



Diplomarbeit: Optimistische Nachrichtenzustellung bei atomarem Multicast mit adaptiver Gruppenzusammensetzung

Im Rahmen des EU-Projekts XtreamOS wird hier in Ulm eine Bibliothek zur Unterstützung *Virtueller Knoten (Virtual Nodes, VN)* entwickelt. Ein Virtueller Knoten repräsentiert dabei eine Gruppe von Prozessen, die jeweils die gleiche Anwendung repliziert ausführen. Durch diesen Mechanismus wird die Anwendung fehlertolerant, da der Ausfall eines oder mehrerer Knoten nicht zum Ausfall der Anwendung führt. Eines der Hauptprobleme von Replikationsmechanismen ist die im Vergleich zur nicht-replizierten Ausführung entstehende Latenz - zumeist bedingt durch Konsistenzanforderungen. Das VN System verwendet einen Paxos Algorithmus um Nachrichten atomar, d.h. total geordnet zuzustellen.

Die eingesetzte Paxos Implementierung arbeitet jedoch auf der Annahme, dass sich die Gruppe der beteiligten Knoten niemals ändert. Sie erlaubt nur temporäre Ausfälle der Gruppenmitglieder. In einem realen Einsatzszenario ist es jedoch durchaus gewünscht, dass ausgefallene Maschinen durch die Hinzunahme neuer Gruppenmitglieder kompensiert werden. Im Zuge dieser Diplomarbeit ist zu evaluieren in wie weit dies möglich ist ohne die durch den Paxos Algorithmus gebotenen Garantien zu verletzen.

Vom Algorithmus an die Anwendung zugestellte Nachrichten werden bearbeitet. Die Bearbeitung wird durch einen sogenannten deterministischen Scheduler gesteuert, der immer dann eingreift, wenn während der Bearbeitung das Sperren einer Mutex angefordert wird. Dieser Sperrvorgang ist die Voraussetzung um den Zustand eines VNs zu verändern. Daraus folgt auch, dass alle vor der ersten Sperre ausgeführten Operationen keinen Einfluss auf den Zustand haben und deswegen noch keine totale Ordnung auf den Nachrichten erforderlich ist.

Daraus ergibt sich folgendes Optimierungsmöglichkeit: Nachrichten werden zunächst in beliebiger Reihenfolge zugestellt. In der Zeit, die bis zum Aufruf der ersten Mutex-Operation vergeht, kann sich der Paxos Algorithmus auf eine Reihenfolge einigen. Im Zuge der Arbeit soll die Schnittstelle zwischen Anwendung und Gruppenkommunikation so erweitert werden, dass diese Art von Interaktion möglich wird. Ebenso soll der Algorithmus so angepasst werden, dass er die vorzeitige Zustellung unterstützt.