Brief Announcement: Agreement Tasks in Fault-Prone Synchronous Networks of Arbitrary Structures

Pierre Fraigniaud
Institut de Recherche en Informatique Fondamentale (IRIF)
CNRS and Université Paris Cité, France

Minh Hang Nguyen
Institut de Recherche en Informatique Fondamentale (IRIF)
CNRS and Université Paris Cité, France

Ami Paz
Interdisciplinaire des Sciences du Numérique (LISN)
CNRS and Université Paris-Saclay, France

Abstract

Consensus is arguably the most studied problem in distributed computing as a whole, and particularly in distributed message-passing settings. Research on consensus has considered various failure types, memory constraints, and much more. Surprisingly, almost all of this work assumes that messages are passed in a complete network, i.e., each process has a direct link to every other process. Set agreement, a relaxed variant of consensus, has also been heavily studied in different settings, yet research on it has also been limited to complete networks. We address this situation by considering consensus and set agreement in general networks, i.e., that can have an arbitrary graph G as their communication graph. We focus on fault-prone networks, where up to t nodes may crash and irrevocably stop communicating, and present upper and lower bounds for such networks. We establish the following collection of results:

- The consensus algorithm by [Castañeda et al., 2023] is optimal for *all* graphs, and not only for symmetric graphs.
- This algorithm can be extended to a generic algorithm for k-set agreement, for every $k \geq 1$. For $k = 1$, our generic algorithm coincides with the existing one for consensus.
- All these algorithms can be extended to the case where the number t of failures exceeds the connectivity κ of the graph, while the existing consensus algorithm assumed that $t < \kappa$.

2012 ACM Subject Classification Theory of computation → Distributed algorithms

Keywords and phrases Consensus, set-agreement, fault tolerance, crash failures

Digital Object Identifier 10.4230/LIPIcs.DISC.2024.47

Funding ANR projects DUCAT (ANR-20-CE48-0006), QuDATA (ANR-18-CE47-0010), and TEMPORAL (ANR-22-CE48-0001); the European Union's Horizon 2020 program H2020-MSCA -COFUND-2019 Grant agreement n° 945332.
Pierre Fraigniaud: Additional support from ANR projects DUCAT (ANR-20-CE48-0006) and QuDATA (ANR-18-CE47-0010).
Minh Hang Nguyen: Additional support from ANR projects DUCAT (ANR-20-CE48-0006), TEMPORAL (ANR-22-CE48-0001), and the European Union's Horizon 2020 program H2020-MSCA -COFUND-2019 Grant agreement n° 945332.

Acknowledgements The authors want to thank Mikaël Rabie for early discussions on the subject.

38th International Symposium on Distributed Computing (DISC 2024).
Editor: Dan Alistarh; Article No. 47; pp. 47:1–47:5
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

The standard *synchronous t-resilient message-passing* model, for $t \geq 0$, assumes $n \geq 2$ nodes labeled from 1 to n and connected as a clique, i.e., as a complete graph K_n. Computation proceeds in synchronous rounds, during which every node can send a message to every other node, receive the message sent by every other node, and perform some local computation. Up to t nodes may crash during the execution of an algorithm, and when a node v crashes at some round $r \geq 1$ it stops functioning after round r and never recovers. Moreover, some (possibly all) of the messages sent by v at round r may be lost. This model has been extensively studied in the literature [2, 7]. In particular, it is known that consensus can be solved in $t+1$ rounds in the t-resilient model [6], and this is optimal for every $t < n-1$ as far as the worst-case complexity is concerned [1, 6]. Similarly, k-set agreement, in which the cardinality of the set of output values decided by the (correct) nodes must not exceed k, is known to be solvable in $\lfloor t/k \rfloor + 1$ rounds, which are also necessary [4].

It is only very recently that the synchronous t-resilient model has been extended to settings in which the complete communication graph K_n is replaced by an arbitrary communication graph G [3, 5]. Specifically, let $\kappa(G)$ denote the node-connectivity of G, which is the smallest number of nodes whose removal disconnects G. If the number of failures is smaller than the connectivity of the graph, i.e., if $t < \kappa(G)$, then consensus in G can be solved in $\text{radius}(G,t)$ rounds in the t-resilient model [3], where $\text{radius}(G,t)$ generalizes the standard notion of graph radius to the scenarios in which up to t nodes may crash. For $t=0$, $\text{radius}(G,0)$ is the standard radius of the graph G, and, for the complete graph K_n, $\text{radius}(K_n,t) = t+1$ for every $0 \leq t < n-1$ (while $\text{radius}(K_n, n-1) = n-1$). Therefore, the $\text{radius}(G,t)$ upper bound for consensus in G in the t-resilient model generalizes the seminal $t+1$ upper bound for consensus in K_n in the same model. The algorithm of [3] is *oblivious*, that is, the output of a node is solely based on the set of pairs (*node-identifier, input-value*) collected by that node during $\text{radius}(G,t)$ rounds (and not, e.g., from whom, when, and how many times it received each of these pairs). In other words, the consensus algorithm of [3] is generic, i.e., it applies to any graph G.

For a fixed graph G, the optimality of the consensus algorithm performing in $\text{radius}(G,t)$ rounds is left as an open question in [3]. It was conjectured there that for every graph G and every $0 \leq t < \kappa(G)$, no oblivious algorithms can solve consensus in G in less than $\text{radius}(G,t)$ rounds, but this was only proved for the specific case of *symmetric* (a.k.a. *vertex-transitive*) graphs. This lower bound does not come entirely as a surprise since all nodes of a symmetric graph have the same eccentricity (i.e., maximum distance to any other node), even when generalized to include crash failures. The fact that all nodes have the same eccentricity implies that they can merely be ordered according to their identifiers for selecting the output value from the received pairs (*node-identifier, input-value*). Instead, if the graph is not symmetric, a node that received a pair (*node-identifier, input-value*) after $\text{radius}(G,t)$ rounds does not known whether all the nodes have received this pair, and thus the choice of the output value from the set of received pairs is more subtle. This not only complicates the design of an upper bound but also makes the determination of a lower bound more involved.

2 The Model

We use the (synchronous) t-resilient model for networks as defined in [3]. Let $G = (V,E)$ be an n-node undirected graph, which is also connected and simple (i.e., no multiple edges nor self-loops). Initially, every node knows the graph G, that is, it knows the identifiers of all nodes and how they are connected. The uncertainty is thus not related to the initial structure of the connections, but is only due to the presence of potential failures. More specifically, computation in G proceeds as a sequence of synchronous rounds, and each node may fail by crashing – when a node crashes, it stops functioning and never recover. However, if a node v crashes at round r it may still succeeds in sending messages to a subset of its set $N(v)$ of neighbors.

For every positive integer $t \geq 0$, the t-resilient model assumes that at most t nodes may crash. A *failure pattern* is thus defined as a set $\varphi = \{(v, F_v, f_v) \mid v \in F\}$ where $F \subset V$, $0 \leq |F| \leq t$, is the set of faulty nodes in φ, and a triplet (v, F_v, f_v) designates that $v \in F$ fails at round f_v and fails to send messages to the nodes of F_v, $\varnothing \neq F_v \subseteq N(v)$, at this round. In any execution of an algorithm in the t-resilient model, the nodes know t, but do not know in advance which failure pattern occurs. The set of all failure patterns in which at most t nodes fail is denoted by $\Phi^{(t)}_{\text{all}}$.

The *eccentricity* of a node v in G with respect to a failure pattern φ, denoted by $\text{ecc}(v, \varphi)$, is the minimum number of rounds for broadcasting a message from v to all *correct* nodes of G under φ. The broadcast protocol is by flooding, i.e., when a node receives a message at round r, it forwards it to all its neighbors at round $r + 1$. Note that $\text{ecc}(v, \varphi)$ might be infinite, in case v cannot broadcast to all correct nodes in G under φ. Let

$$\Phi^\star_v = \{\varphi \in \Phi^{(t)}_{\text{all}} \mid \text{ecc}(v, \varphi) < \infty\}$$

denote the set of failure patterns in the t-resilient model in which v eventually manages to broadcast to all correct nodes. The *t-resilient radius* of G is then defined as

$$\text{radius}(G, t) = \min_{v \in V} \max_{\varphi \in \Phi^\star_v} \text{ecc}(v, \varphi).$$

Castañeda et al. [3] have designed a generic oblivious consensus algorithm which, for every graph G, and every number t of failures with $t < \kappa(G)$, runs in $\text{radius}(G, t)$ rounds. In addition, they have shown that, for every symmetric graph G, and every $t < \kappa(G)$, no oblivious algorithms can solve consensus in G with t crash failures in less than $\text{radius}(G, t)$ rounds.

3 Our Results

We extend the investigation of the t-resilient model in arbitrary graphs, in various complementary directions. The proofs of these results can be found in the full version of the paper.

3.1 Lower Bounds

We establish a general lower bound for consensus in the aforementioned synchronous t-resilient model for network, which states that the oblivious consensus algorithm from [3] is optimal among oblivious algorithms for *every* graph G, and not only for symmetric graphs.

▶ **Theorem 1.** *For every graph G and every $t < \kappa(G)$, consensus in G cannot be solved in less than* $\text{radius}(G, t)$ *rounds by an oblivious algorithm in the t-resilient model.*

3.2 Set-Agreement

We demonstrate the existence of a generic oblivious algorithm for k-set agreement. This algorithm is generic in the sense that it obeys a general structure: (1) flooding the graph with the inputs of a predetermined "core set" of nodes $C(G) \subseteq V$, for $R(G)$ rounds, and (2) after

$R(G)$ rounds, letting every node $v \in V$ pick the input of the node $u \in C(G)$ with smallest identifier among all the nodes in $C(G)$ received by v. We show that for every graph G, every $t < \kappa(G)$, and every $k \geq 1$, k-set agreement is solved in radius(G, t, k) rounds, where radius(G, t, k) extends the standard notion of graph radius to the case in which there are k centers, and whenever up to t nodes can crash. For $t = 0$ and $k = 1$, radius(G, t, k) coincides with the standard radius of G. Moreover, for $k = 1$, radius$(G, t, 1) =$ radius(G, t).

More concretely, like in the k-center problem, we consider broadcast in G from a set $S \subseteq V$ of k nodes by flooding, and radius(G, t, k) essentially denotes the minimum, taken over all sets S of k nodes, of the broadcast time of S, i.e., of the smallest number of rounds sufficient to guarantee that every non-faulty node receives information from at least one node in S. The definition is a bit more subtle though, as the broadcast time of S actually depend on the failure pattern (i.e., which nodes crash and when), and it may even be the case that S cannot broadcast at all for some failure patterns (e.g., whenever all nodes in S crash at the first round without sending any messages to their neighbors). More specifically, for every set $S \subseteq V$ of size at most k, let the eccentricity of S with respect to a failure pattern φ, denoted by ecc(S, φ), be the minimum number of rounds such that whenever every node in S broadcasts information, every correct node of G under φ receives the information sent by at least one of the nodes in S. Let

$$\Phi_S^\infty = \{\varphi \in \Phi_{\text{all}}^{(t)} \mid \text{ecc}(S, \varphi) = \infty\},$$

and let $\Phi_S^\star = \Phi_{\text{all}}^{(t)} \setminus \Phi_S^\infty$. The *$k$-center t-resilient radius* of G is then defined as

$$\text{radius}(G, t, k) = \min_{\substack{S \subseteq V \\ |S| \leq k}} \max_{\varphi \in \Phi_S^\star} \text{ecc}(S, \varphi).$$

▶ **Theorem 2.** *For every graph G, every $k \geq 1$, and every $t < \kappa(G)$, k-set agreement in G can be solved in* radius(G, t, k) *rounds by an oblivious algorithm in the t-resilient model.*

3.3 Beyond the Connectivity Threshold

Finally, inspired by [5], we extend the study of consensus and set agreement in the t-resilient model in arbitrary graphs to the case where the number t of crash failures is arbitrary, i.e., not necessarily lower than the connectivity $\kappa(G)$ of the considered graph G. We show that our generic k-set agreement algorithm, which include the case of consensus for $k = 1$, can be extended to this framework, at the mere cost of relaxing consensus and k-set agreement to impose agreement to hold within each connected component of the graph resulting from removing the faulty nodes from G. Under this somehow unavoidable relaxation, we present extension of the consensus algorithm from [3] in particular, and of our k-set agreement algorithm in general, to t-resilient models for $t \geq \kappa(G)$, and express the round complexities of these algorithms in term of a straightforward extension of the radius notion to disconnected graphs.

4 Discussion

We have completed the picture for consensus in the t-resilient model for arbitrary graphs, by proving that the consensus algorithm in [3] is optimal among oblivious algorithms. Moreover, we have designed a generic (oblivious) algorithm for k-set agreement in arbitrary graph G performing in radius(G, t, k) rounds under the t-resilient model, for $t < \kappa(G)$.

Our results open a vast domain for further investigations. In particular, what could be said for sets of failure patterns Φ other than $\Phi_{\text{all}}^{(t)}$? Another intriguing and potentially challenging area for further research is exploring scenarios where no upper bound on the number of failing nodes is assumed, while concentrating solely on failure patterns that do not result in the disconnection of the graph. Finally, the design of early-stopping algorithms in the t-resilient model for arbitrary graphs is also highly desirable. The algorithms in [5], early stopping and others, are very promising, but their analysis must be refined to a grain finer than the stretches of the failure patterns, by focusing, e.g., on eccentricities and radii.

References

1 Marcos Kawazoe Aguilera and Sam Toueg. A simple bivalency proof that t-resilient consensus requires t+ 1 rounds. *Information Processing Letters*, 71(3-4):155–158, 1999. `doi:10.1016/S0020-0190(99)00100-3`.

2 Hagit Attiya and Jennifer Welch. *Distributed computing: fundamentals, simulations, and advanced topics*, volume 19. John Wiley & Sons, 2004.

3 Armando Castañeda, Pierre Fraigniaud, Ami Paz, Sergio Rajsbaum, Matthieu Roy, and Corentin Travers. Synchronous t-resilient consensus in arbitrary graphs. *Inf. Comput.*, 292:105035, 2023. `doi:10.1016/J.IC.2023.105035`.

4 Soma Chaudhuri, Maurice Herlihy, Nancy A. Lynch, and Mark R. Tuttle. Tight bounds for k-set agreement. *J. ACM*, 47(5):912–943, 2000. `doi:10.1145/355483.355489`.

5 Bogdan S. Chlebus, Dariusz R. Kowalski, Jan Olkowski, and Jedrzej Olkowski. Disconnected agreement in networks prone to link failures. In *SSS*, volume 14310 of *LNCS*, pages 207–222. Springer, 2023. `doi:10.1007/978-3-031-44274-2_16`.

6 Danny Dolev and H. Raymond Strong. Authenticated algorithms for byzantine agreement. *SIAM Journal on Computing*, 12(4):656–666, 1983. `doi:10.1137/0212045`.

7 Maurice Herlihy, Dmitry N. Kozlov, and Sergio Rajsbaum. *Distributed Computing Through Combinatorial Topology.* Morgan Kaufmann, 2013.