Bounds on the Effects of Replication on Availability
(Position Paper)

Larry Raab
Transarc Corporation
The Gulf Tower, 707 Grant Street
Pittsburgh, Pennsylvania 15219

Abstract

In this paper we discuss some theoretical limitations on the potential benefits of replication. In particular, we investigate two fundamental questions: (1) does placing copies of data around a network increase the probability that the data will be available, and (2) does such a technique decrease the mean duration of unavailability of the data. Given that many applications require mutually exclusive access to the data, we show that the potential benefits of replication are rather low with respect to both of these metrics. Although these results are not necessarily surprising, it is interesting that the proofs are protocol independent and, in the case of the availability measure, topology independent. They are useful, therefore, in focusing attention on the replica consistency model and performance measures and away from any particular protocol or network configuration.

1 Introduction

Replication has been proposed and widely studied as a method for increasing the availability of resources in the presence of network component faults. Replication is an enticing idea, since a local or nearby site may be reachable despite site and link failures. This is certainly true for read-mostly resources (e.g., telephone books, encyclopedias); however, difficulties arise when the state of the system is altered by the access (e.g., database updates and missile launches). If we require such systems to behave as they would without replication, some consistency constraints must be enforced.

The most common consistency constraint is mutual exclusion, since inconsistency can arise if the resource is modified by more than one site at a time. In particular, the mutual exclusion constraint is frequently used to enforce the 1-copy serializability of update accesses in replicated database systems. The difficulty is ensuring mutual exclusion in a network subject to component faults and consequent partitioning.

In our work, we have found that the consistency constraint of mutual exclusion nearly eliminates many of the benefits of replication in partitionable networks. This is interesting given the volume of work that has appeared in the literature devoted to maximizing availability while ensuring mutual exclusion. Of course, this is not to say that the limitations have not been recognized, especially by practitioners confronted with the costs of replication (e.g., communication overhead, storage increases, and protocol complexity). What follows is a brief description of some of our work, including simulation results and theorems that confirm these limitations. We conclude with a discussion of system models and consistency requirements different from those assumed here, thereby identifying interesting areas for further investigation. (The results cited here were developed jointly with Donald Johnson while at Dartmouth College.)

2 Model

We model a distributed system as a collection of autonomous sites and communication links. Access requests are submitted at random according to a fixed access request distribution. Requests are either granted or denied instantaneously upon submission, based on a protocol in effect at each site. We assume that sites either execute according to the agreed-upon protocol or cease local processing and communication. Each site and each link failure and recover independently according to a Poisson process.

The percentage of requests granted is termed the availability. This availability metric has been termed site availability, as opposed to system availability, which is the probability that there exists some site from which a request can be granted. It is our view
that site availability reports more nearly the availability as experienced by a user of the system, who typically cannot readily move from site to site or have knowledge a priori of which sites are functioning.

We do not model many system realities, including copy reintegration, communication delays, transactions, data contention, and dependent failures. The reason is simple. We are most interested in the trade-off between potential availability increases due to replication and availability decreases due to the consistency constraint. If replication does not perform well in this simple model, it unlikely to prove useful when these other factors are modeled.

3 Simulation Results

A number of protocols that ensure mutual exclusion in such a system have been proposed in the literature; we chose to study the Majority Consensus protocol[8] and the Dynamic Voting protocol[2, 7], since these protocols guarantee mutual exclusion in the model described above. We compared their performance to that of a single copy and to that of an optimal (unrealizable) protocol that knows the future.

Each of the networks that we study is based on a ring of 101 sites. By varying the number of chords added to the ring, we obtained performance values for a number of different levels of connectivity. Each site and each link is 96% available. We submit access requests uniformly (over the sites) at random. In [3], we justify these values and examine a number of other replica control protocols, number of replicas, and placement strategies.

Generally, our simulations showed that

- the Dynamic Voting protocol performs nearly optimally, and
- the performance of a single copy closely tracks that of the Dynamic Voting protocol.

This first observation implies that existing protocols perform nearly as well as can be expected with respect to the availability metric. In fact, the Dynamic Voting protocol produces optimal availability (given a confidence interval of ±5%) for all reasonably reliable topologies (those with single copy availability over 50%) that we tested. The second observation suggests that there may exist a relationship between the performance of a single copy and the performance of the best possible replication scheme. We describe such a relationship in the next section.

4 Analytical Results

The relationship between the availability provided by a nonreplicated data object and that provided by a replicated data object is given below. The proof can be found in [5], where we also prove that this bound is the tightest possible for single copy availabilities greater than 25%. (Recall that we define availability as the probability that a request will be allowed to succeed when submitted to some site according to a specified access request distribution.)

If a well-placed nonreplicated data object provides availability $A$, then no replication scheme can provide availability greater than $\sqrt{A}$ on the same network.

By “well-placed” in this theorem we mean that the object's location provides availability at least as high as the average that would be provided by the other locations. Although we have proven that this problem and a number of related network reliability problems are $\#P$-complete (and, therefore, likely to be computationally tractable only in very small networks), it is not difficult to find a practical approximation method for this problem in real systems.

As an illustration of this theorem, suppose that a database administrator manages a network in which a single well-placed server provides availability 95%. The above theorem implies that no replication scheme can provide availability greater than $\sqrt{0.95} = 97.5%$. This information is useful both for determining whether to attempt replication and for assessing the performance of replication schemes.

5 Duration of Unavailability Results

Using the same model, we have shown that replication does not, in many cases, substantially improve the mean duration of unavailability[4]. By mean duration of unavailability (MDU), we mean the expectation over all periods of unavailability of the length of time, once an access request has first been denied by the governing replica control protocol, that the access request continues to be denied. Simulations using the model described in Section 3 show that

- the MDU incur by the optimal protocol (with respect to availability) and the Dynamic Voting protocol are roughly the same, and
- the Majority Consensus protocol and an optimal protocol with respect to MDU incur significantly lower MDU.
Unfortunately, the availability provided by the latter two protocols is intolerably low.

We have also examined analytically networks with infallible communication that employ the Available Copies replication control protocol[1, 6]. Assuming infallible communication, it is easy to show the following:

Replication using the Available Copies protocol reduces the mean duration of unavailability by at most a factor of \((1 - p)/(1 + p)\).

Here \(p\) is the reliability of a site; that is, the steady-state probability that a site is operational. With \(p\) very close to 1, as is expected of today's network components, this difference is minimal. For example, if the MDU of a single copy is 5 time units and \(p = .99\), then the MDU using the Available Copies protocol is at least \(5 - 5(1 - p)/(1 + p) = 4.975\). A simple absorbing Markov chain analysis shows similarly small improvements for networks with highly reliable, yet fallible, communication.

6 Conclusions

These results show that the benefits of replication are limited when mutual exclusion must be guaranteed. Although this is not necessarily surprising, we prove that the performance limitations are inherent in the problem as commonly defined, rather than an artifact of existing implementations.

Fortunately, replication may still prove valuable by other metrics, in less restrictive applications, or in the presence of real-world constraints. Concerning other metrics, response time, throughput, communication costs, and storage costs are also pertinent performance measures.

Concerning real-world constraints, as global databases grow to include federations of local databases, it is increasingly likely that two autonomous corporations will desire to store the same piece of information within their respective databases. Each corporation may be willing to allow the other to remotely access their data; neither may be willing to relinquish possession. In this case, replication is required, and the question again becomes one of performance. Similarly, it may not be possible to "well-place" a single data object for reasons of security or existing infrastructure.

Also, not all applications require mutual exclusion. In addition to read accesses, a number of loose-consistency models have been proposed for update accesses. The challenge lies in characterizing the behavior of protocols enforcing these new models and in finding applications for which loose consistency is adequate. Since few applications can tolerate permanent "inconsistency", we will need to develop tools for expressing, detecting, and correcting inconsistency.

Finally, the bounds presented here leave some room for availability improvement via replication, even for operations requiring mutual exclusion. Although the improvement is small and comes at a high cost (in terms of protocol complexity, response time and storage costs), these costs are getting smaller and will continue to do so.

The preceding list emphasizes the practical aspects of replication. As can be seen from the results cited in this paper, we are also interested in theoretic issues. One interesting theoretical question that remains is the existence of an optimal on-line replica control protocol for partionable networks.

References