Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

FS 2015

Prof. Dr. Roger Wattenhofer
P. Bissig, P. Brandes, S. Brandt,
K.T. Förster, B. Keller, M. König,
L. Peer, J. Seidel, D. Stolz, J. Uitto

Exam
Principles of Distributed Computing

Thursday, August 20, 2015
9:00 – 11:00

Do not open or turn until told to by the supervisor!

The exam lasts 120 minutes, and there is a total of 120 points. The maximal number of points for each question is indicated in parentheses. Your answers may be in English or in German. Be sure to always justify your answers. Algorithms can be specified in high-level pseudocode or as a verbal description. You do not need to give every last detail, but the main aspects need to be there. Big-O notation is acceptable when giving algorithmic complexities.

Please write down your name and Legi number (your student ID) in the following fields.

Name	Legi-Nr.

Points

Question Nr.	Achieved Points	Maximal Points
1		20
2		22
3		32
4		30
5		16
Total		120

1 Multiple Choice (20 Points)

Evaluate each of the following statements in terms of correctness. Indicate whether a statement is true or not by ticking the corresponding box. Each correct answer gets 1 point, **each wrong answer gets -1 point**. An unanswered statement gets 0 points. If the sum of collected points is negative, then you get 0 points for this question set.

Statement	**true**	**false**
In the Heavy-Light-Decomposition, every internal node has exactly one heavy child.	☐	☐
Let u and v be two nodes. If h is a hub node that appears in both u's and v's labels, then the distance $d(u, v)$ between u and v is $d(u, h) + d(h, v)$.	☐	☐
If G is connected, then the Naïve-Hub-Labeling adds at least one hub node to the label of each node.	☐	☐
If a graph can be colored with c colors in r rounds, then there exists an MIS algorithm with running time at most $\mathcal{O}(r + c)$.	☐	☐
Mutual exclusion can only be solved if at least one read-modify-write (RMW) operation, such as test-and-set, is available.	☐	☐
Let n be the number of processes in a shared memory system. After at least 3 store operations, the step complexity of the collect operation for any algorithm can be $\Omega(n)$ in the worst case.	☐	☐
If a splitter is accessed by $k > 2$ processes at the same time, at least 1 process exits with value **left**.	☐	☐
The Greedy Sequential algorithm (Algorithm 1) always produces an optimal (minimal number of colors) coloring.	☐	☐
The 6-Color algorithm (Algorithm 4) only terminates after $\mathcal{O}(\log^* n)$ rounds if the nodes initially have the IDs from 1 to n.	☐	☐
Every ring can be colored with 2 colors.	☐	☐
Slotted Aloha is a typical example of a uniform distributed algorithm.	☐	☐
Assume a wireless network where every node has an ID, and the goal is that every node knows all IDs. This requires *always* at least n rounds.	☐	☐
Deterministic leader election is always possible in anonymous and uniform trees with an odd number of nodes.	☐	☐
In one round of the GHS algorithm, at most $\log n$ merge requests are transmitted.	☐	☐
The time complexity of the α synchronizer is the diameter of a spanning tree rooted at the leader node l.	☐	☐
A Half Cleaner with n input wires is a comparison network of depth $n/2$.	☐	☐
There are $2n$ possible binary bitonic sequences of length n.	☐	☐
Algorithms with runtime k can often be made self stabilizing with stabilization time k.	☐	☐
Increasing the parameter α of an augmented grid increases the expected number of directed edges into the local neighborhood on the grid each node has.	☐	☐
Every wireless distributed algorithm assuming no collision detection can also be run in a model with collision detection.	☐	☐

2 Independent Dominating Sets (22 Points)

In the lecture, the notions of independent/dominating set were introduced. Let $G = (V, E)$ be a connected undirected graph, with V being the set of n nodes with unique identifiers and E being the set of m edges. A set of nodes I is a maximal independent set (MIS), if i) no two nodes of I are adjacent, and ii) no other node from V may be added to I without violating condition i). Similarly, a set of nodes S is a dominating set (DS), if every node in V is in S or has a neighbor in S.

Definition (Independent Dominating Set (IDS))**.** *We call a set of nodes S an independent dominating set, if S is a dominating set, and no two nodes of S are adjacent.*

A) [5] Is there an IDS for every graph?

If yes: Show that every graph has an IDS.

If no: Draw/construct a graph that does not have an IDS.

In practice, one wants to find dominating sets that are small, i.e., contain as few nodes as possible. The sets which contain the fewest nodes are called minimum dominating sets (MDS).

We hope that IDS are small, since they avoid unnecessary domination of nodes that are in the dominating set themselves.

B) [6] Show that there exists a graph G, for which *the smallest* IDS contains many more nodes than the MDS. Draw/describe G, the MDS, the IDS, and compare the size of the MDS and the IDS. Graphs with worse asymptotic ratios between IDS and MDS give more points.

For the remainder of this exam question, we study unrooted **trees** where every node has a unique ID of size $\mathcal{O}(\log n)$, and we assume that every node knows its own degree.

C) [8] Give a deterministic algorithm for the synchronous time model that constructs an IDS, where each node only sends $\mathcal{O}(1)$ many messages of at most $\mathcal{O}(\log n)$ size. (Note that sending the same message in the same round to 5 neighbors counts as sending 5 messages.) What is the runtime of your algorithm?

D) [3] Assume that you constructed an IDS, but then, one of the **leaves** crashes. Potentially, some of the nodes must now enter / leave the IDS, to fix the problem. Show how far this "fixing" can propagate at most in the worst case (with the restriction that you want to change as few nodes as possible).

3 Non-Adjacency Labeling Scheme (32 Points)

Let us study labeling schemes for *non-adjacency* in trees. For two nodes u, v, the decoder should output TRUE if u and v are *not* adjacent, and FALSE if u and v are adjacent. We are particularly interested in *efficient* labeling schemes, i.e., in labeling schemes where the label size is small.

A) [4] Devise an efficient labeling scheme for the special case where the depth of the tree is 1, i.e., for star graphs.

B) [8] Show that already for trees of depth 2, the label size must be $\Omega(\log n)$ bits.

Reminder: Analogously to the big-O notation, the notation $f(n) \in \Omega(g(n))$ means that f is bounded from *below* by g asymptotically. More precisely, it means that there is some constant k so that $f(n) \geq k \cdot g(n)$ for all n greater than some constant n_0.

C) [4] Sketch an efficient labeling scheme for arbitrary depth trees. State the label size and explain why your scheme is correct.

D) [8] Hoping to find a scheme with smaller label size, you accept one-sided errors. That is, we are now looking for an *erroneous labeling scheme* in which the decoder may err if the two nodes in question are not adjacent. The decoder should always return a correct answer if the two nodes in question are adjacent.

Design an erroneous labeling scheme for non-adjacency in trees with a label size of 1 bit. In the worst-case tree, what is the probability that your scheme gives an incorrect answer (claiming that non-adjacent nodes are adjacent) for two randomly picked nodes?

E) [8] The error probability from **D)** is too large in practice. How large do the labels need to be to obtain a correct answer with high probability, i.e., with probability $(1 - n^{-c})$ for arbitrary c? Design an efficient erroneous labeling scheme for non-adjacency in trees that is correct with high probability.

4 Very Small Diameter (30 Points)

Let $G = (V, E)$ be a simple connected network with n nodes, diameter D and highest node degree Δ. The nodes do not know the value of n. Each node of G has a unique ID containing at most $\mathcal{O}(\log n)$ bits. All considered algorithms proceed in synchronous rounds. In each round, each node v is allowed to send $\mathcal{O}(\log n)$ bits over each incident edge.

A) [4] Can D be computed in time $\mathcal{O}(\frac{n}{\log n})$ if it is known beforehand that n is even? Justify your answer.

In the lecture notes the following is stated: "Note that one can check whether a graph has diameter 1 by exchanging some specific information such as degree with the neighbors." But how exactly does this work?

B) [12] Give a synchronous algorithm where each node outputs "=" or "$\neq$" such that the following hold:

- If $D = 1$, then all nodes output "=".
- If $D \neq 1$, then at least one node outputs "$\neq$".

Show that your algorithm is correct and analyze its runtime.

The number of points awarded will depend on the (asymptotic) runtime, faster algorithms giving more points, e.g., the maximally possible number of points are

- 12 points for an achieved runtime of $\mathcal{O}(1)$,
- 6 points for an achieved runtime of $\mathcal{O}(\Delta)$,
- 3 points for an achieved runtime of $\mathcal{O}(n)$.

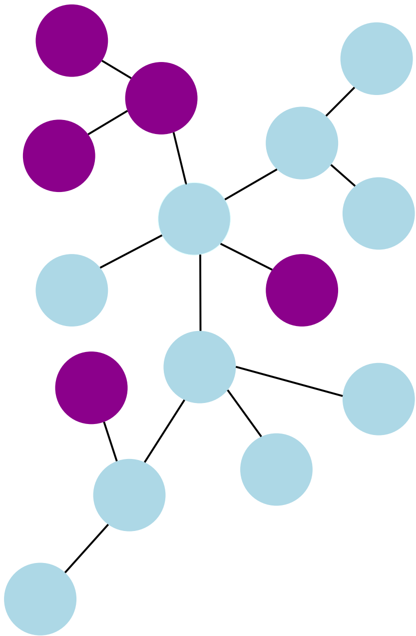
C) [14] Give a synchronous algorithm that determines whether $D = 1$ (i.e. *each* node knows in the end whether $D = 1$). Show that your algorithm is correct and analyze its runtime.

The given task can be solved in time $\mathcal{O}(1)$, but this is not trivial. Thus, again, points will be awarded for slower algorithms, e.g.,

- maximally 14 points for an achieved runtime of $\mathcal{O}(1)$,
- maximally 6 points for an achieved runtime of $\mathcal{O}(\Delta)$,
- maximally 3 points for an achieved runtime of $\mathcal{O}(n)$.

5 Reading Assignment: Questions (16 Points)

We take a closer look at the model *biased preferential attachment* presented in the reading assingment using the following example:



A) [2] Guess from this graph the entry rate r of the minority nodes.

B) [6] The graph was built using $\rho = 0.8$ and $r = 0.3$. Now a minority node enters. What is the probability that it attaches itself to another minority node?

C) [4] Assume now a network is built using the theoretical model from the paper. The three basic assumptions in the model hold, but the start of the network is crooked and the first 100 nodes entering are female. After this the network behaves exactly as described in the paper. How does this affect the outcome of the model after another 10000 nodes joined?

D) [4] Criticize the paper. What is a shortcoming, and why?