



Master-Arbeit Nr. 992

Analyse paralleler Verarbeitungsprozesse durch Task-Graphen Reduktion



Methoden

Parallelverarbeitung
Leistungsbewertung

Themengebiete

Warteschlangentheorie
Task-Graphen-Reduktion

Hintergrund

Die Verarbeitungsgeschwindigkeit von Computersystemen kann durch Parallelarbeit auf Multi-Core Prozessoren erhöht werden, um harten Echtzeitanforderungen gerecht zu werden. Typische Anwendungsfälle dafür sind MapReduce-Algorithmen, Suchvorgänge in grossen Datenbanken sowie Verarbeitungsprozesse in Kommunikationssystemen auf der Basis von Network Function Virtualization (NFV) und Software-Defined Network (SDN). Diese Art von Parallelverarbeitung führt auf die Leistungsbewertung des sog. Fork-Join-Problems, bei dem ein Verarbeitungsprozess an bestimmten Stellen auf mehrere parallel ablaufende Teilprozesse aufgeteilt werden kann, der dann aber erst nach Abschluss aller parallel durchgeführten Teilprozesse seriell fortgeführt werden kann. Dazu werden in der Literatur zwei Grundmodelle unterschieden, mit und ohne Blockierung der beteiligten Server nach Abschluss ihrer Bearbeitung bis zum Re-Synchronisationszeitpunkt, an dem die serielle Prozessverarbeitung erst fortgesetzt werden kann. Die Performance dieser Grundprobleme wurde in der Literatur als Grundproblem mittels Ordnungs-Statistiken beschrieben und approximativ mittels des Network Calculus - Verfahrens quantitativ untersucht, um die Leistungsfähigkeit der Parallelverarbeitung quantitativ zu bewerten im Hinblick auf die erforderliche Verarbeitungsdauer sowie den erzielbaren Maximaldurchsatz.

Aufgabenstellung

Ziel der Masterarbeit ist die Durchführung exakter und approximativer Performance-Analysen auf der Basis des Verfahrens der analytischen Task-Graphen-Reduktion des Betreuers. Das FG-Problem mit Server-Blockierung lässt sich mit dem Verfahren der Task-Graphen Reduktion analytisch exakt lösen für die kumulative Gesamt-Verteilungsfunktion heterogener und beliebig verteilte Prozesse; mittels Zeitdiskretisierung können daraus der Mittelwert und alle gewöhnlichen Momente numerisch berechnet werden, einschliesslich der Bestimmung des Maximaldurchsatzes. Das FG-Problem ohne Server-Blockierung ist wesentlich schwieriger und führt auf ein FG-Problem, bei dem die Teilphasen Durchlaufzeiten (Sojourn Times) von Queuing Modellen des Typs GI/G/1 darstellen; diese sind nur für Queuing-Modelle des Typs M/M/1 und GI/M/1 exakt bekannt und somit berechenbar; für Queuing Modelle des Typs M/G/1 sind jedoch die zweiten Momente der Wartezeiten bekannt für die Queuing-Disziplinen FIFO, RANDOM und LIFO (L. Takacs, BSTJ 1964). Mit Hilfe der ersten und zweiten Momente lassen sich die Wartezeitverteilungen dieser Queuing-Modelle recht genau approximieren. Die Durchlaufzeit folgt dann aus der Summe der voneinander unabhängigen Wartezeit und Bedienzeit. Auf diese Weise soll das FG-Problem ohne Blockierung analytisch exakt bzw. durch die vorgenannte Approximation analysiert werden.

Zur approximativen Bestimmung der Durchlaufzeiten des Wartesystemtyps M/G/1 bieten sich zwei Verfahren an: die Faltung der Verteilungs-Dichtefunktionen aus Bedien- und Wartezeit als Integraloperation oder die Momenten-Methode; bei der Momenten-Methode kann mit den erhaltenen ersten beiden Momenten die Durchlaufzeit-Verteilungsfunktion z.B. mittels der Weibull-Funktion approximiert und damit auf das Problem der Lösung des FG-Problems mit (Ersatz-)Server-Blockierung zurückgeführt werden. Letztere Methode ist vorteilhafter im Hinblick auf die darauf folgende Lösung des entstehenden FG-Problems mit Server-Blockierung. Für beide Fälle sollen dazu Computer Programme entwickelt werden.

Beide Modelle sollen abschließend im Hinblick auf die erweiterten Anwendungen des FG-Problems für heterogene Modelle hinsichtlich der von der Anwendung her interessanten Eigenschaften fallweise untersucht werden für die Bedienungsmodelle des Typs M/G/1 und

GI/M/1 hinsichtlich ihres Maximal-Durchsatzes und insbesondere ihrer Durchlaufzeit-Performance. Ein weiterer Aspekt ist das Verhalten des FJ-Modelles hinsichtlich der Abfertigungsdisziplinen FIFO und LIFO; LIFO kann in zeitkritischen Überlastfällen der Strategie FIFO überlegen sein, da bevorzugt die neuesten Sensorwerte verarbeitet werden, während FIFO im Hochlastfalle im Wesentlichen veraltete Nachrichten verarbeitet. Die anvisierten Ziele sind nach Kenntnis des Betreuers bislang in der Literatur nicht behandelt worden. Die laufenden Arbeiten zu diesen Themen erfolgen in loser Kooperation mit Wissenschaftlern des DFG-SFB MAKI an der TU Darmstadt. Ein Beispielergebnis zu diesem Problemkreis wurde in einer zuvor abgeschlossenen Forschungsarbeit bei einem ITG-Workshop im Dezember 2018 vorgetragen.

Kontakt

Prof. em. Dr.-Ing. Dr. h.c. mult. Paul J. Kühn

Raum 1.341 (ETI II), Telefon 685-68027, E-Mail paul.j.kuehn@ikr.uni-stuttgart.de