Brief Announcement: On the Round Complexity of Distributed Consensus over Synchronous Networks

D.V.S. Ravikant V. Muthuramakrishnan V. Srikanth K. Srinathan* C. Pandu Rangan†

Categories and Subject Descriptors: C.3.2 [Distributed Systems]: Distributed applications
General Terms: Algorithms, Reliability, Theory
Keywords: Consensus, Round Complexity.

In a synchronous network, it is well-known that $t+1$ rounds are necessary and sufficient to achieve distributed consensus tolerating t stopping faults[2]. In this work, we show that in a network consisting of all k-cast channels, the corresponding number of rounds is $\lfloor (t-1)/k \rfloor + 2$.

THEOREM 1. *Consider a synchronous round-based system with n players connected by a network having all k-casts. Suppose that at most t crash-failures can occur with at most k-players crashing in each round.*[1] *If $n > t+k$, there is no algorithm that solves consensus in $\lambda = \lfloor \frac{t-1}{k} \rfloor + 1$ rounds.*

Proof: We assume that there exists a protocol A that achieves consensus in λ rounds and arrive at a contradiction. The proof is based on the standard bivalency argument using forward induction. A particular configuration C of a synchronous system is univalent if there is only one value that the correct players can agree upon. C is said to be bivalent if it is not univalent (either 1-valent or 0-valent). In the following, a l-round partial run r_l denotes the execution of A up to the end of round l. We prove two lemmas similar to [1]. The second one contradicts the first and completes the necessity proof of the theorem.

Lemma: Any $(\lambda-1)$-round run $r_{\lambda-1}$ is univalent.

Proof: Suppose $r_{\lambda-1}$ is bivalent. w.l.g. assume that the λ-round run r^0 obtained by extending $r_{\lambda-1}$ by one round such that no player crashes in round λ is 0-valent. Let r^1 be a 1-valent extension of $r_{\lambda-1}$ where some players crash in round λ. The only difference between r^0 and r^1 is that some messages $\{m_1, m_2, \ldots, m_s\}$ were sent in r^0 but not in r^1. We define runs r^i for all $2 \leq i \leq s+1$, as follows: For every i, $1 \leq i \leq s$, r^{i+1} is identical to r^i, except that the message m_i was sent in round λ. If m_i was sent along the k-cast Δ_i then for every player other than the recipients of Δ_i, r^{i+1} is indistinguishable from r^i. Note that, since $n > t+k$, this includes at least one correct player. This implies that each of these runs is 1-valent. However the view of any correct player c in r^{s+1} is the same as that in r^0, which means that c should decide 0 in r^{s+1}, giving the contradiction.

Lemma: There is a bivalent $(\lambda-1)$-round run $r_{\lambda-1}$.

Proof: We show by induction on l that for each l, $0 \leq l \leq \lambda-1$, there is a bivalent l-round partial run r_l.

From [1], there exists an initial bivalent configuration C_0. Let r_0 be the 0-round partial run ending in C_0. Assume, for contradiction, that every one-round extension of r_l is univalent. Let r^*_{l+1} be the (univalent) partial run obtained by extending r_l by one round such that no new crashes occur. w.l.g. assume that it is 0-valent. Since r_l is bivalent and every one-round extension of r_l is univalent, there is at least one one-round extension r^1_{l+1} of r_l that is 1-valent. Suppose the messages $m_1, \ldots, m_s$ were not sent in round $l+1$ in r^1_{l+1}. The only difference between r^*_{l+1} and r^1_{l+1} is that the messages $m_1, \ldots, m_s$ were sent in r^*_{l+1} but not in r^1_{l+1}. Starting from r^1_{l+1}, we now define $l+1$-round partial runs as follows. For every j, $1 \leq j \leq s$, r^{j+1}_{l+1} is identical to r^j_{l+1}, except that the message m_j was sent in round $l+1$. Note that for every j, $1 \leq j \leq s+1$, r^j_{l+1} is univalent. There are two cases:

1. There is a j, $1 \leq j \leq s$, such that r^j_{l+1} is 1-valent while r^{j+1}_{l+1} is 0-valent. Extend partial runs r^j_{l+1} and r^{j+1}_{l+1} into runs r and r', respectively, by crashing the k recipients of Δ_j at the beginning of round $l+2$, and continuing with no additional crashes. Note that (a) no player except the recipients of Δ_j can distinguish between r and r', and (b) all correct players must decide 1 in r and 0 in r' – a contradiction.
2. $\forall j$, $1 \leq j \leq s+1$, r^j_{l+1} is 1-valent. (like in case 1.) □

To prove the sufficiency condition, we give an optimal protocol. Let P be the set of players. For $1 \leq i \leq n$, let x_i be the initial value of player p_i. A message sent by a player is of the form (p_h, x_h, S) where $x_h \in \{0,1\}$ is the initial value of the player p_h and $S \subset 2^P$. The following protocol is executed by the player p_i.

(1) Set $W_i = x_i$. Send $(p_i, x_i, \{p_i\})$ along all k-casts that p_i can use.
(2) For any round $r > 1$,
(a) If (p_h, x_h, S) was received in round $(r-1)$ through k-cast Δ', send $(p_h, x_h, S \cup \Delta')$ using Δ, for every Δ such that $\Delta \cap S = \emptyset$. If such a k-cast does not exist then use a k-cast that covers $(P - S)$.
(b) for every message (p_h, x_h, S) received, update $W_i = W_i \cup \{x_h\}$.
(3) After $\lambda+1$ rounds, if $W = \{v\}$ finalvalue=v else finalvalue=0.

The proof of correctness of the protocol is sketched below. Let $1 \leq r < \frac{n}{k}$. Any message (p_h, x_h, S) sent during round r has $|S| > rk$. This ensures the following. If p_i and p_j are active players at the end of round r, $1 \leq r \leq \lambda$, and p_i knows the initial value of p_h, and p_j does not, then at least $(r-1)k+1$ players crashed by the end of round r. Thus, if at most $(r-1)k$ players crashed by the end of round r then $W_i = W_j$ for any two active players p_i and p_j, and hence all correct players decide on the same value.

REFERENCES

[1] M.K. Aguilera and S. Toueg. A simple bivalency-based proof that t-resilient consensus requires t+1 rounds. IPL 71(3-4):155–158, 1999.
[2] M. J. Fischer and N. A. Lynch. A lower bound for the time to assure interactive consistency. IPL 4(14):183–186, 1982.

*Financial support from Infosys Tech. Ltd., India, is acknowledged.
†Contact author. All the authors are affiliated to the Department of Computer Science and Engineering, Indian Institute of Technology, Madras, 600036, INDIA. email: rangan@iitm.ernet.in
[1]The lower bound holds even for a restricted failure pattern.

PODC'04, July 25–28, 2004, St. John's, Newfoundland, Canada.
ACM 1-58113-802-4/04/0007.