A Fault-Tolerant Triple Triangular Mesh Protocol for Distributed Mutual Exclusion[1]

Ye-In Chang and Yao-Jen Chang
Dept. of Applied Mathematics
National Sun Yat-Sen University
Kaohsiung, Taiwan
Republic of China
{E-Mail: changyi@math.nsysu.edu.tw}
{Tel: 886-7-5252000 (ext. 3710)}
{Fax: 886-7-5253809}

Abstract

In the paper, we propose a *triple triangular mesh protocol* for mutual exclusion, in which the nodes in the system are organized into a triangular mesh. The quorum size is k that is $O(\sqrt{N})$, where N is the number of nodes in the system and is equal to $\frac{k(k+1)}{2}$. The protocol is fault-tolerant up to $(k-2)$ site failures and communication failures in the worst case, even when such failures lead to network partitioning.

(*Key Words:* Availability, distributed systems, fault tolerance, mutual exclusion, quorum consensus.)

[1]This research was supported by National Science Council of the Republic of China, NCS-81-0408-E-110-508.

1 Introduction

A distributed system consists of a set of distinct nodes and a communication network through which nodes can communicate with each other by sending messages. The system goes on working even a number of nodes failed. Each node is assumed to be fail-stop, i.e., a failed node will not send out messages which will make alive nodes confused. Node or link failures or combinations of both kind may divide the system into disjoint partitions. Nodes can communicate with nodes residing in the same partition but have no way to communicate with nodes belonging to other partitions.

To make distributed mutual exclusion protocols fault-tolerant to node and communication failures, many researches apply the replica control strategies to achieve mutual exclusion. The majority voting protocol [6] and the quorum consensus protocol [3] are such examples. However, these protocols require high communication cost which is $O(N)$ due to the large quorum size.

To reduce the overhead of achieving mutual exclusion while supporting fault tolerance, many protocols imposing a logical structure on the network are proposed [1, 2, 4]. (Note that imposed on the system is a logical structure which does not take into account the real topology of the network). The hierarchical quorum consensus protocol (HQC) [4] requires $O(N^{0.63})$ messages, the tree quorum protocol [1] requires $O(logN)$ messages in the best case and degrades gracefully, and the grid protocol [2] requires $O(\sqrt{N})$ messages. All the quorums constructed from these protocols can be used to replace the set of nodes in Maekawa's protocol [5] to achieve mutual exclusion.

In the paper, we propose a *triple triangular mesh protocol* for mutual exclusion, in which the nodes in the system are organized into a triangular mesh. A quorum contains nodes from some side of each of three subtriangles in the triangular mesh and the quorum size is k that is $O(\sqrt{N})$, where N is the number of nodes in the system and is equal to $\frac{k(k+1)}{2}$. The protocol is fault-tolerant up to $(k-2)$ site failures and communication failures in the worst case, even when such failures lead to network partitioning. From our simulation study, the proposed protocol can have higher availability and less quorum size than the grid protocol. Moreover, the quorum size of the proposed protocol will be less than that in the HQC protocol when N is greater than or equal to 15 and less than that in the tree quorum protocol when node failures exist.

Figure 1: (a) A 6-triangular mesh (b) subtriangles associated with node 7 in a 6-triangular mesh.

2 The Triple Triangular Mesh Protocol

In this section, we present the triple triangular mesh (TTM) protocol, prove the correctness of the protocol, and describe a property of the protocol.

2.1 Definitions

We organize nodes into a triangular mesh. A *k-triangular mesh* consists of a vertex set V and an edge set E. V is defined as a set of $(x,\ y)$-tuples, where x, y are both integers, $0 \leq x \leq k-1$, $0 \leq y \leq k-1$ and $0 \leq x+y \leq k-1$. The y-axis is slanted to the right 30 degree to accommodate the left-hand side of the right triangle. E is defined as a set of vertex pairs (v_1, v_2), where v_1 and v_2 are in V, $v_1 = (x_1, y_1)$, $v2 = (x_2, y_2)$ and x_1, y_1, x_2 and y_2 satisfy one of the following conditions: (1) $x_1 - x_2 = y_2 - y_1 = 1$, (2) $x_2 - x_1 = y_1 - y_2 = 1$, (3) $\mid x_1 - x_2 \mid + \mid y_1 - y_2 \mid = 1$. A *6-triangular mesh* is shown in Figure 1-(a). In this example, node 3 is at $(0,\ 3)$ and node 13 is at $(3,\ 1)$. The left-hand, right-hand and bottom sides of the right triangle are referred to as side 0, side 1 and side 2, respectively.

Given a node *x* residing inside the triangular mesh, we can draw three lines along those edges of the triangular mesh, such that each of them is parallel with one side of the triangular mesh and passes through the given node. Based on these three sides and three lines, three smaller triangles are formed in the triangular mesh. Figure 1-(b) shows such an example, where node 7 is the given center. We define a small triangle as subtriangle *i* such that one of its sides is a subset of side *i* of the triangular mesh, and define subtriangle *i*'s two sides counterclockwisely, which are not subsets of any side of the triangular mesh, as side (a) and side (b), respectively. Figure 2 shows

Figure 2: Sides of subtriangles: (a) subtriangle 0; (b) subtriangle 1; (c) subtriangle 2.

sides of three subtriangles from Figure 1-(b).

Definition 1. *Each quorum consists of a center and subquorum i, for i = 0, 1, 2. Subquorum i contains all the nodes on either side (a) or side (b) of subtriangle i, and we refer to the former as subquorum i-(a) and the latter as subquorum i-(b), respectively. Formally, subquorum* i *is a sequence of nodes,* $(v_0, \ldots, v_m)$*, where* v_0 *is the center of the quorum, and* v_m *is the ending node of the subquorum. Any two adjacent nodes* v_j*,* v_{j+1}*,* $0 \leq j < m$ *in the subquorum must satisfy the condition according to its type:*

1. $x_{j+1} - x_j = -1$ *and* $y_{j+1} = y_j$*, for subquorum 0-(a);*
2. $x_{j+1} - x_j = -1$ *and* $y_{j+1} - y_j = 1$*, for subquorum 0-(b);*
3. $x_{j+1} = x_j$ *and* $y_{j+1} - y_j = 1$*, for subquorum 1-(a);*
4. $x_{j+1} - x_j = 1$ *and* $y_{j+1} = y_j$*, for subquorum 1-(b);*
5. $x_{j+1} - x_j = 1$ *and* $y_{j+1} - y_j = -1$*, for subquorum 2-(a);*
6. $x_{j+1} = x_j$ *and* $y_{j+1} - y_j = -1$*, for subquorum 2-(b).*

Also, we will call a quorum *"associated with"* node x if it uses node x as its center, and a subquorum *"associated with"* node x if it is a subquorum of the quorum associated with node x.

Example 1. For a node residing inside a triangular mesh, there are eight quorums associated with it. Figure 3 shows eight quorums associated with node a residing inside a 6-triangular mesh.

Example 2. For a node residing at one side of a triangular mesh but not at any corners, there are four quorums associated with it. Figure 4 shows of four quorums associated with node a residing at one side of the triangular mesh. (Note that the number of subtriangles generated depends on the location of the center of the quorum.)

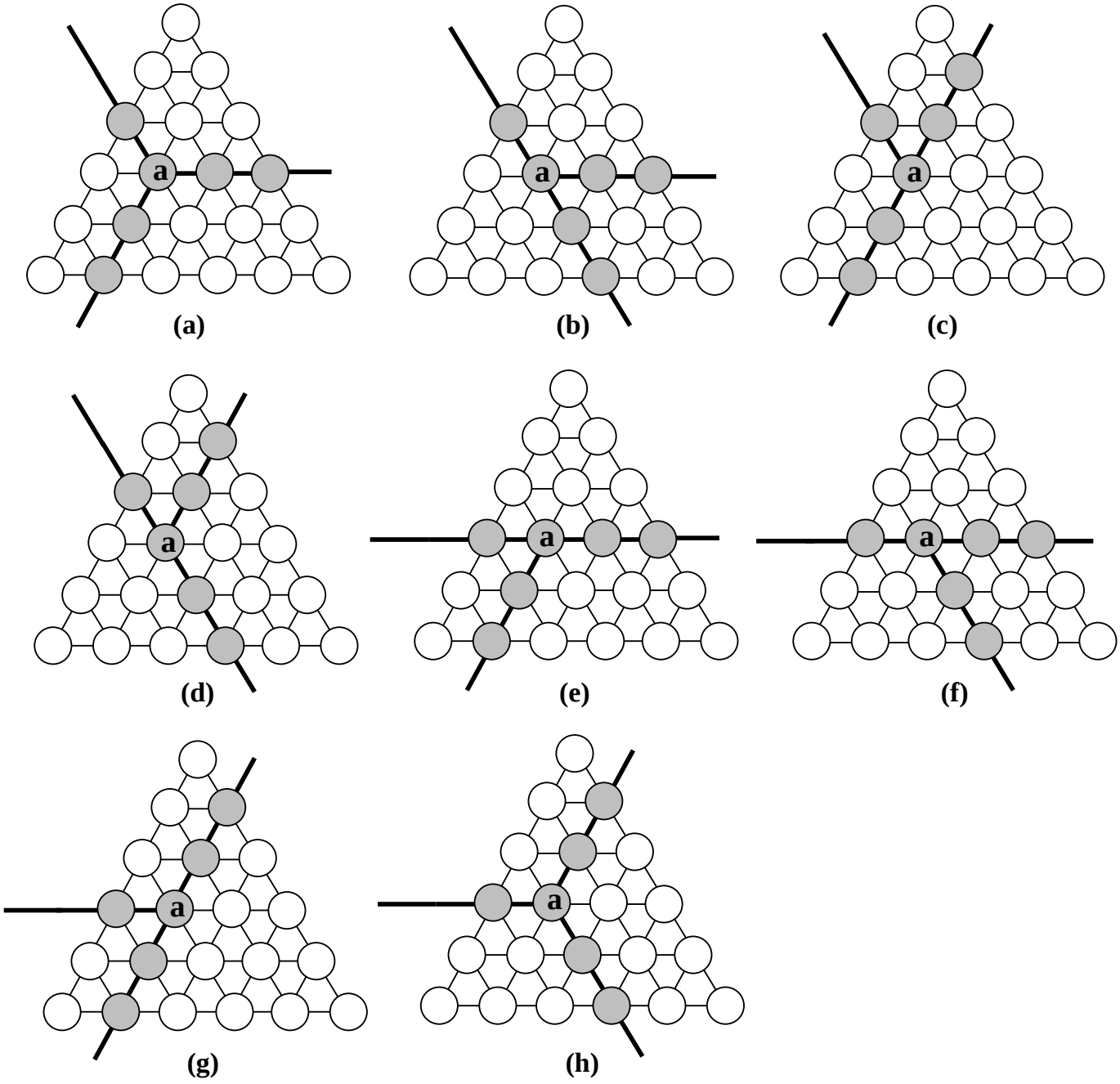

Figure 3: Examples of quorums associated with a node a in a 6-triangular mesh where a resides inside the triangle: (a)-(h).

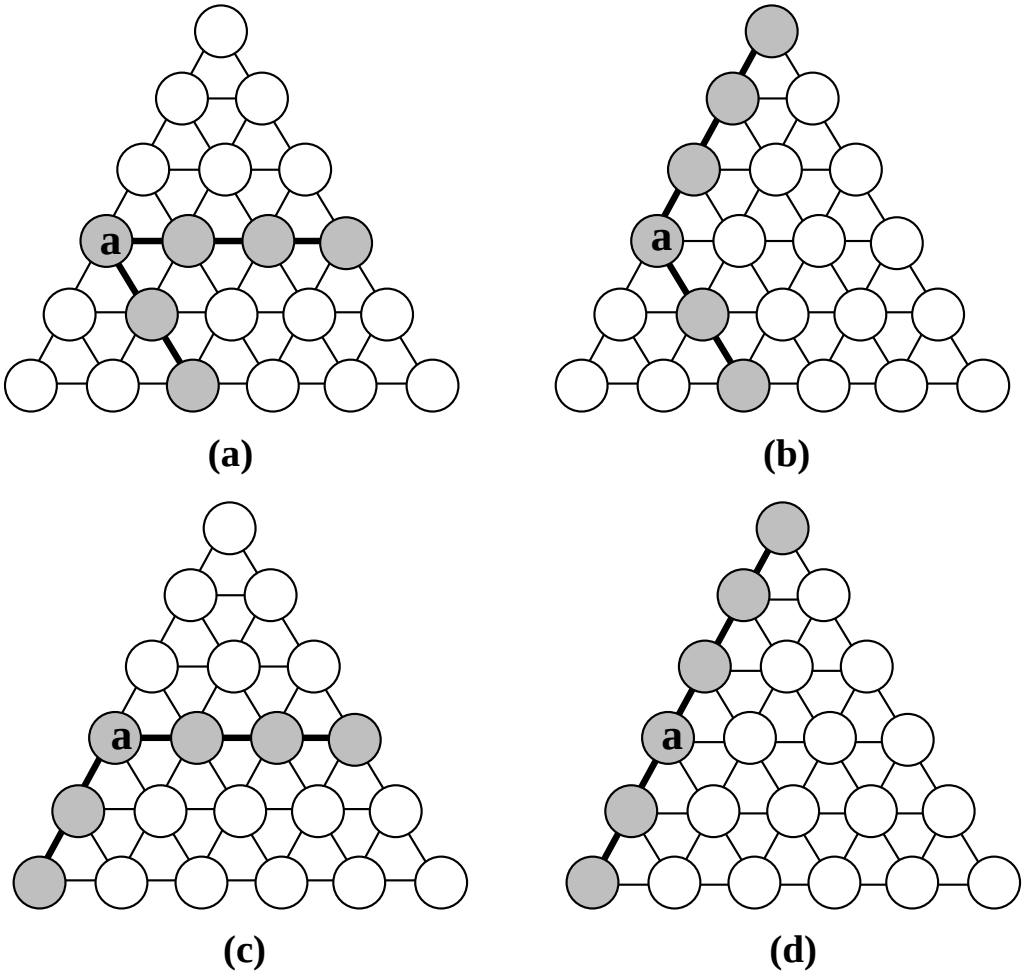

Figure 4: Examples of quorums associated with a node a in a 6-triangular mesh where a resides at one side but is not at a corner: (a)-(d).

2.2 Proof of Correctness

The following theorem is used to show that the proposed protocol can ensure mutual exclusion.

Theorem 1. *Any two quorums of the triple triangular mesh protocol intersect.*

Proof. We prove by first constructing a quorum x, then further construction of a quorum y will intersect some node in quorum x. There are three cases which should be taken into account according to where the center of quorum x is:

(1) If the center of quorum x is a node inside the triangular mesh, quorum x will divide the triangular mesh into three regions and no nodes of quorum x belong to any region. Figure 3 is a demonstration. To prevent from intersecting quorum x, the center of quorum y should reside in any region. Since every region can provide nodes of two sides, only the other two regions can provide nodes of the third side which quorum y needs. The path from the center of quorum y to another region will intersect some node of quorum x since every two regions are separated by quorum x.

(2) If the center of quorum x resides on any of three sides but is not any of three corners, there are three cases should be taken into account: none, one or two of x's subquorums resides at some side of the triangular mesh. When none of x's subquorum resides at some side (ex., Figure 4-(a)), the triangular mesh is divided into three parts by quorum x. In a similar way as in case (1), we can show that no other quorum can be constructed without intersecting quorum x. If only one of x's subquorum resides at some side, the triangle will be divided into two regions (ex., Figures 4-(b)). Since every region can provide nodes of two sides, only the other region can provide nodes of the third side which quorum y needs. The path from the center of quorum y to another region will intersect some node of quorum x. If two of x's subquorums reside at some side of the triangle, quorum x is one side of the triangle (ex., Figure 4-(c)), and quorum y will intersect quorum x at least one node since quorum y must get at least one node from every side.

(3) If the center of quorum x is one of three corners, quorum x will contain all nodes of one side of the triangular mesh, quorum y will intersect quorum x at least one node since quorum y must get at least one node from every side. □

2.3 Property of the Protocol

Theorem 2. *Given a k-triangular mesh, the quorum size of the proposed protocol is k, which is proportional to* $\sqrt{N}$*, where N is the number of nodes in the system.*

Proof. Assume that the $(x,\ y)$-tuple associated with the center of a given quorum is $(x_0,\ y_0)$, subquorum 0 is a sequence of nodes $(v_0, \ldots,\ v_m)$, where v_0 is the center and v_m is the ending node. Take any two adjacent nodes $v_{i+1} = (x_{i+1},\ y_{i+1})$, and $v_i = (x_i,\ y_i)$, the condition $(x_{i+1} - x_i = -1)$ is true for all $i = 0, \ldots,\ m-1$. When we trace from the center to the ending node, we find that the value of x in the $(x,\ y)$-tuple is changed from x_0 to zero; therefore, there are $(x_0 + 1)$ nodes in subquorum 0. Apply the same approach and notation to any two adjacent nodes v_{i+1} and v_i in subquorum 1, the condition $((x_{i+1} + y_{i+1}) - (x_i + y_i) = 1)$ holds. Since the value of $(x + y)$ is changed from $(x_0 + y_0)$ to $(k-1)$, there are $(k - x_0 - y_0)$ nodes in subquorum 1. Similarly, for any two adjacent nodes v_{i+1} and v_i in subquorum 2, the condition $(y_{i+1} - y_i = -1)$ holds. Since the value of y is changed from y_0 to 0, there are $(y_0 + 1)$ nodes in subquorum 2. Since the center is counted three times, the number of nodes in the quorum is $(x_0 + 1) + (k - x_0 - y_0) + (y_0 + 1) - 2 = k$. Given $N = \frac{k(k+1)}{2}$ nodes in a *k-triangular mesh*, we get $k < \sqrt{2N} < k + 1$, i.e., $\sqrt{2N} - 1 < k < \sqrt{2N}$. Apparently, k is of $O(\sqrt{N})$. □

3 The Performance

In this section, some aspects of distributed mutual exclusion protocols imposing logical structures are analyzed: quorum size, availability and fault tolerance. We compare these features of the triple triangular mesh (TTM) protocol with the tree [1], the HQC [4], and the grid protocols [2]. (Note that since the write-write intersection property holds, the write quorums of the grid protocol can be used to control accesses to a shared resource.)

3.1 Quorum Size

The number of messages required to construct a quorum is proportional to the size of the quorums. In the HQC protocol [4], the quorum size is $N^{0.63}$. In the tree protocol [1], the quorum size is $log_2 N$ and is increased up to $\lceil \frac{N+1}{2} \rceil$ as the number of node failures is increased. In the grid

protocol [2], it organizes nodes into a $M_1 \times M_2$ $(= N)$ grid. Based on Theorem 2, the quorum size in our *k-triangular mesh* protocol is k, which is $[\sqrt{2N}]$. Therefore, the quorum size in our triple triangular mesh protocol is less than that in the grid protocol all the time, and is less than that of the HQC protocol when $N \geq 15$. Although the quorum size in the tree protocol is less than that in our protocol, if the node fails starting from the root to the leaf, and from the left to the right, and the number of failed nodes is increased by one at a time, our protocol has less quorum size than the tree protocol as node failure occurs. Take $N = 15$ as an example, where the tree protocol and the triple triangular mesh protocol have a similar topology, when the number of failed nodes is 1, in the worst case, the quorum size in our protocol is 5 and is 6 in the tree protocol. (Note that when the number of failed nodes is greater than 7, in the worst case, there is no quorum that can be constructed in the tree protocol.)

3.2 Availability

The availability is defined as the probability that a quorum can be constructed. We assume that each node is assumed to be independent available with probability p. Since given a number of node failures, the number of nodes prevented from constructing their associated quorums depends on the relative positions among those failed nodes, it is difficult to get a close form of availability in the triple triangular mesh protocol. Therefore, the availability of the triple triangular mesh protocol is computed by first generating all possible combinations for nodes to be available or unavailable, then we check each case to see whether a quorum can be constructed by a simulation study. If a quorum can be constructed, we add the possibility of occurrence of this case into the availability. The availability of the tree protocol and the HQC protocol are computed in the same way. For a $M_1 \times M_2$ grid, we use the formula: $(1-(1-p)^{M_1})^{M_2} - (1-p^{M_1}-(1-p)^{M_1})^{M_2}$ to get its availability [2]. Figure 5 shows the availability of these four protocols when $N = 15$. From this figure, we observe that the proposed protocol has higher availability than the grid protocol all the time, and also can have higher availability than the tree protocol when $p \geq 0.92$.

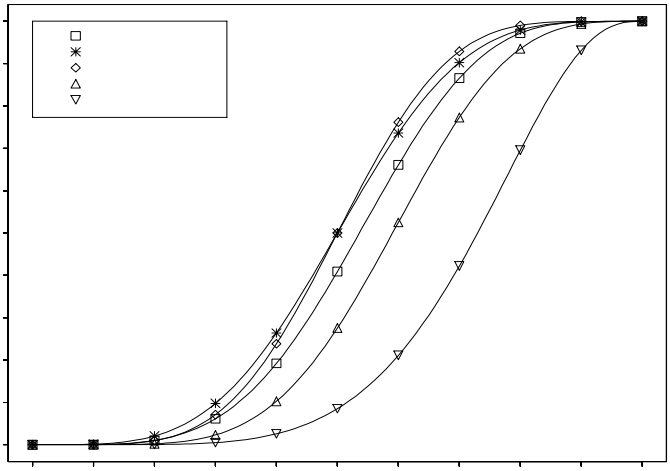

Figure 5: A comparison of availability

3.3 Fault Tolerance

Our simulation reveals that in the worst case, the proposed protocol can tolerate up to $(k-1)$ node failures when $k \leq 4$, and up to $(k-2)$ node failures when $k \geq 5$, where $N = \frac{k(k+1)}{2}$. This is because that there exist some patterns of $(k-1)$ node failures which disable all quorum constructions. For example, in a k-triangular mesh, $k \geq 5$, the set consists of failed nodes whose (x, y)-tuples satisfy one of the following conditions will make all quorums unavailable: for $i = 1, \ldots, k-1$, (1) ($\frac{i+1}{2}$, $\frac{i-1}{2}$), i is odd; (2) ($\frac{i-2}{2}$, $\frac{i+2}{2}$), i is even. For example, the set of 5 failed nodes: $\{$ 4, 6, 9, 12, 16 $\}$ in the 6-triangular mesh shown in Figure 1 will make all quorum constructions impossible.

The above discussion can be summarized in Table 1 based on six criteria, where we consider a $M_1 \times M_2 \, (= N)$ grid. The first two criteria are the quorum sizes in the best and worst cases, respectively. The third criterion is the impact that a single node failure would have on the size of a quorum. In the grid protocol, if the failed node belongs to a column included in the quorum, then another $(M_1 - 1)$ nodes are needed to form another quorum. In the tree protocol, the failure of the root will result in a requirement of another $log_2 N$ nodes to construct a new quorum. In the TTM protocol, in the worst case, we need another $(k-1)$ nodes to construct another quorum.

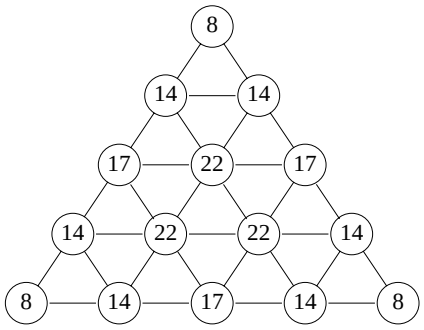

Figure 6: The number of quorums which a node can participate in the TTM protocol when N = 15

The fourth criterion is whether the protocol is a fully distributed one. The tree protocol assigns greater burden to the nodes with smaller level numbers. Therefore, the tree protocol is not a fully distributed protocol. The HQC and the grid protocols are fully distributed ones. In the TTM protocol, obviously, the nodes near the center bear greater burden than the nodes near the sides or the corners as shown in Figure 6; therefore, it is not a fully distributed protocol. (Note that in Figure 6, the total number of valid quorums is 45 when N=15, where the value inside the node denotes the number of quorums which a node can participate.) The last two criteria are the number of failed nodes which does not halt the system in the best and worst case, respectively. While in the best case, all these four protocols can be fault-tolerant up to all node failures except those nodes which have already constructed a quorum. While in the worst case, the tree, the HQC and the grid tree protocol can be fault-tolerant up to $(the\ quorum\ size - 1)$, $(the\ quorum\ size - 1)$ and $(\min\{M_1,\ M_2\} - 1)$ failures, respectively. The TTM protocol tolerates up to $(k - 2)$ node failures according to the simulation results.

4 Conclusion

In this paper, we have proposed a fault-tolerant triple triangular mesh protocol for mutual exclusion. From our simulation study, the triple triangular mesh protocol can have higher availability and less quorum size than the grid protocol. Moreover, the quorum size of the proposed protocol will be less than that in the HQC protocol when N is greater than or equal to 15 and less than that in the tree quorum protocol when node failures exist. Also, the triple triangular mesh protocol can have higher availability than the tree protocol when $p \geq 0.92$ and $N = 15$. How to extend

		HQC	Tree	Grid	TTM
(1)	quorum size (best case)	$N^{0.63}$	$log_2 N$	$M_1 + M_2 - 1$	$\lceil\sqrt{2N}\rceil$
(2)	quorum size (worst case)	$N^{0.63}$	$\lceil\frac{N+1}{2}\rceil$	$M_1 + M_2 - 1$	$\lceil\sqrt{2N}\rceil$
(3)	cost of one node failure (worst case)	1	$log_2 N$	$M_1 - 1$	$\lceil\sqrt{2N}\rceil - 1$
(4)	fully distributed?	yes	no	yes	no
(5)	fault tolerance (best case)	$N - N^{0.63}$	$N - log_2 N$	$N - (M_1 + M_2) + 1$	$N - \lceil\sqrt{2N}\rceil$
(6)	fault tolerance (worst case)	$N^{0.63} - 1$	$log_2 N - 1$	$\min\{M_1,\ M_2\} - 1$	$\lceil\sqrt{2N}\rceil - 2$

Table 1: A comparison of four mutual exclusion algorithms imposing logical structures

the TTM protocol to tolerate even more node failures is the future research direction.

References

[1] D. Agrawal and A. E. Abbadi, "An Efficient and Fault-Tolerant Solution for Distributed Mutual Exclusion," *ACM Transactions on Computer Systems*, Vol. 9, No. 1, pp. 1-20, Feb. 1991.

[2] S. Y. Cheung, M. H. Ammar, and M. Ahamad, "The Grid Protocol: A High Performance Scheme for Maintaining Replicated Data," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 4, No. 6, pp. 582-592, Dec. 1992.

[3] D. K. Gifford, "Weighted Voting for Replicated Data," in *Proc. of the 7th Symposium on Operating Systems Principles*, pp. 150-159, 1979.

[4] A. Kumar, "Hierarchical Quorum Consensus: A New Algorithm for Managing Replicated Data," *IEEE Transactions on Computers*, Vol. 40, No. 9, pp. 996-1004, Sep. 1991.

[5] M. Maekawa, "A $\sqrt{N}$ Algorithm for Mutual Exclusion in Decentralized Systems," *ACM Transactions on Computer Systems*, Vol. 3, No. 2, pp. 145-159, May 1985.

[6] R. H. Thomas, "A Majority Consensus Approach to Concurrency Control for Multiple Copy Databases," *ACM Transactions on Database Systems*, Vol. 4, No. 2, pp. 180-209, Jun. 1979.