Cache Consistency and Concurrency Control in a Client/Server DBMS Architecture†

Yongdong Wang
Lawrence A. Rowe

Computer Science Division-EECS
University of California
Berkeley, CA 94720

ABSTRACT

This paper examines five application cache consistency algorithms in a client/server database system: two-phase locking, certification, callback locking, no-wait locking, and no-wait locking with notification. A simulator was developed to compare the average transaction response time and server throughput for these algorithms under different workloads and system configurations. Two-phase locking and callback locking dominate no-wait locking and no-wait locking with notification when the server or the network is a bottleneck. Callback locking is better than two-phase locking when the inter-transaction locality is high or when inter-transaction locality is medium and the probability of object update is low. When there is no network delay and the server is very fast, no-wait locking with notification and callback locking dominate two-phase and no-wait locking.

1. Introduction

This paper presents the results of a simulation study of database concurrency control and application program cache consistency algorithms in a distributed computing system. This work was motivated by the recent development of persistent programming languages and object-oriented database systems [3-5, 7, 10, 13-18, 21], and client/server database architectures. In a client/server architecture, the database resides on the server. Objects in the database are accessed by application programs running on client workstations. Objects are cached in the application to reduce the time required to access an object. Consequently, several copies of a shared object can exist in more than one application cache at the same time. Mechanisms must be provided to guarantee that concurrent transactions accessing the cached objects and the database do not interfere. The efficiency of the application caching mechanisms is very important to the performance of the applications and the database system.

Concurrency control and cache consistency must be coordinated since caches are invalidated on transaction boundaries. Object updates become permanent when transactions commit. In this study, we extend concurrency control algorithms to include the consistency check of cached objects. We will use the terms *concurrency control* and *cache consistency* interchangeably.

Several researches have investigated the performance of database concurrency control algorithms. Agrawal, Carey, and Livny (ACL) found that in a centralized DBMS two-phase locking outperformed the immediate-restart and certification algorithms for medium to high levels of resource utilization [1, 2]. Carey and Livny investigated distributed concurrency control algorithms [8]. They found that two-phase locking and certification dominated timestamp ordering and wound-wait. When the CPU cost of sending and receiving messages was low, two-phase locking outperformed certification.

Wilkinson and Neimat studied application cache consistency algorithms in a client/server database system [22]. They proposed two algorithms: (1) cache locking and (2) notify locking. They showed that for both short batch and interactive transactions cache locking is never worse than two-phase locking without caching. Furthermore, notify locking is better than cache locking when server CPU utilization is not high, but it is worse than two-phase locking without caching when the server CPU is saturated. However, they did not study other algorithms such as two-phase locking with caching. And, they did not measure transaction response time which we believe is a very important metric for applications. Lastly, their simulation model did not address the cache replacement problem.

In this paper, we distinguish between caching within a transaction (intra-transaction) and between transactions (inter-transaction). The performance of five inter-transaction algorithms are compared: (1) two-phase locking, (2) certification, (3) callback locking, (4) no-wait locking, and (5) no-wait locking with notification. A simulator similar to the one used by ACL was used to compare average response time and throughput under different workloads and system configurations. Three workloads were included in our experiments: small batch transactions, large batch transactions, and interactive transactions. System configurations included a system where the server was a bottleneck, a fast server system, and a fast server and fast network system.

The remainder of the paper is organized as follows. Section 2 describes the algorithms that we studied. Section 3 describes the structure of our simulation model and the simulation parameters. Section 4 describes two experiments that verify the correctness of our simulator. Section 5 describes the results of our initial performance experiments. Finally, section 6 summarizes our findings.

2. Cache Consistency Algorithms

This section describes the algorithms included in our study. The algorithms are extensions or variations of existing concurrency control algorithms for centralized database systems.

Two-phase locking was included in our study because numerous studies conclude that it outperforms all other algorithms unless unrealistic assumptions are made (e.g., infinite physical resources) [2]. Certification was included because it is used by at least one object-oriented DBMS that has a client/server architecture (GemStone [7]). Moreover, some recently proposed algorithms (e.g., cache locking) incorporate ideas from the certification algorithm.

We assume that all locking algorithms use *in-place updates* and that certification algorithms use *deferred updates*. We also assume that an object must be fetched to the client before it is updated. In other words, all updates are executed on the clients first. Updates are sent to the server either when an updated object is swapped out of the client cache or at transaction commit time. Transactions within a client are executed sequentially. There is at most one active transaction, called the *current transaction*, in a client at any time.

† This research was supported by NSF grant MIP-8715557 and NASA grant NAG 2-530.

There are two models of object caching: caching within a transaction, called *intra-transaction caching*, and caching between transactions, called *inter-transaction caching*. In the case of intra-transaction caching, a client assumes that objects in its cache are invalid at the beginning of each transaction. Objects are fetched into the cache when they are first accessed by the transaction. Intra-transaction caching algorithms are easy to implement, but they cannot benefit from inter-transaction reference locality. In the case of inter-transaction caching, a client must check that objects in the cache are valid when they are accessed. This check can be accomplished in two ways: (1) the client can contact the server to verify object validity when the object is accessed (*check-on-access*) or (2) the server can notify clients whenever an object in the cache is updated by other clients (*notify-on-update*).

The two-phase locking and certification algorithms can be easily extended to support both intra- and inter-transaction caching.[1] In the rest of this section, two algorithms that support inter-transaction caching are described: (1) callback locking and (2) no-wait locking. A potential problem with no-wait locking is the high transaction restart rate. Update notification can be incorporated with no-wait locking to reduce transaction restarts. Thus, we also include a third algorithm, (3) no-wait locking with notification.

2.1. Callback Locking

Callback locking was first used in the Andrew File System to maintain the consistency of cached files [12]. It guarantees the validity of cached objects by retaining locks on them even after a transaction terminates. Therefore, there is no need for the client to contact the server to check object validity or to acquire a lock when a transaction accesses a cached object with the appropriate lock. However, if a transaction accesses a cached object without a retained lock or with the wrong lock (e.g., the transaction wants to update an object that has only a read lock), it will need to get the lock from the server. The server sends a message to all clients that have uncompatible locks on this object requesting them to return the locks. A client releases the lock requested by the server immediately if the object has not been accessed by the current transaction on the client. Otherwise, it waits until the current transaction terminates to release the lock. The server cannot grant the requested lock until all uncompatible locks on the object are released.

Both read and write locks can be retained. However, write locks are more likely to cause incompatibility, and thus more likely to be retrieved by the server before it is used by the next transaction. Consequently, we chose to retain only the read locks.

2.2. No-Wait Locking

In two-phase locking, the client application is blocked while waiting for a response from the server that a cached object is valid. An alternative is to assume that all objects in the cache are valid. The application program continues executing without being blocked when accessing cached objects. If the cached object is valid and the requested lock can be granted, the server does not send any response to the client until a transaction commit request is received. If the cached object is invalid or the requested lock cannot be granted because of a deadlock, the server aborts the transaction and the client must restart it. Note that the client must receive a response from the server before it can commit because it needs to get locks even if all objects are valid.

This algorithm was originally proposed by Gerson for Statice [11, 21]. He called it *optimistic locking* because of the optimistic assumption that cached objects are valid and the requested locks can be granted.

2.3. No-Wait Locking with Notification

In no-wait locking, a transaction can be aborted because it has read an invalid object or it is involved in a deadlock. While nothing can be done about transaction aborts due to deadlocks, notification can be integrated into no-wait locking to reduce the transaction aborts caused by reading invalid cached objects. This is achieved by making the server send the updated object to clients when a cached object is updated by a committed transaction. Client transactions will read valid objects at a later time and subsequently commit. Please note that notification cannot eliminate reading invalid objects because a transaction can still read an invalid object before it receives the updates from the server. Instead of sending updates to clients, the server can also send only an invalidation message to the clients.

[1] For a complete description of two-phase locking and certification algorithms, see the book by Bernstein [6].

3. Simulation Model

This section describes our simulation model. It is based on the ACL Model for a centralized DBMS [1, 2] and the Carey and Livny model for a distributed DBMS [8]. Our model has important extensions to and modifications of the ACL Model. First, it models a client/server DBMS, which is different from a centralized or distributed DBMS. Second, it has client cache managers and a buffer manager in the server. Third, it models subobjects shared by multiple complex objects and object clustering. Fourth, it supports an arbitrary number of interleaved object reads and updates. And fifth, it models inter-transaction object reference locality.

There are three main parts to the simulation model: the database model, the transaction model, and the system model. The database model captures the characteristics of the database, such as the database size and the object structures. The transaction model captures the object reference behavior of transactions in a workload. And, the system model captures the characteristics of the system's hardware and software. The physical structure of the modeled system is shown in Figure 1. It is composed of a database server and *n* clients connected by a network. We assume that there is only one application on each client workstation. Thus, the term *client* refers to both the application and the

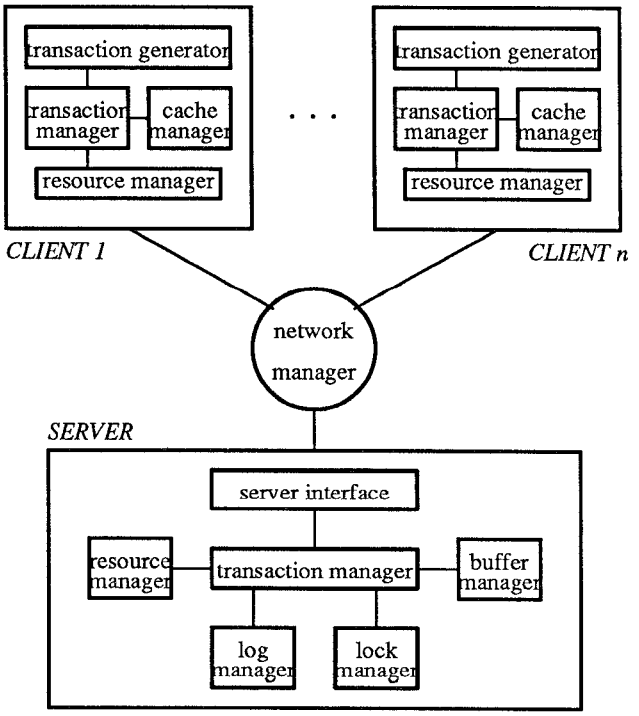


Figure 1. Client/Server DBMS Structure

workstation. A more detailed description of our model can be found in [20].

The simulator is implemented in the simulation language CSIM, which is a C-based simulation language developed at MCC [19]. The core of the simulator consists of about 5000 lines of CSIM code. Each cache consistency algorithm adds an additional 500 to 1000 lines of code, depending on the complexity of the algorithm.

3.1. Database Model

A database is composed of several *classes* (or *relations*). A class is a collection of *objects* (or *tuples*) that have the same attributes. The *size* of an object is the number of pages contained in that object. Multiple objects can share pages. Degrees of object sharing can be modeled by varying the number of pages and the object size in a class. Database parameters are summarized in Table 1. *ClusterFactor* is the probability that consecutive pages in an object are stored sequentially on the disk. Pages are only shared by objects from the same class. Navigation between objects of different classes is represented in the transaction model described in the next section.

3.2. Transaction Model

The transaction model supports the following operations:

(1) BeginXact: start a new transaction,

(2) ReadObject: read an object into the client cache,

(3) UpdateObject: update an object in the client cache,

(4) UserDelay: delay by the application,

(5) CommitXact: commit a transaction,

(6) AbortXact: abort a transaction.

The operation *UserDelay* is included to model non-database processing time by the application and interactive applications.

A transaction is modeled by a finite loop of *ReadObject* and *UpdateObject* operations, as shown in Figure 2. A simulation run can simulate transactions belonging to the same type, or a mix of transactions belonging to different types. Table 2 summarizes parameters that characterize a transaction type. The number of *ReadObject* operations in a transaction is called the *transaction size*, which is uniformly distributed between *MinXactSize* and *MaxXactSize*. The *UpdateObject* operation updates pages of the object read by the preceding *ReadObject* operation. The parameter *ProbWrite* is the probability that each page of the object is updated.[2] The delay parameters (i.e., *UpdateDelay*, *InternalDelay*, and *ExternalDelay*) are exponentially distributed delay times used to model interactive transactions.

To model inter-transaction object reference locality, we introduce a new concept, *InterXactSet*, which contains the last x objects read by the most recent transactions where x is the value of the parameter *InterXactSetSize*. The parameter *InterXactLoc* specifies the probability that an object read by the current transaction is in the *InterXactSet*. The larger the *InterXactLoc* is, the more overlap the read sets of consecutive transactions have, and the higher the inter-transaction object reference locality. We chose *InterXactSetSize* as a simulation parameter so that we can adjust it according to the size of the client cache to make sure that objects in the *InterXactSet* can almost always be found in the cache.

Parameter	Description
NClasses	Number of classes in the database
NPages[i]	Number of pages in class *i*
ObjectSize[i]	Size of the objects in class *i*
ClusterFactor	Probability that an object is clustered

Table 1. Database Parameters

```
BeginXact
        dotimes (transaction_size)
                ReadObject
                UserDelay (UpdateDelay)
                UpdateObject
                UserDelay (InternalDelay)
        end dotimes
CommitXact
```

Figure 2. A General Transaction Model

Parameter	Description
MinXactSize	Min number of ReadObject operations in a transaction
MaxXactSize	Max number of ReadObject operations in a transaction
ProbWrite	Probability that a page in an object is updated
UpdateDelay	Avg delay between a ReadObject and an UpdateObject
InternalDelay	Avg delay for each loop in a transaction
ExternalDelay	Avg delay between two transactions
InterXactSetSize	Number of objects in InterXactSet
InterXactLoc	Prob. that an object read belongs to InterXactSet

Table 2. Parameters for A Transaction Type

3.3. System Model

The system model consists of the network manager, the resource manager, and the other client and server modules. The parameters for all the modules are summaried in Table 3.

The network manager models communication among clients and between clients and the server. Messages are assumed to be transferred in packets, with each packet incurring certain

Parameter	Description
NetDelay	Average message delay time on the network
PacketSize	Maximum number of bytes in a message packet
MsgCost	CPU cost of sending/receiving a message packet
NClients	Number of clients
NClientCPUs	Number of CPU's on a client
ClientMips	Speed of each client CPU
NServerCPUs	Number of CPU's on the server
ServerMips	Speed of each server CPU
NDataDisks	Number of data disks on the server
NLogDisks	Number of log disks on the server
CacheSize	Number of pages in a client cache
BufferSize	Number of pages in the server buffer pool
SeekLow	Minimal disk seek time
SeekHigh	Maximum disk seek time
DiskTran	Transfer time for one disk block
PageSize	Disk block size
InitDiskCost	CPU cost of initiating a disk access
ServerProcPage	CPU cost of processing a page on the server
ClientProcPage	CPU cost of processing a page on the client
MPL	Max number of active transactions allowed

Table 3. System Parameters

[2] This implies that the write set of a transaction is always a subset of its read set.

CPU overhead at both the sending and receiving sites and a certain delay on the network.

The resource managers on the clients and the server manage physical resources. We assume that all clients are diskless. The server has a number of data disks and log disks, all with the same physical characteristics. We separate disk seek time (including rotation time) and data transfer time so we can model sequential disk accesses. Disk seek time is uniformly distributed between *SeekLow* and *SeekHigh*. We assume that classes are uniformly distributed to the data disks. All objects in one class reside on the same disk. *PageSize* is the size of a disk block. For convenience, we assume that it is also the size of memory pages. *InitDiskCost* is the number of instructions executed to initiate a disk access. Disks and CPUs use an FCFS policy.

The client transaction generator generates user transactions. The client cache manager uses an LRU replacement policy. When an object is replaced from the cache, a function in the transaction manager is called to perform algorithm dependent actions. The cache hit ratio is determined by the database size, the client cache size, and the object access pattern. The client transaction manager executes user transactions. If a transaction aborts, it restarts the same transaction again and again until it finally commits. This module is the only client module that is algorithm dependent.

The buffer manager on the server is responsible for managing the buffer pool on the server. It implements an LRU replacement policy. None of the previous simulation models had an explicit buffer manager [2, 8, 22]. We believe that the addition of a buffer manager in the simulator can allow us to model system behavior more accurately. The server transaction manager models the execution of transactions on the server. It is the only server module that is algorithm dependent. The lock manager implements locking for the lock-based algorithms. The log manager models a log-based recovery scheme. The parameter *MPL* is the multiple programming level which is the maximum number of active transactions allowed on the server. It can be varied to model server contention.

4. Simulator Verification Experiments

This section describes two experiments that we performed to verify the simulator. In the first experiment, the simulation parameters were set to those of the ACL experiments and two-phase locking and certification algorithms were simulated [2]. Since our simulator was designed for a client/server DBMS architecture, we had to make a few changes in order to compare our results with their results which were obtained on a centralized DBMS. We found that two-phase locking dominated certification which matches the limited resource case (i.e., 1 CPU and 2 data disks) reported by ACL [2].

In the second experiment, we compared the performance of intra- and inter-transaction caching algorithms for both two-phase locking and certification. This experiment was performed for two reasons. First, we wanted to check that inter-transaction caching algorithms performed better when inter-transaction locality was high. Second, since Wilkinson and Neimat only compared their algorithms against two-phase locking with intra-transaction caching, we wanted to find out the magnitude of difference between the intra- and inter-transaction caching algorithms to determine the effect of their assumption.

Table 4 shows the values of the simulation parameters used for this experiment and the experiments reported in the next section unless noted otherwise. The database has 40 classes, each with 50 pages. Since each page is 4K bytes, the database has 8M bytes of data.[3] All objects contained a single page.[4] The transactions executed were short transaction reading an average of 8 objects. There was no internal delay within transactions. The server had a 400 page buffer pool and each client had a 100 page cache.

Parameter	Value
NClasses	40
NPages[i]	50
ObjectSize[i]	1
ClusterFactor	1.0
MinXactSize	4
MaxXactSize	12
ProbWrite	0.0, 0.2, and 0.5
UpdateDelay	0
InternalDelay	0
ExternalDelay	1 second
InterXactSetSize	20
InterXactLoc	0.05, 0.25, 0.50, and 0.75
NetDelay	2 milliseconds
PacketSize	4096 bytes
MsgCost	5,000 instructions
NClients	2, 10, 30, and 50
NClientCPUs	1
ClientMips	1
NServerCPUs	1
ServerMips	2
NDataDisks	2
NLogDisks	1
CacheSize	100 pages
BufferSize	400 pages
SeekLow	0 milliseconds
SeekHigh	44 milliseconds
DiskTran	2 milliseconds
PageSize	4096 bytes
InitDiskCost	5,000 instructions
ServerProcPage	10,000 instructions
ClientProcPage	20,000 instructions
MPL	50

Table 4. Simulation Parameter Setting

The parameters varied were: *NClients*, *ProbWrite*, and *InterXactLoc*. *NClients* determined the server resource contention rate. *ProbWrite* and *InterXactLoc* modeled the user workload. *ProbWrite* determined the database object contention rate. *InterXactLoc* determined the inter-transaction reference locality, called locality hereafter, for consecutive transactions. Response times and throughput results are given in seconds and transactions committed per second, respectively.

Our experiment results show that there is very little difference between algorithms when the locality and write probability are low. Inter-transaction algorithms dominate the corresponding intra-transaction algorithms when locality is high. Two-phase locking algorithms dominate the certification algorithms when the write probability is high and there are a large number of clients [20]. These results match our expectations.

Wilkinson and Neimat compared other algorithms against two-phase locking with intra-transaction caching. Because two-phase locking with inter-transaction caching is 12% to 30% better when locality is high, we will use two-phase locking with inter-transaction caching, called two-phase locking hereafter, as a benchmark against which other algorithms are compared.

5. Experiments and Results

This section describes a set of experiments that summarize our initial findings. We compared the performance of callback locking, no-wait locking, and no-wait locking with notification against that of two-phase locking under the following conditions: (1) when small transactions were executed, (2) when large transactions were executed, (3) when the server had a very fast CPU, (4) when the server had a very fast CPU and there was no network

[3] This is a small database. However, the results do not depend on database size rather on the degree of resource and data contention. A larger database will allow more clients for the same degree of data contention.

[4] We did not study the impact of large objects or object clustering in our initial experiments.

delay, and (5) when interactive transactions were executed. The experiment parameters that varied from the values in Table 4 are given in the description of the experiments.

5.1. Short Transaction Experiment (SXACT)

In this experiment, short transactions were executed that read an average of 8 objects without any internal delay.

The transaction response time is determined by the following factors: (1) the client CPU cost, (2) the network delay caused by messages exchanged between clients and the server, (3) the server CPU and disk I/O cost, and (4) waiting time due to data contention. In this experiment, the server becomes a bottleneck when there are a large number of clients. Therefore, algorithms that can reduce server load should win. Algorithms that cause more transaction aborts lose.

When the locality is low (*InterXactLoc* set to 0.05), none of the algorithms appears to be better than two-phase locking regardless of the write probability [20]. The four algorithms behave virtually the same when transactions are read-only. This is because they are all variations of two-phase locking, and these variations make no difference in this case. When the write probability is higher, callback locking, no-wait locking, and no-wait locking with notification perform worse than two-phase locking because they cause more network traffic or transaction restarts but cannot benefit from cached objects because of the low locality. Thus, two-phase locking is the best.

Figures 3(a) through 3(c) show the average transaction response time with medium locality (*InterXactLoc* set to 0.25). Callback locking performs a little better than two-phase locking when there are no object updates, as shown in Figure 3(a). In this case, when a client accesses a cached object with a retained lock, it does not need to send a message to the server, which saves client waiting time and reduces the server CPU load which in turn improves the response time of other transactions. But its performance degrades for write probabilities that are higher, as shown in Figure 3(c). Thus, callback locking is best when transactions are read-only. Otherwise two-phase locking is the best.

When the locality is high (*InterXactLoc* set to 0.50), Callback locking outperforms two-phase locking by about 15% when there are no object updates. No-wait locking also outperforms two-phase locking because the server does not need to respond to every client message which reduces the server load. When write probability is 0.50, two-phase and callback locking have roughly the same performance.

Figures 4(a) through 4(c) show the average transaction response time for very high locality transactions (*InterXactLoc* set to 0.75). Callback locking dominates the other three algorithms. When there are no object updates, as shown in Figure 4(a), callback locking performs about 35% better than two-phase locking and 17% better than no-wait locking. No-wait locking also outperforms two-phase locking, because clients do not have to wait for responses from the server when they access cached objects which saves clients waiting time and reduces server CPU contention.

However, the performance of callback and no-wait locking degrades quickly for higher write probabilities, as shown in Figures 4(b) and 4(c). When the write probability is 0.2, callback locking only performs about 15% better than two-phase locking while no-wait locking performs about the same as two-phase locking. When the write probability is 0.5, callback locking is just slightly better than two-phase locking while no-wait locking performs worse than two-phase locking because it causes more transaction aborts. Thus, callback locking is the best when locality is high.

An interesting observation from these figures is that no-wait locking with notification rarely performs better than no-wait locking. When there are no object updates, they are the same, as shown in Figures 3(a) and 4(a). Notification is very sensitive to the server CPU load which is determined by the number of clients. With 2 clients, notification does not make a difference because there is only one other client to notify and there are few transaction aborts. When there are a large number of clients (30 to 50), the server is a bottleneck, so sending updates to clients causes more contention which makes notification not worthwhile even though it can reduce the number of transaction aborts. This is most obvious when locality is low, as shown in Figures 3(b) and 3(c). When the locality is very high (*InterXactLoc* set to 0.75), clients access cached objects most of the time. Consequently, the server becomes less loaded and more transaction aborts can be avoided by notification. Even when there are a large number of clients, no-wait locking with notification is only slightly worse than no-wait locking, as shown in Figures 4(b) and 4(c),

We also measured transaction throughput rate for all cases. Figures 5(a) and 5(b) show the transaction throughput rate for medium and very high locality, both with medium write probability. The corresponding response time is shown in Figures 3(b) and 4(b). Transaction throughput results show similar relative performance of the algorithms.

In summary, we conclude that in the setup of our experiments where server CPU is the bottleneck and the workload is short transactions without internal delay:

(1) Two-phase locking and callback locking dominate no-wait locking and no-wait locking with notification.

(2) Callback locking is better than two-phase locking when locality is high (greater than 0.5), or, it is better when locality is medium (between 0.25 and 0.5) and write probability is low (less than 0.2).

(3) No-wait locking is better than two-phase locking when locality is high (greater than 0.5) and write probability is low (less than 0.2). However, callback locking is better than both of them in these cases.

(4) No-wait locking with notification is slightly better than no-wait locking only when the locality is very high (greater than 0.5) and there are a small number of clients. It does not help under other circumstances.

Figure 6 summarizes our findings. In the upper-left corner, it doesn't make any difference which algorithm is used. In the lower-left area, callback locking should be used. In the remaining area, two-phase locking should be used.

5.2. Large Transaction Experiment (LXACT)

In this experiment, we studied the impact of large transactions on the performance of cache consistency algorithms. *MinXactSize* was set to 20 and *MaxXactSize* to 60. Each transaction read an average of 40 objects.

Compared to the previous experiment, the server became saturated with even less clients because there were more operations within a transaction. Transaction aborts became more

Locality \ WriteProb	0.00	0.20	0.50
0.05	any algorithm	two-phase locking	two-phase locking
0.25	callback locking	two-phase locking	two-phase locking
0.50	callback locking	callback locking	two-phase locking
0.75	callback locking	callback locking	callback locking

Figure 6. Summary of Algorithm Performance

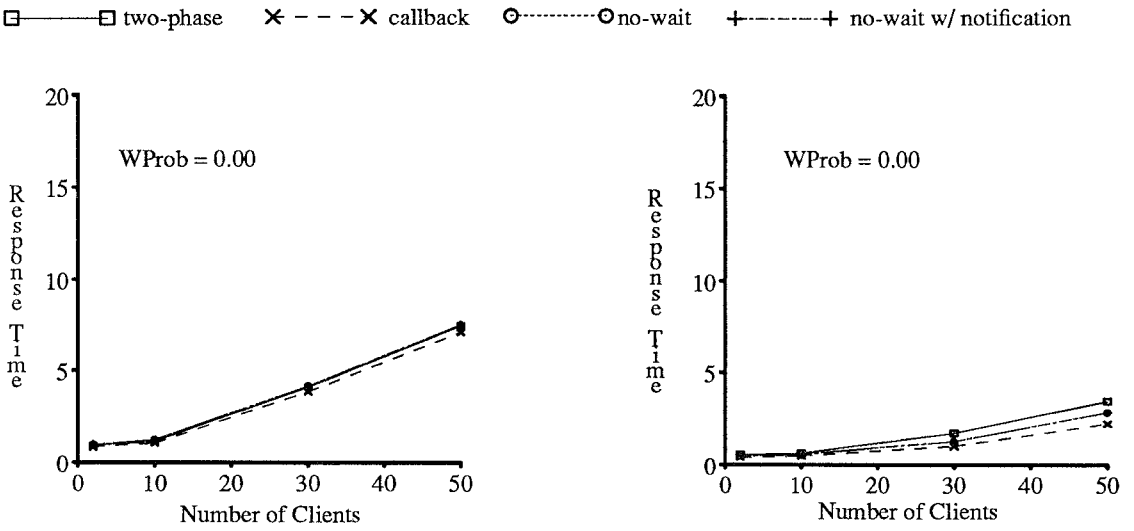


Figure 3(a). SXACT (Loc = 0.25)

Figure 4(a). SXACT (Loc = 0.75)

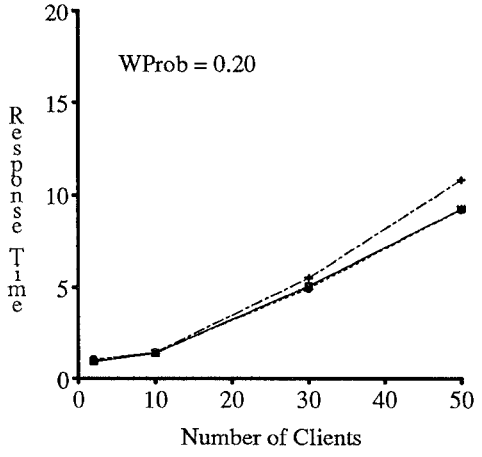


Figure 3(b). SXACT (Loc = 0.25)

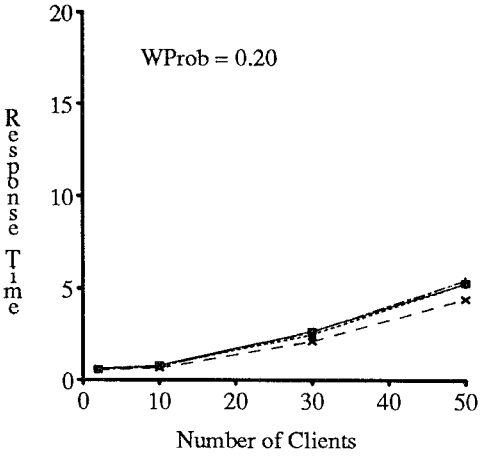


Figure 4(b). SXACT (Loc = 0.75)

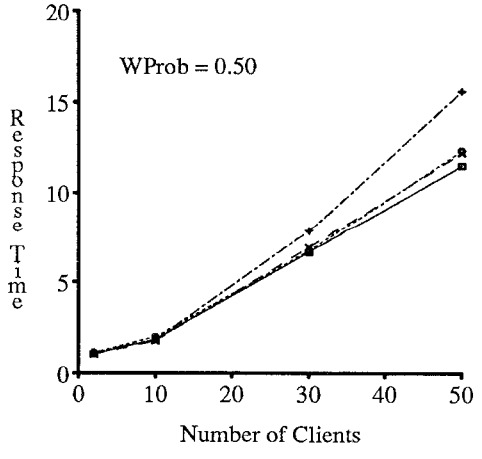


Figure 3(c). SXACT (Loc = 0.25)

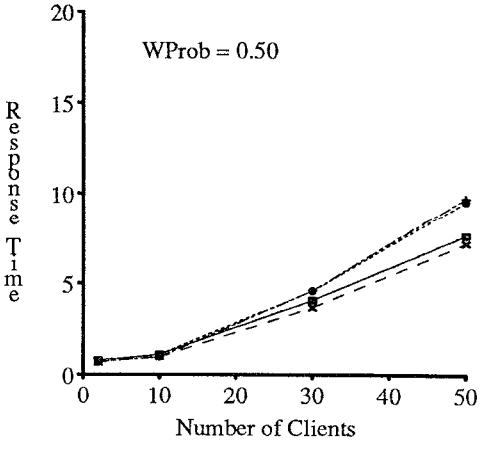


Figure 4(c). SXACT (Loc = 0.75)

expensive due to the larger transaction size and the server load. Figures 7(a) and 7(b) show the transaction response time for very high locality (*InterXactLoc* set to 0.75).

There are three interesting observations. First, results of this experiment are similar to those of the previous experiment because the server is still a bottleneck. Second, no-wait locking degrades more rapidly for higher write probabilities than in the short transaction experiment because it causes more transaction aborts that are more expensive. When there are no object updates, both callback and no-wait locking outperform two-phase locking when locality is above medium. But with high write probabilities, as shown in Figures 7(b), callback locking became slightly worse than two-phase locking, and no-wait locking became much worse than two-phase locking. Third, notification helps no-wait locking, as shown in Figure 7(b), because it reduces the number of transaction aborts. The savings of the reduced transaction aborts more than compensate for the cost of notification messages. However, both no-wait locking and no-wait locking with notification are still dominated by two-phase and callback locking.

5.3. Fast Server Experiment (FSERVER)

The server was the bottleneck in the previous two experiments. In this experiment, we set the server CPU speed to 20 mips, 10 times faster than in the previous experiments. All other parameters were the same as reported in Table 4. The workload was short transactions without internal delay.

In this experiment, we found that the system bottleneck shifted from the server CPU to the network. The network became a bottleneck when the number of clients reached 30. Therefore, algorithms that can reduce the number of messages exchanged between clients and the server should have better performance. Figures 8(a) and 8(b) show the transaction response time for very high locality (*InterXactLoc* set to 0.75). They are very close to the corresponding figures for the short transaction experiment (4(b) and 4(c)). The only obvious difference is that no-wait locking with notification performs more poorly with a large number of clients due to the large number of notification messages.

This experiment gave virtually the same results as the short transaction experiment because there is a close relationship between the number of messages and the server load. Both the network and the server are resources for which clients compete. More messages increase both network and server contention. Therefore, algorithms that perform poorly in the short transaction experiment because of server contention perform equally poorly in this experiment because of network contention.

5.4. Fast Network and Fast Server Experiment (FNET)

In this experiment, we wanted to find out what would happen if the network bottleneck was removed. We assumed that the network was infinitely fast and set *NetDelay* to 0. Server CPU speed was set to 20 mips. All other parameters were the same as in Table 4. A short transaction workload without internal delay was executed.

In this experiment, there was no bottleneck. The resource with the highest degree of contention was the server disks which reached a utilization rate of about 80% with 50 clients. Figures 9(a) and 9(b) show the transaction response time for medium locality (*InterXactLoc* set to 0.25). Figures 10(a) and 10(b) show the transaction response time for very high locality (*InterXactLoc* set to 0.75).

There are substantial differences between the results of this experiment and those of the short transaction experiment. Comparing Figures 9(b) and 10(b) to their corresponding Figures 3(c) and 4(b), we find that no-wait locking with notification performs significantly better. It dominates other algorithms in both cases. In previous experiments, notification performed poorly because it sent more messages than all other algorithms and messages were expensive. The cost of messages was more than the savings due to the reduced transaction aborts. However, in this experiment, with *NetDelay* set to 0 and a very fast server CPU, messages are very cheap while disk I/O's become relatively more expensive. Sending updates to clients can reduce the number of transaction aborts and the number of disk accesses because clients do not need to fetch invalid objects from the server which may require disk reads. Therefore, no-wait locking with notification outperforms the other three algorithms.

When locality is high and no object updates, as shown in Figure 10(a), callback performs slightly better than the other algorithms (this is not obvious in the figure because of the scale). This is because notification does not make a difference in this case.

In summary, when locality and write probability are low, all algorithms behave virtually the same. When locality is high and write probability is low, callback locking performs best. In other situations, no-wait locking with notification performs the best.

5.5. Interactive Transaction Experiment (INTER)

In this experiment, we studied the impact of interactive transactions on the performance of the algorithms. *UpdateDelay* was set to 5 seconds and *InternalDelay* to 2 seconds. All other parameters were the same as reported in Table 4.

Because of the long internal delay and the limited number of clients, all resources are lightly used. Time saved from less communication with the server or not waiting for responses from the server is trivial compared with the internal delay time. Therefore, differences in transaction response time are mainly caused by data contention, or the number of restarts per transaction.

Figure 11(a) and 11(b) shows the transaction response time for medium locality (*InterXactLoc* set to 0.25). When there are no object updates, as shown in Figure 11(a), the response times for all algorithms are flat because they are all essentially determined by the internal delays within transactions which average 56 seconds (7 seconds for each object read and each transaction reads an average of 8 objects). However, with high write probability (*WriteProb* set to 0.50), algorithms that cause more transaction aborts perform poorly, as shown in Figure 11(b). The poor performance of callback and no-wait locking shown in Figure 11(b) are also related to the way they are implemented in our simulator. In both algorithms, clients receive asynchronous messages from the server. In the current implementation, these messages are not processed during the internal delay time.

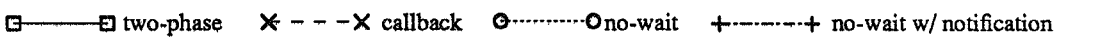
In summary, when there are no object updates, the algorithms behave virtually the same. When the write probability is non-zero, two-phase locking is best because it has the least number of transaction aborts.

6. Conclusions

Five algorithms for inter-transaction caching were described and a simulation experiment was performed to compare their performance under various conditions. We found that two-phase locking and certification with inter-transaction caching almost always outperform the corresponding intra-transaction algorithms. Among the inter-transaction algorithms, when the network or the server is a bottleneck, two-phase locking and callback locking dominate no-wait locking and no-wait locking with notification. Callback locking is better than two-phase locking when the inter-transaction locality is high or when the inter-transaction locality is medium and the probability of object update is low.

When there is no network delay and the server is very fast, no-wait locking with notification and callback locking dominate two-phase and no-wait locking. Callback locking is better than no-wait locking with notification when inter-transaction locality is high and write probability is low. Otherwise no-wait locking with notification is better. For interactive transactions, two-phase locking is the best.

Current database applications typically have low inter-transaction locality and low to medium probability of object update, thus two-phase locking is used. It remains to be seen



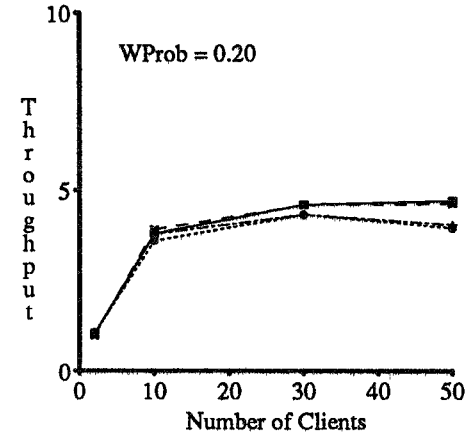


Figure 5(a). SXACT (Loc = 0.25)

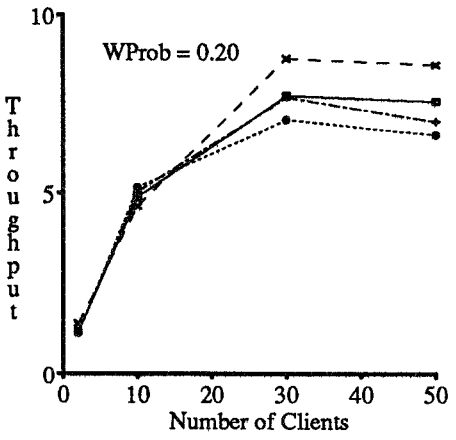


Figure 5(b). SXACT (Loc = 0.75)

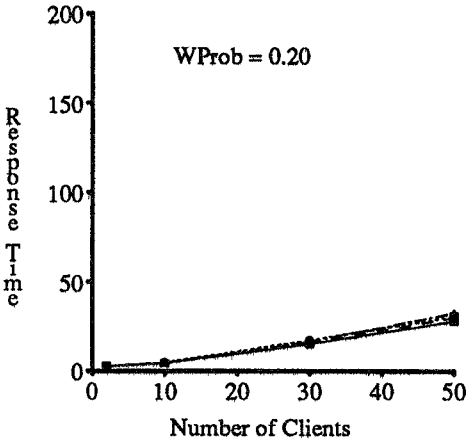


Figure 7(a). LXACT (Loc = 0.75)

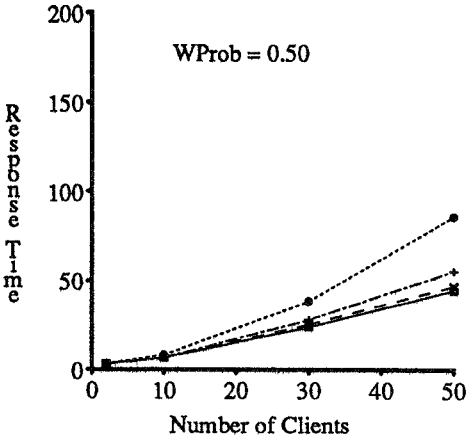


Figure 7(b). LXACT (Loc = 0.75)

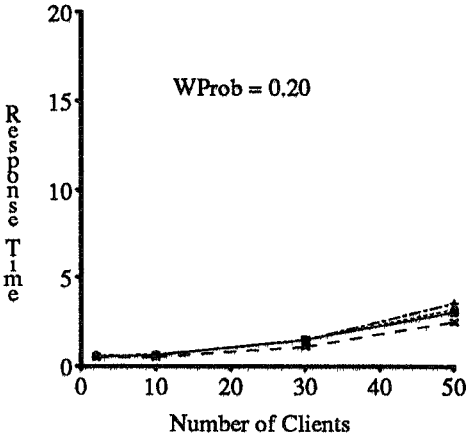


Figure 8(a). FSERVER (Loc = 0.75)

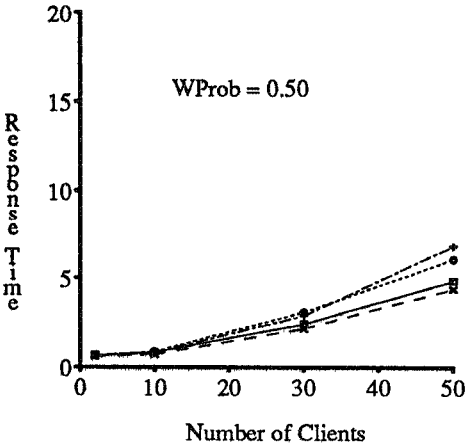
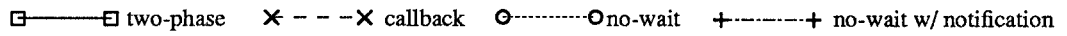


Figure 8(b). FSERVER (Loc = 0.75)



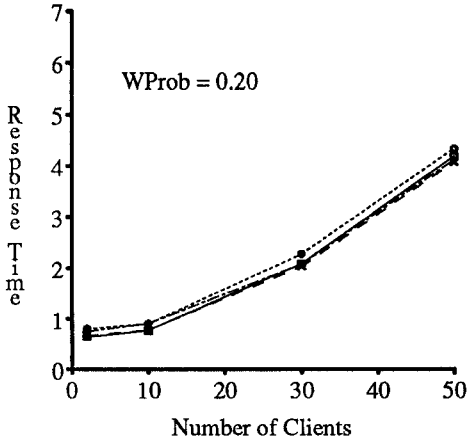


Figure 9(a). FNET (Loc = 0.25)

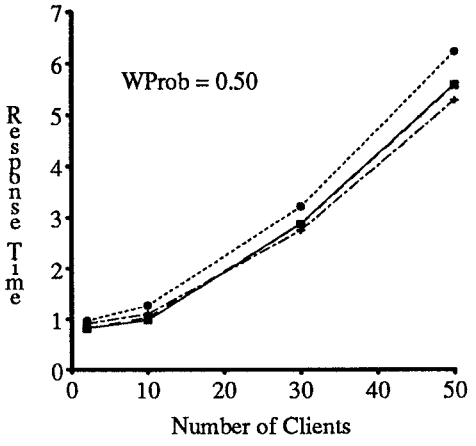


Figure 9(b). FNET (Loc = 0.25)

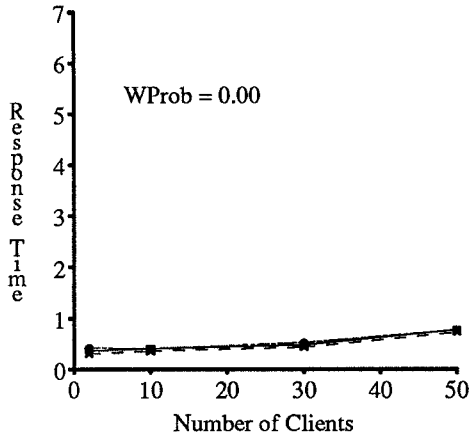


Figure 10(a). FNET (Loc = 0.75)

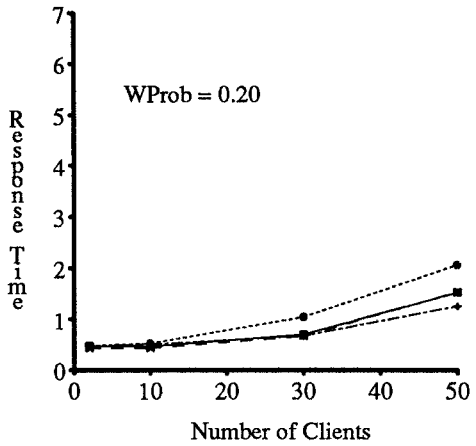


Figure 10(b). FNET (Loc = 0.75)

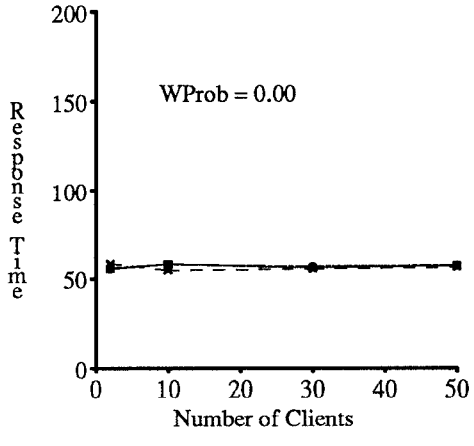


Figure 11(a). INTER (Loc = 0.25)

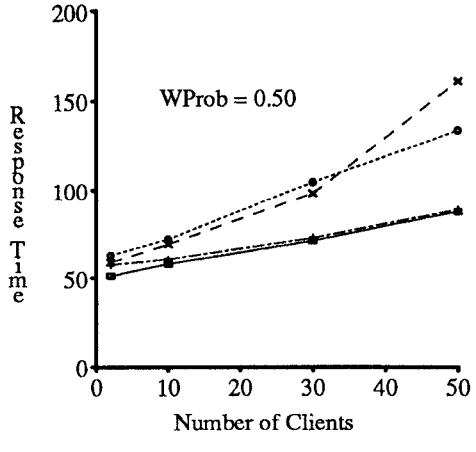


Figure 11(b). INTER (Loc = 0.25)

whether object-oriented DBMS applications have either high inter-transaction locality or medium inter-transaction locality and low write probability to justify changing to callback locking, or networks and server processing speed become fast enough to justify changing to no-wait locking with notification.

Another factor that has to be taken into account is how difficult it is to implement these algorithms. Two-phase locking with inter-transaction caching requires that the server maintain a version number for each object. The version number is cached with the object in each client cache. A client also needs to remember which cached objects have been locked by the current transaction. All messages between clients and the server are synchronous (i.e., all messages are initiated by clients and clients always wait for responses from the server). Therefore, the communication protocol should be easy to implement.

Callback locking requires more modifications. It is still necessary to have a version number for each object.[5] A client must remember which cached objects have been locked and whether they are locked by the current transaction. It also must process asynchronous messages from the server requesting the release of locks. The server lock manager needs to maintain a potentially much larger lock table. Deadlock detection becomes more difficult because if all retained locks are considered part of the wait-for graph, potentially many more deadlocks can happen. On the other hand, if retained locks are not considered part of the wait-for graph, when the server cannot get back a retained lock, it has to decide whether to continue waiting for the client to release it or to abort the client transaction.

No-wait locking also requires more modification than two-phase locking. It requires that a client remember which cached objects it has asked the server to lock but has not received a negative response. It also must handle asynchronous messages from the server requesting that a transaction be restarted. Notification requires that the server remember which objects have been cached by which clients if it sends updates to individual clients instead of broadcasting them to all clients.

It is possible for a system to provide more than one cache consistency algorithms and dynamically switch between them depending on the characteristics of the applications. However, if a single algorithm is desired, two-phase locking seems to be a reasonable choice. It provides stable performance for different workloads and system configurations. In cases where it is outperformed, it is not much worse than the best performer.

Acknowledgement

We want to thank Mike Stonebraker for his advice on the simulation model and experiments, John Ousterhout for suggesting that we study callback locking, Luis Miguel, Wei Hong, Young-Chul Shim, Kevin Wilkinson for their constructive comments on our work, and Dan Gerson and Dan Weinreb for elaborating on the no-wait locking algorithm in private communication.

References

1. Agrawal, R., Carey, M. J. and Livny, M., "Models for Studying Concurrency Control Performance: Alternatives and Implications", *Proceedings of SIGMOD*, 1985.
2. Agrawal, R., Carey, M. J. and Livny, M., "Concurrency Control Performance Modeling: Alternatives and Implications", *TODS 12*, 4 (December 1987).
3. Agrawal, R. and Gehani, N. H., "ODE (Object Database and Environment): The Language and the Data Model", *Proceedings of SIGMOD*, 1989.
4. Andrews, T. and Harris, C., "Combining Language and Database Advances in an Object-Oriented Development Environment", *Proceedings of OOPSLA*, 1987, 430-440.
5. Bancilhon, F., et. al., "The Design and Implementation of O2, an Object-Oriented Database System", *Proceedings of the 2nd Intl. Workshop on Object-Oriented Database Systems*, September, 1988.
6. Bernstein, P. A., Hadzilacos, V. and Goodman, N., *Concurrency Control and Recovery in Database Systems*, Addison-Wesley, 1987.
7. Bretl, R., et. al., "The GemStone Data Management System", in *Object-Oriented Concepts, Databases, and Applications*, Kim, W. and Lochovsky, F. H. (editor), 1989, ACM Press.
8. Carey, M. J. and Livny, M., "Distributed Concurrency Control Performance: A Study of Algorithms, Distribution, and Replication", *Proceedings of VLDB*, 1988.
9. Date, C. J., *An Introduction to Database Systems, Volumn I, 5th Edition*, Addison-Wesley, 1990.
10. Deux, O., et. al., "The Story of O2", *IEEE Transactions on Knowledge and Data Engineering 2*, 1 (March 1990), 91-108.
11. Gerson, D., Personal Communication, March 1989.
12. Howard, J. H., et. al., "Scale and Performance in a Distributed File System", *ACM Transactions on Computer Systems 6*, 1 (February 1988), 51-81.
13. Kempf, J. and Snyder, A., "Persistent Objects on a Database", STL-86-12, HP Labs, September, 1986.
14. Kim, W., et. al., "Architecture of the ORION Next-Generation Database System", *IEEE Transactions on Knowledge and Data Engineering 2*, 1 (March 1990), 109-124.
15. Object Design, Inc., ObjectStore DBMS, 1990.
16. Objectivity, Inc., Objectivity/DB, 1990.
17. Paepcke, A., "PCLOS: A Flexible Implementation of CLOS Persistence", *European Conference on Object-Oriented Programming*, 1988.
18. Richardson, J. E. and Carey, M. J., "Programming Constructs for Database System Implementation in EXODUS", *Proceedings of SIGMOD*, 1987.
19. Schwetman, H. D., "CSIM: A C-Based, Process-Oriented Simulation Language", *Proceedings of the 1986 Winter simulation Conference*, December 1986, 387-396.
20. Wang, Y., "Cache Consistency and Concurrency Control in a Client/Server DBMS Architecture", ERL M90/120, UC Berkeley, December 1990.
21. Weinreb, D., et. al., "An Object-Oriented Database System to Support an Integrated Programming Environment", *IEEE Database Engineering Bulletin 11*, 2 (June 1988), 33-43.
22. Wilkinson, K. and Neimat, M., "Maintaining Consistency of Client-Cached Data", *Proceedings of VLDB*, 1990.

[5] The version number is not necessary if a client assumes that all unlocked objects are invalid and always refetches them.