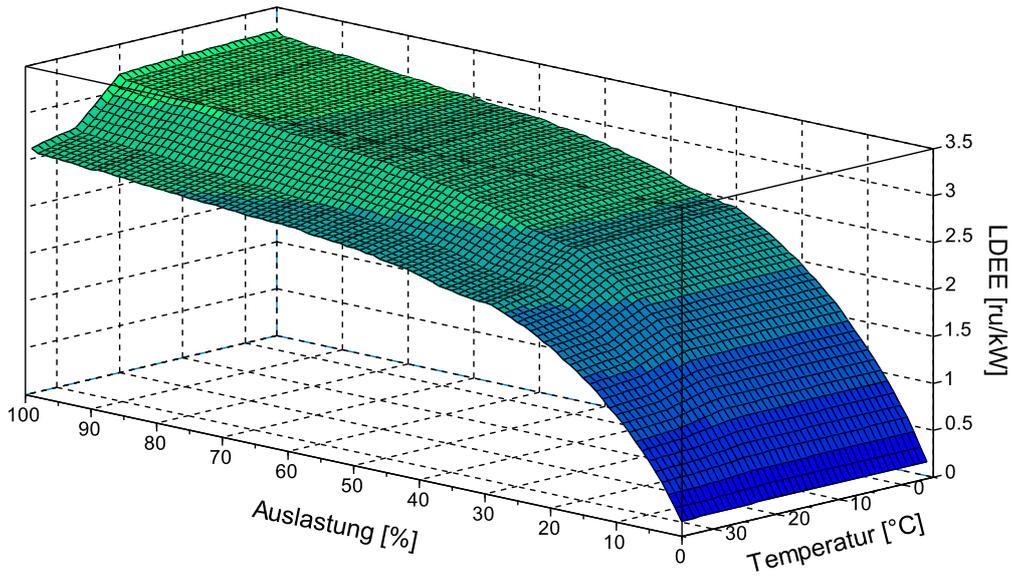


Abbildung 6.9: Auslastungs- und temperaturabhängiger Verlauf der LDEE des RZ A im dynamischen Betrieb

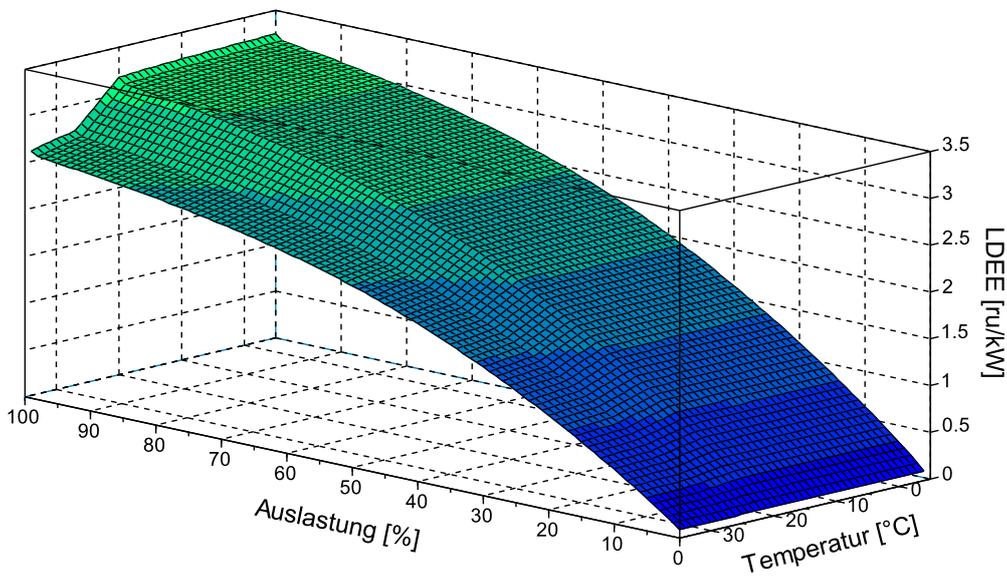


Abbildung 6.10: Auslastungs- und temperaturabhängiger Verlauf der LDEE des RZ A im statischen Betrieb

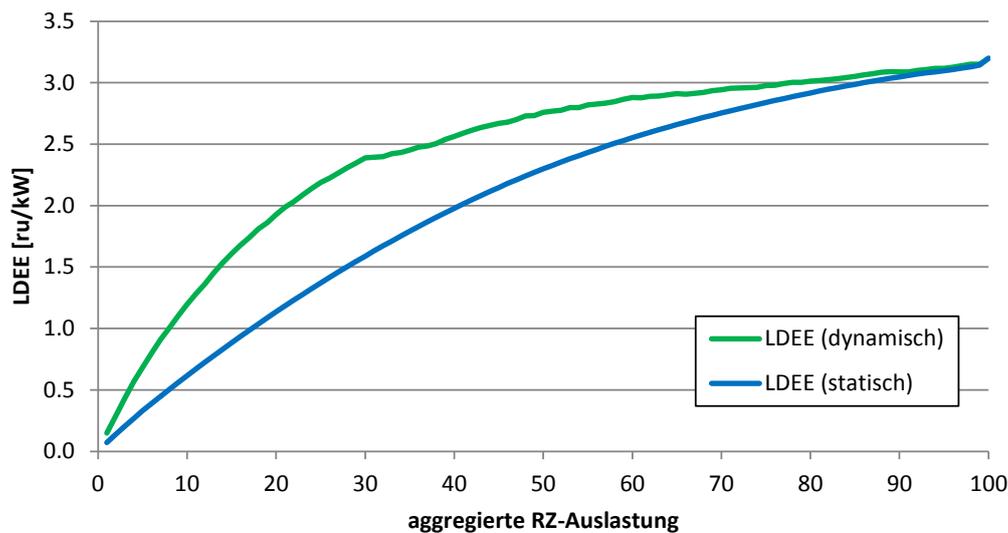


Abbildung 6.11: Vergleich der messunabhängigen LDEE für den dynamischen und den statischen Betrieb bei einer Außentemperatur von 15°C

Aus dem Energieeffizienzverlauf erhalten Anwender Informationen über die möglichen Betriebspunkte des Rechenzentrums, die bei der Entscheidungsfindung helfen, ohne dass konkrete Messdaten vorliegen müssen. So kann beispielsweise aus dem Verlauf abgeleitet werden, ob es bei einem aktuell ineffizienten Betrieb sinnvoller ist direkt HW-Komponenten auszutauschen oder zunächst die Betriebsparameter anzupassen (z.B. durch Erhöhung der Auslastung oder Optimierung der Klimaregelung). Ebenso liefert der Verlauf Informationen zu der Skalierbarkeit bzgl. aggregierter RZ-Auslastung und Temperaturen, so dass das RZ-Verhalten bereits vor einer Anpassung der Parameter abgeschätzt werden kann.

Flexibilität (R 3.9) Die Anforderung der Flexibilität verlangt von einer Metrik, dass sie nicht nur starr auf einer konkreten Installation angewandt werden kann, sondern dass sie die Bewertung von potentiellen Änderungen erlaubt. Die LDEE erfüllt diese Anforderung, in dem die Performance- und Powermodelle der installierten Hardwarekomponenten mit vorhandenen Modellen von potentiell neuer Hardware ausgetauscht werden können. Am Beispiel des RZ C wird dieser Anwendungsfall demonstriert. So werden hypothetisch 50 der nach $LDEE_r$ ineffizientesten Server (HP ProLiant DL380 G5 mit $4 \times 1,6$ GHz von 2009, vgl. Tabelle 6.5) direkt durch ein aktuelleres Servermodell (Dell PowerEdge R820 mit $32 \times 2,3$ GHz von 2012) ausgetauscht und die LDEE für die resultierende theoretische Konstellation ermittelt. Die Ergebnisse des Vergleichs zwischen dem ursprünglichen Rechenzentrum C mit dem hypothetischen RZ C_a sind in Tabelle 6.6 aufgeführt.

	RZ C	RZ C_a
Server	200	200
VMs	1400	1400
RZ-Gesamtauslastung	0,19	0,11
RZ-Leistungsaufnahme [kW] (dynamisch)	14,3	9,4
IT-Leistungsaufnahme [kW] (dynamisch)	12,7	8,2
LDEE _r [ru/kW] (dynamisch)	0,72	1,63
RZ-Leistungsaufnahme [kW] (statisch)	27,3	17,7
IT-Leistungsaufnahme [kW] (statisch)	24,2	15,7
LDEE _r [ru/kW] (statisch)	0,49	1,18

Tabelle 6.6: Vergleich der Auswirkungen eines potentiellen Servertauschs am Beispiel des RZ C

Beim Betrieb der gleichen VM-Menge ist die aggregierte Auslastung des RZ C_a deutlich geringer, da durch das aktualisierte Servermodell mehr Systemressourcen zur Verfügung stehen als durch die VM tatsächlich benötigt werden. Eine niedrigere Auslastung verringert zwar die LDEE, durch die Kombination mit der nun deutlich gestiegenen Performance wird dies jedoch wieder ausgeglichen. Da zudem die Leistungsaufnahme der IT und damit des gesamten Rechenzentrums durch den Serveraustausch sinkt, verbessert sich die repräsentative LDEE_r insgesamt um einen Faktor von 2,26 für den dynamischen und 2,41 für den statischen Betrieb. Im statischen Betrieb ist die Verbesserung noch etwas höher, da im dynamischen Betrieb bereits bevorzugt die ineffizientesten Server – und damit die hypothetisch ausgetauschten – deaktiviert wurden. Ein potentieller Austausch würde neben der erhöhten Rechenkapazität eine Energieeinsparung von ca. 35% erzielen.

Die Flexibilität der LDEE ermöglicht nicht nur den Austausch von Komponentenmodellen zur Untersuchung theoretischer RZ-Anpassungen, sondern kann auch zur Exploration von Konfigurationsänderungen genutzt werden. Obligatorisch ist die bereits demonstrierte Untersuchung von veränderten Arbeitslasten und klimatischen Bedingungen, die als direkte Parameter in die LDEE einfließen. Ein weiteres Beispiel wurde ebenfalls bereits im Rahmen dieser Evaluation gezeigt, indem unterschiedliche Ansätze des Last- und Powermanagements angewandt wurden (statisch/dynamisch). Durch Anpassungen der Parameter des IT-Lastmodells (vgl. Abschnitt 4.3) lassen sich hier noch weitere Varianten bewerten.

7 Fazit

Zur Unterstützung eines energieeffizienten Rechenzentrumsbetriebs wurde im Rahmen dieser Arbeit die Load Dependent Energy Efficiency als Bewertungs- und Vergleichsmetrik der Energieeffizienz von Rechenzentren entwickelt. Im Gegensatz zu verbreiteten Metriken wie der Power Usage Effectiveness oder der Data Center Energy Productivity setzt die LDEE dabei keine im Produktivbetrieb gemessenen Daten voraus, sondern basiert auf vorcharakterisierten, auslastungsabhängigen Komponentenmodellen, die öffentlich verfügbare Daten nutzen. Dadurch kann die LDEE auch für weitere Anwendungsfälle eingesetzt werden als nur einen Überblick über die Energieeffizienz zu liefern. So ist ein objektiver, vollständiger Vergleich der Energieeffizienz zwischen Rechenzentren unabhängig der konkreten Arbeitslasten möglich, indem die LDEE jeweils für die identische, aggregierte Auslastung angewandt wird. Bei typischen Metriken fließen hingegen durch die Nutzung von Messdaten die im RZ betriebenen Anwendungen mit ein, so dass eine isolierte Bewertung der Hardware nicht möglich ist. Durch die Basis der Komponentenmodelle lässt sich die Energieeffizienz mit der LDEE zudem auf einzelne Komponenten herunterbrechen, so dass die Quellen für Ineffizienzen bis auf Komponentenebene identifizierbar sind. Des Weiteren ist durch den Austausch von Modellen die Abschätzung der Energieeffizienz bzgl. potentieller Veränderungen am Rechenzentrum möglich, sofern entsprechende Modelle vorhanden sind.

Die grundsätzliche Ausrichtung der LDEE – die Metrik aus auslastungsabhängigen Komponentenmodellen zusammensetzen – ergibt sich aus den Anforderungen, die von den beschriebenen Anwendungsfällen abgeleitet wurden. Von zentraler Bedeutung sind dabei die Anforderungen der Auslastungsabhängigkeit (R 3.2) und der Messunabhängigkeit (R 3.8), die die Ausrichtung des modellbasierten Ansatzes wesentlich beeinflussen. Jede RZ-Komponente wird dabei separat bzgl. ihrer Leistungsaufnahme (Power) und die IT-Hardware zusätzlich bzgl. ihrer Performance modelliert und anhand von öffentlich verfügbaren Daten wie Benchmarkergebnissen oder Datenblättern charakterisiert. Die jeweiligen Modelle liefern dann für eine gegebene Auslastung, deren Verteilung über ein IT-Lastmodell für das gesamte Rechenzentrum abgeschätzt wird, die entsprechende Power oder Performance. In Evaluationen hatten die Modelle mit Ausnahme der Freiluftkühlung jeweils eine Abweichung von unter 5%. Die Einzelergebnisse der Komponentenmodelle werden dann geeignet kombiniert um die Power bzw. die mit der Auslastung skalierte – also die genutzte – Performance des gesamten Rechenzentrums darzustellen. Anschließend werden diese aggregierten Ergebnisse zur Ermittlung der Energieeffizienz zueinander in Relation gesetzt. Durch diesen

Ansatz können sowohl funktionale Anforderungen wie die Vergleichbarkeit (R 3.5) oder die Betrachtung mehrerer Systemebenen (R 3.7) als auch nichtfunktionale Anforderungen wie die Flexibilität (R 3.9) oder die Anwendbarkeit (R 3.10) umgesetzt werden. Die jeweiligen Umsetzungen der Anforderungen wurden im Rahmen von Evaluationen untersucht und diskutiert.

Da vor allem im Infrastrukturbereich nicht gewährleistet werden kann, dass alle benötigten Daten ohne direkte Messungen vorliegen, ist der optionale Einsatz einer Ersatzkennzahl möglich. Hierzu wird die Infrastructure Power Adaptability Metrik, die ebenso im Rahmen dieser Arbeit entwickelt wurde, genutzt. Die IPA ermittelt aus dem PUE-Verlauf den Grad der Adaptivität der Infrastruktur gegenüber der IT. In Kombination mit der PUE kann sie dann für eine gegebene Auslastung die Leistungsaufnahme der Infrastruktur abschätzen. In einer Evaluation lag die Abweichung der Ersatzkennzahl gegenüber der Nutzung von Infrastrukturmodellen bei 1,1%. Eine Anwendung der IPA erfordert allerdings die Ermittlung eines PUE-Verlaufs, so dass mindestens die gesamte Energieaufnahme und die der IT in einer geeigneten Auflösung gemessen werden müssen.

Die Anforderung der Anwendbarkeit (R 3.10) setzt neben der möglichen Anwendung im Produktivbetrieb auch eine einfache Umsetzbarkeit voraus. Da sich die LDEE aus vielen Einzelmodellen zusammensetzt, steht eine manuelle Anwendung der Metrik dem allerdings entgegen. Entsprechend wurde im Rahmen der Arbeit eine Integration der LDEE in die Unternehmensprozesse zur Überwachung und Verwaltung von Rechenzentren untersucht. Der Ansatz orientiert sich dabei an einer direkten Nutzung von verbreiteten Managementwerkzeugen wie Data Center Infrastructure Management und/oder Cloud Computing Managern über die angebotenen Schnittstellen. Die konkrete Umsetzung unterscheidet sich zwischen den genutzten Systemen, das Grundprinzip bleibt allerdings gleich und lässt sich vom beschriebenen Szenario ableiten.

Ausblick

In der beschriebenen Variante liegt der Fokus einer Anwendung der LDEE auf der Energieeffizienz. Wie in Abschnitt 3.1.6 beschrieben wurde, sind für RZ-Verantwortliche neben der Effizienz allerdings auch die direkten Auswirkungen auf die Kosten von Interesse. Diese wurden in der präsentierten Form der LDEE nicht definiert, da zum einen der Fokus auf der Energieeffizienz liegen sollte und da zum anderen Informationen zu den Kapital- und Betriebskosten den Anwendern der Metrik nicht zwangsläufig vorliegen oder die Informationen zumindest manuell in die Metrik integriert werden müssten. Wenn diese Daten ebenso wie die Betriebsdaten über die gängigen Managementplattformen gesammelt werden, ist eine optionale Erweiterung der LDEE zur Ermittlung der Kosteneffizienz sinnvoll. Damit ließen sich potentielle Optimierungsmaßnahmen direkt hinsichtlich ihrer Kostenersparnis untersuchen. Unter Betrachtung der notwendigen Investitionen wäre dann auch der Return on Investment ableitbar.

Mit dem Server Efficiency Rating Tool [Sta13] hat SPEC ein Werkzeug zur Bewertung von Servern veröffentlicht, welches bewusst nicht als Benchmark bezeichnet wird. Der Fokus dieses Werkzeugs liegt auf einer allgemeinen Energieeffizienzbewertung anhand unterschiedlicher Arbeitslasten ohne Berücksichtigung des installierten Betriebssystems. Dadurch soll vermieden werden, dass Ergebnisoptimierungen durch Feinabstimmungen auf der Softwareseite durchgeführt werden, wie sie bei typischen Benchmarks gebräuchlich sind. Für die Vergleichbarkeit stellt die Feinoptimierung der Benchmarkdurchläufe zwar kein Problem dar, da sie von allen Parteien durchgeführt wird. Nichtsdestotrotz würde die Integration der SERT das Spektrum der Arbeitslasten erweitern und die Abhängigkeiten zwischen Performance und Power könnten noch besser durch das Serverpowermodell abgebildet werden. Bislang werden allerdings noch keine Ergebnisse des SERT veröffentlicht.

Auf der Infrastrukturseite gibt es bislang noch keine zentralen, öffentlich zugänglichen Informationsquellen für die Betriebsparameter. Diese müssen aktuell häufig aus Datenblättern abgeleitet werden, wenn sie nicht direkt gemessen werden. Eine Initiative zur zentralen Vermessung entsprechender Komponenten analog zu Performancebenchmarks der IT-Systeme würde den Bewertungsprozess deutlich vereinfachen und außerdem zu genaueren Ergebnissen führen. Die Durchführung einer solchen Initiative wäre zum Beispiel durch Instanzen wie die US Environmental Protection Agency (EPA)¹ (z.B. im Rahmen des Energy Star-Programms) oder das Institute for Energy and Transport (IET)² der Europäischen Kommission (z.B. im Rahmen der Codes of Conduct) denkbar.

¹<http://www3.epa.gov/> (letzter Zugriff: 26.11.2015)

²<http://iet.jrc.ec.europa.eu/energyefficiency/> (letzter Zugriff: 26.11.2015)

Literaturverzeichnis

- [AAF⁺09] ALLALOUF, MIRIAM, YURIY ARBITMAN, MICHAEL FACTOR, RONEN I KAT, KALMAN METH und DALIT NAOR: *Storage modeling for power estimation*. In: *Proceedings of SYSTOR 2009: The Israeli Experimental Systems Conference*, Seite 3. ACM, 2009.
- [AAF12] AZEVEDO, DAN, VICTOR AVELAR und ALAN FRENCH: *PUE: a Comprehensive Examination of the Metric*. White Paper #49 – The Green Grid, 2012. http://www.thegreengrid.org/~media/WhitePapers/WP49-PUE%20A%20Comprehensive%20Examination%20of%20the%20Metric_v6.pdf?lang=en (letzter Zugriff: 21.01.2016).
- [AFLV08] AL-FARES, MOHAMMAD, ALEXANDER LOUKISSAS und AMIN VAHDAT: *A scalable, commodity data center network architecture*. ACM SIGCOMM Computer Communication Review, 38(4):63–74, 2008.
- [AHR08] AHRI: *210/240 – Standard for Performance Rating of Unitary Air Conditioning and Air-Source Heat Pump Equipment*. Air-Conditioning, Heating, and Refrigeration Institute, 2008. http://www.ahrinet.org/App_Content/ahri/files/standards%20pdfs/ANSI%20standards%20pdfs/ANSI.AHRI%20Standards%20210.240%20with%20Addenda%201%20and%202.pdf (letzter Zugriff: 21.01.2016).
- [Alb72] ALBERT, ARTHUR: *Regression and the Moore-Penrose pseudoinverse*. Elsevier, 1972.
- [Alf04] ALFA LAVAL: *The theory behind heat transfer – Plate heat exchangers*. White Paper, 2004. <http://www.rybsa.com.mx/pdf/ALFA%20LAVAL/The%20theory%20behind%20heat%20transfer.pdf> (letzter Zugriff: 21.01.2016).
- [AMC⁺08] ANDERSON, D., P. MORRIS, T. CADER, A. RAWSON, T. DARBY, F. RAWSON et al.: *A Framework for Data Center Energy Productivity*. White Paper #13 – The Green Grid, 2008. <http://www.thegreengrid.org/~media/WhitePapers/WhitePaper13FrameworkforDataCenterEnergyProductivity5908.ashx?lang=en> (letzter Zugriff: 21.01.2016).

- [AP03] ARREGOCES, MAURICIO und MAURIZIO PORTOLANI: *Data center fundamentals*. Cisco Press, 2003.
- [ASH12] ASHRAE TECHNICAL COMMITTEE 9.9: *Thermal Guidelines for Data Processing Environments*. American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), 2012.
- [BAN⁺11] BASMADJIAN, ROBERT, NASIR ALI, FLORIAN NIEDERMEIER, HERMANN DE MEER und GIOVANNI GIULIANI: *A methodology to predict the power consumption of servers in data centres*. In: *Proceedings of the 2nd international conference on energy-efficient computing and networking*, Seiten 1–10. ACM, 2011.
- [BCH13] BARROSO, LUIZ ANDRÉ, JIMMY CLIDARAS und URS HÖLZLE: *The datacenter as a computer: An introduction to the design of warehouse-scale machines*. *Synthesis lectures on computer architecture*, 8(3):1–154, 2013.
- [BD08] BEAN, JOHN und KEVIN DUNLAP: *Energy efficient cooling for data centers: a close-coupled row solution*. *ASHRAE Journal*, 1(10):34–42, 2008.
- [BDF⁺03] BARHAM, P., B. DRAGOVIC, K. FRASER, S. HAND, T. HARRIS, A. HO, R. NEUGEBAUER, I. PRATT und A. WARFIELD: *Xen and the art of virtualization*. In: *Proceedings of the nineteenth ACM symposium on Operating systems principles*, Seiten 164–177. ACM, 2003.
- [BH07] BARROSO, L.A. und U. HÖLZLE: *The case for energy-proportional computing*. *Computer*, 40(12):33–37, 2007.
- [BK12] BAEHR, HANS DIETER und STEPHAN KABELAC: *Thermodynamik: Grundlagen und technische Anwendungen*. Springer-Verlag, 2012.
- [BKR⁺13] BALLANI, H., T. KARAGIANNIS, A. ROWSTRON, P. COSTA und V. JALAPARTI: *Predicting datacenter performance to improve provisioning*, August 22 2013. US Patent App. 13/400,579.
- [BKST13] BRADY, GEMMA A, NIKIL KAPUR, JONATHAN L SUMMERS und HARVEY M THOMPSON: *A case study and critical assessment in calculating power usage effectiveness for a data centre*. *Energy Conversion and Management*, 76:155–161, 2013.
- [BKZ⁺13] BILAL, KASHIF, SAMEE U KHAN, LIMIN ZHANG, HONGXIANG LI, KHIZAR HAYAT, SAJJAD A MADANI, NASRO MIN-ALLAH, LIZHE WANG, DAN CHEN, MAJID IQBAL et al.: *Quantitative comparisons of the state-of-the-art data center architectures*. *Concurrency and Computation: Practice and Experience*, 25(12):1771–1783, 2013.

- [BR10] BROWN, DAVID J und CHARLES REAMS: *Toward energy-efficient computing*. Communications of the ACM, 53(3):50–58, 2010.
- [Bri07] BRILL, K.G.: *Data center energy efficiency and productivity*. White Paper – The Uptime Institute, 2007. <https://connect.ufl.edu/cns/DC0/ecdc/ECDC%20Construction%20Project/Whitepapers/Data%20Center%20Energy%20Efficiency%20%28Uptime%20Inst%29.pdf> (letzter Zugriff: 21.01.2016).
- [BRPC07] BELADY, C., A. RAWSON, J. PFLUEGER und T. CADER: *The Green Grid Data Center Power Efficiency Metrics: PUE and DCiE*. White Paper #6 – The Green Grid, 2007. http://www.premiersolutionsco.com/wp-content/uploads/TGG_Data_Center_Power_Efficiency_Metrics_PUE_and_DCiE.pdf (letzter Zugriff: 21.01.2016).
- [Bru14] BRUNEO, DARIO: *A stochastic model to investigate data center performance and qos in iaas cloud computing systems*. Parallel and Distributed Systems, IEEE Transactions on, 25(3):560–569, 2014.
- [BWP+10] BREEN, THOMAS J, EJ WALSH, JEFF PUNCH, AMIP J SHAH und CULLEN E 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*, Seiten 1–10. IEEE, 2010.
- [BYV+09] BUYYA, RAJKUMAR, CHEE SHIN YEO, SRIKUMAR VENUGOPAL, JAMES BROBERG und IVONA BRANDIC: *Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility*. Future Generation computer systems, 25(6):599–616, 2009.
- [CB13] CHABAREK, JOSEPH und PAUL BARFORD: *Energy Audit: Monitoring power consumption in diverse network environments*. In: *Green Computing Conference (IGCC), 2013 International*, Seiten 1–10. IEEE, 2013.
- [CFH+05] CLARK, CHRISTOPHER, KEIR FRASER, STEVEN HAND, JACOB GORM HANSEN, ERIC JUL, CHRISTIAN LIMPACH, IAN PRATT und ANDREW WARFIELD: *Live migration of virtual machines*. In: *Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation – Volume 2*, Seiten 273–286. USENIX Association, 2005.

- [Cis15] CISCO SYSTEMS INC.: *The Zettabyte Era: Trends and Analysis*. White Paper, 2015. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.pdf (letzter Zugriff: 21.01.2016).
- [CKH07] CANTNER, U., J. KRÜGER und H. HANUSCH: *Produktivitäts- und Effizienzanalyse: Der nichtparametrische Ansatz*. Springer-Lehrbuch. Springer Berlin Heidelberg, 2007.
- [CSB⁺08] CHABAREK, JOSEPH, JOEL SOMMERS, PAUL BARFORD, CRISTIAN ESTAN, DAVID TSIANG und STEVE WRIGHT: *Power awareness in network design and routing*. In: *INFOCOM 2008. The 27th Conference on Computer Communications*. IEEE, 2008.
- [DBSP⁺11] DU BOIS, KRISTOF, TIM SCHAEPS, STIJN POLFLIET, FREDERICK RYCKBOSCH und LIEVEN EECKHOUT: *SWEEP: evaluating computer system energy efficiency using synthetic workloads*. In: *Proceedings of the 6th International Conference on High Performance and Embedded Architectures and Compilers*, Seiten 159–166. ACM, 2011.
- [DH09] DARROW, KEN und BRUCE HEDMAN: *Opportunities for combined heat and power in data centers*. *Cogeneration and Distributed Generation Journal*, 24(3):78, 2009.
- [Die05] DIEDERICHS, HENNER: *Komplexitätsreduktion in der Softwareentwicklung: ein systemtheoretischer Ansatz*, Band 3. BoD – Books on Demand, 2005.
- [DJ13] DUKARIC, ROBERT und MATJAZ B JURIC: *Towards a unified taxonomy and architecture of cloud frameworks*. *Future Generation Computer Systems*, 29(5):1196–1210, 2013.
- [DR14] DUNLAP, KEVIN und NEIL RASMUSSEN: *Choosing between room, row, and rack-based cooling for data centers*. White Paper 130 – APC, 2014. <http://iisgroup11c.com/wp-content/uploads/2013/02/APC-White-Paper-130-Choosing-Between-Room-Row-and-Rack-based-Cooling-for-Data-Centers.pdf> (letzter Zugriff: 21.01.2016).
- [Eba13] EBAY INC.: *Digital Service Efficiency*. White Paper, 2013. <http://tech.ebay.com/sites/default/files/publications/Ebay-DSE-130523.pdf> (letzter Zugriff: 21.01.2016).
- [EI09] ELLSWORTH, MICHAEL J und MADHUSUDAN K IYENGAR: *Energy efficiency analyses and comparison of air and water cooled high*

- performance servers. In: *ASME 2009 InterPACK Conference collocated with the ASME 2009 Summer Heat Transfer Conference and the ASME 2009 3rd International Conference on Energy Sustainability*, Seiten 907–914. American Society of Mechanical Engineers, 2009.
- [Eme09] EMERSON NETWORK POWER: *Energy logic: Reducing data center energy consumption by creating savings that cascade across systems*. White Paper – Emerson Electric Co, 2009. <http://www.emersonnetworkpower.com/documentation/en-us/latest-thinking/edc/documents/white%20paper/energylogicreducingdatacenterenergyconsumption.pdf> (letzter Zugriff: 21.01.2016).
- [Eme15] EMERSON NETWORK POWER: *Liebert EFC, the Highly Efficient Indirect Evaporative Freecooling Solution*. Product Brochure – Emerson Network Power, 2015. http://www.emersonnetworkpower.com/en-EMEA/Products/PrecisionCooling/Evaporative-Freecooling/Documents/Liebert-EFC/Liebert%20EFC300-PER%20STAMPA_LR.pdf (letzter Zugriff: 21.01.2016).
- [ERKR06] ECONOMOU, DIMITRIS, SUZANNE RIVOIRE, CHRISTOS KOZYRAKIS und PARTHA RANGANATHAN: *Full-system power analysis and modeling for server environments*. International Symposium on Computer Architecture-IEEE, 2006.
- [Eva12] EVANS, TONY: *The different technologies for cooling data centers*. White Paper 59 – APC, 2012. http://www.apcmedia.com/salestools/VAVR-5UDTU5/VAVR-5UDTU5_R2_EN.pdf?sdirect=true (letzter Zugriff: 21.01.2016).
- [FE10] FURHT, BORIVOJE und ARMANDO ESCALANTE: *Handbook of cloud computing*, Band 3. Springer, 2010.
- [FM05] FURLONG, JAMES W und FRANK T MORRISON: *Optimization of water-cooled chiller-cooling tower combinations*. CTI JOURNAL, 26(2):14, 2005.
- [FW86] FLEMING, PHILIP J und JOHN J WALLACE: *How not to lie with statistics: the correct way to summarize benchmark results*. Communications of the ACM, 29(3):218–221, 1986.
- [FWB07] FAN, XIAOBO, WOLF-DIETRICH WEBER und LUIZ ANDRE BARROSO: *Power provisioning for a warehouse-sized computer*. In: *ACM SIGARCH Computer Architecture News*, Band 35, Seiten 13–23. ACM, 2007.

- [Gen14] GENG, HWAIYU: *Data Center Handbook*. Wiley, 2014.
- [GHMP08] GREENBERG, ALBERT, JAMES HAMILTON, DAVID A MALTZ und PARVEEN PATEL: *The cost of a cloud: research problems in data center networks*. ACM SIGCOMM computer communication review, 39(1):68–73, 2008.
- [Gre12] GREEN IT PROMOTION COUNCIL: *New Data Center Energy Efficiency Evaluation Index DPPE (Datacenter Performance per Energy) Measurement Guidelines*. Richtlinie – Green IT Promotion Council (GIPC), 2012. http://home.jeita.or.jp/greenit-pc/topics/release/pdf/dppe_e_DPPE_Measurement_Guidelines.pdf (letzter Zugriff: 21.01.2016).
- [GRU04] GRUNDFOS: *Pump Handbook*. Handbuch – GRUNDFOS Management A/S, 2004. http://net.grundfos.com/doc/webnet/mining/_downloads/pump-handbook.pdf (letzter Zugriff: 21.01.2016).
- [GVM00] GIBSON, GARTH A und RODNEY VAN METER: *Network attached storage architecture*. Communications of the ACM, 43(11):37–45, 2000.
- [GWT⁺08] GUO, CHUANXIONG, HAITAO WU, KUN TAN, LEI SHI, YONGGUANG ZHANG und SONGWU LU: *Dcell: a scalable and fault-tolerant network structure for data centers*. ACM SIGCOMM Computer Communication Review, 38(4):75–86, 2008.
- [HA14] HILTY, L. und B. AEBISCHER: *ICT Innovations for Sustainability*. Advances in Intelligent Systems and Computing. Springer International Publishing, 2014.
- [HC14] HINTEMANN, RALPH und JENS CLAUSEN: *Rechenzentren in Deutschland: Eine Studie zur Darstellung der wirtschaftlichen Bedeutung und der Wettbewerbssituation*. Studie im Auftrag des Bundesverbandes Informationswirtschaft, Telekommunikation und neue Medien e. V. (Bitkom) – Borderstep, 2014. <https://www.bitkom.org/Publikationen/2014/Studien/Studie-zu-Rechenzentren-in-Deutschland-Wirtschaftliche-Bedeutung-und-Wettbewerbssituation/Borderstep-Institut-Studie-Rechenzentren-in-Deutschland-05-05-20141.pdf> (letzter Zugriff: 21.01.2016).
- [HDC⁺05] HEATH, TALIVER, BRUNO DINIZ, ENRIQUE V CARRERA, WAGNER MEIRA JR und RICARDO BIANCHINI: *Energy conservation in heterogeneous server clusters*. In: *Proceedings of the tenth ACM*

- SIGPLAN symposium on Principles and practice of parallel programming*, Seiten 186–195. ACM, 2005.
- [Her13a] HERZOG, CHRISITAN: *Betriebssicheres Rechenzentrum – Leitfaden*. BITKOM Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V., 2013. <https://www.bitkom.org/Bitkom/Publicationen/Betriebssicheres-Rechenzentrum.html> (letzter Zugriff: 21.01.2016).
- [Her13b] HERZOG, CHRISITAN: *Planungshilfe betriebssicheres Rechenzentrum*. BITKOM Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V., 2013. https://www.bitkom.org/Bitkom/Publicationen/Publication_1349.html (letzter Zugriff: 21.01.2016).
- [Her15] HERZOG, CHRISTIAN: *Energieeffizienz im Rechenzentrum – Leitfaden*. BITKOM Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V., 2015. <https://www.bitkom.org/Publicationen/2015/Leitfaden/LF-Energieeffizienz-in-Rechenzentren/150911-LF-Energieeffizienz-in-RZ.pdf> (letzter Zugriff: 21.01.2016).
- [Hew09] HEWLETT-PACKARD DEVELOPMENT COMPANY, L.P.: *Turning Cost Centers into Dynamic Technology Assets*. White Paper – Hewlett Packard, 2009. http://whitepaper.heise.de/_misc/download.cfm?filepath=/fileserver/whitepaperheise/vzones_tandardcontents/393.pdf&filename=Turning%20cost%20centres%20into%20dynamic%20technology%20assets_4AA2-7740EEW.hires&filetype=pdf (letzter Zugriff: 21.01.2016).
- [HFS10] HINTEMANN, RALPH, K FICHTER und L STOBBE: *Materialbestand der Rechenzentren in Deutschland – Eine Bestandsaufnahme zur Ermittlung von Ressourcen- und Energieeinsatz*. Umweltbundesamt – Texte, 55, 2010. <http://www.uba.de/uba-info-medien/4037.html> (letzter Zugriff: 21.01.2016).
- [HLB12] HUPPLER, KARL, KLAUS-DIETER LANGE und JOHN BECKETT: *SPEC: Enabling Efficiency Measurement*. In: *Proceedings of the 3rd ACM/SPEC International Conference on Performance Engineering*, ICPE '12, Seiten 257–258, New York, NY, USA, 2012. ACM.
- [HMP⁺09] HAAS, J., M. MONROE, J. PFLUEGER, J. POUCHET, P. SNELLING, A. RAWSON und F. RAWSON: *Proxy proposals for measuring data center productivity*. White Paper #17 – The Green Grid, 2009. <http://www.thegreengrid.org/~media/WhitePapers/Wh>

- ite%20Paper%2017%20-%20Proxies%20Proposals%20for%20Measuring%20Data%20Center%20Efficiencyv2.pdf (letzter Zugriff: 21.01.2016).
- [HRC13] HANKENDI, CAN, SHERIEF REDA und AYSE K COSKUN: *vCap: Adaptive power capping for virtualized servers*. In: *Proceedings of the International Symposium on Low Power Electronics and Design*, Seiten 415–420. IEEE Press, 2013.
- [HSSN11] HOYER, M., K. SCHRÖDER, D. SCHLITT und W. NEBEL: *Proactive dynamic resource management in virtualized data centers*. In: *Proceedings of the 2nd International Conference on Energy-Efficient Computing and Networking*, Seiten 11–20. ACM, 2011.
- [Inc11] INCROPERA, FRANK P: *Fundamentals of heat and mass transfer*. John Wiley & Sons, 2011.
- [KCR13] KREIDER, JAN F, PETER S CURTISS und ARI RABL: *Heating and cooling of buildings: design for efficiency*. CRC Press, 2013.
- [KFK08] KAPLAN, J., W. FORREST und N. KINDLER: *Revolutionizing data center energy efficiency*. White Paper – McKinsey & Company, 2008. http://www.sallan.org/pdf-docs/McKinsey_Data_Center_Efficiency.pdf (letzter Zugriff: 21.01.2016).
- [Kin12] KINGSTON TECHNOLOGY CORPORATION: *Memory Module Specifications - KVR1333D3N9/8G*. Data Sheet, 2012. http://www.kingston.com/dataSheets/KVR1333D3N9_8G.pdf (letzter Zugriff: 21.01.2016).
- [Kin15] KINGSTON TECHNOLOGY CORPORATION: *Technical Brief - Power Benchmark*, 2015. http://www.kingston.com/us/business/server_solutions/power_benchmark# (letzter Zugriff: 21.01.2016).
- [Koo07] KOOMEY, J.G.: *Estimating total power consumption by servers in the US and the world*. Final Report, 2007. http://ccsl.iccip.net/koomey_estimating_technicalnote.pdf (letzter Zugriff: 21.01.2016).
- [Koo11] KOOMEY, JONATHAN: *Growth in data center electricity use 2005 to 2010*. A report by Analytical Press, completed at the request of The New York Times, 2011. http://www.co.twosides.info/download/Koomey_Johnathon_G_-_Growth_In_Data_Center_Electricity_Use_2005_to_2010_2011.pdf (letzter Zugriff: 21.01.2016).
- [Kop13] KOPPY, JENNIFER: *IDC MarketScape: Worldwide Datacenter Infrastructure Management 2013 Vendor Analysis*. Survey – IDC

- Corporate USA, 2013. <http://assets.fiercemarkets.net/public/sites/energy/reports/idcdatareport.pdf> (letzter Zugriff: 21.01.2016).
- [KSL⁺10] KIM, WONHO, PUNEET SHARMA, JEONGKEUN LEE, SUJATA BANNERJEE, JEAN TOURRILHES, SUNG-JU LEE und PRAVEEN YALAGANDULA: *Automated and scalable QoS control for network convergence*. Proc. INM/WREN, 10:1–1, 2010.
- [KSS⁺13] KOLLER, PETER-JULIAN, SANDRA SCHULZ, HARALD A. SUMMA, MICHAEL OPITZ, LARS RIEGEL, NICOLAI SCHÄTTGEN, KAREN SCHELB, NORBERT POHLMANN, MICHAEL SPARENBERG und CLAUDIO KERST: *Die deutsche Internetwirtschaft 2012-2016 – Zahlen, Trends und Thesen*. Studie – eco / Arthur D. Little, 2013. https://www.eco.de/wp-content/blogs.dir/eco_adl_report2013_web1.pdf (letzter Zugriff: 21.01.2016).
- [Lan09a] LANGE, K.D.: *Identifying shades of green: The SPECpower benchmarks*. Computer, 42(3):95–97, 2009.
- [Lan09b] LANGE, KLAUS-DIETER: *The Next Frontier for Power/Performance Benchmarking: Energy Efficiency of Storage Subsystems*. In: *Computer Performance Evaluation and Benchmarking*, Seiten 97–101. Springer, 2009.
- [LCB⁺12] LIU, ZHENHUA, YUAN CHEN, CULLEN BASH, ADAM WIERMAN, DANIEL GMACH, ZHIKUI WANG, MANISH MARWAH und CHRIS HYSER: *Renewable and cooling aware workload management for sustainable data centers*. In: *ACM SIGMETRICS Performance Evaluation Review*, Band 40, Seiten 175–186. ACM, 2012.
- [Liu11] LIU, HUAN: *A measurement study of server utilization in public clouds*. In: *Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference on*, Seiten 435–442. IEEE, 2011.
- [LT11] LANGE, KLAUS-DIETER und MICHAEL G TRICKER: *The design and development of the server efficiency rating tool (SERT)*. In: *ACM SIGSOFT Software Engineering Notes*, Band 36, Seiten 145–150. ACM, 2011.
- [LWL⁺09] LIU, LIANG, HAO WANG, XUE LIU, XING JIN, WEN BO HE, QING BO WANG und YING CHEN: *GreenCloud: a new architecture for green data center*. In: *Proceedings of the 6th international conference industry session on Autonomic computing and communications industry session*, Seiten 29–38. ACM, 2009.

- [LWQ⁺10] LIU, ZHUO, FEI WU, XIAO QIN, CHANGSHENG XIE, JIAN ZHOU und JIANZONG WANG: *TRACER: A trace replay tool to evaluate energy-efficiency of mass storage systems*. In: *Cluster Computing (CLUSTER), 2010 IEEE International Conference on*, Seiten 68–77. IEEE, 2010.
- [Mah14] MAHMOOD, Z.: *Cloud Computing: Challenges, Limitations and R&D Solutions*. Computer Communications and Networks. Springer International Publishing, 2014.
- [Mas11] MASSE, MARK: *REST API design rulebook*. O'Reilly Media, Inc., 2011.
- [MCRS05] MOORE, JUSTIN D, JEFFREY S CHASE, PARTHASARATHY RANGANATHAN und RATNESH K SHARMA: *Making Scheduling "Cool": Temperature-Aware Workload Placement in Data Centers*. In: *USENIX annual technical conference, General Track*, Seiten 61–75, 2005.
- [MG08] MARQUARD, ULRICH und CLARISSA GÖTZ: *SAP Standard Application Benchmarks – IT Benchmarks with a Business Focus*. In: *Performance Evaluation: Metrics, Models and Benchmarks*, Seiten 4–8. Springer, 2008.
- [MSBB10] MORAN, M.J., H.N. SHAPIRO, D.D. BOETTNER und M.B. BAILEY: *Fundamentals of Engineering Thermodynamics*. Wiley, 2010.
- [MSBR09] MAHADEVAN, PRIYA, PUNEET SHARMA, SUJATA BANERJEE und PARTHASARATHY RANGANATHAN: *A power benchmarking framework for network devices*. In: *NETWORKING 2009*, Seiten 795–808. Springer, 2009.
- [MW11] MEISNER, DAVID und THOMAS F WENISCH: *Does low-power design imply energy efficiency for data centers?* In: *Proceedings of the 17th IEEE/ACM international symposium on Low-power electronics and design*, Seiten 109–114. IEEE Press, 2011.
- [Nap05] NAPHON, PAISARN: *Study on the heat transfer characteristics of an evaporative cooling tower*. *International communications in heat and mass transfer*, 32(8):1066–1074, 2005.
- [ND08] NILES, SUZANNE und PATRICK DONOVAN: *Virtualization and cloud computing: Optimized power, cooling, and management maximizes benefits*. White Paper 118 – APC, 2008. http://www.schneider-electric.com/solutions/it/it/m/d/4664738/application/pdf/1240_snis-7aulcp_r3_en.pdf (letzter Zugriff: 21.01.2016).

- [NLLS12] NEWCOMBE, L., Z. LIMBUWALA, P. LATHAM und V. SMITH: *Data centre Fixed to Variable Energy Ratio metric DC-FVER*. White Paper – BCS Data Centre Specialist Group, 2012. http://www.bcs.org/upload/pdf/dc_fver_metric_v1.0.pdf (letzter Zugriff: 21.01.2016).
- [Pat08] PATTERSON, DAVID A: *The data center is the computer*. Communications of the ACM, 51(1):105–105, 2008.
- [Peh10] PEHNT, MARTIN: *Energieeffizienz: Ein Lehr- und Handbuch*. Springer-Verlag, 2010.
- [Pfl08] PFLUEGER, J.: *Re-defining the 'green' data center*. White Paper – DELL, 2008. <http://www.gobiztech.com/docs/Dell-Redefining-Green-Datacenter.pdf> (letzter Zugriff: 21.01.2016).
- [PMSR09] PATNAIK, DEBPRAKASH, MANISH MARWAH, RATNESH SHARMA und NAREN RAMAKRISHNAN: *Sustainable Operation and Management of Data Center Chillers Using Temporal Data Mining*. In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, Seiten 1305–1314, New York, NY, USA, 2009. ACM.
- [PMWV09] PELLEY, S., D. MEISNER, T.F. WENISCH und J.W. VANGILDER: *Understanding and abstracting total data center power*. In: *Workshop on Energy-Efficient Design*, 2009.
- [PPH⁺13] PATTERSON, MICHAEL K, STEPHEN W POOLE, CHUNG-HSING HSU, DON MAXWELL, WILLIAM TSCHUDI, HENRY COLES, DAVID J MARTINEZ und NATALIE BATES: *TUE, a new energy-efficiency metric applied at ORNL's jaguar*. In: *Supercomputing*, Seiten 372–382. Springer, 2013.
- [Pul08] PULVERMÜLLER, PATRICK: *Bestandsaufnahme effiziente Rechenzentren in Deutschland*. Studie – eco – Verband der deutschen Internetwirtschaft e.V., 2008. <https://www.eco.de/wp-content/blogs.dir/bestandsaufnahme-effiziente-rechenzentren-in-deutschland.pdf> (letzter Zugriff: 21.01.2016).
- [Ras07a] RASMUSSEN, NEIL: *Calculating total cooling requirements for data centers*. White Paper 25 – APC, 2007. http://apcmedia.com/sales_tools/nran-5te6he/nran-5te6he_r3_en.pdf?sdirect=true (letzter Zugriff: 21.01.2016).
- [Ras07b] RASMUSSEN, NEIL: *Electrical efficiency measurement for data centers*. White Paper 154 – APC, 2007. <http://>

- [//www.apcmedia.com/salestools/NRAN-72754V/NRAN-72754V_R2_EN.pdf?sdirect=true](http://www.apcmedia.com/salestools/NRAN-72754V/NRAN-72754V_R2_EN.pdf?sdirect=true) (letzter Zugriff: 21.01.2016).
- [Ras12] RASMUSSEN, NEIL: *The Different Types of Air Distribution for IT Environments*. White Paper 55 – APC, 2012. http://www.apcmedia.com/salestools/NRAN-5TN9QM/NRAN-5TN9QM_R3_EN.pdf?sdirect=true (letzter Zugriff: 21.01.2016).
- [RPE11] RYCKBOSCH, FREDERICK, STIJN POLFLIET und LIEVEN ECKHOUT: *Trends in server energy proportionality*. *Computer*, (9):69–72, 2011.
- [RRK08] RIVOIRE, SUZANNE, PARTHASARATHY RANGANATHAN und CHRISTOS KOZYRAKIS: *A Comparison of High-Level Full-System Power Models*. *HotPower*, 8:3–3, 2008.
- [RSR⁺07] RIVOIRE, S., M.A. SHAH, P. RANGANATHAN, C. KOZYRAKIS und J. MEZA: *Models and metrics to enable energy-efficiency optimizations*. *Computer*, 40(12):39–48, 2007.
- [RSRK07] RIVOIRE, SUZANNE, MEHUL A SHAH, PARTHASARATHY RANGANATHAN und CHRISTOS KOZYRAKIS: *JouleSort: a balanced energy-efficiency benchmark*. In: *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, Seiten 365–376. ACM, 2007.
- [SAP11] SAP AG: *SAP Power Benchmarks Specification*. Specifications, 2011. http://global.sap.com/solutions/benchmark/pdf/Specification_SAP_Power_Benchmarks_V12.pdf (letzter Zugriff: 21.01.2016).
- [SEZ13] SCHÖDWELL, BJÖRN, KORAY EREK und RÜDIGER ZARNEKOW: *Data Center Green Performance Measurement: state of the art and open research challenges*. 2013.
- [SJS⁺14] SCHLITT, DANIEL, STEFAN JANACEK, GUNNAR SCHOMAKER, KIRIL SCHRÖDER, BERND HANSTEIN, THORSTEN HEIMBERG, STEFFEN LEUKROTH, GERALD KIELMANN, EIKE LÜERSSEN, SIMON OBERTHÜR, RALPH HINTEMANN und KLAUS FICHTER: *Abschlussbericht des Verbundprojekts AC4DC ‚Adaptive Computing for green Data Centers‘*, 2014. <https://www.tib.eu/de/suchen/id/TIBKAT%3A833375288/Adaptive-Computing-for-green-Data-Centers-AC4DC/> (letzter Zugriff: 21.01.2016).
- [SKL14] STANIK, ALEXANDER, MARC KOERNER und LEONIDAS LYMBEROPOULOS: *SLA-driven Federated Cloud Networking: Quality of*

- Service for Cloud-based Software Defined Networks*. *Procedia Computer Science*, 34:655–660, 2014.
- [Sku12] SKURK, HOLGER: *Speichervirtualisierung – Leitfaden*. BITKOM Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V., 2012. https://www.bitkom.org/Publicationen/2012/Leitfaden/Speichervirtualisierung/121119_Speichervirtualisierung.pdf (letzter Zugriff: 21.01.2016).
- [SLB07] STOESS, JAN, CHRISTIAN LANG und FRANK BELLOSA: *Energy Management for Hypervisor-Based Virtual Machines*. In: *USENIX annual technical conference*, Seiten 1–14, 2007.
- [SPX13] SPX: *Free Cooling - Minimizing Energy Costs*. White Paper – SPX Cooling Technologies, Inc., 2013. <http://spxcooling.com/pdf/H-002A.pdf> (letzter Zugriff: 21.01.2016).
- [SSJ13] STOYAN, D., H. STOYAN und U. JANSEN: *Umweltstatistik: Statistische Verarbeitung und Analyse von Umweltdaten*. Teubner – Reihe Umwelt. Vieweg+Teubner Verlag, 2013.
- [SSN14] SCHLITT, DANIEL, GUNNAR SCHOMAKER und WOLFGANG NEBEL: *Gain More from PUE: Assessing Data Center Infrastructure Power Adaptability*. In: *Energy Efficient Data Centers*, Seiten 152–166. Springer, 2014.
- [Sta08] STANDARD PERFORMANCE EVALUATION CORPORATION (SPEC): *SPECsfs2008 User’s Guide*. Specifications, 2008. <http://www.spec.org/sfs2008/docs/usersguide.pdf> (letzter Zugriff: 21.01.2016).
- [Sta11] STANDARD PERFORMANCE EVALUATION CORPORATION (SPEC): *SPEC CPU2006 Run and Reporting Rules*. Specifications, 2011. <http://www.spec.org/cpu2006/Docs/runrules.html> (letzter Zugriff: 21.01.2016).
- [Sta12] STANDARD PERFORMANCE EVALUATION CORPORATION (SPEC): *SPECpower_ssj2008 Design Document SSJ Workload*. Specifications, 2012. http://www.spec.org/power/docs/SPECpower_ssj2008-Design_ssj.pdf (letzter Zugriff: 21.01.2016).
- [Sta13] STANDARD PERFORMANCE EVALUATION CORPORATION (SPEC): *Server Efficiency Rating Tool (SERT) Design Document*. Specifications, 2013. http://www.spec.org/sert/docs/SERT-Design_Document.pdf (letzter Zugriff: 21.01.2016).

- [Sto13] STORAGE PERFORMANCE COUNCIL (SPC): *SPC Benchmark 1 (SPC-1) Official Specification*. Specifications, 2013. http://www.storageperformance.org/specs/SPC-1_SPC-1E_v1.14.pdf (letzter Zugriff: 21.01.2016).
- [Tec08] TECHNISCHE UNIVERSITÄT BERLIN – INNOVATIONSZENTRUM ENERGIE: *Konzeptstudie zur Energie- und Ressourceneffizienz im Betrieb von Rechenzentren*. Studie, 2008. https://www.energie.tu-berlin.de/uploads/media/IZE_Konzeptstudie_Energieeffizienz_in_Rechenzentren.pdf (letzter Zugriff: 21.01.2016).
- [TKBL13] TIGHE, MICHAEL, GASTÓN KELLER, MICHAEL BAUER und HANAN LUTFIYYA: *A distributed approach to dynamic vm management*. In: *2013 9th International Conference on Network and Service Management (CNSM)*, Seiten 166–170. IEEE, 2013.
- [Tor13] TORELL, W: *High efficiency indirect air economizer-based cooling for data centers*. White Paper 136 – APC, 2013. http://www.apcmedia.com/salestools/WTOL-8DNR9U/WTOL-8DNR9U_R1_EN.pdf?sdirect=true (letzter Zugriff: 21.01.2016).
- [TSRB06] TURNER, W PITT, JOHN H SEADER, VINCE RENAUD und KENNETH G BRILL: *Tier classification define site infrastructure performance*. White Paper – The Uptime Institute, 2006. <http://www.fibertown.com/pdf/uptimeinstitutetierysystem.pdf> (letzter Zugriff: 21.01.2016).
- [TVCA10] TSCHUDI, BILL, OTTO VANGEET, J COOLEY und D AZEVEDO: *ERE: A metric for measuring the benefit of reuse energy from a data center*. White Paper #29 – The Green Grid, 2010. http://www.thegreengrid.org/~media/WhitePapers/ERE_WP_101510_v2.pdf (letzter Zugriff: 21.01.2016).
- [VAN08] VERMA, AKSHAT, PUNEET AHUJA und ANINDYA NEOGI: *pMapper: power and migration cost aware application placement in virtualized systems*. In: *Middleware 2008*, Seiten 243–264. Springer, 2008.
- [VLDWF12] VON LASZEWSKI, GREGOR, JAVIER DIAZ, FUGANG WANG und GEOFFREY C FOX: *Comparison of multiple cloud frameworks*. In: *Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on*, Seiten 734–741. IEEE, 2012.
- [Vog10] VOGELSANG, THOMAS: *Understanding the energy consumption of dynamic random access memories*. In: *Proceedings of the 2010*

43rd Annual IEEE/ACM International Symposium on Microarchitecture, Seiten 363–374. IEEE Computer Society, 2010.

- [Wes12] WESKE, MATHIAS: *Business process management: concepts, languages, architectures*. Springer Science & Business Media, 2012.
- [YCN11] YOUNG, ERIK, PAUL CAO und MIKE NIKOLAIEV: *First TPC-energy benchmark: lessons learned in practice*. In: *Performance Evaluation, Measurement and Characterization of Complex Systems*, Seiten 136–152. Springer, 2011.

Abbildungsverzeichnis

1.1	Entwicklung der Serverenergieeffizienz	12
1.2	Aufteilung der Kosten von Rechenzentren	13
2.1	Tier I Rechenzentrum	23
2.2	Tier IV Rechenzentrum	24
2.3	Aufteilung des Energiebedarfs auf RZ-Komponenten	25
2.4	Funktionalität DCIM vs. IT Service Management	33
3.1	Verlauf der LDEE eines nicht optimierten Betriebs	68
3.2	Verlauf der LDEE bei Nutzung von LPM	69
4.1	Kombiniertes RZ-Performancemodell	75
4.2	Kombiniertes RZ-Powermodell	76
4.3	Aufbau des IT-Lastmodells	78
4.4	Maxima der abgeschätzten Planungszeiträume des AC4DC-LPM	82
4.5	Modellierung der Auslastung eines IT-Systems	84
4.6	Beispiel für Abschätzung der CPU-Auslastung	85
4.7	Ablauf der Performancemodellierung	89
4.8	Beispiel für Skalentransformation bzgl. Benchmarkresultate	91
4.9	Beispiel für Zusammensetzung der Referenzmaschine	104
4.10	Beispiel für Anwendung des Serverpowermodells	108
4.11	Anwendung des Serverpowermodells inkl. RAM-Erweiterung	110
4.12	Leistungsaufnahmen von Storagesystemen aus SPC-1	111
4.13	Elektrische Leistung einer Beispiel-USV	114
4.14	Leistungsaufnahmen verschiedener Klimaketten	117
4.15	Beispiel für Anwendung des CRAH-Powermodells	118
4.16	Beispiel für Anwendung des Chillerpowermodells	120
4.17	Betriebsmodi für Chiller und Freiluftkühlung	124
4.18	Evaluation der Abschätzung der Anzahl aktiver Server	126
4.19	Evaluation der Abschätzung der Serverauswahl	127
4.20	Evaluation der Auslastungsabschätzung	128
4.21	Beispielhafte Verläufe der simulierten CPU-Auslastung	131
4.22	Untersuchung der Auslastungsabhängigkeit	132
4.23	Verlauf der genutzten Performance	133
5.1	Unterteilung einer Zeitreihe zur Anwendung der PVar	140
5.2	Auszug der Leistungsaufnahme eines Testrechenzentrums	141

Abbildungsverzeichnis

5.3	Beispiel für Häufigkeiten von Leistungsaufnahmen eines RZ	145
5.4	Auszug der Leistungsaufnahmen eines Produktivrechenzentrums .	146
5.5	Beispielanwendung der PUE Scalability	150
5.6	Datensätze A bis C zur Evaluation der IPA	151
5.7	Datensätze D und E zur Evaluation der IPA	153
5.8	Datensätze F und G zur Evaluation der IPA	154
6.1	Komponentendiagramm zur Automatisierung der LDEE	158
6.2	Pakete der LDEE-Komponente aus Abbildung 6.1	159
6.3	Klassenhierarchie für RZ-Equipment in der StruxureWare API . .	161
6.4	ER-Diagramm zur Speicherung der Komponentendaten	162
6.5	ER-Diagramm zur Integration von Benchmarks und Modellen . .	165
6.6	Vergleich der Metriken für dynamischen Betrieb	174
6.7	Vergleich der Metriken für statischen Betrieb	175
6.8	Relation der LDEE zur Auslastung der aktiven IT-Systeme . . .	178
6.9	Verlauf der LDEE für RZ A im dynamischen Betrieb	183
6.10	Verlauf der LDEE für RZ A im statischen Betrieb	183
6.11	Vergleich der LDEE für dynamischen und statischen Betrieb . . .	184

Tabellenverzeichnis

2.1	Übersicht bekannter Energieeffizienzmetriken	42
3.1	Vergleich bekannter Energieeffizienzmetriken	57
4.1	Vergleich unterschiedlicher Ansätze zur Performanceabschätzung	87
4.2	Performance und Produktivität für unterschiedliche Workloads .	131
4.3	Performance und Produktivität für zwei Testszenarien	134
5.1	Untersuchung der Leistungsaufnahmen aus Abbildung 5.2	141
5.2	Ergebnisse der IPA für unterschiedliche Datensätze	152
5.3	Abweichung durch Nutzung der IPA in der LDEE	155
6.1	In der Evaluation betrachtete Metriken	173
6.2	Resultate der EE-Metriken im Vergleich	174
6.3	Korrelationskoeffizienten der EE-Metriken zur Auslastung	177
6.4	Vergleich unterschiedlicher Formate der LDEE	180
6.5	Anwendung der LDEE auf Serverebene	181
6.6	Anwendung der LDEE bei einem beispielhaften Serveraustausch .	185

Abkürzungsverzeichnis

AC4DC	Adaptive Computing for green Data Centers
API	Application Programming Interface
BHKW	Blockheizkraftwerk
CADE	Corporate Average Data Center Efficiency
CIO	Chief Information Officer
COP	Coefficient of Performance
CRAH	Computer Room Air Handler
CRM	Customer Relationship Management
CTO	Chief Technology Officer
DC	Data Center
DCeP	Data Center Energy Productivity
DCIM	Data Center Infrastructure Management
DIMM	Dual Inline Memory Module
DSE	Digital Service Efficiency
EE	Energieeffizienz
EER	Energy Efficiency Ratio
ERP	Enterprise Resource Planning
EVU	Energieversorgungsunternehmen
FC	Fibre Channel
FLOPS	Floating Point Operations per Second
HDD	Hard Disk Drive
HPC	High-Performance-Computing
I/O	Input/Output

Abkürzungsverzeichnis

IaaS	Infrastructure-as-a-Service
IPA	Infrastructure Power Adaptability
LAN	Local Area Network
LDEE	Load Dependent Energy Efficiency
LPM	Last- und Powermanagement
MSHV	Mittelspannungshauptverteilung
NAS	Network-Attached Storage
NEA	Netzersatzanlage
NFS	Network File System
NSHV	Niederspannungshauptverteilung
OpEx	Operational Expenditure
PaaS	Platform-as-a-Service
PDU	Power Distribution Unit
PUE	Power Usage Effectiveness
PVar	Power Variability
QoS	Quality of Service
RAID	Redundant Array of Independent Disks
REST	Representational State Transfer
ru	Reference Units
RZ	Rechenzentrum
SaaS	Software-as-a-Service
SAN	Storage Area Network
SLA	Service-Level-Agreements
SMB	Server Message Block
SNMP	Simple Network Management Protocol
SOAP	Simple Open Access Protocol
SSD	Solid State Drive

TCO	Total Cost of Ownership
UPS	Uninterruptible Power Supply
USV	Unterbrechungsfreie Stromversorgung
UV	Stromunterverteilung
VM	Virtuelle Maschine
ZUCS	Zero-U-Space-Cooling-System

Ehrenwörtliche Erklärung zu meiner Dissertation mit dem Titel „Entwicklung einer auslastungsabhängigen Energieeffizienzmetrik für Rechenzentren“

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbstständig verfasst habe und keine anderen als die angegebenen Hilfsmittel genutzt habe. Alle wörtlich oder inhaltlich übernommenen Stellen habe ich als solche gekennzeichnet.

Oldenburg, 06.07.2016



Daniel Schlitt