Distributed Computing Column 67

Review of 2016 BIRS CMO Workshop on Complexity and Analysis of Distributed Algorithms

Jennifer L. Welch
Department of Computer Science and Engineering
Texas A&M University, College Station, TX 77843-3112, USA
welch@cse.tamu.edu

You may be familiar with the Banff International Research Station (BIRS), a conference center located in Alberta, Canada, that hosts workshops in the mathematical sciences. Recently, BIRS has expanded to include a location in Oaxaca, Mexico, called the Casa Matemática Oaxaca (CMO) and workshops started at CMO in 2015. In this column, I provide a review of the CMO Workshop on Complexity and Analysis of Distributed Algorithms, organized by Hagit Attiya, Sergio Rajsbaum, and Philipp Woelfel, which I was fortunate enough to attend in 2016.

Call for contributions: I welcome suggestions for material to include in this column, including news, reviews, open problems, tutorials and surveys, either exposing the community to new and interesting topics, or providing new insight on well-studied topics by organizing them in new ways.

Review of 2016 BIRS CMO Workshop on Complexity and Analysis of Distributed Algorithms

The BIRS-CMO workshop on Complexity and Analysis of Distributed Algorithms ran from November 27 to December 2, 2016, and was organized by Hagit Attiya, Sergio Rajsbaum, and Philipp Woelfel. The objective of the workshop was to bring together researchers to exchange ideas on how to advance the design and analysis of algorithms for problems in distributed systems. Topics covered included theory of shared memory algorithms, networks and message-passing systems, applications of combinatorics and topology to concurrency. The attendees were a mix of theoreticians and systems researchers.

Figure 1: Attendees. https://www.birs.ca/cmo-workshops/2016/16w5152/groupphoto.sized.jpg

Below is a brief summary of the presentations[1]. In addition to the thought-provoking scientific program and lively discussions, attendees enjoyed the perfect weather and features of the fascinating location in Oaxaca, Mexico: colorful colonial buildings, art galleries selling crafts made by indigenous artisans, amazing food, historic museums, and much more.

[1]Slides are available for many of them at `https://www.birs.ca/cmo-workshops/2016/16w5152/files/` and videos of the presentations are available at `https://www.birs.ca/events/2016/5-day-workshops/16w5152/videos`.

Monday, November 28, 2016

Dan Alistarh: Data Structures of the Future: Concurrent, Optimistic, and Relaxed. Dan pointed out inherent limitations when trying to parallelize classic data structures and presented two ways to circumvent them. The first, a software approach, is to relax the semantics of the data structures so that approximate and randomized techniques can be applied; specifically a relaxed priority queue implementation called SprayList was presented. The second approach is to modify the hardware to allow more parallelism, using a technique called Lease/Release.

Michel Raynal: t-Resilient k-Immediate Snapshot and its Relation with Agreement Problems. Michel reminded us about the definition of an immediate snapshot object, which supports operations by which processes can write individual values and can obtain a set of written values with certain containment properties. Previous work has focused on wait-free implementations of this object using registers (i.e., in asynchronous systems in which any number of processes can crash). Michel's presentation described the landscape of computability between k-immediate snapshot objects and x-set agreement in systems that are subject to t crashes, for different relationships between k, x, and t.

Fabian Kuhn: Complexity of Distributed Graph Algorithms in the LOCAL Model. Fabian explained that there is an exponential gap between the round complexities of the best known randomized and best known deterministic algorithms for many graph problems in the LOCAL model (a popular message-passing model for networks). He presented a complexity-theoretic approach to understanding more about this problem and some preliminary results on solving it.

Faith Ellen: Deterministic Objects: Life Beyond Consensus. Faith's presentation concerned the conjecture that any deterministic object with consensus number $m \geq 2$ can be implemented in a wait-free manner using registers and objects with consensus number m no matter how many processes are in the system, as long as it is finite. She showed that this conjecture is false, by demonstrating a few new kinds of objects and showing, using valency arguments, that they cannot be implemented.

Wojciech Golab: Supporting and Analyzing Probabilistic Consistency in Distributed Storage Systems. Wojciech gave an extensive tutorial on the use of randomized data consistency algorithms for storage systems that circumvent impossibility results. The results surveyed were both theoretical and systems-oriented. His own ongoing work includes a mathematical model of eventual consistency and improved ways to tune the consistency-latency tradeoff.

Armando Castañeda: Asynchronous Robot Gathering. Armando gave a tutorial on applying combinatorial topology to analyze problems in distributed computing. The distributed computing tasks to which combinatorial topology has been applied in the past have primarily been variations of consensus. In contrast, the running example that Armando used was the problem of asynchronous robot gathering. Now the problem can be considered under more adversarial models that include asynchrony and failures.

Maurice Herlihy: A Tutorial on Applying Combinatorial Topology to Byzantine Tasks. Maurice continued the topic of combinatorial topology with a tutorial on how to apply it to tasks in the presence of Byzantine failures. It turns out that the concepts extend naturally to this case, and the only "superpower" that Byzantine processes have is to lie about their inputs.

Figure 2: Workshop locale. Photo credit: Hagit Attiya.

Tuesday, November 29, 2016

Dahlia Malkhi: PODC – Practice of Distributed Computing. Dahlia began by stating that real-world distributed applications that need to store state efficiently are a mess and gave a scary (and humorous) outline of how companies try to deal with this, using variants of state machine replication, frequently based on Paxos. During a review of Paxos and related algorithms, she pointed out confusions that arise from mixing logs and conflating durability and scale-out. She presented a new version of Paxos in which only quorums across phases have to intersect. She closed with a call for action to our community to provide better distributed services in practice.

Avery Miller: Communicating with Beeps. Avery presented a new randomized algorithm that solves the global synchronization (or wake-up) problem in a network where processes communicate solely by beeps. The algorithm tolerates channels that fail with some probability. The algorithm can be applied to solve consensus.

Fabian Reiter: Asynchrony and Least Fixpoint Logic. Fabian showed connections between distributed computing and formal logic. His running example involved finite-memory automata on the one hand, and least fixpoint logic on the other. The main result is that quasi-acyclic asynchronous distributed automata are equivalent, on finite directed graphs, to least fixpoint fragment of (backward) μ-calculus.

Michael Bender: Three Backoff Dilemmas. Michael showed us asymptotic analyses of the classical exponential backoff algorithm and more efficient alternatives. The three dilemmas are how to maximize throughput, minimize energy usage, and tolerate jamming or other failures. The talk highlighted recent work on designing and analyzing backoff algorithms when requests arrive dynamically (instead of in an initial batch).

Eli Gafni: 3 Original Sins +. Eli started by posing the question: how to do election in an asynchronous shared memory system with initial failures? The answer is to run any mutual exclusion algorithm and let the process that enters the critical section first write its id. The original sins of the distributed computing community are related to objects versus tasks, modeling failures, and defining progress conditions. He closed with the observation that iterated snapshots are equivalent to communication closed layers; an open question is to iterate other tasks, such as weak symmetry breaking.

Erez Petrank: Memory Management for Lock-Free Data Structures. Erez pointed out that although lock-free data structures have many advantages, providing efficient memory-management support for them is challenging. He described a new scheme for doing so that is optimistic, and a bit in the style of garbage collection.

Leqi Zhu: Space Lower Bound for Consensus. Leqi presented his new result showing that every randomized wait-free (or obstruction-free) consensus algorithm for n processes must use at least $n-1$ registers. This result almost closes the space complexity gap for this problem, as there exist algorithms that use only n registers. Given how long this problem has been open and how complex the prior lower bounds have been, it is striking how simple this proof is.

Rati Gelashvili: Space-based Hierarchy, Buffered Read-Write and Multi-Assignment. Rati pointed out that there is a difference between objects and instructions with respect to consensus number. He proposes a new hierarchy that is with respect to instructions instead of objects and considers what happens with different combinations of instructions. There are many intriguing related open questions.

Wednesday, November 30, 2016

Idit Keidar: Dynamic Memory – It's Time to Talk about Complexity. Idit began by connecting Bitcoin to dynamic storage reconfiguration as motivation. She overviewed the evolution of theoretical research on dynamic memory, which first focused on safety properties, then liveness properties, and more recently abstractions (algorithm building blocks) and complexity (how efficient are these algorithms). Abstractions included configurations (and how to update them), static registers, and speculating snapshots. These abstractions lead to faster algorithms, which ultimately can help speed up Bitcoin transactions.

Anne-Marie Kermarrec: Sketching Your Way Toward Greater Efficiency. Anne-Marie introduced her current work on improving recommender systems, by speeding up the core computation of finding k-closest neighbors in large data graphs. She is focusing on reducing the cost of the bottleneck similarity computation using sketches, which apply several hash functions. She showed some experimental results comparing her approach to others. Ongoing work is to find the right number of sketches and other heuristic parameters.

Seth Gilbert: Gossip, Latency, and Weighted Conductance. Seth's topic was the gossip problem in graphs whose edges have latencies. Prior work in unweighted graphs showed that a graph property called "conductance" is important. Seth presented ways to generalize conductance to weighted graphs, and then discussed upper and lower bounds on the time for gossip.

Jennifer Welch: Message-Passing Implementations of Shared Data Structures. I presented some recent results on the time complexity of operations on linearizable shared objects that are implemented on top of a partially synchronous message-passing system. Upper and lower bounds are presented for general classes of operations, defined axiomatically, that are tight on some cases.

For k-relaxed queues, the worst-case time is almost as bad as that for an unrelaxed queue; however, the amortized time is shown to inherently decrease as k increases.

Excursion: The excursion was to Monte Alban, the ruins of the ancient Zapotec capital, which is one of the most important archeological sites in Latin America.

Figure 3: Monte Alban.

Figure 4: Faith Ellen and Lisa Higham.

Thursday, December 1, 2016

Rotem Oshman, CONGEST Lower Bounds: Beyond Two-Party Reductions. The CONGEST model captures networks with bandwidth restrictions on the edges, which makes it challenging to solve global problems fast. Rotem gave a tutorial on how to prove time lower bounds in the CONGEST model by reductions from results about two-party communication complexity. She then discussed some limitations of such lower bounds and described ways they might be overcome using other techniques from information theory. An open question is how to apply these approaches to shared memory lower bounds, which will require incorporating asynchrony and failures.

Rachid Guerraoui: The Atomic Commit Problem: A Brief History. Rachid surveyed results going back to the 1980's on the computability and complexity of the atomic commit problem. He discussed what failure detectors are necessary and sufficient for different versions of the problem and how many rounds and messages are needed. He concluded with a pointer to his new work on this topic, which focuses on failure-free executions.

Lisa Higham: Test and Set in Optimal Space. Lisa described her work on how to implement a test-and-set object using registers in asynchronous systems subject to crashes, which is impossible to accomplish deterministically in a wait-free manner. The first algorithm is obstruction-free and uses $O(\log n)$ registers, which is optimal. The second algorithm is randomized and wait-free; it also uses $O(\log n)$ registers and has $O(\log^* n)$ expected step complexity against an oblivious adversary. Her talk highlighted four "gadgets" which are combined to produce the algorithms.

George Giakkoupis: Randomized Adversary Models. There is a gap between the theoretically predicted behavior of shared memory algorithms with an adversarial scheduler and the observed behavior, which motivates the need for new models of schedulers. George described his work on a "randomized adversary" that models a family of schedulers between the standard adversarial one and the uniform stochastic scheduler. He presented simple deterministic algorithms for test-and-set and other problems that work with the new scheduler.

Danny Hendler: The Backtracking Covering Proof Technique. Danny described a technique called backtracking covering for proving time lower bounds in shared memory systems. It has been used since 2005 and evolved from covering arguments that were originally used to prove space lower bounds. The idea is first to construct a sequence of L executions such that each one ends when it is about to take a step to an object not yet covered and this step is visible to a fixed process; next to devise a progress function; and then to apply the backtracking covering argument to get an $\Omega(\min(\log L, n))$ time lower bound.

Valerie King: Byzantine Agreement with Full Information: Towards a Practical Algorithm. With the advent of Bitcoin, finding more practical algorithms for solving Byzantine agreement in asynchronous systems with a full-information adversary is especially timely. After reviewing the history of this problem, Valerie focused on her work on algorithms for implementing a fair coin in polynomial time. Most recently, she and an undergraduate have brought the resilience down from about a factor of 6 billion to about a factor of 30.

Petr Kuznetsov: Concurrency as an Iterated Affine Task. Petr considered the asynchronous model in which processes have access to shared k-set-agreement objects. He showed that this model is equivalent to that with shared read-write registers in which at most k processes are active concurrently. The method is to use the techniques from topology, particularly to show that each model can be characterized by an affine task.

Eric Ruppert: Analysing the Average Time Complexity of Lock-Free Data Structures. Eric gave several examples of how to compute the average (amortized) time of operations on lock-free data structures, starting with a stack implemented with a linked list that uses a CAS to swing the top pointer, and then moving on to a more sophisticated example of a single-writer snapshot algorithm using a potential function. For more complicated data structures, where an operation might need, say, two successful CAS's to terminate, he described a way to limit the amount of blame that can be put on non-concurrent operations, so we just need to focus on the contention. He pointed out that bounding the amortized time implies lock-freedom.

Figure 5: Baskets in the market.

Figure 6: Oaxacan cuisine.

1 Friday, December 2, 2016

Yoram Moses: Indistinguishability, Duality and Necessary Conditions. Yoram discussed the notion of indistinguishability, which he called the most fundamental notion in distributed computing, and

pointed out that knowledge is the dual of indistinguishability. He gave a short tutorial on knowledge-based analyses, using consensus as a running example. He closed with a recent result showing how knowledge-based analysis allowed the development of an optimal algorithm for majority consensus.

Ohad Ben Baruch: Lower Bound on the Step Complexity of Anonymous Binary Consensus. Ohad followed up on Danny Hendler's talk on Thursday with an application of the backtracking covering technique to prove a lower bound of $\Omega(\log n)$ on the solo step complexity for solving obstruction-free binary consensus when processes do not have ids. As per the method, first a space lower bound is shown and then it is converted into a time lower bound.

Zahra Aghazadeh: Boundless Tagging with Applications to Wait-Free Memory Reclamation. Zahra presented a new technique for memory reclamation in asynchronous shared memory algorithms. The goal is to identify an object (or a value or a tag, etc.) that no process currently has a reference to, so that it can be reused. She introduced the notion of a taggable register and showed how to use it to implement, in bounded space, a long-lived test-and-set object, a writeable object, and a taggable-register array.