CARL VON OSSIETZKY universität OLDENBURG

Fakultät II – Informatik, Wirtschafts- und Rechtswissenschaften
Department für Informatik

Probabilistic Quorum Systems for Dependable Distributed Data Management

Dissertation zur Erlangung des Grades eines
Doktors der Naturwissenschaften

vorlegt von

M.Sc. CS Kinga Kiss Iakab

Gutachter:

Prof. Dr.-Ing. Oliver Theel
Jun.-Prof. Dr. Daniela Nicklas

Tag der Disputation: 20. Juli 2012

Abstract

Among failure-prone and dynamic distributed systems there is a significant class of systems that strive for high availability and can function with inconsistent data. Examples include flight reservation systems which allow overbooking or emergency ambulance systems which return informative responses to time-critical queries.

Data replication is a well-known technique for tolerating failures and dependably managing data in distributed systems. For this purpose quorums are used for executing the basic operations: writing new data and reading previously written data. Strict quorum systems rely on a strict consistency notion called sequential consistency by ensuring the mutual exclusion between read and write operations as well as write and write operations. The guarantee of this strict consistency limits their availability. Probabilistic quorum systems increase the availability of operations by relaxing the previously mentioned mutual exclusions. This relaxation requires the mutual exclusion and therefore the intersections of quorums to hold only with high probability for read and write as well as write and write probabilistic operations.

The first contribution of this work is the construction of probabilistic quorum systems based on strict quorum systems as input. The generation, selection, and integration of quorums are identified as steps of the construction. The selection is driven by consistency by putting emphasis on the intersections with the previous operation's quorums. Additionally, they preserve beneficent characteristics of the original underlying strict quorum systems (e.g., operation availability, communication costs, etc.). Furthermore in the integration step, different priorities are considered when combining strict and probabilistic quorums to obtain different resulting probabilistic quorum systems. These combination methods are called integration strategies.

The second contribution of this thesis is the analysis of different probabilistic quorum system constructions with respect to the trade-off between data consistency and operation availabilities. By means of a Markov chain analysis, qualitative and quantitative aspects of the trade-off are identified. The empirical results strongly indicate that there is a total order among the three introduced integration strategies with respect to data consistency independent of the particular investigated replication strategy.

The third contribution is the optimization of probabilistic quorum systems in terms of data consistency and operation availabilities. Concepts and methods from the area of strict quorum systems are extended for the area of probabilistic quorum systems. In this context, it is proven that only a single non-dominated probabilistic quorum system exists and that only the write-write intersection of so-called availability-symmetric probabilistic quorum systems can be relaxed. Additionally, a graphical data consistency measure is presented. Although this measure is more abstract than the one used in the Markov chain analysis, it allows to identify availability-symmetric quorum systems that exhibit maximal data consistency with respect to another quorum system and with respect to a particular, fixed operation.

The closing contribution of the work is the general analysis of data consistency with respect to the integrations of probabilistic quorum systems for an arbitrary number of processes in a distributed system. This analysis formally proves and generalizes the results from the previous Markov chain analysis.

Zusammenfassung

Unter fehleranfälligen, dynamischen, verteilten Systemen gibt es eine bedeutende Klasse von Systemen, die nach hoher Verfügbarkeit streben und mit inkonsistenten Daten umgehen können. Beispiele dafür sind Flugreservierungssysteme, die Überbuchungen erlauben oder Rettungsdienstsysteme, die nicht immer korrekte – aber für eine Entscheidung ausreichende – Antworten auf zeitkritische Anfragen geben.

Datenreplikation ist eine bekannte Technik in verteilten Systemen, um Fehler zu tolerieren und Daten zuverlässig zu verwalten. Dafür werden oftmals Quoren benutzt, um Basisoperationen (neue Daten schreiben und vorher geschriebene Daten lesen) durchzuführen. Strikte Quorensysteme basieren auf einem strikten Konsistenzbegriff, der sogenannten sequentiellen Konsistenz. Dieser stellt den gegenseitigen Ausschluss konkurrierender Lese- und Schreiboperationen, beziehungsweise zweier konkurrierender Schreiboperationen sicher. Die Gewährleistung der strikten Konsistenz beschränkt die Verfügbarkeit der Operationen. Probabilistische Quorensysteme erhöhen die Verfügbarkeit der Operationen durch die Abschwächung der vorher genannten gegenseitigen Ausschlüsse. Diese Abschwächung bezieht sich auf die Wahrscheinlichkeit der Überschneidung von Lese- und Schreibquoren, beziehungsweise Schreib- und Schreibquoren: sie werden nur noch mit hoher Wahrscheinlichkeit gewährleistet.

Der erste Beitrag dieser Arbeit liegt in der Konstruktion probabilistischer Quorensysteme, die auf strikten Quorensystemen als Eingabe basieren. Die Erzeugung, die Auswahl und die Integration der Quoren werden als Schritte der Konstruktion identifiziert. Die Auswahl ist konsistenzgetrieben durch Bevorzugung der Quoren, die Überschneidungen mit den Quoren der vorher ausgeführten Operation haben. Zusätzlich erhalten die probabilistischen Quoren die nützlichen Eigenschaften der strikten, ursprünglichen Eingabequoren, wie zum Beispiel die Operationsverfügbarkeit und die Kommunikationskosten. Weiterhin werden im Integrationsschritt unterschiedliche Prioritäten für die Kombination strikter und probabilistischer Quoren betrachtet, um unterschiedliche probabilistische Quorensysteme als Ergebnis zu erzielen. Diese Kombinationsarten der Quoren werden Integrationsstrategien genannt.

Der zweite Beitrag der Arbeit ist die Analyse verschiedener probabilistischer Quorensysteme, resultierend aus den Konstruktionen, bezüglich des Trade-offs zwischen Datenkonsistenz und Operationsverfügbarkeiten. Mit Hilfe einer Markowkettenanalyse werden qualitative und quantitative Aspekte des Trade-offs bestimmt. Die empirischen Ergebnisse der Analyse deuten stark darauf hin, dass bezüglich der Datenkonsistenz eine totale Ordnung zwischen den drei eingeführten Integrationsstrategien existiert, die unabhängig von der spezifischen untersuchten Replikationsstrategie ist.

Der dritte Beitrag ist die Optimierung probabilistischer Quorensysteme bezüglich Datenkonsistenz und Operationsverfügbarkeiten. Konzepte und Methoden aus dem Bereich der strikten Quorensysteme werden in den Bereich der probabilistischen Quorensysteme übertragen und erweitert. In diesem Kontext wird gezeigt, dass es nur ein einziges nicht-dominiertes probabilistisches Quorensystem gibt und dass nur die Schreib-Schreib-Überschneidung der sogenannten verfügbarkeitssymmetrischen probabilistischen Quorensysteme abgeschwächt werden kann. Zusätzlich wird ein graphisches Maß für die Datenkonsistenz eingeführt. Obwohl dieses Maß abstrakter ist als jenes, welches in der Markowkettenanalyse verwendet wurde, es ermöglicht die Identifizierung verfügbar-

keitssymmetrischer Quorensystemen, die eine maximale Datenkonsistenz bezüglich eines anderen Quorensystems und einer vorgegebenen, festen Operation haben.

Der die Arbeit abschließende Beitrag ist die allgemeine Analyse der Auswirkung der verwendeten Integrationsstrategien bei der Konstruktion probabilistischer Quorensysteme auf die Datenkonsistenz für eine beliebige Anzahl von Prozessen in verteilten Systemen. Diese Analyse beweist formal die Ergebnisse bezüglich der Integrationsstrategien aus der vorherigen Markowkettenanalyse und verallgemeinert sie.

Contents