Protocol Refinement for Maintaining Replicated Data in Distributed Systems

David Shou and Sheng-De Wang

Department of Electrical Engineering
National Taiwan University
Taipei, Taiwan 106, R.O.C.

Abstract

Data can be replicated to improve availability and performance in distributed systems. To ensure one copy serializability, many protocols have been proposed. Most of these protocols are based on the quorum set approach. However, some of them are dominated. Hence, these protocols can be further improved. We devise a paradigm for refining protocols for managing replicated data based on theoretical analysis. Examples are presented to illustrate the usefulness of our methodology.

1. Introduction

Replicated data management has been intensively studied over decades. Recently, several structured quorums have been successfully proposed to reduce the quorum size significantly [1][7][8]. Most of them take the quorum size and the data availability as the main performance measures. In the recent work [3][4], the authors not only present a new quorum-based scheme but also investigate another crucial measure: response time. By using simulation study, they claim that the Grid protocol has significant achievements in load sharing and response time for higher arrival rates as compared to that of voting schemes. Therefore, the Grid protocol is a high performance scheme. Nevertheless, we prove that there exist quorum schemes dominating the Grid quorum scheme.

On the basis of theoretical analysis, we show that dominating schemes exist over the Grid protocol. The similar theorem drawn from the domination relation between coteries [2][5] is used to prove that some members in the quorum sets defined for the Grid scheme are not necessary to the correctness of the protocol. That is, the Grid protocol can be improved. This refinement paradigm can help designer to improve protocols. We also proposed two possible refined quorum set pairs.

2. Basic Concepts

The resources such as data, files, etc., can be replicated in distributed systems to enhance availability as well as response time. However, a protocol must be used to maintain data consistency. The most famous approach is weighted voting proposed by Gifford [6]. Carcia-Molina and Barbara summarized it into a general notion of quorums—quorum agreement [2]. Theoretically, quorum agreements should be used to maintain replicated data because its minimality and completeness.

To make this paper self-contained, we review related concepts of quorum agreement [2]. For convenience, a set of nodes (copies) will be referred to as a group.

Definition 1[2]. Coterie. Let U be the set of copies of a data object. A set of groups, S is a coterie under U iff

1. $G \in S \Rightarrow G \subseteq U \wedge G \neq \emptyset$.
2. (Intersection property) $\forall G, H \in S: G \cap H \neq \emptyset$.
3. (Minimality property) $\forall G, H \in S: G \not\subset H$. ❑

By their intersection property, the members in a coterie can be used as quorums to guarantee mutual exclusion in a distributed system. Now we review the domination relation of quorum sets. Interested readers can refer to [8][9] for more details.

Definition 2[2]. Let S be a coterie under U. S is C-dominated iff there exists a group G such that $U \supseteq G$ and
(i) G is not a superset of any group in S.
(ii) G has the intersection property. That is, $\forall H \in S: G \cap H \neq \emptyset$. ❑

Note that we use the term C-dominated in order to distinguish it from Q-dominated that will be defined later.

Terminology

Quorum set. A quorum set is the set of groups that can complete a critical operation. More formally, let U be the set of copies of a data object. A set of groups, Q, is a quorum set under U iff

1. $G \in Q \Rightarrow G \subseteq U \wedge G \neq \emptyset$.
2. (Minimality property) $\forall G, H \in Q: G \not\subset H$.

Complementary quorum set. Given a quorum set Q, a complementary quorum set Q^c is another quorum set , such that every group in one intersects every group in the other.

Quorum set pair. The set $q=\{Q, Q^c\}$ is called the quorum set pair.

Transversal. A transversal of a quorum set Q under U is defined to be a set $T \subset U$ such that for every group $G \in Q$, $G \cap T \neq \emptyset$. A minimal transversal is a transversal such that no proper subset of it is a transversal.

Antiquorum set. An antiquorum set Q^{-1} is the set of all

1063-6374/93 $03.00 © 1993 IEEE

minimal transversals of a quorum set Q.

Quorum agreement. The set $q=\{Q, Q^{-1}\}$ is called the quorum agreement.

We rewrite Definition 2 as the following theorem so that two rules derived allow us to easily check whether a coterie is C-dominated.

Theorem 1. A coterie S under U is C-dominated iff there exists a group $G \subseteq U$ such that for all $H \in S$, $G \notin S$, $G \cap H \neq \emptyset$ and at least one of the following conditions holds.
(i) $G \subset H$ (Elimination Rule)
(ii) $G \not\subset H$ and $H \not\subset G$. (Addition Rule)

Proof: (i) There are one or more $H_1, H_2, H_3, ..., H_n \in S$ such that $G \subset H_1, H_2, H_3, ..., H_n$. Then $R=(S-H_1-H_2-H_3-...-H_n) \cup \{G\}$ is a coterie. (ii) $R = S \cup \{G\}$ is a coterie. R dominates S.

If S is dominated, there exists R dominating S; that is, $R \neq S$ and, for each $H \in S$, there is a $H \in R$ such that $G \subseteq H$. Two conditions should be considered for $R \neq S$: $S \subset R$ and $S \not\subset R$. $S \subset R$ implies there exists $G \in (R-S)$; condition (ii) holds. $S \not\subset R$ means $S-(S \cap R) \neq \emptyset$ and $R-(S \cap R) \neq \emptyset$(if $R-(S \cap R)=\emptyset$, $R \subset S$ which is contradicted to R dominates S). Let $H_1 \in (S-(S \cap R))$. There must exist $G \in (R-(S \cap R))$ such that $G \subset H_1$. The two conditions in this theorem are necessary and sufficient for dominated coteries. ❑

Definition 3[2]. Domination for Quorum Sets. Let R, S be quorum sets under U. R Q-dominates S iff for each $H \in S$ there is a $G \in R$ such that $G \subseteq H$. ❑

Theorem 2[2]. Given a quorum set Q, any complementary quorum set is dominated by Q^{-1} ❑

In the Grid protocol, the write quorum set is a coterie. However, the read quorum set is only a complementary quorum set of the write quorum set, not a antiquorum set. This observation leads us to investigate how to improve the protocol.

3. Implementation Considerations

Not every quorum agreement is practical to be implemented. The brute force approach of quorum agreements represents the quorum sets as lists. In this scenario, Request_Set_Generating is to select one quorum (or quorums) among all quorums to request their permissions and Quorum_Containment_Test searches the quorum sets to see if reply set contains a quorum. This approach is hardly practical to be implemented because the memory requirement is extremely high even the number of copies is only moderate large. Moreover, the complexity of determining whether there exists a quorum that is subset of reply set is high. Consequently, specific algorithms and data structures that can be efficiently implemented are devised. Some of them seem to be devised with minor consideration of theoretical (quorum level) efficiency constraint— no dominated quorum sets should be used. As a result, the schemes derived may be dominated. Due to this reason, any schemes that are not derived from a sound logical design should be checked to see if there is room to improve.

4. Quorum Set Pairs Refinement

Quorum set pairs of practical protocols may be improved when they are not quorum agreements. That is, they are dominated.

Definition 4. Domination for Quorum set pairs. Given two quorum set pairs, $p = \{W_p, R_p\}$ and $q=\{W_q, R_q\}$ where W_p, W_q are coteries, p dominates q iff $p \neq q$, W_p Q-dominates W_q and R_p Q-dominates R_q. ❑

Notice that Definition 3 may be further relaxed so that W_p and W_q are not required to be coteries. We impose this restriction to reflect the basis of the Grid protocol.

Theorem 3. Let q be a quorum set pair, $p = \{W_p, R_p\}$ where W_p is a coterie. p is dominated iff
(I) Refinement of Wp
There exists a group $G \subseteq U$ such that for all $H \in W_p$, $G \notin W_p$, $G \cap H \neq \emptyset$, and for all $I \in R_p$, $G \cap I \neq \emptyset$, and at least one of the following conditions holds.
(i) $G \subset H$ where there exists an $H \in W_p$ (Elimination Rule)
(ii) $G \not\subset H$ AND $H \not\subset G$ for all $H \in W_p$ (Addition Rule)
(II) Refinement of Rp
There exists a group $G \subseteq U$ such that for all $H \in W_p$, $G \notin R_p$, $G \cap H \neq \emptyset$ and at least one of the following conditions holds.
(i) $G \subset I$ where there exists an $I \in R_p$ (Elimination Rule)
(ii) $G \not\subset I$ AND $I \not\subset G$ for all $I \in R_p$ (Addition Rule)
Proof: It is similar to the proof of Theorem 1. ❑

By using Theorem 3, we can check whether a quorum set pair is dominated as well as whether improvement is possible. Two rules can guide us a possible refinement quorum set pair. One (Elimination Rule) is to explore if there is any unnecessary member in a group; that is, even eliminating that member the quorum sets still remains the correctness. When this is found, we can modify a protocol so that all aspects of the protocol— availability, communication cost, and response time could be enhanced. The other (Addition Rule) is to explore whether there is a nonsuperset group that can be added into the quorum set. This can enhance the availability of quorums but some complexity will be introduced if the new added quorum can not be easily incorporated into the protocol. From the exploring complexity of view, the former rule is relatively easily to perform than the latter one. Hence, we highly recommend that the designer of a quorum-based protocol, especially those derived from implementation view, to check at least the first rule.

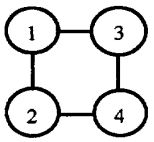

Fig. 1. A two by two grid structure.

Take the quorum set pair of the Grid protocol as an example. Four nodes are organized into a two by two grid structure as Figure 1. The W_p={{1, 2, 3}, {1, 2, 4}, {1, 3, 4}, {2, 3, 4}} and R_p={{1, 3}, {2, 3}, {1, 4}, {2, 4}}. Using the first rule, we check the first group in W_p can be reduced to {1, 2} while the correctness is preserved. Consequently, the quorum set pair can be improved. Now the second group {1, 2, 4} is a superset of {1, 2}. So we can modify the original protocol (Quorum_Containment_Test) such that only when those copies of {1, 2} reply, the update of data object is allowed. Moreover, we modify the original protocol (Request_Set_Generating) such that the protocol sent requests to only {1, 2} instead of {1, 2, 3} and {1, 2, 4}. In this way, all measures such as the availability, the communication cost, and the response time could be enhanced. On the same original example, if we apply the second rule (Addition Rule) to the R_p, we can find that additional quorums {1, 2} and {3, 4} are valid for read quorums. They can be incorporated into the original protocol.

The demonstration above shows that checking a small instance of a class of quorum set pairs allows us to gain insight on the possibility of refining protocols.

5. Two Proposed Refinements

This section proposes two possible refinements of Grid quorum sets. The first one introduces the hierarchy concept. The second one keeps the responsibility of every node as equally as possible. Both quorum set pairs dominate the Grid quorum set pair.

5.1. Review of Grid Quorum Set Pair

The Grid quorum scheme proposed by Cheung [2][3] arranged the nodes storing copies of the data in a logical grid. By randomly choosing nodes to form a quorum, the grid quorum method can effectively share the transaction processing among the nodes storing the data copies and hence provides high data availability as well as low response time. All these benefits in the Grid quorum scheme can be applied to the scheme that uses our proposed quorum set pairs.

To describe the quorum set pairs, we organize the nodes into an M by N grid structure as Figure 2. We denote the columns with a sequence of numbers from left to right. The set of all nodes in column i is denoted as S_i.

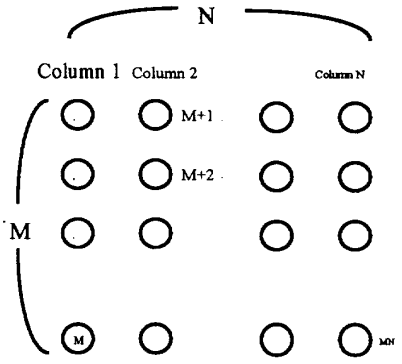

Fig. 2. A grid structure.

Definition 5. C-Cover(i, j). A set G of nodes is a C-Cover(i, j) if from column i to column j each columns intersects with G. ❑

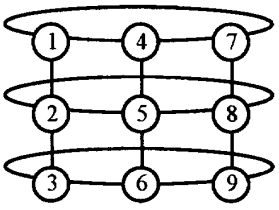

Fig. 3. A wrap-around mesh structure.

For example, in Fig. 3, the set {1, 4, 8} is a C-Cover(1, 3) and the set {6, 7} is a C-Cover(2, 3).

In our notations, the Grid protocol defines a write quorum and a read quorum as a member of the set Q_{rw} and Q_r respectively as follows.

Q_{rw}={C-Cover(1, N) $\cup$ S_i | $1 \le i \le N$}

Q_r={C-Cover(1, N)}

where N denotes the number of columns in the system. For example, the set {1, 2, 3, 4, 7} is a write quorum where the set {1, 5, 8} is a read quorum.

5.2. The Level Quorum Set Pair

A quorum set pair, named as Level quorum pair, dominating the Grid one is described as follows.

Definition 6. The Level quorum set pair is defined as q={Q_{rw}, Q_r} where

Q_{rw}={C-Cover(1, i-1) $\cup$ S_i | $1 \le i \le N$},

Q'_r={C-Cover(1, N)},

$Q_r = Q'_r \cup Q_{rw}$. ❑

It is obvious that the Level quorum set pair dominates the Grid quorum pair. We prove the correctness of the Level quorum set pair.

Theorem 4. The Level quorum set pair meets the requirement of the quorum set pair. (Proof Omitted) ❑

Theorem 5. In Level quorums, when M≥2, $q=\{Q_{rw}, Q_r\}$ is a quorum agreement. (Proof Omitted) ❑

5.3. Wrap-Around Mesh Quorum Set Pair

Another quorum set pair, named as Wrap-Around Mesh quorum pair, dominating the Grid one is described as follows. To describe the new quorum pair, we define another set as LC-cover(l, i) where l denotes how many columns should be covered directly *before* column i. Note that nodes are organized as a horizontally-wrap-around mesh.

Take Fig. 3 as an example. The set {2, 8} is LC-Cover(2, 2). The set {3, 4} is LC-Cover(2, 3). The set {9} is LC-Cover(1, 1}.

Definition 7. The Wrap-Around Mesh quorum set pair is defined as $q=\{Q_{rw}, Q_r\}$ where

$Q_{rw}=\{\text{LC-Cover}((N-1)/2, i) \cup S_i \mid 1 \le i \le N\}$,
$Q'_r=\{\text{C-Cover}(1, N)\}$,
$Q_r= Q'_r \cup Q_{rw}$. ❑

Take Fig. 3 as example, the Q_{rw}={{1, 2, 3, 7}, {1, 2, 3, 8}, {1, 2, 3, 9}, {1, 4, 5, 6}, {2, 4, 5, 6}, {3, 4, 5, 6}, {4, 7, 8, 9}, {5, 7, 8, 9}, {6, 7, 8, 9}}. Any C-Cover(1, N) is a member of Q_r such as {1, 4, 7}, {2, 6, 7}.

Theorem 6. The Wrap-Around Mesh quorum set pair meets the requirement of a quorum set pair.

Proof: (Write-Write Intersection) Those quorums including S_i intersect with those ones including S_j where the difference between i and j is within (N-1)/2 and j is in the left of i due to the C-Cover part. They intersect the remainder quorums due to their S_i part. Hence, the write-write intersection property holds.

(Read-Write Intersection) Because each write quorum includes S_i part and C-Cover(1, N) intersects every S_i, together with write-write intersection the read-write intersection property holds. ❑

When N is odd, N≥3, and M≥2, we believe that the Wrap-Around Mesh quorum set pair is a quorum agreement. If the mesh structure is not the case, some further improvement could be made. Whatever, the Wrap-Around Mesh quorum set pair dominates the Grid quorum set pair.

Theorem 7. The Wrap-Around Mesh quorum set pair dominates the Grid quorum set pair.
Proof: Straight forward. ❑

As can be seen from the above discussion, the level quorums and the Wrap-Around Mesh quorums perform better than the grid quorum method in view of the data availability, communication costs, and response time. Hence, we can conclude that replicated data management protocols derived from implementation view may be further improved by theoretical investigation such as the paradigm shown in section 4.

6. Summary

In summary, we have proposed a paradigm to improve the quorum-based replicated data management schemes. The alternative schemes for replicated data management are derived and their performance is even higher than a previous high performance scheme— the Grid protocol. A theorem based on the domination relation between coteries is used to prove that some members in the quorum sets corresponding to some protocols are not necessary to correctness of the protocols. Useful insights gained in this paper can help designer to improve their quorum-based replicated data management schemes.

References

[1] D. Agrawal and A. El Abbadi, "An efficient and fault-tolerant algorithm for distributed mutual exclusion," ACM Transaction on Computer System, vol. 9, no. 1, pp. 1-20, Feb. 1991.

[2] D. Barbara and H. Garcia-Monina, "Mutual exclusion in partitioned distributed systems," Distributed Computing, vol. 1, pp. 119-132, 1986.

[3] S. Y. Cheung, M. H. Ammar and M. Ahmad, "The grid protocol: a high performance scheme for maintaining replicated data," IEEE Trans. Knowledge and Data Engineering, vol. 4, no. 6, pp. 582-592, Dec. 1992.

[4] S. Y. Cheung, M. H. Ammar and M. Ahmad, "The grid protocol: a high performance scheme for maintaining replicated data," in Proc. Sixth IEEE International Conference on Data Engineering, Los Angeles. CA, pp. 438-445, Feb. 1990.

[5] H. Garcia-Molina and D. Barbara, "How to assign votes in a distributed system," ACM Journal, vol. 32, no. 4, pp. 841-860, Oct. 1985.

[6] D. K. Gifford. Weight voting for replicated data. in Proceeding of 7th ACM SIGOPS Symposium on Operating System Principles, Pacific Grove, CA, 150-159, Dec. 1979.

[7] A. Kumar, "Heirarchical quorum consensus: A new algorithm for managing replicated data," IEEE Transation on Computer, vol. 40, no. 9, pp. 996-1004, Sept. 1991.

[8] M. L. Neilsen and M. Mizuno, "Coterie join algorithm," IEEE Transaction on Parallel and Distributed Systems, vol.3, no. 5, pp. 582-590, Sept. 1992.

[9] D. Shou and S. D. Wang, "A new transformation method for nondominated coterie design," (To appear) Information Sciences: An International Journal.