



Evaluation eines Scheduling- Algorithmus für eine elastische Cloud mit rekonfigurierbaren Hardwarebeschleunigern

Abschlusspräsentation zur Projektarbeit ET - INF STA

Valentin Kandetzki

Valentin.Kandetzki@mailbox.tu-dresden.de

Dresden, 07.09.17

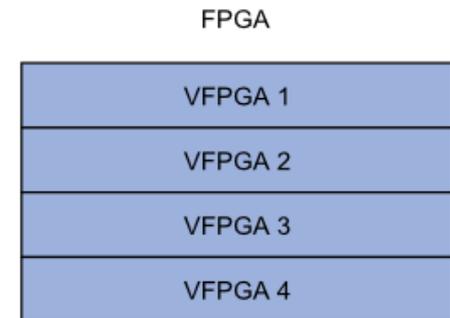


Gliederung

- 1. Grundlagen
- 2. Simulator
- 3. Auswertung
 - 3.1 Virtualisierung
 - 3.2 Migration
 - 3.3 Nodes zeitgleich starten
 - 3.4 Nodestart nach Warteschlangenlänge
- 4. Zusammenfassung und Ausblick
- 5. Quellen

1. Grundlagen

- Virtualisierung
 - Beschreibt Abstraktion von Hardwareressourcen zur virtuellen Maschine
 - Hardware wird in mehrere virtuelle Maschinen unterteilt (Partitionierung)
 - Virtual Machine Monitor (VMM) ermöglicht zeitgleiche Ausführung mehrerer virtueller Maschinen [3]
- Migration - Livemigration
 - Virtuelle Maschine wird im laufenden Betrieb vom Host auf einen anderen geschoben
 - Im Simulator: Migration eines Jobs mit Platzbedarf 1, bei FPGA-Auslastung von 25% [4]



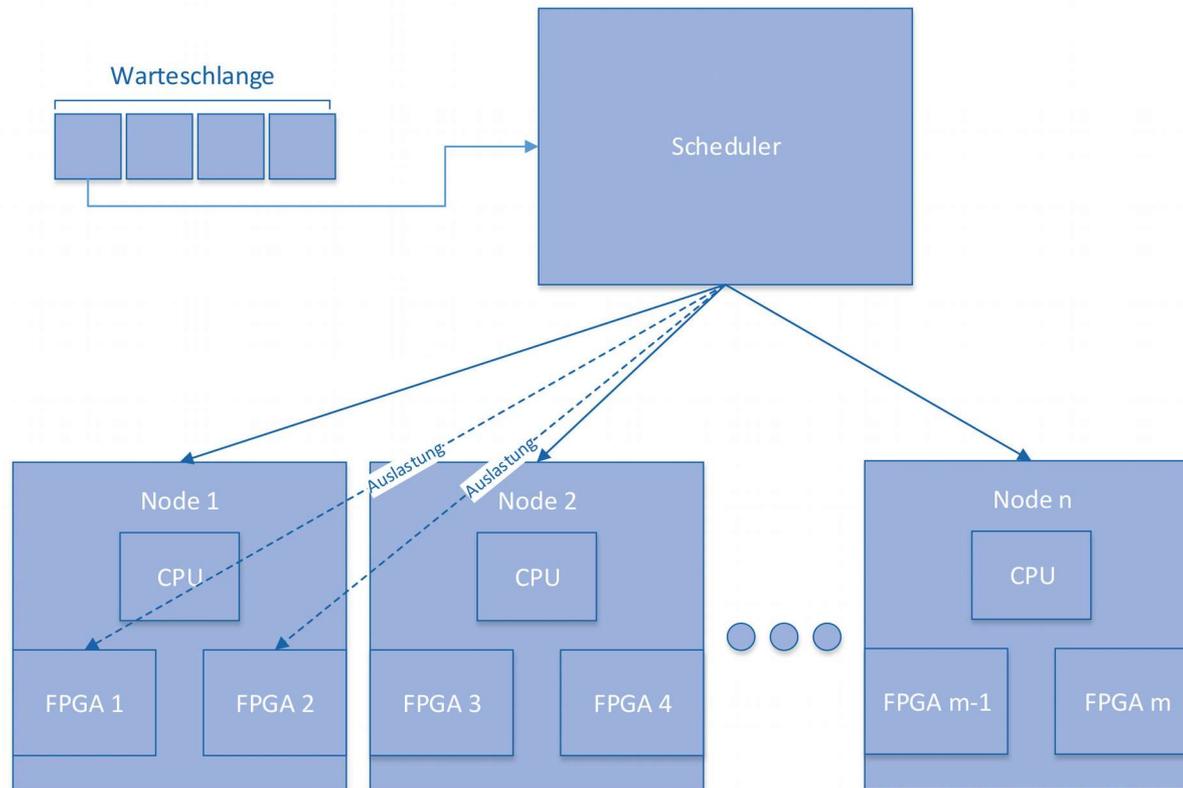
1. Grundlagen

- Simulation
 - Zeitgesteuert
 - Simulationsablauf ist in mehrere Zeitschritte unterteilt
 - In diesen führen Modellkomponenten Zustandsänderungen durch
 - Keine Schritte können übersprungen werden [5]
- Skalierung
 - Geschwindigkeit (Wechsel von unterversorgtem Zustand in optimal- oder überversorgten Zustand)
 - Präzision (Unterschied geforderter und bereitgestellter Ressourcen)
 - Horizontale Skalierung – Allokieren und Freigeben mehrerer Ressourcen
 - Vertikale Skalierung – Ersetzen einer vorhandenen Ressource durch eine Größere oder Kleinere [6]

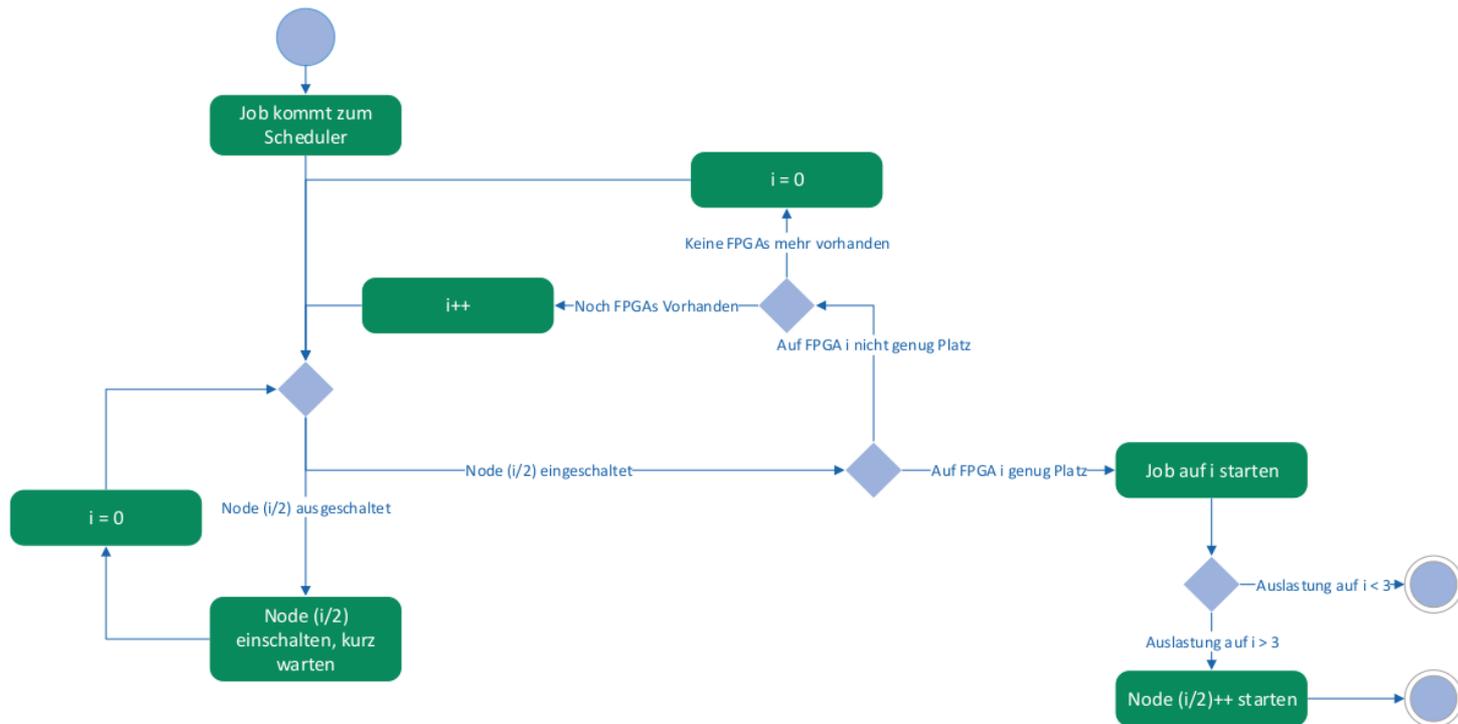
1. Grundlagen

- Service Level Agreement
 - Beschreibt die Qualität des Cloud Service
 - Wird meist vom Cloudanbieter herausgegeben [6]
 - Im Simulator liegt eine SLA-Verletzung vor, wenn die Zeit zwischen Ankunft eines Jobs in der Warteschlange und der Beginn der Ausführung dieses Jobs einen bestimmten Wert überschreitet.
 - Prozentuale Verletzungen = $(\text{Anzahl Verletzungen} / \text{Gesamtzahl Jobs}) * 100$

2. Simulator - Aufbau



2. Simulator - Funktion



2. Simulator - Berechnung Leistungsbedarf

- FPGA 19 Watt im idle, 23 Watt unter Vollast
(Xilinx Virtex-7 - VC707 board)
[1]

$$\text{Leistungsbedarf } f = \frac{\sum_{FPGA[0]}^{FPGA[n]} t_{an} + \sum_{Job[1]}^{Job[m]} \text{Job.Ausfuehrzeit} * \text{Job.Platzbedarf}}{\text{Gesamtlaufzeit}}$$

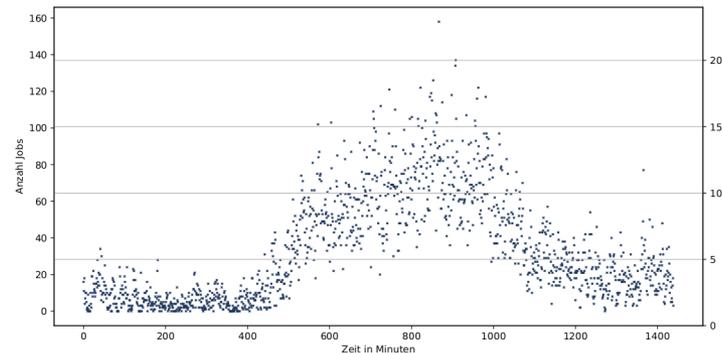
n = Maximale Anzahl vorhandener FPGAs

m = Letzter Job der Simulation

t_{an} = Zeit, welche der entsprechende FPGA eingeschaltet ist *19

2. Simulator - Eingangsdaten und Generierung von Jobs

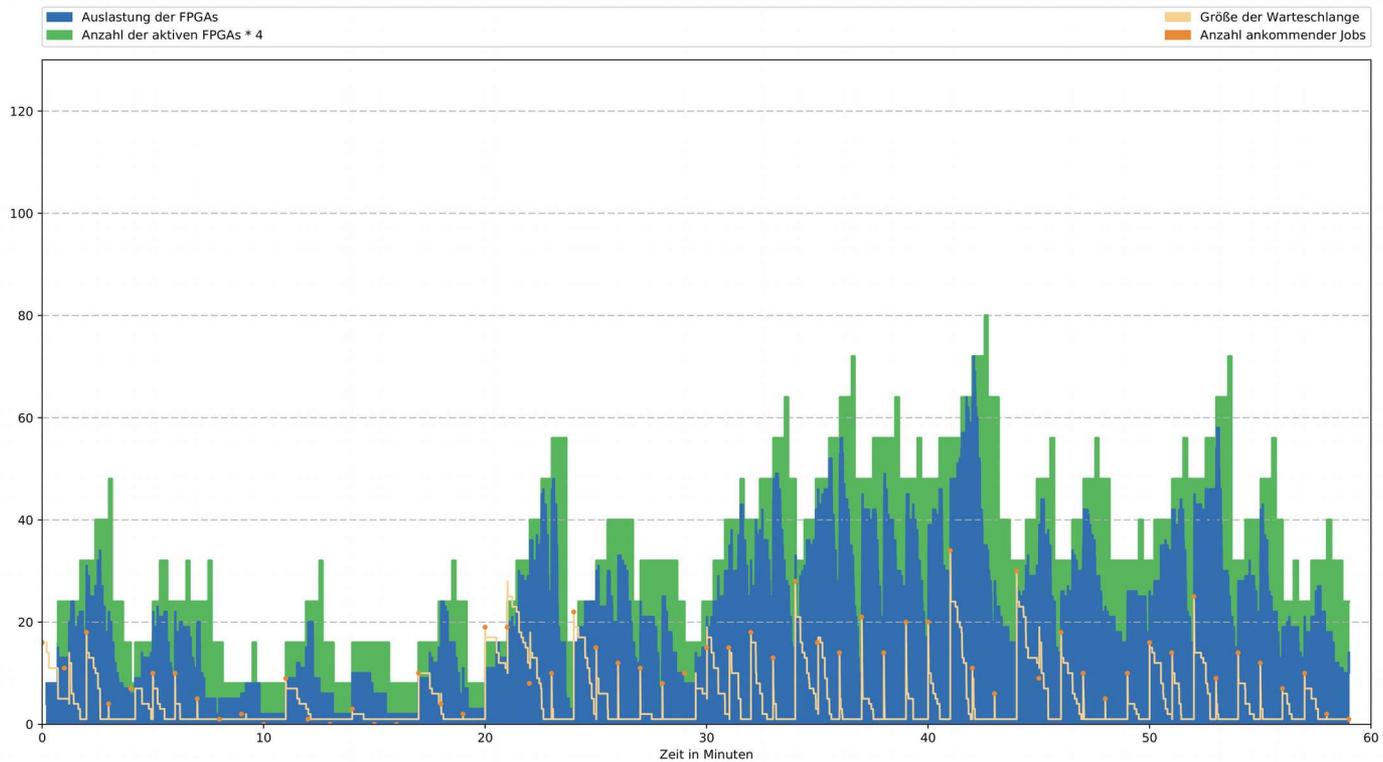
- Tageswert aller Anfragen an einen Webserver in New York als Grundlage [2]
- Jobs generieren
 - Platzbedarf
 $1 \leq x \leq 4 \wedge x \in \mathbb{N}$
 - Bearbeitungszeit
 $1 \leq x \leq 10 \wedge x \in \mathbb{N}$
 - Generierte Jobs in Textdatei gespeichert
pro Zeiteinheit der Warteschlange entsprechend Abb. Jobs hinzufügen
 - Gesamtzahl der Jobs: 47748



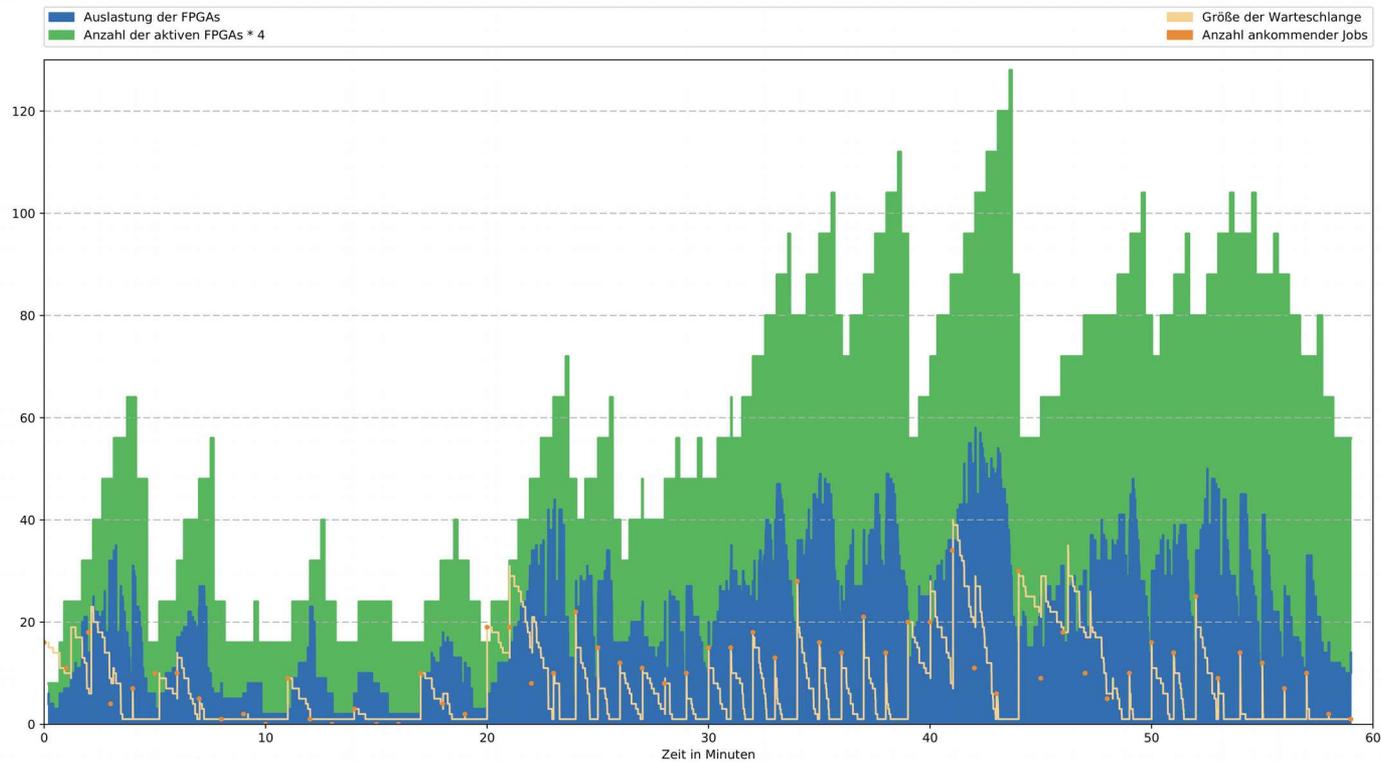
2. Simulator - Parameter

- Zeitspanne zwischen eingehenden Jobs
 - 10 Sekunden
- Migrationsgrenze - nachfolgende Grafik ohne Jobmigration
- Auslastungsgrenze für Nodestart
 - 87,5%
- Anzahl kleiner Jobs mit langer Laufzeit
 - Von 10% der Jobs mit Platzbedarf eins wurde die Laufzeit um das 100-fache erhöht
- Maximale Anzahl an Rechenknoten
 - Unbegrenzt
- SLA-Verletzung
 - Liegt nach 8 Sekunden vor
- Benötigte Zeit für Migration - nachfolgende Grafik ohne Jobmigration

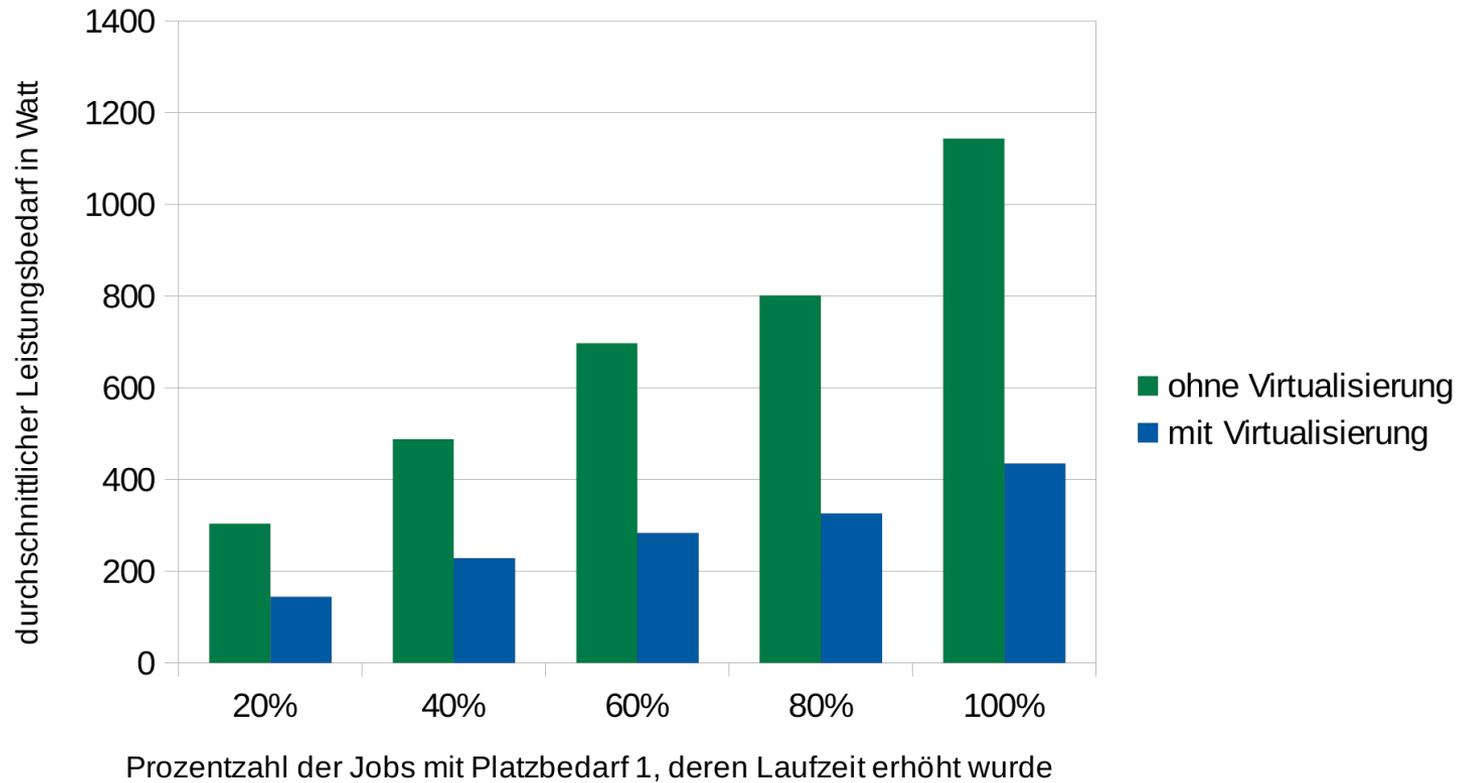
3.1 Auswertung - Virtualisierung



3.1 Auswertung - Virtualisierung



3.1 Auswertung - Virtualisierung



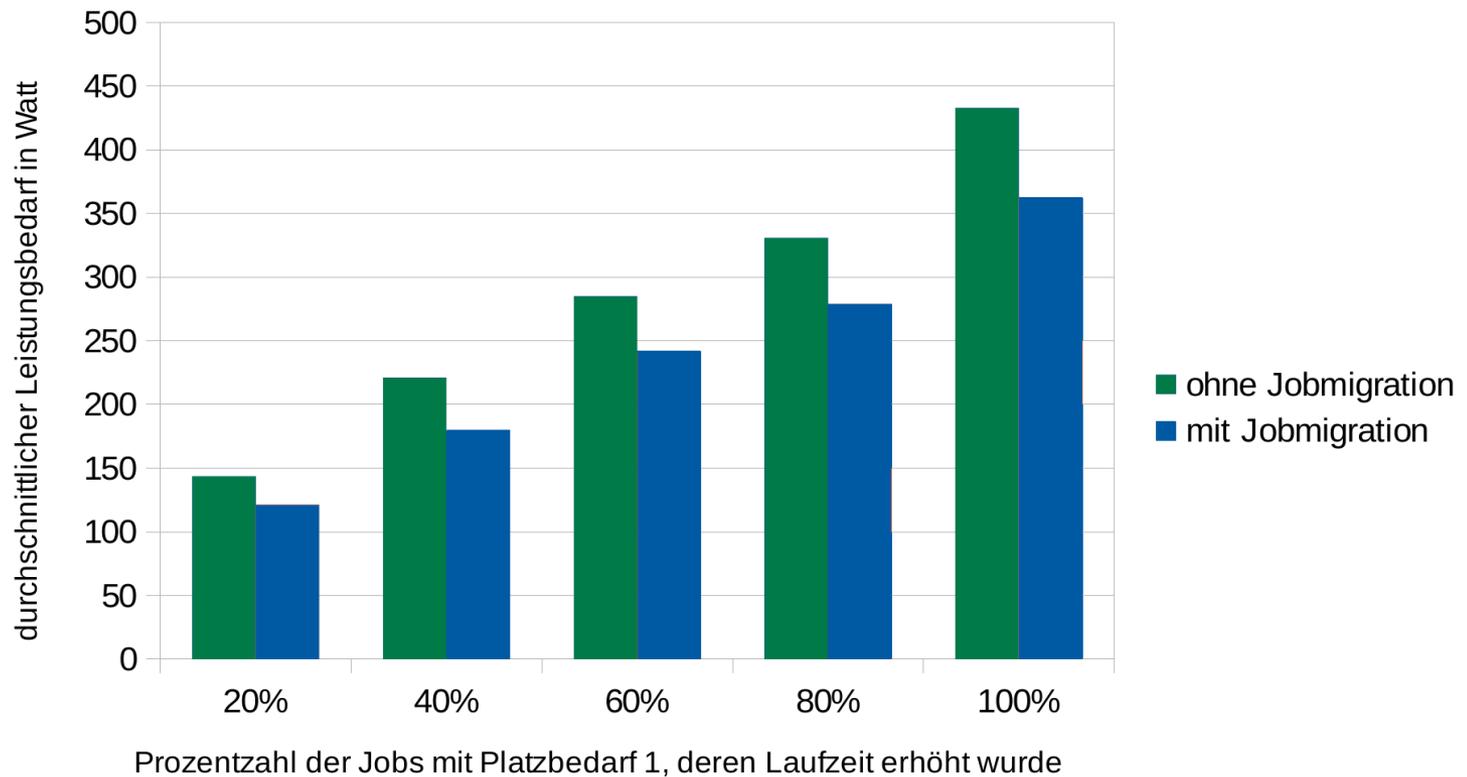
3.2 Auswertung – Migration

- Migration wenn
 - Nur ein Job mit Platzbedarf eins auf einem FPGA läuft (Auslastung 25%)
 - Der Job wurde bereits 11 Sekunden bearbeitet
- Speichern der aktuellen Systemzeit wenn ein Job gestartet wird
 - Ermittlung, wie lange ein Job bereits bearbeitet wurde
- Simulation der Migration
 - Freigeben des entsprechenden Platzes des Jobs
 - Einfügen eines Jobs mit entsprechender Restbearbeitungszeit an erster Stelle in der Jobwarteschlange
 - Berechnung der Restbearbeitungszeit:
 - Bearbeitungszeit Job - (aktuelle Systemzeit - Systemzeit bei Jobstart)

3.2 Auswertung – Migration – verwendete Parameter

- Zeitspanne zwischen eingehenden Jobs
 - 20 Sekunden
- Anzahl verfügbarer Rechenknoten
 - unbegrenzt
- Migrationsgrenze
 - Migration von Jobs, welche mindestens 11 Sekunden bearbeitet wurden
- Benötigte Zeit für Migration
 - 5 Sekunden

3.2 Auswertung - Migration



3.3 Auswertung - Nodes zeitgleich starten

```
aktive_fpgas=0
for h in fpga_zustand:
    if h == 1:
        aktive_fpgas +=1
for m in range((aktive_fpgas//10)+1):
    if(((i+m*2)<(fpgaanzahl-1)) and fpga_zustand[i+(m*2)+1]==0 and
    fpga_zustand[i+1]==0 and node_gestartet[(i+(m*2)+1)]==0):
        node_gestartet[(i+(m*2)+1)] = 1
        print("node an stelle", (i+(m*2)+1), "starten")
        b = Process(target=node_starten, args=(fpga_zustand, i+(m*2),))
        b.start()
```

- Durchschnittlicher Leistungsbedarf: 240 Watt
- SLA-Verletzungen: 5%

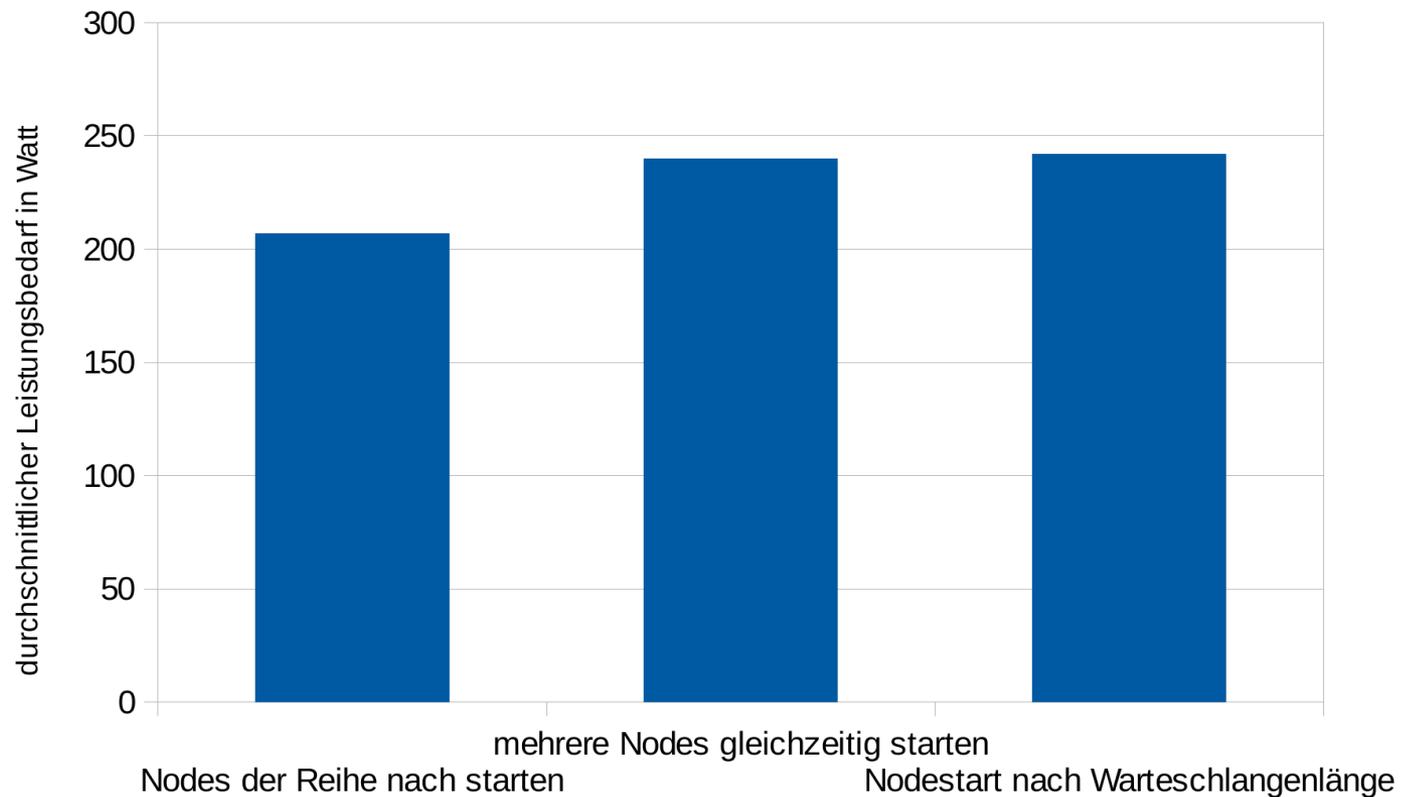
3.4 Auswertung - Nodestart nach Warteschlangenlänge

```
for n in range(Nodeanzahl):  
    if len(Warteschlange) > (Maximale_Warteschlangenlaenge//Nodezahl)*n  
        && Node[n] == abgeschaltet && Nodestart_noch_nicht_begonnen:  
            Node[n] starten()
```

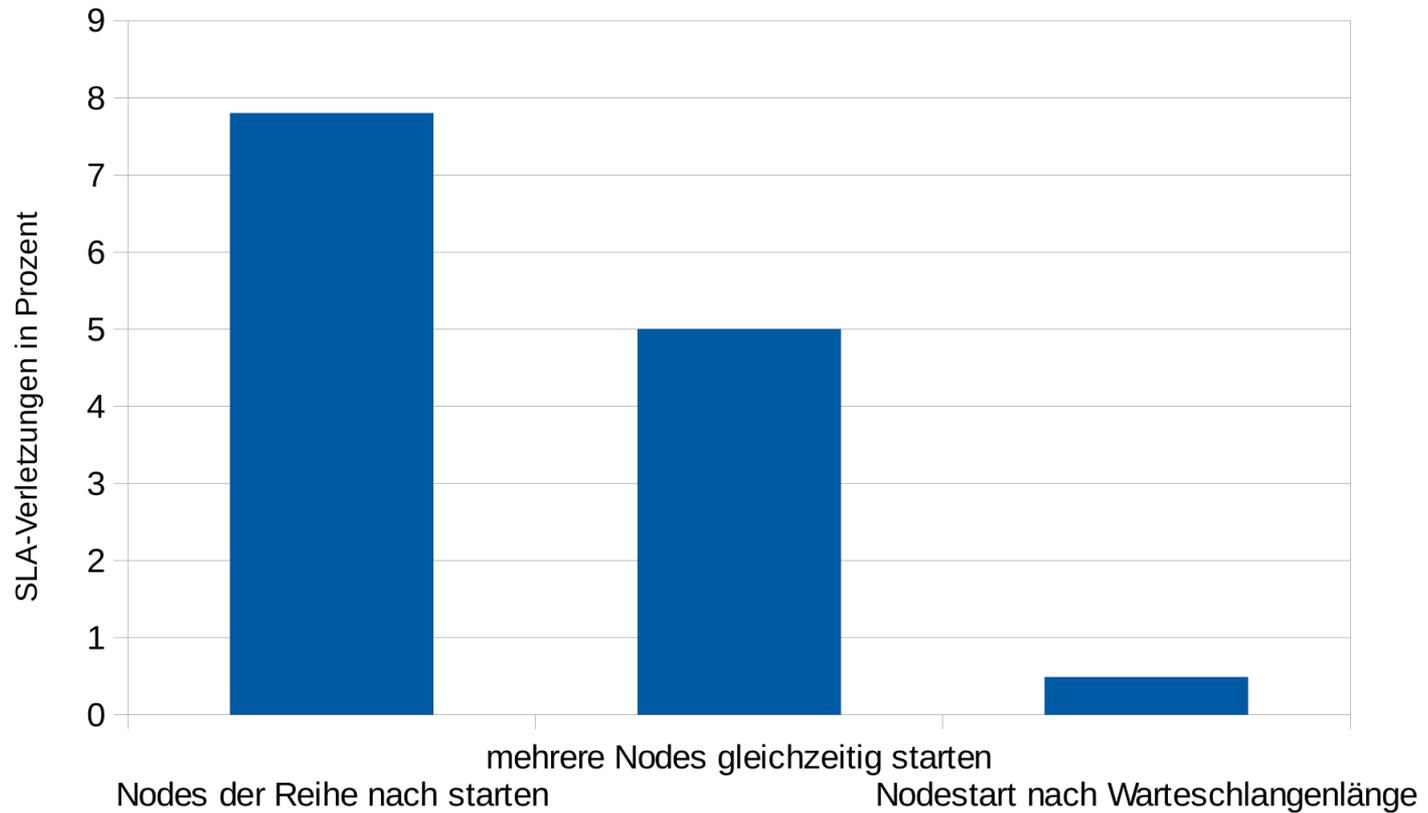
- Maximale Warteschlangenlänge: 170 Jobs
- Maximale FPGA-Zahl: 80

- Durchschnittlicher Leistungsbedarf: 242 Watt
- SLA-Verletzungen: 0,49%

Auswertung - Algorithmen für den Nodestart



Auswertung - Algorithmen für den Nodestart



Zusammenfassung und Ausblick

- Untersucht wurde:
 - Simulation der der parallelen Abarbeitung von Jobs mittels Multiprocessing
 - Effekt der Virtualisierung von FPGAs
 - Effekt der Jobmigration
 - Unterschiedliche Algorithmen zur effizienten Ressourcenverwaltung in einer Cloud
- Weitere Untersuchungsmöglichkeiten:
 - Migration größerer Jobs, Jobs mit geringer Laufzeit, Änderung der Migrationsgrenze
 - Auslagerung der konkreten Jobverteilung (auf die Virtuellen FPGAS) auf die einzelnen Nodes

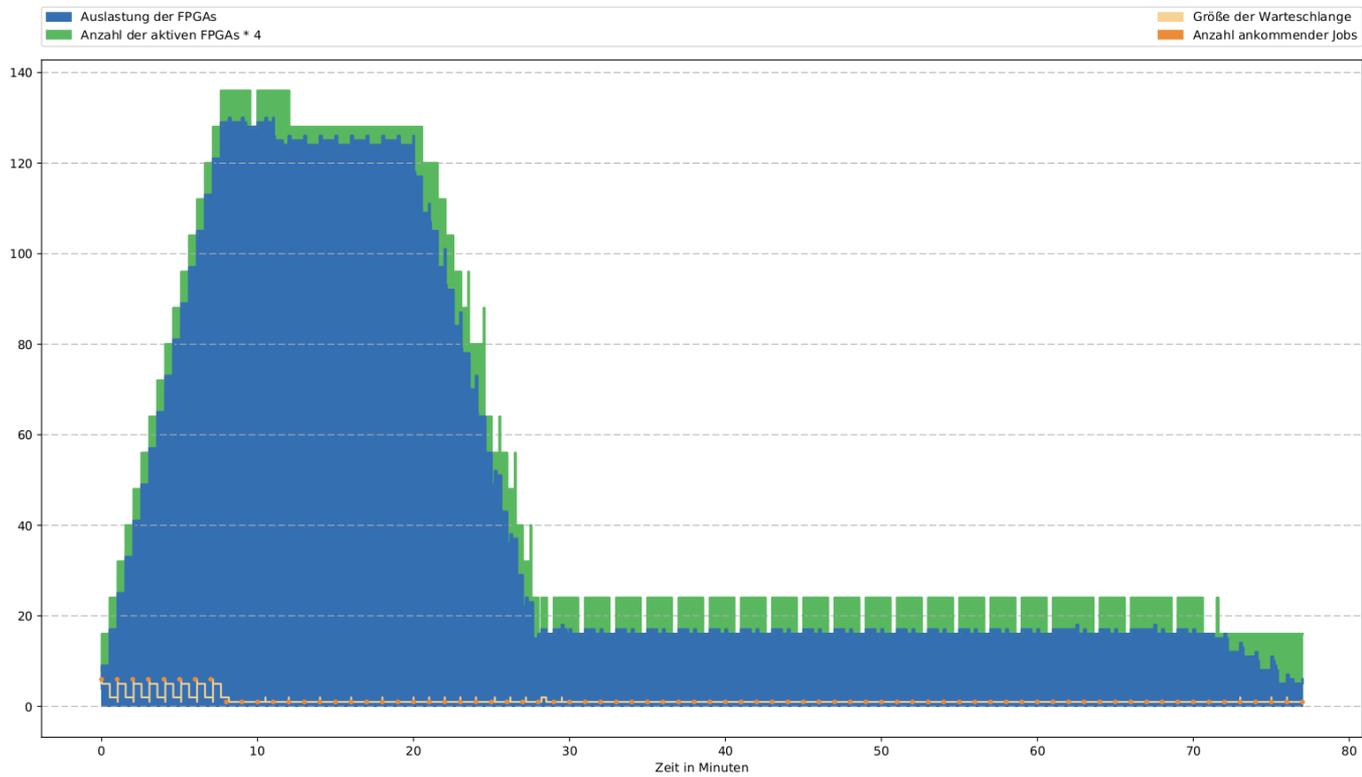
5. Quellen

- [1] Oliver Knodel, Patrick Lehmann und Rainer G Spallek. RC3E: Reconfigurable Accelerators in Data Centres and their Provision by Adapted Service Models. In: Cloud Computing (CLOUD), 2016 IEEE 9th International Conference on. IEEE. 2016, S. 1926.
- [2] Tageswert aller HTTP-Anfragen an den EPA WWW Server im RTP, NC .
URL: <http://ita.ee.lbl.gov/html/contrib/EPA-HTTP.html> .
- [3] James Edward Smith und Ravi Nair. Virtual machines . Bd. 1. Morgan Kaufmann San Francisco, Calif, USA, 2005.
- [4] KNUT HILDEBRAND. IT-Integration und Migration: Konzepte und Vorgehensweisen. In: Dpunkt, Heidelberg (2007)
- [5] Friedmann Mattern. „Modellbildung und Simulation“. In: Perspektiven der Informatik. 1993, S. 56-64.
- [6] Thomas Erl, Ricardo Puttini und Zaigham Mahmood. Cloud computing: concepts, technology & architecture. Pearson Education, 2013.

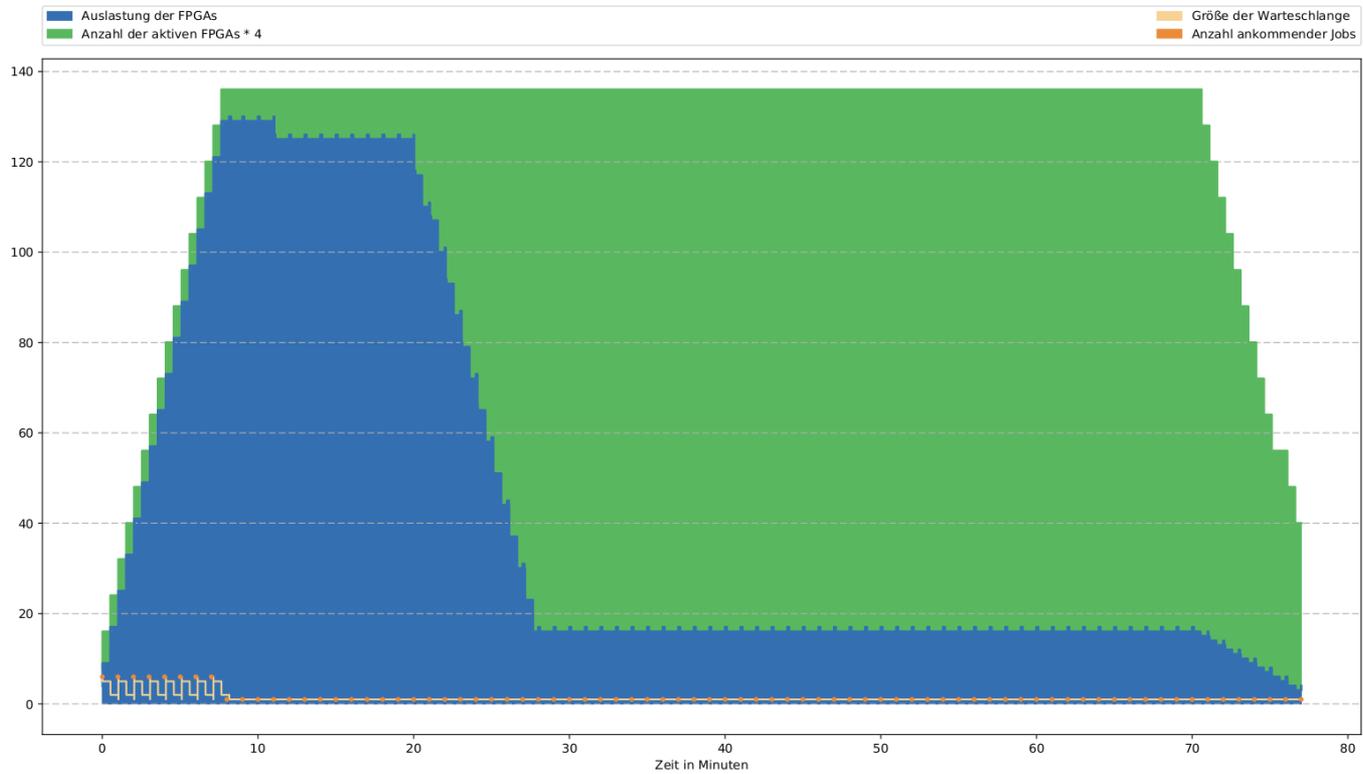


»Wissen schafft Brücken.«

Mit Jobmigration



Ohne Jobmigration



Nodestart nach Warteschlangenlänge

