

DISS. ETH NO. 20083  
DISS. TIK NO. 129

**Decentralized Coordination:  
Methods and Applications**

A dissertation submitted to

ETH ZURICH

for the degree of

Doctor of Sciences

presented by

JOHANNES SCHNEIDER

MSc ETH INFK, ETH Zürich

born 10.10.1979

citizen of

Zurich (ZH)

accepted on the recommendation of

Prof. Roger Wattenhofer, examiner

Prof. Uzi Vishkin, co-examiner

Prof. Rachid Guerraoui, co-examiner

2011

## Abstract

We look at several problems in the field of distributed and parallel computing. Part I presents symmetry breaking techniques in the message passing model, i.e. each entity in a network can communicate through exchanging a distinct message with each of its neighbors in a synchronized round without failures. We introduce MULTI-TRIALS, a new technique for symmetry breaking for distributed algorithms and apply it to various problems in general graphs. For instance, the technique is demonstrated by giving three randomized algorithms for distributed (vertex or edge) coloring and a deterministic algorithm for computing ruling sets. Additionally, we present a coloring algorithm whose running time and number of colors needed depends on the chromatic number of the graph. We also generalize the deterministic coin tossing technique from rings to general graphs. The optimality of our algorithm is proven for the maximal independent set problem for geometric graphs such as unit disc graphs. Furthermore, we show how to trade among different complexity measures with a focus on different coloring and MIS algorithms. Finally, we derive techniques and bounds on trading among different complexity measures, such as the number of exchanged bits, messages and the number of communication rounds.

Part II investigates various problems related to wireless networks. In our theoretical investigation we look at different communication models and primarily focus on geometric graphs. The models share the assumption that the network topology is unknown and nodes wake up asynchronously. Furthermore, in a (synchronized) communication round the same message is broadcast to all neighbors but might be lost due to collisions of concurrent transmitters. We look at a model where nodes can detect collisions, e.g. through carrier sensing, and a model, where nodes are (almost) completely unaware of the number of transmitting neighbors. We derive several lower and upper bounds in both models for problems such as coloring, maximal independent set and broadcast. We show that the asymptotic gain in the time complexity when using collision detection depends heavily on the task at hand (and the maximal degree of the graph), i.e. it ranges from exponential down to no asymptotic gain at all.

In the practical part we derive (and evaluate on actual sensor nodes) a form of pulse-position modulation, i.e. message-position modulation (MPM), to reduce the payload of a message by encoding parts of the message through its transmission time. MPM allows to conserve energy and channel usage, thus reducing the risk of collisions and increasing available bandwidth.

Additionally, we present a technique called “Three Plane Localization” to improve accuracy of range based and range free localization schemes. The key

idea is to intentionally create interference at a node by scheduling concurrent transmissions of nearby nodes.

Part III deals with parallel shared memory systems, e.g. a conventional (multi-core) PC. We discuss aspects of transactional memory, (tree) data structures and graph based (parallel) algorithms. To improve throughput in transactional memory systems we analyze and present different strategies for contention management as well as load adaption mechanisms. Our contention management policies try to optimize the throughput by making the right decision in case of conflicts, i.e. decide which transaction has to abort, wait or can continue. However, new transactions are allowed to start (independent of overall system contention) and transactions are (usually) not delayed for a long time. We present and analyze two new algorithms for contention management.

We also look at load adaption strategies, i.e. how to optimize throughput by (temporarily) keeping computational resources, i.e. cores, idle. Opposed to prior work our load adapting schemes are simple and fully distributed, while maintaining the same throughput rate. Our experimental results show a substantial overall improvement for our best performing strategies compared to the best existing contention management policies (without load adaption).

For data structures representable by directed acyclic graphs, i.e. rooted trees, we show how to partition them to allow for efficient complex operations, which lie beyond inserts, deletes and finds. The approach potentially improves the performance of any operation modifying more than one element of the data structure. It covers common data structures implementable via linked lists or trees such as sets and maps.

Finally, we give a graph decomposition technique that creates entirely independent subproblems for graph problems such as coloring and dominating sets. The solutions of the subproblems can be computed as well as composed to an overall solution without synchronization on a shared memory system. The technique allows to trade performance for solution quality.

## Zusammenfassung

Wir untersuchen einige Probleme im Bereich des verteilten und parallelen Rechnens. Teil I fokussiert auf Symmetriebrechung im Message-Passing-Modell, d.h. jede Einheit in einem Netzwerk kann durch den Austausch einer Nachricht mit jedem seiner Nachbarn in einer synchronisierten Runde ohne Fehler kommunizieren. Wir stellen die Multi-Trials Technik vor und wenden sie auf verschiedene Probleme in allgemeinen Graphen an. Die Technik dient beispielsweise als Grundlage für drei randomisierte Algorithmen für verteiltes Färben eines Graphen und für ein deterministisches Verfahren zur Berechnung von Ruling Sets. Zu dem diskutieren wir einen Algorithmus zum Färben von Graphen in Abhängigkeit der chromatischen Zahl des Graphens.

Wir verallgemeinern die deterministische “Coin-Tossing” Technik von Ringen auf allgemeine Graphen. Wir beweisen die Optimalität des Algorithmus für das maximal unabhängige Mengen Problem für geometrische Graphen wie Einheitskreis Graphen. Des weiteren geben wir Austauschbeziehungen zwischen verschiedenen Komplexitätsmassen für verteilte Algorithmen an, wobei unser Fokus auf Symmetriebrechungsalgorithmen liegt.

Teil II untersucht verschiedene Probleme im Zusammenhang mit drahtlosen Netzwerken. Wir betrachten verschiedene Kommunikationsmodelle und geometrische Graphen. Die Modelle teilen die Annahme, dass die Netzwerk-Topologie unbekannt ist und Knoten asynchron aufwachen. Ferner wird in einer (synchronisierten) Kommunikationsrunde die gleiche Nachricht an alle Nachbarn gesendet. Diese könnte aufgrund von Kollisionen von mehreren gleichzeitigen Sendern jedoch verloren gehen. Wir betrachten ein Modell, bei dem Knoten Kollisionen erkennen können, z. B. durch Carrier Sensing, und ein Modell, in dem Knoten pro Runde (fast) keine Information über die Anzahl der sendenden Nachbarn erhalten. Wir leiten mehrere untere und obere Schranken in beiden Modellen für Probleme wie Farbgebung, maximal unabhängige Mengen und Broadcast. Wir zeigen, dass der asymptotische Gewinn in der Zeitkomplexität bei der Verwendung von Kollisionserkennung stark von der Aufgabenstellung (und des maximale Grades des Graphen) abhängt. Er reicht von exponentiell bis kein asymptotischer Gewinn.

Im praktischen Teil beschreiben wir (und evaluieren mit echten Sensorknoten) eine Form von Puls-Position-Modulation, dh Message-Position-Modulation (MPM), um die gesendete Nachricht durch Verschlüsselung eines Teils der Nachricht über seine Übertragungszeit zu verkleinern. MPM reduziert die Energie- und Kanalnutzung, wodurch die Gefahr von Kollisionen reduziert wird und sich die verfügbare Bandbreite erhöht.

Darüber hinaus präsentieren wir eine Technik namens “Three Plane Localization”, um die Genauigkeit der Lokalisierung von Sensorknoten zu verbessern.

Teil III beschäftigt sich mit Parallelrechnern, welche über gemeinsam genutzten Speicher kommunizieren, z. B. ein konventioneller Mehrkern PC. Wir diskutieren Aspekte von Transactional Memory, (Baum-)Datenstrukturen und Graphen- basierten (parallelen) Algorithmen. Zur Verbesserung des Durchsatzes in Transactional Memory Systemen untersuchen wir verschiedene Strategien für Contention Management als auch Last-Adaptions Mechanismen. Diese Strategien müssen entscheiden was mit zwei Transaktionen geschieht, welche im Konflikt stehen. Eine Transaktion kann abgebrochen oder verzögert werden oder fortfahren.

Für Datenstrukturen, welche als gerichtete azyklische Graphen darstellbar sind, zeigen wir, wie diese zu partitionieren sind, um komplexe Operationen, die über das Einfügen, Löschen und Suchen hinaus gehen, effizient zu berechnen.

Schließlich präsentieren wir einen Technik, die völlig unabhängig Teilprobleme für Graphen Probleme wie Farbgebung erzeugt. Die Lösung der Subprobleme als auch das Zusammenfügen der Teillösungen zu einer Gesamtlösung kann ohne Synchronisation erfolgen. Die Technik erlaubt es Rechenzeit für Lösungsqualität einzutauschen.