NASA/TM-2015-218789



Achieving Agreement In Three Rounds With Bounded-Byzantine Faults

Mahyar R. Malekpour
Langley Research Center, Hampton, Virginia

August 2015

NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counter-part of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov

- E-mail your question to help@sti.nasa.gov

- Phone the NASA STI Information Desk at 757-864-9658

- Write to:
  NASA STI Information Desk
  Mail Stop 148
  NASA Langley Research Center
  Hampton, VA 23681-2199

NASA/TM-2015-218789



Achieving Agreement In Three Rounds With Bounded-Byzantine Faults

Mahyar R. Malekpour
Langley Research Center, Hampton, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

August 2015

The use of trademarks or names of manufacturers in this report is for accurate reporting and does not constitute an official endorsement, either expressed or implied, of such products or manufacturers by the National Aeronautics and Space Administration.

Available from:

NASA STI Program / Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199
Fax: 757-864-6500

Abstract

A three-round algorithm is presented that guarantees agreement in a system of $K \geq 3F+1$ nodes provided each faulty node induces no more than F faults and each good node experiences no more than F faults, where, F is the maximum number of simultaneous faults in the network. The algorithm is based on the Oral Message algorithm of Lamport *et al.* and is scalable with respect to the number of nodes in the system and applies equally to the traditional node-fault model as well as the link-fault model. We also present a mechanical verification of the algorithm focusing on verifying the correctness of a bounded model of the algorithm as well as confirming claims of determinism.

Keywords: Oral Message, Agreement, Byzantine, fault tolerant, synchronization, distributed, model checking

Table of Contents

ABSTRACT ..... ERROR! BOOKMARK NOT DEFINED.

TABLE OF CONTENTS ..... **II**

1. INTRODUCTION **1**

2. FAULT MODELS **3**

3. SYSTEM OVERVIEW **3**

3.1. COMMUNICATION DELAY 4

3.2. THE *SYNC* MESSAGE AND ITS VALIDITY 4

4. 3ROM **5**

4.1. PROOF OF THE *3ROM* ALGORITHM FOR LINK-FAULT MODEL 6

4.2. PROOF OF THE *3ROM* ALGORITHM FOR NODE-FAULT MODEL 7

4.3. MESSAGE OBSERVATION WINDOW, AGREEMENT WITHIN A TIME BOUND 10

4.4. COMPLEXITY OF THE *3ROM* ALGORITHM 11

5. MODEL CHECKING **11**

5.1. MODEL CHECKED PROPOSITIONS 12

6. CONCLUSIONS **13**

REFERENCES **13**

1. Introduction

Distributed systems have become an integral part of safety-critical computing applications, necessitating system designs that incorporate complex fault-tolerant resource management functions to provide globally coordinated operations with ultra-reliability. As a result, robust clock synchronization has become a required fundamental component of fault-tolerant safety-critical distributed systems. Synchronization has practical significance as a fundamental service for higher-level algorithms that solve other problems. For example, in safety-critical TDMA (Time Division Multiple Access) architectures [1, 2, 3], synchronization is the most crucial element of these systems. Typically, the assumed topology is a regular graph such as a fully connected graph or a ring since they provide a base case to solve the distributed synchronization problem.

A fundamental property of a robust distributed system is the capability of tolerating and potentially recovering from failures (loss of service due to a fault) that are not predictable in advance. A *fault* is a defect or flaw in a system component resulting in an incorrect state [2, 4]. In the context of fault-tolerant distributed systems, a fault presenting different symptoms to different observers is known as a Byzantine (arbitrary) fault. We assume that there are a maximum of F simultaneous faults in the network. The requirement to handle faults adds a new dimension to the complexity of the synchronizing distributed systems.

We call an approach to solving the clock synchronization problem *direct* if it relies solely on local (node level) detection and filtering of faults. This approach is primarily limited to detecting timing and/or value faults of a node's incoming messages. In contrast, we call an approach *indirect* if it relies on the network level detection and filtering of faults independent of, and in addition to, the local detection and filtering of the faults. This approach however requires coordination at the network level.

Thus far, there is no verifiable solution for the general case of the clock synchronization problem, where the topology is arbitrary and any number of various types of faults are tolerated. Furthermore, most attempts have been in trying to solve this problem directly, although there are some approaches to solve this problem indirectly using authenticated (signed) messages [5]. Driscoll *et al.* in [6] however argues that "while the arguments of unforgeable signed messages make sense in the context of communicating generals, the validity of necessary assumptions in a digital processing environment is not supportable. In fact, the philosophical approach of utilizing cryptography to address the problem within the real world of digital electronics makes little sense. The assumptions required to support the validity of unbreakable signatures are equally applicable to simpler approaches (such as appending a simple source ID or a CRC to the end of a message). It is not possible to prove such assumptions analytically for systems with failure probability requirements near 10^{-9}/hr." Furthermore, addressing network element imperfections, such as oscillator drift with respect to real time and differences in the lengths of the physical communication media, is necessary to make a solution applicable to realizable systems.

The main issue in solving the clock synchronization problem is a lack of symmetric view (agreement) in the system at the participating good nodes in the sense that two good nodes may

disagree on the message sent. However, there are a number of ways of achieving message symmetry across the system. In [5, 7] various ideas for overcoming failures in a robust distributed system are addressed that include tolerating Byzantine faults. In solving the consensus problem, which is the ability of a set of nodes to agree on a single value despite failures, Schmid *et al.* argue in [8] that: "A fully-fledged n-process consensus algorithm is obtained by using a separate instance of a Byzantine agreement algorithm (with n-1 receivers) for disseminating any process's local value, and using a suitable choice function (majority) for the consensus result[*]." The consensus problem, and hence, the proposed idea by Schmid, is based on inherent assumption of synchrony among the good nodes, and so is not applicable to solving the clock synchronization problem.

Other methods include using variety of engineering practices, e.g., using a self-checking pair at the node level [9, 10] or central guardians at the system level [11, 12]. However, as Driscoll *et al.* reported in [6], correctness of claims of these approaches may not be verifiable. Furthermore, we believe that to be generally useful, algorithms that guarantee agreement must be able to handle non-authenticated messages. Thus, the crux of our idea, as proposed in [13], is to solve this problem indirectly by first converting any message to a symmetric message, and then use a verified protocol based on the symmetry assumption to solve the synchronization problem.

The Oral Message (OM) algorithm of Lamport *et al.* [5] that solves the Byzantine Agreement (BA) problem [14] is also an indirect approach, and is meant to reliably transform a message from a single source to a symmetric message (an agreement). The OM algorithm has been proven to reach agreement at the network level for a given source [5, 14, 15] and does not require initial synchrony among the good nodes. The OM requires $F+1$ rounds of exchanges and, with a message complexity of $O(K^F)$, the number of exchanged messages grows exponentially as F grows linearly. Therefore, the use of the OM algorithm for $F > 2$ is very costly and impractical.

In this paper, we present an alternative for achieving agreement, hereafter referred to as *3ROM* (3 Rounds using OM) algorithm that is based on the OM algorithm. The *3ROM* assumes each node N_i, $i = 1..K$, either induces up to F faults if it is a faulty node, or experiences no more than F faults if it is a good node. It further assumes that the maximum number of simultaneous faults in the network is limited to F. We indicate the number of faults associated with N_i by f_i, thus, $F = \sum f_i$, and $f_i \leq F$. The *3ROM* algorithm is independent of the fault model (node-fault or link-fault model), and as the name implies, achieves agreement in three rounds. Thus, it is independent of the number of faults (in terms of number of required rounds, not the amount of messages). The algorithm has a message complexity of $O(K^3)$, and is also scalable with respect to K. We also present the model checking results of a bounded model of the algorithm to verify its correctness.

This paper is organized as follows. We describe the fault models in Section 2. In Section 3 we provide a system overview. We present the *3ROM* algorithm and its formal proof in Section 4. In Section 5, we present the model checking efforts toward verification of correctness of a bounded model of the algorithm and the results of that effort. Finally, we present concluding remarks in Section 6.

[*] Since we use N_i to address a node, we use K here instead of n as is traditionally used in the literature.

2. Fault Models

In synchronous message-based distributed systems, a fault is typically defined as a message that was not transmitted when it was expected or a message that was transmitted but not received or received but not accepted, i.e., deemed invalid by a receiver. Thus, the fault is either associated with the source node of the message, the corresponding link between the source node and the destination node, or the destination node. Consequently, there are two viewpoints, *node-centric* and *link-centric*, and thus, there are two ways of modeling faults. In the node-centric model, which we refer to as the *node-fault model*, the faults are associated with the source node of the message and all fault manifestations between the source and the destination nodes for the messages from that source count as a single fault, which is specially the case when the faults are associated with a Byzantine faulty node [5, 6, 16, 17]. In this model all links are assumed to be good. Miner *et al.* [16] for instance, model the absence of a link as a link fault and even though both nodes and links failures are considered, they abstractly model link failures as failures of the source node.

In the link-centric model, which we refer to as the *link-fault model*, a fault is associated with the communication means connecting the source node to the destination node. In this model, all nodes are assumed to be good and an invalid message at the receiving node is counted as a single fault for the corresponding input link. Thus, from the global perspective, a Byzantine faulty node manifests as one or more link failures.

A link-fault model introduced by Schmid *et al.* [18] is called *perception-based hybrid fault model*, where faults are viewed from the perspective of the receiving nodes. Faults are associated with their input links, and all nodes are assumed to be good. They argued that since F faulty nodes can produce at most F faulty perceptions in any node, the link-fault model is compatible with the traditional node-fault model and so, all existing lower bound and impossibility results remain valid.

"In the perception-based model, the system-wide number of faults is replaced by the number of faults that are observable in the nodes' local 'perceptions' of the system. Formally, node r's *perception vector* $V_r = (V_r^1, V_r^2, \ldots, V_r^K,)$ is considered, where every *perception* $V_r^s \in V_r$ represents the message node r received from node s in some specific round; type and value(s) depend upon the particular algorithm considered" [18]. In that paper, Schmid *et al.* present a solution for synchronous deterministic consensus problem, where all nodes are expected to achieve agreement on a single value, in synchronous distributed systems with link faults.

3. System Overview

We considered "synchronous" message-passing distributed systems and modeled the system as a graph with a set of nodes (vertices) that communicate with each other by sending messages via a set of communication links (edges) that represent the nodes' interconnectivity. The underlying topology considered is a fully connected network of K nodes that exchange messages through a

set of communication links. We leave the generalization to other topologies to future works. The system consists of a set of good nodes and a set of faulty nodes. A good node is assumed to be an active participant and correctly execute the algorithms. A faulty node is either benign (detectably bad), symmetrically faulty, or arbitrarily (Byzantine) faulty. However, in this paper our primary focus is Byzantine faults.

The communication links are point-to-point and unidirectional, each connecting a source to a destination node. Thus, the fully connected graph consists of $K(K\text{-}1)$ unidirectional links. A good link is assumed to correctly deliver a message from its source node to its destination node within a bounded communication delay time. A faulty link does not deliver the message, delivers a corrupted message, or delivers a message outside the expected communication delay time.

The nodes communicate with each other by exchanging broadcast messages. Broadcast of a message by a node is realized by transmitting the message, at the same time, to all nodes that are directly connected to it. The communication network does not guarantee any relative order of arrival of a broadcast message at the receiving nodes, that is, a consistent delivery order of a set of messages does not necessarily reflect the temporal or causal order of the message transmissions [1]. A maximum of F faults are assumed to be present in the system, where $F \geq 0$. We assume $K \geq 3F+1$ and define the minimum number of good nodes in the system, G, by $G = K\text{-}F$ nodes. The minimum number of nodes needed to maintain synchrony is well established to be $3F+1$ [6, 14, 19].

3.1. Communication Delay

The communication delay between directly connected (adjacent) nodes is expressed in terms of the minimum event-response delay, D, and network imprecision, d. These parameter are measured at the network level. A message broadcast by a node at real time t is expected to arrive at its directly connected adjacent nodes, be processed, and subsequent messages to be generated by those nodes within the time interval $[t+D, t+D+d]$. Communication between independently-clocked nodes is inherently imprecise. The network imprecision, d, is the maximum time difference among all receivers of a message from a transmitting node with respect to real time. The imprecision is due to many factors including, but not limited to, the drift of the oscillators with respect to real time, jitter, discretization error, temperature effects and differences in the lengths of the physical communication media. These parameters are assumed to be bounded, $D > 0$, $d \geq 0$, and both have units of real-time clock ticks and their values known in the network. The communication delay, denoted γ, is expressed in terms of D and d, is defined as $\gamma = D+d$, and has units of real-time clock ticks. In other words, we assume synchronous communication and bound the communication delay between any two directly connected adjacent nodes by $[D, \gamma]$. However, for simplicity of notation, in the remainder of this paper we assume that the messages arrive *logically* at the same time at the destination nodes.

3.2. The *Sync* Message And Its Validity

In order to achieve and maintain desired synchrony, the nodes communicate by exchanging *Sync* messages, where *synchrony* is defined as a measure of the relative imprecision of the good

nodes. A *Sync* message from a given source is *valid* if it arrives at or after one *D* of an immediately preceding *Sync* message from that source, that is, the message validity in the value domain, i.e., valid *Sync* messages are rate-constrained. Assuming physical-layer error detection is dealt with separately, the reception of a *Sync* message is indicative of its validity in the value and time domains.

4. 3ROM

In a synchronous distributed system, the Oral Message algorithm of Lamport *et al.* [5] solves the Byzantine Agreement (BA) problem [14] by reliably transforming a message, in the presence of faults, to a symmetric message at the network level, whereby the good nodes reach an agreement and collectively either accept or reject the message. The OM algorithm is recursive and every iteration of the execution of the algorithm constitutes a step (round) of exchange of messages by the nodes. An instance of the OM algorithm starts with the source node broadcasting a message, the first round, followed by other nodes (except the source of the message) recursively rebroadcasting (relaying) the messages they receive to others in subsequent rounds. For a fully connected graph of *K* nodes, the OM algorithm requires *F+1* communication rounds. At the end of the *F+1* rounds, the nodes vote and reach agreement.

In this section we present a three-round algorithm, similar to the OM algorithm, that achieves agreement among the good nodes, independent of *F*, provided that $K \geq 3F+1$ nodes, where *K* is the number of nodes, $F = \sum f_i$, for $i = 1..K$, where f_i is the number of faults associated with N_i. For the node-fault model, the faulty nodes act arbitrary provided their behavior is bounded by the assumption, i.e., at any round a faulty node broadcasts a valid message to at least *K-F* other nodes. Similarly, for the link-fault model, at any round no more than *F* link faults are perceived at a receiving good node. We will also show that this three-round algorithm applies equally to the node-fault and link-fault models. To simplify presentation of this algorithm, we assume broadcast messages arrive logically at the same time at their destination good nodes. We later remove this simplifying assumption and show that agreement is reached at the good nodes within a time bound. We use two types of messages: *Sync* and *Relay*, and extend the message validity argument of Section 3.2 to both messages.

The *3ROM* Algorithm

The algorithm depends on two positive parameters α and β, which are used in determining the final acceptance or rejection of a *Sync* message from a source node. Hence, the algorithm described is actually *3ROM(α, β)*. The algorithm consists of three rounds and a vote.

Round 1 – The source node broadcasts a *Sync* message to all other nodes, effectively saying "*I'm here*." A node does not physically send a message to itself even though it uses its own message. The nodes that receive the message record that the message was received.

Round 2 – All good nodes that received a *Sync* message in *Round 1* broadcast a *Relay* message to all other nodes, essentially saying "*I've got a message*." The good nodes that do not receive the *Sync* message do nothing. Note that since the source node uses own message, if it is a good node, it too participates in this round.

Round 3 – All good nodes that received at least α messages (*Sync* or *Relay*) in *Round 2* broadcast a vector of K messages containing what they have received from all other nodes in *Rounds 1* and *2*; effectively saying "*This is what I've received from others*." Note that, if the algorithm assumptions hold, all good nodes participate in this round.

Vote – At the end of *Round 3*, each node locally constructs a KxK network-level matrix M of received messages, where entries $M(i, j) = \{s, r, 0\}$, $i, j = 1..K$, where '*s*' indicates having received a *Sync* message, '*r*' indicates having received a *Relay* message and '*0*' for not receiving any messages, i.e., a fault. A column c_j of the matrix M corresponds to the messages perceived as transmitted by N_j and a row r_i of the matrix M corresponds to the messages received by N_i. Let, for $i, j = 1..K$, $X_i = 1$ if $\sum c_j > \alpha$ and $X_i = 0$, otherwise, where $\sum c_j$ is the sum of the non-zero entries in column c_j, i.e., treating '*s*' and '*r*' entries equally. Finally, the node votes "accept" if $\sum X_i > \beta$, i.e., the node accepts the message from the source node if more than β columns of M have more than α non-zero entries each.

4.1. Proof Of The *3ROM* Algorithm For Link-Fault Model

The proof of correctness of the OM algorithm, and consequently, the proof of correctness of the *3ROM* algorithm, is centered on the following two properties.

AP (Agreement Property): If receiver nodes p and q are nonfaulty, then they agree on the value ascribed to the transmitter.

VP (Validity Property): If the transmitter is nonfaulty, then every nonfaulty receiver computes the correct value.

In addition to the network-level matrix M that each node constructs at the end of *Round 3*, the proofs to follow rely on another related matrix M_{global} that can be constructed at the end of *Round 2*. Following *Round 2*, each node has a vector of received messages that describes the messages it received from each node. The M_{global}, with entries related to the matrix M, i.e., $M_{global}(i, j) = \{s, r, 0\}$, reflects a global view of the network at the end of *Round 2*. While it is inaccessible to any nodes, it is related to the matrices M built by each node.

Theorem 1. *For a fully connected graph with $K > 3F$ nodes and the link-fault model, i.e., $f_i \leq F$, 3ROM(K/3, 2K/3) guarantees agreement at the good nodes.*

Proof. The source node is a good node (all nodes are good in the link-fault model).

Round 1 – N_S, the source node, broadcasts a *Sync* message and it is received correctly (valid) by at least K-F nodes. Let H be this set of nodes.

Round 2 – Each node in H broadcasts a *Relay* message to all other nodes and at least K-F other nodes will receive the *Relay* message correctly. Hence, for each node in H, its corresponding column in M_{global} has at least K-F non-zero entries, i.e., '*s*' and '*r*'.

Round 3 – Since at the end of *Round 2*, all good nodes receive messages from the nodes in H, at least $K\text{-}2F = F+1$ messages, all good nodes participate in this round. Each good node broadcasts its vector of received messages to all other nodes and at least K-F nodes will receive it correctly. At the end of this round, a node constructs its matrix M, which is similar to, but likely different from, M_{global}. Since the rows of M are the messages in *Round 3*, at most F rows can be different from M_{global}. Thus, the sum of non-zero entries

in any column in of M differs from the sum of the same column in M_{global} by at most F entries. Therefore, the columns in M corresponding to nodes in H have at least K-F-F non-zero entries. Since $K > 3F$, (equivalently, $K/3 > F$), there are at least K-$F > 2K/3$ columns in M (the nodes in H), with at least K-$2F > K/3$ non-zero entries. Thus, each node votes "accept" for *3ROM(K/3, 2K/3)*. □

Table 1 is an example of the network-level matrix at the end of *Round 3* for $F = 2$ and $K = 7$. The grayed cells along the diagonal in this matrix are the messages a node sends to itself, thus, cannot be faulty. An "~~*s*~~ *r*" entry indicates that a *Sync* message was received in *Round 2* and was replaced by a *Relay* message from the same node in *Round 3*.

Table 1. An example of matrix of received messages at the end of *Round 3*.

N_i	1	2	3	4	5	6	7
1	~~*s*~~ *r*	0	0	r	r	0	0
2	~~*s*~~ *r*	r	0	0	r	0	0
3	~~*s*~~ *r*	r	r	0	0	0	0
4	~~*s*~~ *r*	0	r	r	0	0	0
5	*s*	r	r	r	r	0	0
6	~~*0*~~ *r*	r	r	r	r	0	0
7	0	r	r	r	r	0	0
X_i	1	1	1	1	1	0	0

Table 2 shows the matrices at N_1 and N_2, as examples, at the end of *Round 3*. The grayed rows indicate the effects of link faults at the two nodes in *Round 3*.

Table 2. An example of matrix of received messages at N_1 and N_2 at the end of *Round 3*.

N_1	1	2	3	4	5	6	7	
1	~~*s*~~ *r*	0	0	r	r	0	0	
2	~~*s*~~ *r*	r	0	0	r	0	0	
3	~~*s*~~ *r*	r	r	0	0	0	0	
4	~~*s*~~ *r*	0	r	r	0	0	0	
5	*s*	r	r	r	r	0	0	
6	0	0	0	0	0	0	0	
7	0	0	0	0	0	0	0	
X_i	1	1	1	1	1	0	0	V=1

N_2	1	2	3	4	5	6	7	
1	0	0	0	0	0	0	0	
2	~~*s*~~ *r*	r	0	0	r	0	0	
3	~~*s*~~ *r*	r	r	0	0	0	0	
4	~~*s*~~ *r*	0	r	r	0	0	0	
5	*s*	r	r	r	r	0	0	
6	~~*0*~~ *r*	r	r	r	r	0	0	
7	0	0	0	0	0	0	0	
X_i	1	1	1	1	1	0	0	V=1

4.2. Proof Of The *3ROM* Algorithm For Node-Fault Model

In the classic node-fault model, a node's message may be perceived as faulty by many other nodes. However, we assume there are up to F simultaneous Byzantine faulty nodes present in the network and they behave arbitrarily but are limited to inducing no more than F faults at each round, i.e., $f_i \leq F$.

Theorem 2. *For a fully connected graph with $K \geq 3F+1$ nodes and the node-fault model, i.e., $f_i \leq F$, 3ROM(K/3, 2K/3) guarantees agreement at the good nodes.*
Proof. Note that a bounded-Byzantine faulty node can be modeled by a link-fault model, where the faulty links are exactly those where the received messages are invalid. Thus, whether the source node is good or Byzantine faulty, since the maximum number of bounded-Byzantine faulty nodes, F, is less than a third of K and since the link-fault model allows each node up to F faults per round, the proof of Theorem 1 applies in this case. □

Table 3. An example of matrix of received messages at the good nodes N_i, $i = 1..5$, at the end of *Round 3*. N_6 and N_7 are the Byzantine faulty nodes and N_6 is the source node.

N_i	1	2	3	4	5	6	7	
1	r	r	r	0	0	~~s~~ r	0	
2	r	r	r	0	0	~~s~~ r	r	
3	r	r	r	0	0	s	r	
4	r	r	r	0	0	~~0~~ r	0	
5	r	r	r	0	0	0	r	
6	-	-	-	-	-	~~s~~ r	r	
7	-	-	-	-	-	~~s~~ r	r	
X_i	1	1	1	0	0	1	1	V = 1

Table 3 is an example of the matrix after *Round 3* at N_i. We would like to point out that, unlike the Table 2 for the link-fault model, this matrix is the same at all good nodes except for the rows and columns corresponding to the faulty nodes; c_6, c_7, and r_6, r_7, respectively. A '-' entry in the matrix means *don't care*.

Theorem 3 (Agreement). *For any F and K, for a fully connected graph with $K \geq 3F+1$ nodes and $f_i \leq F$, the 3ROM algorithm satisfies AP at the good nodes.*
Proof. It follows from Theorems 1 and 2 that, if the assumptions are met, the *3ROM* algorithm always guarantees agreement at the good nodes. □

Theorem 4 (Validity). *For any F and K, for a fully connected graph with $K \geq 3F+1$ nodes and $f_i \leq F$, the 3ROM algorithm satisfies AP and VP.*
Proof. It follows from Theorems 1 and 2 that, if the assumptions are met, the *3ROM* algorithm satisfies the agreement and validity properties at the good nodes. □

Corollary 5. *The number of rounds required by 3ROM algorithm is independent of F.*
Proof. It follows from Theorem 3 that the *3ROM* algorithm always guarantees agreement at the good nodes in three rounds and regardless of a particular value of F. □

Theorem 6. *For a fully connected graph with $K \geq 3F+1$ nodes and $f_i \leq F$, the node-fault model subsumes the link-fault model.*
Proof. Given the assumptions that $f_i \leq F$, i.e., a node either is faulty and induces up to F faults per round or it is good and experiences no more than F faults per round. Given the link-fault model, the *3ROM* algorithm converts any message from a node N_i to a symmetric message in three rounds. With the link-fault model, the nodes are considered to be good even though the

faults are manifested on their links. Thus, given up to F faults per node (i.e., the maximum number of outgoing faulty links per node), a maximum of KF faults per round are tolerated. With the node-fault model, a maximum of F faulty nodes are assumed to be present with up to F faults per outgoing links of a faulty node, thus, a total of F^2 faults per round are tolerated. Since for $F > 0$, $F^2 < KF$, the node-fault model subsumes the link-fault model. □

Thus far, we assumed that the Byzantine faulty nodes behave arbitrarily but are limited to inducing no more than F faults at any round. We now weaken the assumption of $f_i \leq F$ so that a faulty node behaves fully arbitrary in *Round 2* and/or *Round 3*. The *3ROM* algorithm still achieves agreement, but, the voting criteria needs to be adjusted to accommodate this weakened assumption, i.e., *3ROM(K/3, K/3+1)*. One manifestation of a faulty behavior is for the node to not broadcast anything during *Round 2* and/or *Round 3*. Note that when a node fails crash-silent, $f_i = K$ and it can readily be detected from the network-level matrix at the end of *Round 3* since its corresponding column will have at least K-F zeroes. This diagnosis information can potentially be used at the network level. We now show that the *3ROM* algorithm still achieves agreement when the source is a Byzantine faulty node.

Theorem 7. *For a fully connected graph with $K > 3F$ nodes and the node-fault model, i.e., $f_i \leq F$, when the source node is a Byzantine faulty node, 3ROM(K/3, K/3+1) guarantees agreement at the good nodes.*

Proof. The source node is a Byzantine faulty node.

Round 1 – The source node broadcasts a *Sync* message to at least K-F nodes where at least K-$F^{\dagger}$-$F^{\ddagger}$ are good nodes and receive the message correctly (valid). Let H be this set of good nodes.

Round 2 – Each node in H broadcasts a *Relay* message to all other nodes with at least K-F other nodes receiving the *Relay* message correctly. Hence, for each node in H, its corresponding column in M_{global} has at least K-F non-zero entries.

Round 3 – Since at the end of *Round 2*, all good nodes receive messages from the nodes in H, at least $K\text{-}2F = F+1$ messages, all good nodes participate in this round. Each good node broadcasts its vector of received messages to all other nodes and at least K-F nodes will receive it correctly. At the end of this round, a node constructs its matrix M, which is similar to, but likely different from, M_{global}. The matrix M at a good node N_i is different from M_{global} in at most F rows and F columns corresponding to the Byzantine faulty nodes. Also, the matrix M at a good node N_i is identical to the matrix M at other good nodes N_j, $j = 1..G$ and $j \neq i$, except in the same rows and columns corresponding to the Byzantine faulty nodes. Therefore, the columns in M corresponding to nodes in H have at least K-F non-zero entries. Furthermore, the column corresponding to the faulty source node, has at least $K\text{-}2F$ non-zero (the nodes in H) entries in M. Since $K > 3F$, (equivalently, $K/3 > F$), there are at least $K\text{-}2F+1 > K/3+1$ columns in M (the nodes in H plus the faulty source), with at least $K\text{-}2F > K/3$ non-zero entries. Therefore, each node votes "accept" for *3ROM(K/3, K/3+1)*. □

Note that when the source is a good node (node-fault model), since the set H consists of at least $2K/3$ good nodes, this weaker assumption holds and Theorems 2 and 7 apply. Also, although

---

† Up to F good nodes do not receive the message.

‡ Up to F simultaneous faulty nodes.

this weaker assumption does not apply to the link-fault model (all nodes are good) and Theorems 1 still holds, nevertheless, since with this weaker assumption, $\beta = K/3+1$ and $K/3+1 < 2K/3$, Theorem 7 applies to both models.

4.3. Message Observation Window, Agreement Within A Time Bound

Earlier in this paper we stated that to simplify the explanation of the problem and our proposed solution, a transmitted message from a single source arrived at the receiving nodes *logically* at the same time. In this section we visit this assumption and justify this rationality. In an implementation, as we have explained in Section 3.1, a given message from a single source arrives at the receiving nodes within d units of each other. Figure 1 is a depiction of a message through three rounds of the *3ROM* algorithm. N_S is the source node, N_i and N_j represent the nodes that receive the message at the two extremes of the communication latencies, i.e., D and $D+d = \gamma$, respectively, $i \neq j \neq S$. Thus, unless proper measures are taken, consequent relaying of messages at subsequent rounds widens message arrivals at the nodes for every round by an additional d. In this figure, '↑' indicates broadcasting a message, '↓' indicates receiving a message, and the labels on these arrows, s, i, and j, correspond to N_S, N_i, and N_j, respectively, as the initiators of the messages.

For the following lemmas, N_S is the source node initiating *Round 1*, N_i is the node that receives the message at the earliest time, i.e., D, and N_j is the node that receives the message at the latest time, i.e., $D+d = \gamma$.

Lemma 8. *Message observation window for Round 2 is [γ-2d, γ+d].*

Proof. *Round 2* begins by N_i and N_j relaying the messages they received from N_S. For this round, the nodes relay the messages as soon as they receive it, i.e., within at most d of each other. Thus, at the end of *Round 2*, the messages arrive at the nodes within $2d$ of each other. Since a node does not physically send a message to itself, but uses its own message, to account for the worst case message delivery time, its message is assumed to arrive at itself at γ. At the end of *Round 2*, the earliest a message can arrive is at N_j and from the first node that started its *Round 2*, i.e., N_i. As shown on the timeline of activities, this message arrives at the longest delay minus the accumulated drift for two rounds, i.e., γ-$2d$. Similarly, the latest a message can arrive is at N_i and from the node that started its *Round 2* last, i.e., N_j, and at the longest delay plus the initial drift between the nodes from the previous round, i.e., $\gamma+d$. Thus, the window of observation for message arrival for *Round 2* is $[\gamma\text{-}2d, \gamma+d]$. □

Lemma 9. *All good nodes participating in Round 2 finish Round 2 and start Round 3 within d of each other.*

Proof. From Figure 1, from the start of *Round 1* to the end of *Round 2*, the earliest message (EM) and latest message (LM) arrival time at N_S at $EM_S = \gamma+D-d$ and $LM_S = \gamma+\gamma$, respectively. Similarly, for N_i, $EM_i = D+d+D$, $LM_i = D+d+\gamma$, and for N_j, $EM_j = \gamma+D-d$, $LM_j = \gamma+\gamma$. Simple algebraic manipulation gives $\Delta_{EM} = d$ and $\Delta_{LM} = 0$ for any two nodes, i.e., the nodes finish *Round 2* and start *Round 3* within d of each other. □

Lemma 10. *Message observation window for Round 3 is [γ-2d, γ+d].*

Proof. It follows from Lemma 9 that the nodes start *Round 3* within *d* of each other. In a similar argument as Lemma 8, the observation window for message arrival for *Round 3* is [γ-2*d*, γ+*d*]. □

It follows from Lemmas 8 through 10 that at the end of each round the nodes are within *d* of each other, thus, justifying rationality of our assumption of logical timing of arrival of messages.

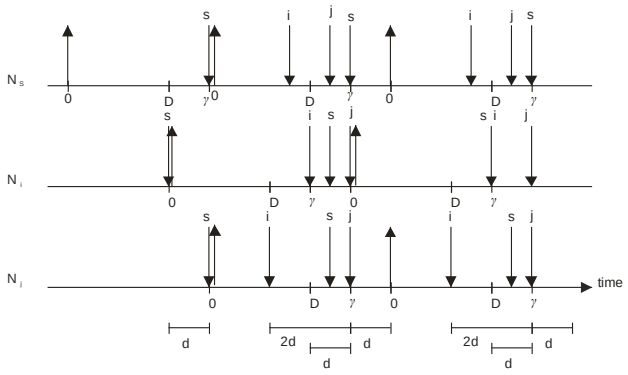


Fig. 1. Message observation window = [-2*d*, *d*] from γ.

4.4. Complexity Of The *3ROM* Algorithm

In *3ROM* algorithm, since a node does not send a message to itself, the number of transmitted messages per node and for each round is (*K*-1). For the worst case analysis, all nodes participate in *Rounds 2* and *3*. Thus, the total number of messages transmitted per round is (*K*-1), *K*(*K*-1), and K^2(*K*-1), for *Rounds 1, 2* and *3*, respectively. Although in *Round 3* a node broadcasts a vector of messages, for complexity analysis purposes, we count each individual message separately. Therefore, the total number of exchanged messages is (*K*-1)+*K*(*K*-1)+K^2(*K*-1) and the message complexity for the *3ROM* algorithm is $O(K^3)$. However, if a message is indeed physically broadcast to all, e.g., when the communication means is wireless, then the number of broadcast messages per node for each round is 1. Thus, the total number of messages broadcast per round is 1, (*K*-1), and *K*(*K*-1), for *Rounds 1, 2* and *3*, respectively, the total number of exchanged messages is 1+(*K*-1)+*K*(*K*-1), and the message complexity for the *3ROM* algorithm is $O(K^2)$. However, the message complexity for the OM algorithm for the above two scenarios is $O(K^F)$ and $O(K^{F-1})$, respectively. We would like to emphasize that the number of rounds of exchanged messages for the *3ROM* algorithm is independent of *F*.

5. Model Checking

In this section we present a mechanical verification of the *3ROM* algorithm using the model checking approach for its ease, feasibility, and quick examination of the problem space, to verify correctness of our formal proof of the algorithm. The Symbolic Model Verifier (SMV) [20] was used in the modeling of this algorithm. SMV's language description and modeling capability provide relatively easy translation from the pseudo-code. SMV semantics are synchronous composition, where all assignments are executed in parallel and synchronously. Thus, a single step of the resulting model corresponds to a step in each of the components.

A number of cases for each fault model were model checked. In particular, for the node-fault model, scenarios with $F = 0..3$ and $K = 4..10$, respectively, were model checked with the weaker assumptions, i.e., $\sum c_j \geq F+1$ and $\sum X_i \geq F+2$. Model checking of the link-fault model requires a specific number of link faults being considered. Two cases with $F = 2$, $K = 7$, and $F = 3$, $K = 10$, were model checked. Model checking of larger graphs and with more number of node and link faults can readily be accommodated. The SMV models are listed in Appendices A and B.

5.1. Model Checked Propositions

Computational tree logic (CTL), a temporal logic, is used to express properties of a system. In CTL formulas are composed of **path quantifiers**, *E* and *A*, and **temporal operators**, *X*, *F*, *G*, and *U* [21]. *A* means "All" and has to hold on all paths starting from the current state. *F* means "Finally" and eventually has to hold (somewhere on the subsequent path). In this section the claims of agreement at the good nodes and at the end of the third round is examined. The node-fault and link-fault models are model checked separately for $F = 1$, 2, and 3, while the same CTL proposition is used to verify agreement has been reached at all good nodes for both models.

For model checking of each scenario, a particular node is instructed to be the source and scheduled to initiate broadcast of a *Sync* message at a particular time. Since the 3*ROM* is deterministic, the final vote time, *VotingResultTime*, is set to the end of the 3rd round after the broadcast of the initial *Sync* message. Validation of the CTL proposition requires examination of an underlying proposition. In particular, the variable *VoteTime* is used in these properties and is defined here.

$$VoteTime = (GlobalClock \geq VotingResultTime) ;$$

The *GlobalClock* is a measure of elapsed time from the beginning of the operation with respect to the real time, i.e., external view. The *VoteTime* is indicative of the *GlobalClock* reaching its target value of *VotingResultTime* and the *GlobalAgreement* is defined as the conjunction of voting results at all good nodes.

Proposition *SystemLiveness*: *AF (VoteTime)*
This property addresses the liveness property of the system and whether time advances and the amount of time elapsed, *VoteTime*, has advanced beyond the broadcast of the message and the three rounds to reach agreement on that message.

Proposition *GlobalAgreement*: *AF (VoteTime & GlobalAgreement)*
This proposition encompasses the criteria for the agreement property as well as the claim of determinism. The proposition specifies whether or not the system will reach agreement in the three rounds after the message was initially broadcast. This property is expected to hold.

In satisfying the Agreement Property, and its related *GlobalAgreement* proposition, both node-fault and link-fault models had to be included, i.e., faulty transmitters and faulty links were model checked. Since this proposition includes the Validity Property (i.e., nonfaulty transmitter), a separate proposition was not needed.

The model checking results of the bounded model of the algorithm have verified the correctness of the algorithm for fully connected networks with $K \geq 3F + 1$ nodes, for both node-fault and link-fault models, and for the following scenarios; $F = 0, 1, 2, 3$ simultaneous faults and $K = 4, 4, 7$, and 10, respectively. In addition, the results have confirmed the claims of determinism and independence of the algorithm from F.

6. Conclusions

Distributed systems have become an integral part of safety-critical computing applications, necessitating system designs that incorporate complex fault-tolerant resource management functions to provide globally coordinated operations with ultra-reliability. As a result, robust clock synchronization has become a required fundamental component of fault-tolerant safety-critical distributed systems. The main issue in solving the clock synchronization problem for the general case is a lack of a symmetric view in the system at the participating good nodes. We first enumerated several ways of achieving message symmetry across the system, and then presented an alternative, referred to as the *3ROM* algorithm, that guarantees agreement in a system in three rounds. The *3ROM* assumes each node N_i, $i = 1..K$, either induces up to F faults if it is a faulty node, or experiences no more than F faults if it is a good node, and in addition, the maximum number of simultaneous faults in the network is limited to F. The algorithm is based on the Oral Message algorithm of Lamport *et al.*, is scalable with respect to the number of nodes in the system, and applies equally to the traditional node-fault model as well as the link-fault models. The *3ROM* is independent of the fault model (node-fault or link-fault model), and is independent of the number of faults (in terms of number of required rounds, not the amount of messages), and has a message complexity of $O(K^3)$. We also presented a mechanical verification of the algorithm for up to three simultaneous Byzantine faults. The model-checking effort was focused on verifying the correctness of a bounded model of the algorithm as well as confirming claims of determinism. The underlying topology in this paper was a fully connected graph. We leave the generalization of our solution to other topologies, including an arbitrary graph that meets the minimum requirements of number of nodes and connectivity, to future works.

References

1. Kopetz, H: *Real-Time Systems, Design Principles for Distributed Embedded Applications*, Kluwar Academic Publishers, ISBN 0-7923-9894-7, 1997.
2. Torres-Pomales, W; Malekpour, M.R.; Miner, P.S.: *ROBUS-2: A fault-tolerant broadcast communication system*, NASA/TM-2005-213540, pp. 201, March 2005.
3. Torres-Pomales, W.; Malekpour, M.R.; Miner, P.S.: *Design of the Protocol Processor for the ROBUS-2 Communication System*, NASA/TM-2005-213934, pp. 252, November 2005.
4. Butler, R.: *A primer on architectural level fault tolerance*, NASA/TM-2008-215108, February 2008.
5. Lamport, L.; Shostak, R.; Pease, M.: *The Byzantine General Problem*, ACM Transactions on Programming Languages and Systems, 4(3), pp. 382-401, July 1982.
6. Driscoll, K.; Hall, B.; Sivencrona, H.; Zumsteg, P.:, *Byzantine Fault Tolerance, from Theory to Reality*, LNCS, 22nd International Conference on Computer Safety, Reliability and Security, pp. 235-248, September 2003.

7. Lamport, L.; Melliar-Smith, P.M.: *Synchronizing clocks in the presence of faults*, J. ACM, vol. 32, no. 1, pp. 52-78, 1985.
8. Schmid, U.; Weiss, B.; Keidar, I.: *Impossibility Results And Lower Bounds For Consensus Under Link Failures*, in SIAM Journal on Computing 38(5), p. 1912-1951, January 2009.
9. Hoyme, K.; Driscoll, K.: *SAFEbusTM*, 11th AIAA/IEEE Digital Avionics Systems Conference, pages 68–73, October 1992.
10. Aeronautical Radio, Inc., Annapolis, MD. *ARINC Specification 659: Backplane Data Bus*, December 1993. Prepared by the Airlines Electronic Engineering Committee.
11. Kopetz, H.; Grünsteidl, G.: *TTP – a time-triggered protocol for fault-tolerant real-time systems*, Fault Tolerant Computing Symposium 23, pages 524–533, June 1993.
12. Bauer, G.; Kopetz, H.; and Steiner, W.: *The central guardian approach to enforce fault isolation in a time-triggered system*, Proc. of 6th International Symposium on Autonomous Decentralized Systems (ISADS 2003), pp. 37–44, April 2003.
13. Malekpour, M.R.: *A Self-Stabilizing Hybrid-Fault Tolerant Synchronization Protocol*, NASA/TM-2014-218285, July 2014.
14. Pease, M.; Shostak, R.; and Lamport, l.: *Reaching agreement in the presence of faults*, Journal of the ACM, 27(2): 228-234, April 1980.
15. Lincoln, P.; Rushby J.: *A Formally Verified Algorithm for Interactive Consistency Under a Hybrid Fault Model*, Proceedings of the Fault-Tolerant Computing Symposium, FTCS 23, June 1993.
16. Miner, P.S.; Geser, A.; Pike, L.; Maddalon, J: *A Unified Fault-tolerance Protocol*, In Yassine Lakhnech and Sergio Yovine, editors, Formal Techniques, Modeling and Analysis of Timed and Fault-Tolerant Systems, volume 3253, pp. 167-182, Springer, 2004.
17. Malekpour, M.R.: *A Self-Stabilizing Byzantine-Fault-Tolerant Clock Synchronization Protocol*, NASA/TM-2009-215758, June 2009.
18. Schmid, U.; Weiss, B.; Rushby, J.: *Formally Verified Byzantine Agreement in Presence of Link Faults*, in 22nd International Conference on Distributed Computing Systems (ICDCS'02), pp. 608–616, July 2002.
19. Dolev, D.; Halpern, J.Y.; Strong, R.: *On the Possibility and Impossibility of Achieving Clock Synchronization*, proceedings of the 16th Annual ACM STOC, pp. 504-511, 1984.
20. http://www-2.cs.cmu.edu/~modelcheck/smv.html
21. Clarke, E.M.; Emerson, E.A.: *Design and synthesis of synchronization skeletons using branching time temporal logic*, In Logic of Programs: Workshop, LNCS 131, Springer, May 1981.

Appendix A

```
-------------------------------------------------------------------------------
-- File Name:       3ROM_NodeFault_K10.smv
--                  This file is for a system of K good nodes, where F > 1.
--                  Note that the graph is fully connected.
--
--
-- Environment :    SMV
-- Organization:    NASA Langley Research Center
-- Project:         Self-Stablization
-- Authors:         Malekpour, Mahyar
--
--                  NASA Langley Research Center
--                  Hampton, VA 23681-2199
-- Creation Date:   9/2/2014
-------------------------------------------------------------------------------
--
-- This SMV description is property of the National Aeronautics and Space
-- Administration.  Unauthorized use or duplication of this VHDL description is
-- strictly prohibited.  Authorized users are subject to the following
-- restrictions:
--
--    . Neither the author, their corporation, nor NASA is responsible for any
--      consequence of the use of this SMV description.
--
--    . The origin of this SMV description must not be misrepresented either
--      by explicit claim or by omission.
--
--    . Altered versions of this SMV description must be plainly marked as
--      such.
--
--
--    . This notice may not be removed or altered.
--
-------------------------------------------------------------------------------
--
--    Modified on:  6/ 3/2014
--    by:           Mahyar Malekpour
--
-------------------------------------------------------------------------------
--
-- Global Constants:
--

#define Node_Id1                1
#define Node_Id2                2
#define Node_Id3                3
#define Node_Id4                4
#define Node_Id5                5
#define Node_Id6                6
#define Node_Id7                7
#define Node_Id8                8
#define Node_Id9                9
#define Node_Id10               10

#define SourceNodeId            (Node_Id1)

-- Drift in units of Gamma
#define DriftP                  (5)

-- Network Size
#define K                       10
#define G                       7
#define F                       3
#define FPlusOne                (F + 1)
#define FPlusTwo                (F + 2)
#define TwoFPlusOne             (2 * F + 1)
```

```
-- Topology = fully connected graph
#define P                       (20 + DriftP)

#define GlobalClockMax          (P + 7)

-- P of SourceNodeId + 3 rounds
#define TimeToVote              (P + 3)

-- Drift at the local level
#define P_N1                    (P - 0)
#define P_N2                    (P - 1)
#define P_N3                    (P - 2)
#define P_N4                    (P - 3)
#define P_N5                    (P - 4)
#define P_N6                    (P + 1)
#define P_N7                    (P + 2)
#define P_N8                    (P + 3)
#define P_N9                    (P + 4)
#define P_N10                   (P + 5)

-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
--
--
--
MODULE main

VAR
  Global_Clock : 0 .. GlobalClockMax ;

  FaultyNode_1 : FaultyNode (Node_Id8, P_N8, Global_Clock, SourceNodeId,
              Node_1.MessageOut,
              Node_2.MessageOut,
              Node_3.MessageOut,
              Node_4.MessageOut,
              Node_5.MessageOut,
              Node_6.MessageOut,
              Node_7.MessageOut,
--              Node_8.MessageOut,
--              Node_9.MessageOut,
--              Node_10.MessageOut) ;
              FaultyNode_1.MessageOut_8,
              FaultyNode_2.MessageOut_9,
              FaultyNode_3.MessageOut_10) ;
  FaultyNode_2 : FaultyNode (Node_Id9, P_N9, Global_Clock, SourceNodeId,
              Node_1.MessageOut,
              Node_2.MessageOut,
              Node_3.MessageOut,
              Node_4.MessageOut,
              Node_5.MessageOut,
              Node_6.MessageOut,
              Node_7.MessageOut,
--              Node_8.MessageOut,
--              Node_9.MessageOut,
--              Node_10.MessageOut) ;
              FaultyNode_1.MessageOut_8,
              FaultyNode_2.MessageOut_9,
              FaultyNode_3.MessageOut_10) ;
  FaultyNode_3 : FaultyNode (Node_Id10, P_N10, Global_Clock, SourceNodeId,
              Node_1.MessageOut,
              Node_2.MessageOut,
              Node_3.MessageOut,
              Node_4.MessageOut,
              Node_5.MessageOut,
              Node_6.MessageOut,
              Node_7.MessageOut,
--              Node_8.MessageOut,
--              Node_9.MessageOut,
--              Node_10.MessageOut) ;
              FaultyNode_1.MessageOut_8,
```

```
                FaultyNode_2.MessageOut_9,
                FaultyNode_3.MessageOut_10) ;

    ------------------------------------------------------------------------
    Node_1 : Node (Node_Id1, P_N1, Global_Clock, SourceNodeId,
                Node_1.MessageOut,
                Node_2.MessageOut,
                Node_3.MessageOut,
                Node_4.MessageOut,
                Node_5.MessageOut,
                Node_6.MessageOut,
                Node_7.MessageOut,
--                Node_8.MessageOut,
--                Node_9.MessageOut,
--                Node_10.MessageOut) ;
                FaultyNode_1.MessageOut_1,
                FaultyNode_2.MessageOut_1,
                FaultyNode_3.MessageOut_1,
                Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
                Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
--                Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;
                FaultyNode_1.MsgVector, FaultyNode_2.MsgVector, FaultyNode_3.MsgVector) ;

    Node_2 : Node (Node_Id2, P_N2, Global_Clock, SourceNodeId,
                Node_1.MessageOut,
                Node_2.MessageOut,
                Node_3.MessageOut,
                Node_4.MessageOut,
                Node_5.MessageOut,
                Node_6.MessageOut,
                Node_7.MessageOut,
--                Node_8.MessageOut,
--                Node_9.MessageOut,
--                Node_10.MessageOut) ;
                FaultyNode_1.MessageOut_2,
                FaultyNode_2.MessageOut_2,
                FaultyNode_3.MessageOut_2,
                Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
                Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
--                Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;
                FaultyNode_1.MsgVector, FaultyNode_2.MsgVector, FaultyNode_3.MsgVector) ;

    Node_3 : Node (Node_Id3, P_N3, Global_Clock, SourceNodeId,
                Node_1.MessageOut,
                Node_2.MessageOut,
                Node_3.MessageOut,
                Node_4.MessageOut,
                Node_5.MessageOut,
                Node_6.MessageOut,
                Node_7.MessageOut,
--                Node_8.MessageOut,
--                Node_9.MessageOut,
--                Node_10.MessageOut) ;
                FaultyNode_1.MessageOut_3,
                FaultyNode_2.MessageOut_3,
                FaultyNode_3.MessageOut_3,
                Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
                Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
--                Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;
                FaultyNode_1.MsgVector, FaultyNode_2.MsgVector, FaultyNode_3.MsgVector) ;

    Node_4 : Node (Node_Id4, P_N4, Global_Clock, SourceNodeId,
                Node_1.MessageOut,
                Node_2.MessageOut,
                Node_3.MessageOut,
                Node_4.MessageOut,
                Node_5.MessageOut,
                Node_6.MessageOut,
                Node_7.MessageOut,
--                Node_8.MessageOut,
```

```
--              Node_9.MessageOut,
--              Node_10.MessageOut) ;
              FaultyNode_1.MessageOut_4,
              FaultyNode_2.MessageOut_4,
              FaultyNode_3.MessageOut_4,
              Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
              Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
--              Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;
              FaultyNode_1.MsgVector, FaultyNode_2.MsgVector, FaultyNode_3.MsgVector) ;

  Node_5 : Node (Node_Id5, P_N5, Global_Clock, SourceNodeId,
              Node_1.MessageOut,
              Node_2.MessageOut,
              Node_3.MessageOut,
              Node_4.MessageOut,
              Node_5.MessageOut,
              Node_6.MessageOut,
              Node_7.MessageOut,
--              Node_8.MessageOut,
--              Node_9.MessageOut,
--              Node_10.MessageOut) ;
              FaultyNode_1.MessageOut_5,
              FaultyNode_2.MessageOut_5,
              FaultyNode_3.MessageOut_5,
              Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
              Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
--              Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;
              FaultyNode_1.MsgVector, FaultyNode_2.MsgVector, FaultyNode_3.MsgVector) ;

  Node_6 : Node (Node_Id6, P_N6, Global_Clock, SourceNodeId,
              Node_1.MessageOut,
              Node_2.MessageOut,
              Node_3.MessageOut,
              Node_4.MessageOut,
              Node_5.MessageOut,
              Node_6.MessageOut,
              Node_7.MessageOut,
--              Node_8.MessageOut,
--              Node_9.MessageOut,
--              Node_10.MessageOut) ;
              FaultyNode_1.MessageOut_6,
              FaultyNode_2.MessageOut_6,
              FaultyNode_3.MessageOut_6,
              Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
              Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
--              Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;
              FaultyNode_1.MsgVector, FaultyNode_2.MsgVector, FaultyNode_3.MsgVector) ;

  Node_7 : Node (Node_Id7, P_N7, Global_Clock, SourceNodeId,
              Node_1.MessageOut,
              Node_2.MessageOut,
              Node_3.MessageOut,
              Node_4.MessageOut,
              Node_5.MessageOut,
              Node_6.MessageOut,
              Node_7.MessageOut,
--              Node_8.MessageOut,
--              Node_9.MessageOut,
--              Node_10.MessageOut) ;
              FaultyNode_1.MessageOut_7,
              FaultyNode_2.MessageOut_7,
              FaultyNode_3.MessageOut_7,
              Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
              Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
--              Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;
              FaultyNode_1.MsgVector, FaultyNode_2.MsgVector, FaultyNode_3.MsgVector) ;

  Node_8 : Node (Node_Id8, P_N8, Global_Clock, SourceNodeId,
              Node_1.MessageOut,
              Node_2.MessageOut,
```

```
                Node_3.MessageOut,
                Node_4.MessageOut,
                Node_5.MessageOut,
                Node_6.MessageOut,
                Node_7.MessageOut,
                Node_8.MessageOut,
                Node_9.MessageOut,
                Node_10.MessageOut,
--              FaultyNode_1.MessageOut_7,
--              FaultyNode_2.MessageOut_7,
--              FaultyNode_3.MessageOut_7,
                Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
                Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
                Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;

  Node_9 : Node (Node_Id9, P_N9, Global_Clock, SourceNodeId,
                Node_1.MessageOut,
                Node_2.MessageOut,
                Node_3.MessageOut,
                Node_4.MessageOut,
                Node_5.MessageOut,
                Node_6.MessageOut,
                Node_7.MessageOut,
                Node_8.MessageOut,
                Node_9.MessageOut,
                Node_10.MessageOut,
--              FaultyNode_1.MessageOut_7,
--              FaultyNode_2.MessageOut_7,
--              FaultyNode_3.MessageOut_7,
                Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
                Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
                Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;

  Node_10 : Node (Node_Id10, P_N10, Global_Clock, SourceNodeId,
                Node_1.MessageOut,
                Node_2.MessageOut,
                Node_3.MessageOut,
                Node_4.MessageOut,
                Node_5.MessageOut,
                Node_6.MessageOut,
                Node_7.MessageOut,
                Node_8.MessageOut,
                Node_9.MessageOut,
                Node_10.MessageOut,
--              FaultyNode_1.MessageOut_7,
--              FaultyNode_2.MessageOut_7,
--              FaultyNode_3.MessageOut_7,
                Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
                Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
                Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;


  Agreement : boolean ;


DEFINE

  ------------------------------------------------------------------------
  VoteTime       := (Global_Clock >= TimeToVote) ;

  -- For all good nodes and good links.
  GlobalAgreement := (Global_Clock = TimeToVote) &
                  (Node_1.VoteResult & Node_2.VoteResult &
                   Node_3.VoteResult & Node_4.VoteResult &
                   Node_5.VoteResult & Node_6.VoteResult &
                   Node_7.VoteResult & Node_8.VoteResult &
                   Node_9.VoteResult & Node_10.VoteResult) ;

  -- For only the good nodes with the node-fault model, i.e, faulty nodes but no faulty links.
  GlobalAgreementNodeFault :=
```

```
                (Global_Clock = TimeToVote) &
                (Node_1.VoteResult & Node_2.VoteResult &
                 Node_3.VoteResult & Node_4.VoteResult &
                 Node_5.VoteResult & Node_6.VoteResult &
                 Node_7.VoteResult) ;

ASSIGN

  ---------------------------------------------------------------------------
  --
  init (Global_Clock) := 0 ;
  next (Global_Clock) :=
    case
      (Global_Clock < GlobalClockMax) : Global_Clock + 1 ;
      1 : Global_Clock ;
    esac ;

  ---------------------------------------------------------------------------
  ---------------------------------------------------------------------------


SPEC

  ---------------------------------------------------------------------------
  --
  -- Proposition #1:
  --
--   AF (VoteTime)

  ---------------------------------------------------------------------------
  --
  -- Proposition #2:
  --
  AF (VoteTime & GlobalAgreementNodeFault) -- true

  ---------------------------------------------------------------------------

-- end of main

---------------------------------------------------------------------------
---------------------------------------------------------------------------
--
-- Faulty node.
--
--
MODULE FaultyNode (Node_Id, MyP, Global_Clock, SourceNode,
                N1_Msg, N2_Msg, N3_Msg, N4_Msg, N5_Msg, N6_Msg, N7_Msg, N8_Msg, N9_Msg, N10_Msg)

VAR

  MessageOut_1  : {NONE, Sync, Relay} ;
  MessageOut_2  : {NONE, Sync, Relay} ;
  MessageOut_3  : {NONE, Sync, Relay} ;
  MessageOut_4  : {NONE, Sync, Relay} ;
  MessageOut_5  : {NONE, Sync, Relay} ;
  MessageOut_6  : {NONE, Sync, Relay} ;
  MessageOut_7  : {NONE, Sync, Relay} ;
  MessageOut_8  : {NONE, Sync, Relay} ;
  MessageOut_9  : {NONE, Sync, Relay} ;
  MessageOut_10 : {NONE, Sync, Relay} ;

  MsgVector  : array 1..10 of {0, 1} ;
  VoteResult : boolean ;

DEFINE

  ---------------------------------------------------------------------------
  Count_Sync :=
      ((N1_Msg = Sync) +
```

```
      (N2_Msg = Sync) +
      (N3_Msg = Sync) +
      (N4_Msg = Sync) +
      (N5_Msg = Sync) +
      (N6_Msg = Sync) +
      (N7_Msg = Sync) +
      (N8_Msg = Sync) +
      (N9_Msg = Sync) +
      (N10_Msg = Sync)) ;

  ------------------------------------------------------------------------------
  ------------------------------------------------------------------------------

ASSIGN

  init (MessageOut_1) := NONE ;
  init (MessageOut_2) := NONE ;
  init (MessageOut_3) := NONE ;
  init (MessageOut_4) := NONE ;
  init (MessageOut_5) := NONE ;
  init (MessageOut_6) := NONE ;
  init (MessageOut_7) := NONE ;
  init (MessageOut_8) := NONE ;
  init (MessageOut_9) := NONE ;
  init (MessageOut_10) := NONE ;

  ------------------------------------------------------------------------------
  next (MessageOut_1) :=
    case
      (Global_Clock = MyP) & (SourceNode = Node_Id) : Sync ;
      !(SourceNode = Node_Id) & (Count_Sync > 0) : {NONE, Relay} ;
      1 : NONE ;
    esac ;

  next (MessageOut_2) :=
    case
      (Global_Clock = MyP) & (SourceNode = Node_Id) : Sync ;
      !(SourceNode = Node_Id) & (Count_Sync > 0) : {NONE, Relay} ;
      1 : NONE ;
    esac ;

  next (MessageOut_3) :=
    case
      (Global_Clock = MyP) & (SourceNode = Node_Id) : Sync ;
      !(SourceNode = Node_Id) & (Count_Sync > 0) : {NONE, Relay} ;
      1 : NONE ;
    esac ;

  next (MessageOut_4) :=
    case
      (Global_Clock = MyP) & (SourceNode = Node_Id) : Sync ;
      !(SourceNode = Node_Id) & (Count_Sync > 0) : {NONE, Relay} ;
      1 : NONE ;
    esac ;

  next (MessageOut_5) :=
    case
      (Global_Clock = MyP) & (SourceNode = Node_Id) : NONE ;
      !(SourceNode = Node_Id) & (Count_Sync > 0) : {NONE, Relay} ;
      1 : NONE ;
    esac ;

  next (MessageOut_6) :=
    case
      (Global_Clock = MyP) & (SourceNode = Node_Id) : NONE ;
      !(SourceNode = Node_Id) & (Count_Sync > 0) : {NONE, Relay} ;
      1 : NONE ;
    esac ;

  next (MessageOut_7) :=
```

```
    case
      (Global_Clock = MyP) & (SourceNode = Node_Id) : NONE ;
      !(SourceNode = Node_Id) & (Count_Sync > 0) : {NONE, Relay} ;
      1 : NONE ;
    esac ;

  next (MessageOut_8) :=
    case
      (Global_Clock = MyP) & (SourceNode = Node_Id) : NONE ;
      !(SourceNode = Node_Id) & (Count_Sync > 0) : {NONE, Relay} ;
      1 : NONE ;
    esac ;

  next (MessageOut_9) :=
    case
      (Global_Clock = MyP) & (SourceNode = Node_Id) : NONE ;
      !(SourceNode = Node_Id) & (Count_Sync > 0) : {NONE, Relay} ;
      1 : NONE ;
    esac ;

  next (MessageOut_10) :=
    case
      (Global_Clock = MyP) & (SourceNode = Node_Id) : NONE ;
      !(SourceNode = Node_Id) & (Count_Sync > 0) : {NONE, Relay} ;
      1 : NONE ;
    esac ;

  ------------------------------------------------------------------------------

-- end of FaultyNode

------------------------------------------------------------------------------
------------------------------------------------------------------------------
--
-- Good node, for node-fault model.
--
--
MODULE Node (Node_Id, MyP, Global_Clock, SourceNode,
          N1_Msg, N2_Msg, N3_Msg, N4_Msg, N5_Msg, N6_Msg, N7_Msg, N8_Msg, N9_Msg, N10_Msg,
          N1_MsgVector, N2_MsgVector, N3_MsgVector, N4_MsgVector, N5_MsgVector,
          N6_MsgVector, N7_MsgVector, N8_MsgVector, N9_MsgVector, N10_MsgVector)

VAR

  RoundNum   : 0 .. 3 ;
  MessageOut : {NONE, Sync, Relay} ;
  -- Messages recieved (Rx) from the nodes are saved in this vector and will be
  -- broadcast/used during round 3.
  MsgVector  : array 1..10 of {0, 1} ;
  VoteResult : boolean ;

DEFINE

  ------------------------------------------------------------------------------
  Count_Sync :=
      ((N1_Msg = Sync) +
       (N2_Msg = Sync) +
       (N3_Msg = Sync) +
       (N4_Msg = Sync) +
       (N5_Msg = Sync) +
       (N6_Msg = Sync) +
       (N7_Msg = Sync) +
       (N8_Msg = Sync) +
       (N9_Msg = Sync) +
       (N10_Msg = Sync)) ;

  Count_Relay :=
      ((N1_Msg = Relay) +
       (N2_Msg = Relay) +
       (N3_Msg = Relay) +
```

```
      (N4_Msg = Relay) +
      (N5_Msg = Relay) +
      (N6_Msg = Relay) +
      (N7_Msg = Relay) +
      (N8_Msg = Relay) +
      (N9_Msg = Relay) +
      (N10_Msg = Relay)) ;

Count_MyVector :=
    (MsgVector [1] +
     MsgVector [2] +
     MsgVector [3] +
     MsgVector [4] +
     MsgVector [5] +
     MsgVector [6] +
     MsgVector [7] +
     MsgVector [8] +
     MsgVector [9] +
     MsgVector [10]) ;

Count_MatrixColumn_1 :=
    (N1_MsgVector [1] +
     N2_MsgVector [1] +
     N3_MsgVector [1] +
     N4_MsgVector [1] +
     N5_MsgVector [1] +
     N6_MsgVector [1] +
     N7_MsgVector [1] +
     N8_MsgVector [1] +
     N9_MsgVector [1] +
     N10_MsgVector [1]) ;

Count_MatrixColumn_2 :=
    (N1_MsgVector [2] +
     N2_MsgVector [2] +
     N3_MsgVector [2] +
     N4_MsgVector [2] +
     N5_MsgVector [2] +
     N6_MsgVector [2] +
     N7_MsgVector [2] +
     N8_MsgVector [2] +
     N9_MsgVector [2] +
     N10_MsgVector [2]) ;

Count_MatrixColumn_3 :=
    (N1_MsgVector [3] +
     N2_MsgVector [3] +
     N3_MsgVector [3] +
     N4_MsgVector [3] +
     N5_MsgVector [3] +
     N6_MsgVector [3] +
     N7_MsgVector [3] +
     N8_MsgVector [3] +
     N9_MsgVector [3] +
     N10_MsgVector [3]) ;

Count_MatrixColumn_4 :=
    (N1_MsgVector [4] +
     N2_MsgVector [4] +
     N3_MsgVector [4] +
     N4_MsgVector [4] +
     N5_MsgVector [4] +
     N6_MsgVector [4] +
     N7_MsgVector [4] +
     N8_MsgVector [4] +
     N9_MsgVector [4] +
     N10_MsgVector [4]) ;

Count_MatrixColumn_5 :=
    (N1_MsgVector [5] +
```

```
      N2_MsgVector [5] +
      N3_MsgVector [5] +
      N4_MsgVector [5] +
      N5_MsgVector [5] +
      N6_MsgVector [5] +
      N7_MsgVector [5] +
      N8_MsgVector [5] +
      N9_MsgVector [5] +
      N10_MsgVector [5]) ;

Count_MatrixColumn_6 :=
    (N1_MsgVector [6] +
      N2_MsgVector [6] +
      N3_MsgVector [6] +
      N4_MsgVector [6] +
      N5_MsgVector [6] +
      N6_MsgVector [6] +
      N7_MsgVector [6] +
      N8_MsgVector [6] +
      N9_MsgVector [6] +
      N10_MsgVector [6]) ;

Count_MatrixColumn_7 :=
    (N1_MsgVector [7] +
      N2_MsgVector [7] +
      N3_MsgVector [7] +
      N4_MsgVector [7] +
      N5_MsgVector [7] +
      N6_MsgVector [7] +
      N7_MsgVector [7] +
      N8_MsgVector [7] +
      N9_MsgVector [7] +
      N10_MsgVector [7]) ;

Count_MatrixColumn_8 :=
    (N1_MsgVector [8] +
      N2_MsgVector [8] +
      N3_MsgVector [8] +
      N4_MsgVector [8] +
      N5_MsgVector [8] +
      N6_MsgVector [8] +
      N7_MsgVector [8] +
      N8_MsgVector [8] +
      N9_MsgVector [8] +
      N10_MsgVector [8]) ;

Count_MatrixColumn_9 :=
    (N1_MsgVector [9] +
      N2_MsgVector [9] +
      N3_MsgVector [9] +
      N4_MsgVector [9] +
      N5_MsgVector [9] +
      N6_MsgVector [9] +
      N7_MsgVector [9] +
      N8_MsgVector [9] +
      N9_MsgVector [9] +
      N10_MsgVector [9]) ;

Count_MatrixColumn_10 :=
    (N1_MsgVector [10] +
      N2_MsgVector [10] +
      N3_MsgVector [10] +
      N4_MsgVector [10] +
      N5_MsgVector [10] +
      N6_MsgVector [10] +
      N7_MsgVector [10] +
      N8_MsgVector [10] +
      N9_MsgVector [10] +
      N10_MsgVector [10]) ;
```

```
  VoteResult :=
       (RoundNum = 3) &
       (((Count_MatrixColumn_1 >= FPlusOne) +
         (Count_MatrixColumn_2 >= FPlusOne) +
         (Count_MatrixColumn_3 >= FPlusOne) +
         (Count_MatrixColumn_4 >= FPlusOne) +
         (Count_MatrixColumn_5 >= FPlusOne) +
         (Count_MatrixColumn_6 >= FPlusOne) +
         (Count_MatrixColumn_7 >= FPlusOne) +
         (Count_MatrixColumn_8 >= FPlusOne) +
         (Count_MatrixColumn_9 >= FPlusOne) +
--         (Count_MatrixColumn_10 >= FPlusOne)) >= TwoFPlusOne) ; -- For faulty nodes participating in rounds 2 and 3.
        (Count_MatrixColumn_10 >= FPlusOne)) >= FPlusTwo) ; -- For faulty nodes going silent in rounds 2 and 3.

  ---------------------------------------------------------------------------

ASSIGN

  ---------------------------------------------------------------------------
  init (RoundNum)   := 0 ;
--  init (MessageOut)    := {NONE, Sync, Relay} ;
  init (MessageOut) := NONE ;

  init (MsgVector [1]) := 0 ;
  init (MsgVector [2]) := 0 ;
  init (MsgVector [3]) := 0 ;
  init (MsgVector [4]) := 0 ;
  init (MsgVector [5]) := 0 ;
  init (MsgVector [6]) := 0 ;
  init (MsgVector [7]) := 0 ;
  init (MsgVector [8]) := 0 ;
  init (MsgVector [9]) := 0 ;
  init (MsgVector [10]) := 0 ;

  ---------------------------------------------------------------------------
  next (RoundNum) :=
    case
      (RoundNum = 3) : 0 ;
      (Global_Clock = MyP) & (SourceNode = Node_Id) : 1 ;
      (Count_Sync > 0) : 2 ;
      -- If at least F Relays and one Sync, from the source of course, then Round = 3.
      ((Count_Relay + MsgVector [SourceNode]) >= FPlusOne) : 3 ;
      -- Or, alternatively, the following will do.
--       (Count_Relay + Count_MyVector >= FPlusOne) : 3 ;
      1 : RoundNum ;
    esac ;

  ---------------------------------------------------------------------------
  next (MessageOut) :=
    case
      (Global_Clock = MyP) & (SourceNode = Node_Id) : Sync ;
      (Count_Sync > 0) : Relay ;
      1 : NONE ;
    esac ;

  ---------------------------------------------------------------------------
  next (MsgVector [1]) :=
    case
      (N1_Msg = Sync) | (N1_Msg = Relay) : 1 ;
      1 : MsgVector [1] ;
    esac ;

  next (MsgVector [2]) :=
    case
      (N2_Msg = Sync) | (N2_Msg = Relay) : 1 ;
      1 : MsgVector [2] ;
    esac ;

  next (MsgVector [3]) :=
    case
```

```
        (N3_Msg = Sync) | (N3_Msg = Relay) : 1 ;
        1 : MsgVector [3] ;
      esac ;

    next (MsgVector [4]) :=
      case
        (N4_Msg = Sync) | (N4_Msg = Relay) : 1 ;
        1 : MsgVector [4] ;
      esac ;

    next (MsgVector [5]) :=
      case
        (N5_Msg = Sync) | (N5_Msg = Relay) : 1 ;
        1 : MsgVector [5] ;
      esac ;

    next (MsgVector [6]) :=
      case
        (N6_Msg = Sync) | (N6_Msg = Relay) : 1 ;
        1 : MsgVector [6] ;
      esac ;

    next (MsgVector [7]) :=
      case
        (N7_Msg = Sync) | (N7_Msg = Relay) : 1 ;
        1 : MsgVector [7] ;
      esac ;

    next (MsgVector [8]) :=
      case
        (N8_Msg = Sync) | (N8_Msg = Relay) : 1 ;
        1 : MsgVector [8] ;
      esac ;

    next (MsgVector [9]) :=
      case
        (N9_Msg = Sync) | (N9_Msg = Relay) : 1 ;
        1 : MsgVector [9] ;
      esac ;

    next (MsgVector [10]) :=
      case
        (N10_Msg = Sync) | (N10_Msg = Relay) : 1 ;
        1 : MsgVector [10] ;
      esac ;

    ------------------------------------------------------------------------------
    ------------------------------------------------------------------------------

-- end of Node

------------------------------------------------------------------------------
------------------------------------------------------------------------------
```

Appendix B

```
--------------------------------------------------------------------------------
-- File Name:       3ROM_LinkFault_k10.smv
--                  This file is for a system of K good nodes, where F > 1.
--                  Note that the graph is fully connected.
--
--
-- Environment :    SMV
-- Organization:    NASA Langley Research Center
-- Project:         Self-Stablization
-- Authors:         Malekpour, Mahyar
--
--                  NASA Langley Research Center
--                  Hampton, VA 23681-2199
-- Creation Date:   9/4/2014
--------------------------------------------------------------------------------
--
-- This SMV description is property of the National Aeronautics and Space
-- Administration.  Unauthorized use or duplication of this VHDL description is
-- strictly prohibited.  Authorized users are subject to the following
-- restrictions:
--
--    . Neither the author, their corporation, nor NASA is responsible for any
--      consequence of the use of this SMV description.
--
--    . The origin of this SMV description must not be misrepresented either
--      by explicit claim or by omission.
--
--    . Altered versions of this SMV description must be plainly marked as
--      such.
--
--
--    . This notice may not be removed or altered.
--
--------------------------------------------------------------------------------
--
--    Modified on:  9/ 4/2014
--    by:           Mahyar Malekpour
--
--------------------------------------------------------------------------------
--
-- Global Constants:
--

#define Node_Id1                1
#define Node_Id2                2
#define Node_Id3                3
#define Node_Id4                4
#define Node_Id5                5
#define Node_Id6                6
#define Node_Id7                7
#define Node_Id8                8
#define Node_Id9                9
#define Node_Id10               10

#define SourceNodeId            (Node_Id1)

-- Drift in units of Gamma
#define DriftP                  (5)

-- Network Size
#define K                       10
#define G                       7
#define F                       3
#define FPlusOne                (F + 1)
#define FPlusTwo                (F + 2)
#define TwoFPlusOne             (2 * F + 1)
```

```
-- Topology = fully connected graph
#define P                       (20 + DriftP)

#define GlobalClockMax          (P + 5)

-- P of SourceNodeId + 3 rounds
#define TimeToVote              (P + 3)

-- Drift at the local level
#define P_N1                    (P - 0)
#define P_N2                    (P - 1)
#define P_N3                    (P - 2)
#define P_N4                    (P - 3)
#define P_N5                    (P - 4)
#define P_N6                    (P + 1)
#define P_N7                    (P + 2)
#define P_N8                    (P + 3)
#define P_N9                    (P + 4)
#define P_N10                   (P + 5)

-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
--
--
--
MODULE main

VAR
  Global_Clock : 0 .. GlobalClockMax ;

  -----------------------------------------------------------------------------
  Node_1 : Node (Node_Id1, P_N1, Global_Clock, SourceNodeId,
              Node_1.MessageOut,
              Node_2.MessageOut,
              Node_3.MessageOut,
              Node_4.MessageOut,
              Node_5.MessageOut,
              Node_6.MessageOut,
              Node_7.MessageOut,
--              Node_8.MessageOut,
--              Node_9.MessageOut,
--              Node_10.MessageOut,
              NONE,
              NONE,
              NONE,
              Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
              Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
              Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;

  Node_2 : Node (Node_Id2, P_N2, Global_Clock, SourceNodeId,
              Node_1.MessageOut,
              Node_2.MessageOut,
              Node_3.MessageOut,
              Node_4.MessageOut,
              Node_5.MessageOut,
              Node_6.MessageOut,
--              Node_7.MessageOut,
              NONE,
              Node_8.MessageOut,
--              Node_9.MessageOut,
--              Node_10.MessageOut,
              NONE,
              NONE,
              Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
              Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
              Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;

  Node_3 : Node (Node_Id3, P_N3, Global_Clock, SourceNodeId,
              Node_1.MessageOut,
              Node_2.MessageOut,
```

```
            Node_3.MessageOut,
            Node_4.MessageOut,
--            Node_5.MessageOut,
--            Node_6.MessageOut,
            NONE,
            NONE,
            Node_7.MessageOut,
--            Node_8.MessageOut,
            NONE,
            Node_9.MessageOut,
            Node_10.MessageOut,
            Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
            Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
            Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;

  Node_4 : Node (Node_Id4, P_N4, Global_Clock, SourceNodeId,
            Node_1.MessageOut,
            Node_2.MessageOut,
            Node_3.MessageOut,
            Node_4.MessageOut,
--            Node_5.MessageOut,
            NONE,
            Node_6.MessageOut,
--            Node_7.MessageOut,
            NONE,
            Node_8.MessageOut,
            Node_9.MessageOut,
--            Node_10.MessageOut,
            NONE,
            Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
            Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
            Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;

  Node_5 : Node (Node_Id5, P_N5, Global_Clock, SourceNodeId,
            Node_1.MessageOut,
            Node_2.MessageOut,
--            Node_3.MessageOut,
            NONE,
            Node_4.MessageOut,
            Node_5.MessageOut,
--            Node_6.MessageOut,
            NONE,
            Node_7.MessageOut,
            Node_8.MessageOut,
--            Node_9.MessageOut,
            NONE,
            Node_10.MessageOut,
            Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
            Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
            Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;

  Node_6 : Node (Node_Id6, P_N6, Global_Clock, SourceNodeId,
            Node_1.MessageOut,
            Node_2.MessageOut,
--            Node_3.MessageOut,
--            Node_4.MessageOut,
            NONE,
            NONE,
            Node_5.MessageOut,
            Node_6.MessageOut,
            Node_7.MessageOut,
--            Node_8.MessageOut,
            NONE,
            Node_9.MessageOut,
            Node_10.MessageOut,
            Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
            Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
            Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;

  Node_7 : Node (Node_Id7, P_N7, Global_Clock, SourceNodeId,
```

```
                Node_1.MessageOut,
--              Node_2.MessageOut,
                NONE,
                Node_3.MessageOut,
--              Node_4.MessageOut,
                NONE,
                Node_5.MessageOut,
--              Node_6.MessageOut,
                NONE,
                Node_7.MessageOut,
                Node_8.MessageOut,
                Node_9.MessageOut,
                Node_10.MessageOut,
                Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
                Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
                Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;

  Node_8 : Node (Node_Id8, P_N8, Global_Clock, SourceNodeId,
--              Node_1.MessageOut,
--              Node_2.MessageOut,
                NONE,
                NONE,
                Node_3.MessageOut,
                Node_4.MessageOut,
                Node_5.MessageOut,
                Node_6.MessageOut,
--              Node_7.MessageOut,
                NONE,
                Node_8.MessageOut,
                Node_9.MessageOut,
                Node_10.MessageOut,
                Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
                Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
                Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;

  Node_9 : Node (Node_Id9, P_N9, Global_Clock, SourceNodeId,
--              Node_1.MessageOut,
                NONE,
                Node_2.MessageOut,
                Node_3.MessageOut,
--              Node_4.MessageOut,
--              Node_5.MessageOut,
                NONE,
                NONE,
                Node_6.MessageOut,
                Node_7.MessageOut,
                Node_8.MessageOut,
                Node_9.MessageOut,
                Node_10.MessageOut,
                Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
                Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
                Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;

  Node_10 : Node (Node_Id10, P_N10, Global_Clock, SourceNodeId,
--              Node_1.MessageOut,
--              Node_2.MessageOut,
--              Node_3.MessageOut,
                NONE,
                NONE,
                NONE,
                Node_4.MessageOut,
                Node_5.MessageOut,
                Node_6.MessageOut,
                Node_7.MessageOut,
                Node_8.MessageOut,
                Node_9.MessageOut,
                Node_10.MessageOut,
                Node_1.MsgVector, Node_2.MsgVector, Node_3.MsgVector, Node_4.MsgVector,
                Node_5.MsgVector, Node_6.MsgVector, Node_7.MsgVector,
                Node_8.MsgVector, Node_9.MsgVector, Node_10.MsgVector) ;
```

```
DEFINE

  ---------------------------------------------------------------------------
  VoteTime       := (Global_Clock >= TimeToVote) ;

  GlobalAgreement := (Global_Clock = TimeToVote) &
                     (Node_1.VoteResult & Node_2.VoteResult &
                      Node_3.VoteResult & Node_4.VoteResult &
                      Node_5.VoteResult & Node_6.VoteResult &
                      Node_7.VoteResult & Node_8.VoteResult &
                      Node_9.VoteResult & Node_10.VoteResult) ;

ASSIGN

  ---------------------------------------------------------------------------
  init (Global_Clock) := 0 ;
  next (Global_Clock) :=
    case
      (Global_Clock < GlobalClockMax) : Global_Clock + 1 ;
      1 : Global_Clock ;
    esac ;

  ---------------------------------------------------------------------------
  ---------------------------------------------------------------------------

SPEC

  ---------------------------------------------------------------------------
  --
  -- Proposition #1:
  --
--   AF (VoteTime)

  ---------------------------------------------------------------------------
  --
  -- Proposition #2:
  --
  AF (VoteTime & GlobalAgreement) -- true

  ---------------------------------------------------------------------------

-- end of main

---------------------------------------------------------------------------
---------------------------------------------------------------------------
--
-- Good node, for link-fault model.
--
--
MODULE Node (Node_Id, MyP, Global_Clock, SourceNode,
             N1_Msg, N2_Msg, N3_Msg, N4_Msg, N5_Msg, N6_Msg, N7_Msg, N8_Msg, N9_Msg, N10_Msg,
             N1_MsgVector, N2_MsgVector, N3_MsgVector, N4_MsgVector, N5_MsgVector,
             N6_MsgVector, N7_MsgVector, N8_MsgVector, N9_MsgVector, N10_MsgVector)

VAR

  RoundNum      : 0 .. 3 ;
  MessageOut    : {NONE, Sync, Relay} ;

  -- Messages recieved (Rx) from the nodes are saved in this vector and will be
  -- broadcast/used during round 3.
  MsgVector     : array 1..10 of {0, 1} ;
  VoteResult    : boolean ;

DEFINE

  ---------------------------------------------------------------------------
  Count_Sync :=
      ((N1_Msg = Sync) +
```

```
      (N2_Msg = Sync) +
      (N3_Msg = Sync) +
      (N4_Msg = Sync) +
      (N5_Msg = Sync) +
      (N6_Msg = Sync) +
      (N7_Msg = Sync) +
      (N8_Msg = Sync) +
      (N9_Msg = Sync) +
      (N10_Msg = Sync)) ;

Count_Relay :=
    ((N1_Msg = Relay) +
      (N2_Msg = Relay) +
      (N3_Msg = Relay) +
      (N4_Msg = Relay) +
      (N5_Msg = Relay) +
      (N6_Msg = Relay) +
      (N7_Msg = Relay) +
      (N8_Msg = Relay) +
      (N9_Msg = Relay) +
      (N10_Msg = Relay)) ;

Count_MyVector :=
    (MsgVector [1] +
      MsgVector [2] +
      MsgVector [3] +
      MsgVector [4] +
      MsgVector [5] +
      MsgVector [6] +
      MsgVector [7] +
      MsgVector [8] +
      MsgVector [9] +
      MsgVector [10]) ;

Count_MatrixColumn_1 :=
    (N1_MsgVector [1] +
      N2_MsgVector [1] +
      N3_MsgVector [1] +
      N4_MsgVector [1] +
      N5_MsgVector [1] +
      N6_MsgVector [1] +
      N7_MsgVector [1] +
      N8_MsgVector [1] +
      N9_MsgVector [1] +
      N10_MsgVector [1]) - F ;

Count_MatrixColumn_2 :=
    (N1_MsgVector [2] +
      N2_MsgVector [2] +
      N3_MsgVector [2] +
      N4_MsgVector [2] +
      N5_MsgVector [2] +
      N6_MsgVector [2] +
      N7_MsgVector [2] +
      N8_MsgVector [2] +
      N9_MsgVector [2] +
      N10_MsgVector [2]) - F ;

Count_MatrixColumn_3 :=
    (N1_MsgVector [3] +
      N2_MsgVector [3] +
      N3_MsgVector [3] +
      N4_MsgVector [3] +
      N5_MsgVector [3] +
      N6_MsgVector [3] +
      N7_MsgVector [3] +
      N8_MsgVector [3] +
      N9_MsgVector [3] +
      N10_MsgVector [3]) - F ;
```

```
Count_MatrixColumn_4 :=
    (N1_MsgVector [4] +
     N2_MsgVector [4] +
     N3_MsgVector [4] +
     N4_MsgVector [4] +
     N5_MsgVector [4] +
     N6_MsgVector [4] +
     N7_MsgVector [4] +
     N8_MsgVector [4] +
     N9_MsgVector [4] +
     N10_MsgVector [4]) - F ;

Count_MatrixColumn_5 :=
    (N1_MsgVector [5] +
     N2_MsgVector [5] +
     N3_MsgVector [5] +
     N4_MsgVector [5] +
     N5_MsgVector [5] +
     N6_MsgVector [5] +
     N7_MsgVector [5] +
     N8_MsgVector [5] +
     N9_MsgVector [5] +
     N10_MsgVector [5]) - F ;

Count_MatrixColumn_6 :=
    (N1_MsgVector [6] +
     N2_MsgVector [6] +
     N3_MsgVector [6] +
     N4_MsgVector [6] +
     N5_MsgVector [6] +
     N6_MsgVector [6] +
     N7_MsgVector [6] +
     N8_MsgVector [6] +
     N9_MsgVector [6] +
     N10_MsgVector [6]) - F ;

Count_MatrixColumn_7 :=
    (N1_MsgVector [7] +
     N2_MsgVector [7] +
     N3_MsgVector [7] +
     N4_MsgVector [7] +
     N5_MsgVector [7] +
     N6_MsgVector [7] +
     N7_MsgVector [7] +
     N8_MsgVector [7] +
     N9_MsgVector [7] +
     N10_MsgVector [7]) - F ;

Count_MatrixColumn_8 :=
    (N1_MsgVector [8] +
     N2_MsgVector [8] +
     N3_MsgVector [8] +
     N4_MsgVector [8] +
     N5_MsgVector [8] +
     N6_MsgVector [8] +
     N7_MsgVector [8] +
     N8_MsgVector [8] +
     N9_MsgVector [8] +
     N10_MsgVector [8]) - F ;

Count_MatrixColumn_9 :=
    (N1_MsgVector [9] +
     N2_MsgVector [9] +
     N3_MsgVector [9] +
     N4_MsgVector [9] +
     N5_MsgVector [9] +
     N6_MsgVector [9] +
     N7_MsgVector [9] +
     N8_MsgVector [9] +
     N9_MsgVector [9] +
```

```
        N10_MsgVector [9]) - F ;

    Count_MatrixColumn_10 :=
        (N1_MsgVector [10] +
         N2_MsgVector [10] +
         N3_MsgVector [10] +
         N4_MsgVector [10] +
         N5_MsgVector [10] +
         N6_MsgVector [10] +
         N7_MsgVector [10] +
         N8_MsgVector [10] +
         N9_MsgVector [10] +
         N10_MsgVector [10]) - F ;

    VoteResult :=
         (RoundNum = 3) &
         (((Count_MatrixColumn_1 >= FPlusOne) +
           (Count_MatrixColumn_2 >= FPlusOne) +
           (Count_MatrixColumn_3 >= FPlusOne) +
           (Count_MatrixColumn_4 >= FPlusOne) +
           (Count_MatrixColumn_5 >= FPlusOne) +
           (Count_MatrixColumn_6 >= FPlusOne) +
           (Count_MatrixColumn_7 >= FPlusOne) +
           (Count_MatrixColumn_8 >= FPlusOne) +
           (Count_MatrixColumn_9 >= FPlusOne) +
--           (Count_MatrixColumn_10 >= FPlusOne)) >= TwoFPlusOne) ;
          (Count_MatrixColumn_10 >= FPlusOne)) >= FPlusTwo) ;

    ---------------------------------------------------------------------------

ASSIGN

    ---------------------------------------------------------------------------
    init (RoundNum)   := 0 ;
--  init (MessageOut) := {NONE, Sync, Relay} ;
    init (MessageOut) := NONE ;

    init (MsgVector [1]) := 0 ;
    init (MsgVector [2]) := 0 ;
    init (MsgVector [3]) := 0 ;
    init (MsgVector [4]) := 0 ;
    init (MsgVector [5]) := 0 ;
    init (MsgVector [6]) := 0 ;
    init (MsgVector [7]) := 0 ;
    init (MsgVector [8]) := 0 ;
    init (MsgVector [9]) := 0 ;
    init (MsgVector [10]) := 0 ;

    ---------------------------------------------------------------------------
    next (RoundNum) :=
      case
        (RoundNum = 3) : 0 ;
        (Global_Clock = MyP) & (SourceNode = Node_Id) : 1 ;
        !(SourceNode = Node_Id) & (Count_Sync > 0) : 2 ;
        -- If at least F Relays and one Sync, from the source of course, then Round = 3.
        ((Count_Relay + MsgVector [SourceNode]) >= FPlusOne) : 3 ;
        -- Or, alternatively, the following will do.
--        (Count_Relay + Count_MyVector >= FPlusOne) : 3 ;
        1 : RoundNum ;
      esac ;

    ---------------------------------------------------------------------------
    next (MessageOut) :=
      case
        (Global_Clock = MyP) & (SourceNode = Node_Id) : Sync ;
        (Count_Sync > 0) : Relay ;
        1 : NONE ;
      esac ;

    ---------------------------------------------------------------------------
```

```
  next (MsgVector [1]) :=
    case
      (N1_Msg = Sync) | (N1_Msg = Relay) : 1 ;
      1 : MsgVector [1] ;
    esac ;

  next (MsgVector [2]) :=
    case
      (N2_Msg = Sync) | (N2_Msg = Relay) : 1 ;
      1 : MsgVector [2] ;
    esac ;

  next (MsgVector [3]) :=
    case
      (N3_Msg = Sync) | (N3_Msg = Relay) : 1 ;
      1 : MsgVector [3] ;
    esac ;

  next (MsgVector [4]) :=
    case
      (N4_Msg = Sync) | (N4_Msg = Relay) : 1 ;
      1 : MsgVector [4] ;
    esac ;

  next (MsgVector [5]) :=
    case
      (N5_Msg = Sync) | (N5_Msg = Relay) : 1 ;
      1 : MsgVector [5] ;
    esac ;

  next (MsgVector [6]) :=
    case
      (N6_Msg = Sync) | (N6_Msg = Relay) : 1 ;
      1 : MsgVector [6] ;
    esac ;

  next (MsgVector [7]) :=
    case
      (N7_Msg = Sync) | (N7_Msg = Relay) : 1 ;
      1 : MsgVector [7] ;
    esac ;

  next (MsgVector [8]) :=
    case
      (N8_Msg = Sync) | (N8_Msg = Relay) : 1 ;
      1 : MsgVector [8] ;
    esac ;

  next (MsgVector [9]) :=
    case
      (N9_Msg = Sync) | (N9_Msg = Relay) : 1 ;
      1 : MsgVector [9] ;
    esac ;

  next (MsgVector [10]) :=
    case
      (N10_Msg = Sync) | (N10_Msg = Relay) : 1 ;
      1 : MsgVector [10] ;
    esac ;

  ------------------------------------------------------------------------
  ------------------------------------------------------------------------

-- end of Node

------------------------------------------------------------------------
------------------------------------------------------------------------
```

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01- 08 - 2015	Technical Memorandum	

4. TITLE AND SUBTITLE

Achieving Agreement In Three Rounds With Bounded-Byzantine Faults

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Malekpour, Mahyar R.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

999182.02.50.07.02

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

NASA Langley Research Center
Hampton, VA 23681-2199

8. PERFORMING ORGANIZATION REPORT NUMBER

L-20585

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSOR/MONITOR'S ACRONYM(S)

NASA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

NASA-TM-2015-218789

12. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified - Unlimited
Subject Category 62
Availability: NASA STI Program (757) 864-9658

13. SUPPLEMENTARY NOTES

14. ABSTRACT

A three-round algorithm is presented that guarantees agreement in a system of K = 3F+1 nodes provided each faulty node induces no more than F faults and each good node experiences no more than F faults, where, F is the maximum number of simultaneous faults in the network. The algorithm is based on the Oral Message algorithm of Lamport et al. and is scalable with respect to the number of nodes in the system and applies equally to the traditional node-fault model as well as the link-fault model. We also present a mechanical verification of the algorithm focusing on verifying the correctness of a bounded model of the algorithm as well as confirming claims of determinism.

15. SUBJECT TERMS

Agreement; Byzantine; Distributed; Fault tolerant; Model checking; Oral message; Synchronization

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Help Desk (email: help@sti.nasa.gov)
					19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	42	(757) 864-9658