Differential Synchronization

by Neil Fraser, January 2009

 Read this as a DocEng paper.

 Watch this as a Google Tech Talk.

Keeping two or more copies of the same document synchronized with each other in real-time is a complex challenge. This paper describes the differential synchronization algorithm. Differential synchronization offers scalability, fault-tolerance, and responsive collaborative editing across an unreliable network.

---

1 Conventional Strategies

The three most common approaches to synchronization are ownership, event passing and three-way merges. These methods are conceptually simple, but all have drawbacks.

Locking is the simplest technique. In its most basic form, a shared document may only be edited by one user at a time. A familiar example is Microsoft Word's behaviour when opening a document on a networked drive.[?] The first user to open the document has global write access, while all others have read-only access. This does not allow for real-time collaboration by multiple users.

A refinement would be to dynamically lock and release subsections of the document. However this still prevents close collaboration. Subsection locking also restricts editability when the document is small. Furthermore, support for fine-grained locking would have to be explicitly built into the application. Finally, locking is ill-suited for operation in environments with unreliable connectivity since the lock or unlock signals can get lost, leaving no owner.

Event passing is also a simple technique. It relies on capturing all user actions and mirroring them across the network to other users. Algorithms based on Operation Transformation[?] are currently popular for implementing edit-based collaborative systems. Obtaining a snapshot of the state is usually trivial, but capturing edits is a different matter altogether. A practical challenge with event passing synchronization is that all user actions must be captured. Obvious ones include typing, but edits such as cut, paste, drag, drop, replacements and autocorrect must also be caught. The richness of modern user interfaces can make this problematic, especially within a browser-based environment.

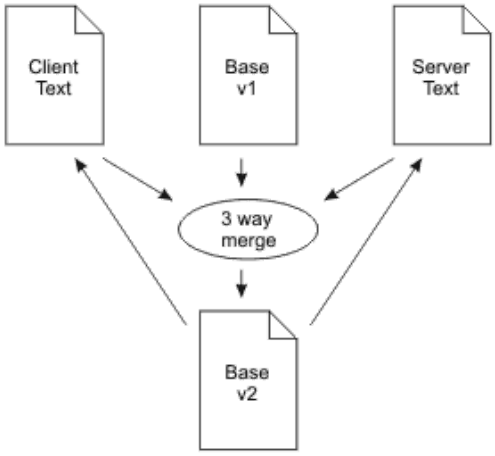
Any failure during edit passing results in a fork. Since each edit changes the location of subsequent edits, one lost edit may cause subsequent edits to be applied incorrectly, thus increasing the gap between the two versions. This is further complicated by the best-effort nature of most networking systems. If a packet gets lost or significantly delayed, the system must be able to recover gracefully. Google Wave[?] is an example of a multi-user application which uses event passing as its synchronization strategy.

Event passing is not naturally convergent.

Three-way merges are found in Subversion,[?] the Mjølner Project[?] and many other products. An overview of the process is:

1. The client sends the contents of the document to the server.
2. The server performs a three-way merge to extract the user's changes and merge them with changes from other users.
3. The server sends a new copy of the document to the client.

If the user made any changes to the document during the time this synchronization was in flight, the client is forced to throw the newly received version away and try again later. This is a half-duplex system: as long as one is typing, no changes are arriving. Shortly after one stops typing, the input from other users is integrated and either appears, or else a dialog pops up to let one know that there was a collision.



This system could be compared to an automobile with a windshield which becomes opaque while driving. Look at the road ahead, then drive blindly for a bit, then stop and look again. Major collisions become commonplace when everyone else on the road has the same type of "look xor drive" cars.

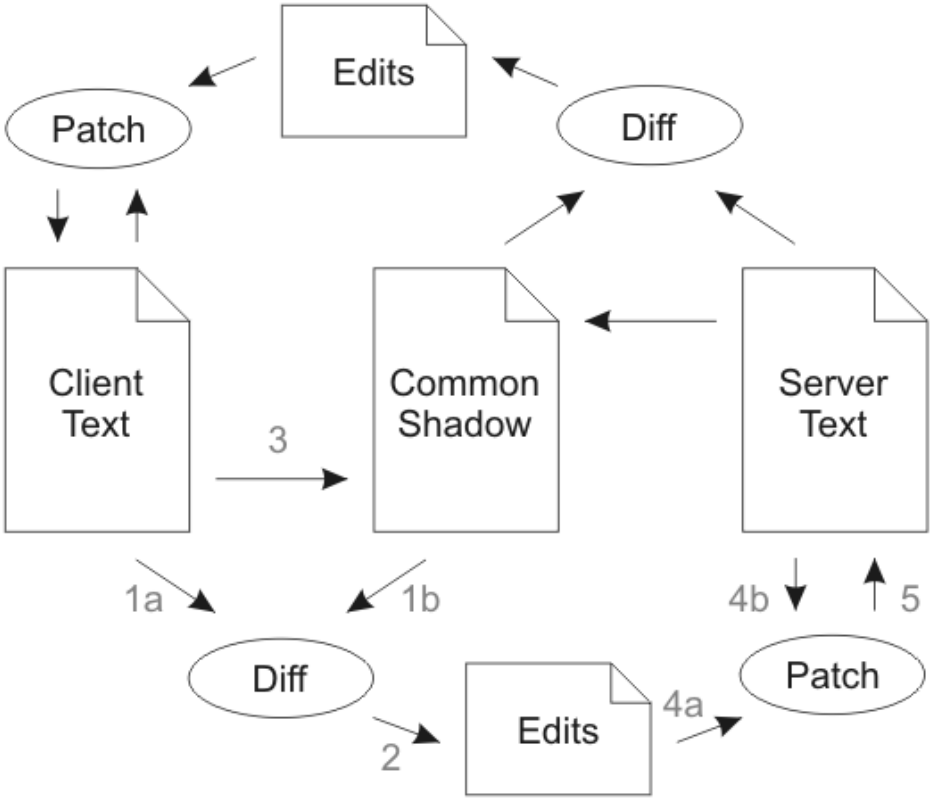
Three-way merges are not a good solution for real-time collaboration across a network with latency.

---

2 Differential Synchronization Overview

Differential synchronization is a symmetrical algorithm employing an unending cycle of background difference (diff) and patch operations. There is no requirement that "the chickens stop moving so we can count them" which plagues server-side three-way merges.

Below is an idealized data flow diagram for differential synchronization. It assumes two documents (misleadingly called Client Text and Server Text) which are located on the same computer with no network.



The following walk-through starts with Client Text, Common Shadow and Server Text all being equal. Client Text and Server Text may be edited at any time. The goal is to keep these two texts as close as possible with each other at all times.

1. Client Text is diffed against the Common Shadow.
2. This returns a list of edits which have been performed on Client Text.

3. Client Text is copied over to Common Shadow. This copy must be identical to the value of Client Text in step 1, so in a multi-threaded environment a snapshot of the text should have been taken.
4. The edits are applied to Server Text on a best-effort basis.
5. Server Text is updated with the result of the patch. Steps 4 and 5 must be atomic, but they do not have to be blocking; they may be repeated until Server Text stays still long enough.

The process now repeats symmetrically in the other direction. This time the Common Shadow is the same as Client Text was in the previous half of the synchronization, so the resulting diff will return modifications made to Server Text, not the result of the patch in step 5.

The enabling feature is that the patch algorithm is fuzzy, meaning patches may be applied even if the document has changed. Thus if the client has typed a few keystrokes in the time that the synchronization took to complete, the patches from the server are likely to have enough recognizable context that they may still be applied successfully. However, if some or all of the patches fail in step 4, they will automatically show up negatively in the following diff and will be patched out of the Client Text. Here's an example of actual data flow.

a. Client Text, Common Shadow and Server Text start out with the same string: "Macs had the original point and click UI."
b. Client Text is edited (by the user) to say: "Macintoshes had the original point and click interface." (edits underlined)
c. The Diff in step 1 returns the following two edits:

```
@@ -1,11 +1,18 @@
 Mac
+intoshe
 s had th
@@ -35,7 +42,14 @@
 ick
-UI
+interface
 .
```

d. Common Shadow is updated to also say: "Macintoshes had the original point and click interface."
e. Meanwhile Server Text has been edited (by another user) to say: "Smith & Wesson had the original point and click UI." (edits underlined)
f. In step 4 both edits are patched onto Server Text. The first edit fails since the context has changed too much to insert "intoshe" anywhere meaningful. The second edit succeeds perfectly since the context matches.
g. Step 5 results in a Server Text which says: "Smith & Wesson had the original point and click interface."
h. Now the reverse process starts. First the Diff compares Server Text with Common Shadow and returns the following edit:

```
@@ -1,15 +1,18 @@
-Macintoshes
+Smith & Wesson
 had
```

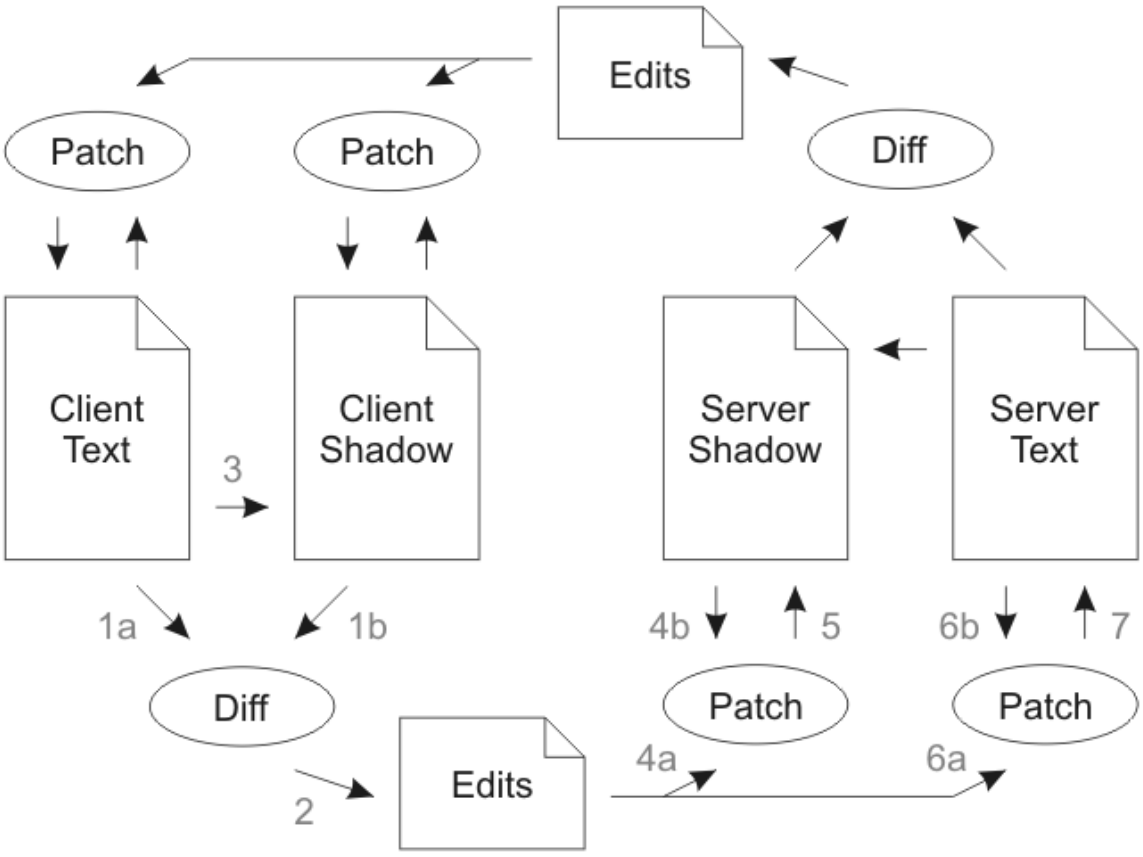
i. Finally this patch is applied to Client Text, thus backing out the failed "Macs" -> "Macintoshes" edit and replacing it with "Smith & Wesson". The "UI" -> "interface" edit is left untouched. Any changes which have been made to Client Text in the mean time will be patched around and incorporated into the next synchronization cycle.

A live demo of diff and patch algorithms for plain text may be used here:
http://neil.fraser.name/software/diff_match_patch/demo_patch.html

---

3 Dual Shadow Method

The method described above is the simplest form of differential synchronization, but it will not work on client-server systems since the Common Shadow is, well, common. In order to execute on two systems, the shadow needs to be split in two and updated separately. Conceptually this is the same algorithm.



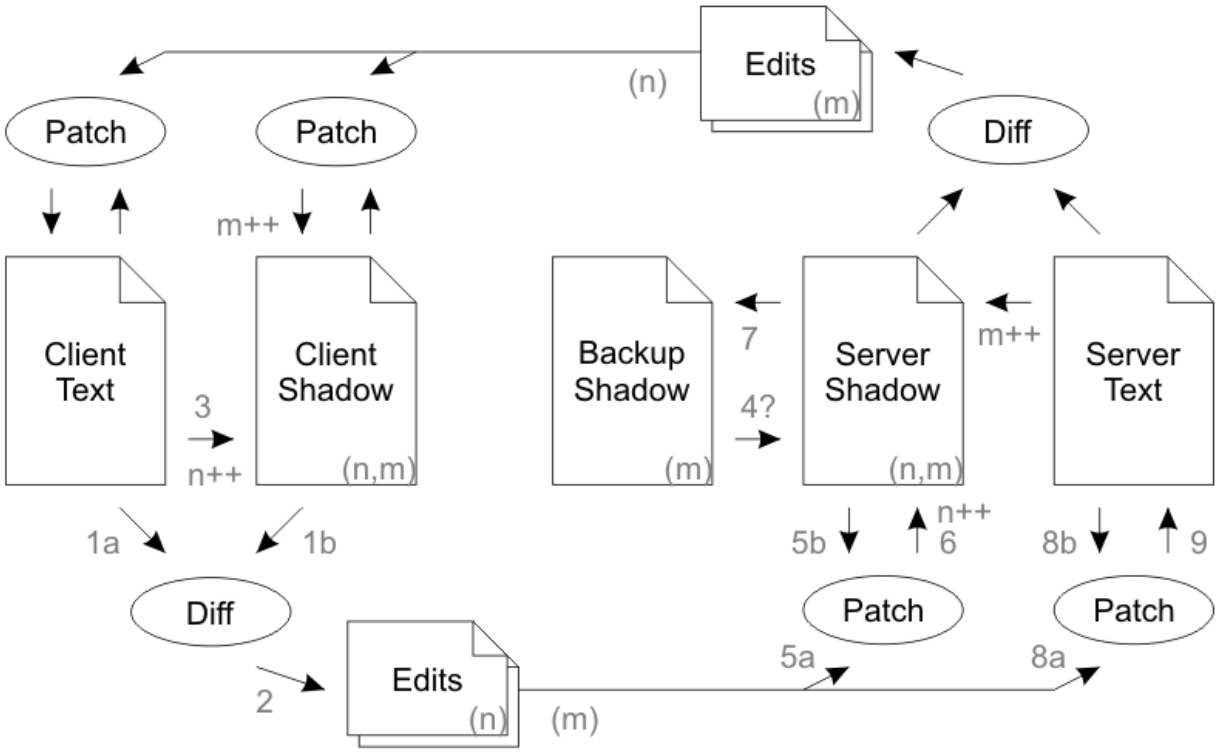
Client Text and Server Shadow (or symmetrically Server Text and Client Shadow) must be absolutely identical after every half of the synchronization. This should be the case since "(v1 Diff v2) Patch v1 == v2". Thus assuming the system starts in a consistent state, it should remain in a consistent state. Note that the patches on the shadows should fit perfectly, thus they may be fragile patches, whereas the patches on the texts are best-effort fuzzy patches.

However, on a network with best-effort delivery, nothing is guaranteed. Therefore a simple checksum of Client Shadow ought to be sent along with the Edits and compared to Server Shadow after the patches have been applied. If the checksum fails to match, then something went wrong and one side or the other must transmit the whole body of the text to get the two parties back in sync. This will result in data loss equal to one synchronization cycle.

---

4 Guaranteed Delivery Method

In the event of a transitory network failure, an outbound or a return packet may get lost. In this case the client might stop synchronizing for a while until the connection times out. When the connection is restored on the following synchronization, the shadows will be out of sync which requires a transmission of the full text to get back in sync. This will destroy all changes since the previous successful synchronization. If this form of data-loss is unacceptable, a further refinement adds guaranteed delivery.



In a nutshell: Normal operation works identically to the Dual System Method described above. However in the case of packet loss, the edits are queued up in a stack and are retransmitted to the remote party on every sync until the remote party returns an acknowledgment of receipt. The server keeps two copies of the shadow, "Server Shadow" is the most up to date copy, and "Backup Shadow" is the previous version for use in the event that the previous transmission was not received.

Normal operation: Client Text is changed by the user. A Diff is computed between Client Text and Client Shadow to obtain a set of edits which were made by the user. These edits are tagged with a client version number ('n') relating to the version of Client Shadow they were created from. Client Shadow is updated to reflect the current value of Client Text, and the client version number is incremented. The edits are sent to the server along with the client's acknowledgment of the current server version number ('m') from the previous connection. The server's Server Shadow should match both the provided client version number and the provided server version number. The server patches the edits onto Server Shadow, increments the client version number of Server Shadow and takes a backup of Server Shadow into Backup Shadow. Finally the server then patches the edits onto Server Text. The process then repeats symmetrically from the server to the client, with the exception that the client doesn't take a backup shadow. During the return communication the server will inform the client that it received the edits for version 'n', whereupon the client will delete edits 'n' from the stack of edits to send.

Duplicate packet: The client appears to send Edits 'n' to the server twice. The first communication is processed normally and the response sent. Server Shadow's 'n' is incremented. The second communication contains an 'n' smaller than the 'n' recorded on Server Shadow. The server has no interest in edits it has already processed, so does nothing and sends back a normal response.

Lost outbound packet: The client sends Edits 'n' to the server. The server never receives it. The server never acknowledges receipt of the edit. The client leaves the edits in the outbound stack. After the connection times out, the client takes another diff, updates the 'n' again, and sends both sets of edits to the server. The stack of edits transmitted keeps increasing until the server eventually responds with acknowledgment that it got a certain version.

Lost return packet: The client sends Edits 'n' to the server. The server receives it, but the response is lost. The client leaves the edits in the outbound stack. After the connection times out, the client takes another diff, updates the 'n' again, and sends both sets of edits to the server. The server observes that the server version number 'm' which the client is sending does not match the server version number on Server Shadow. But both server and client version numbers do match the Backup Shadow. This indicates that the previous

response must have been lost. Therefore the server deletes its edit stack and copies the Backup Shadow into Shadow Text (step 4). The server throws away the first edit because it already processed (same as a duplicate packet). The normal workflow is restored: the server applies the second edit, then computes and transmits a fresh diff to the client.

Out of order packet: The server appears to lose a packet, and one (or both) of the lost packet scenarios is played out. Then the lost packet arrives, and the duplicate packet scenario is played out.

Data corruption in memory or network: There are too many potential failure points to list, however if the shadow checksums become out of sync, or one side's version number skips into the future, the system will reinitialize itself. This will result in data loss for one side, but it will never result in an infinite loop of polling.

Asymmetry

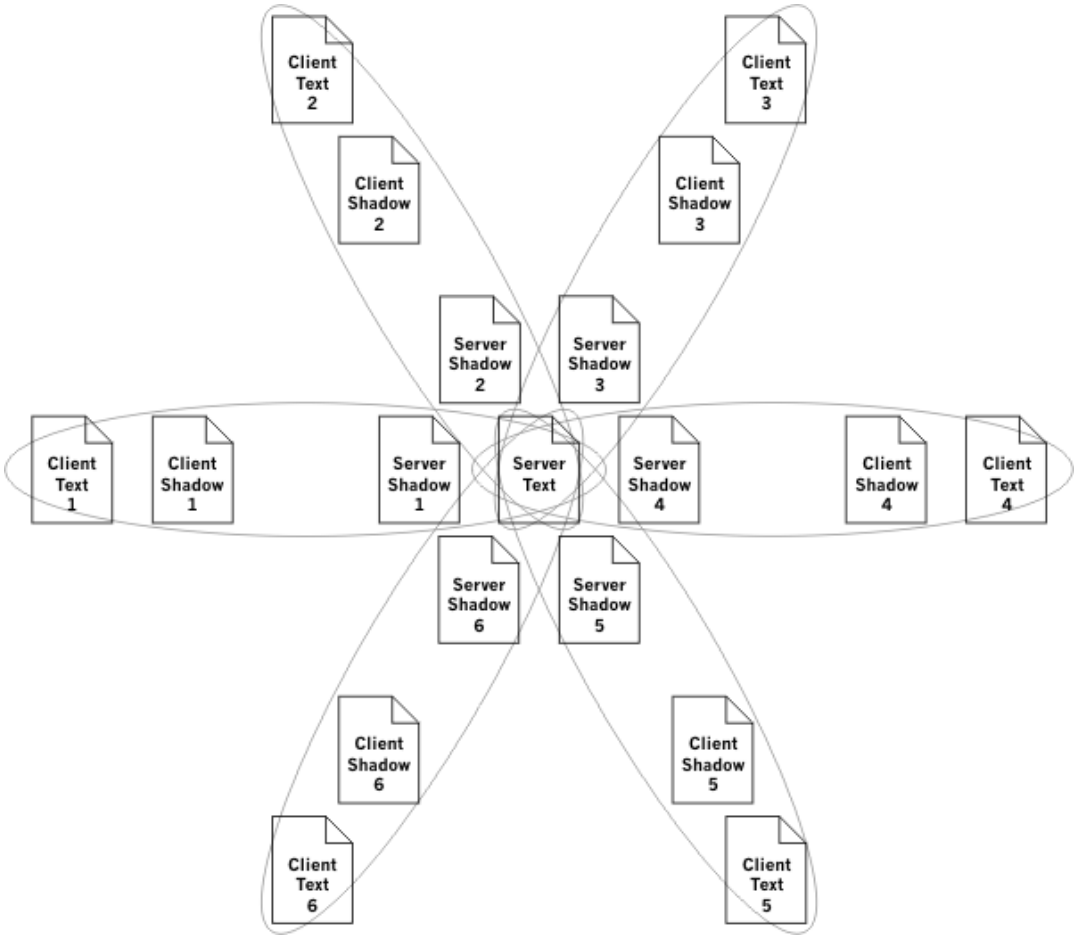
An obvious question is that given the otherwise perfect symmetry between client and server, why does the server have a Backup Shadow whereas the client does not? The source of this asymmetry is the asymmetrical nature of the connections. In a web-based client-server configuration, the client is the only entity which can initiate a connection. Depending on data losses, there are only three possible outcomes: 1) client sends data which is lost before reaching the server, 2) client sends data to server, but server's response is lost before reaching client, 3) client and server complete a successful round-trip. Notably missing is the possibility that the client's data is lost but the server's data is received. Every time the server sends information to the client, that implies a successful connection must have been established from the client to the server. Thus the server cannot get into a situation where it repeatedly sends packets to the client which don't arrive -- while not obtaining any packets from the client.

The client could implement a Backup Shadow, but it would never get used when run on a web-based client-server architecture. For symmetrical architectures (e.g. peer-to-peer or server-to-server) where either side can initiate a connection to the other, then a Backup Shadow would be required on both sides.

---

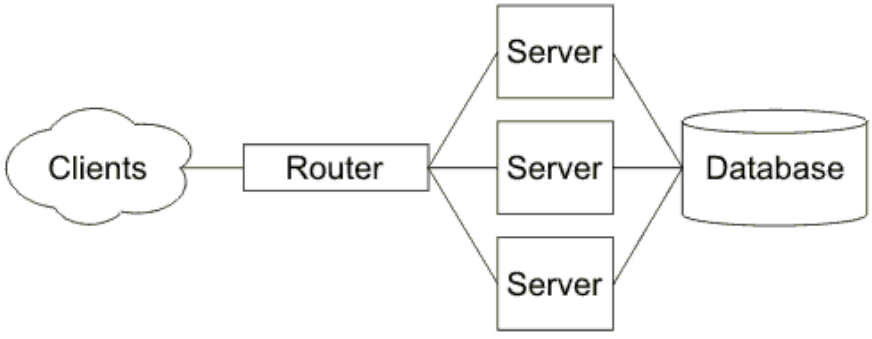
5 Topology

The above diagrams demonstrate synchronization between two parties, either a user and a server, or a pair of users. This same synchronization strategy can be multiplied to service any number of additional clients in a server-centric network. The Server Text for each synchronization loop is common with all the other loops. When Client 1 changes his document, Server Text is updated upon the next synchronization cycle, and those changes are passed on to all other clients on the following cycle.

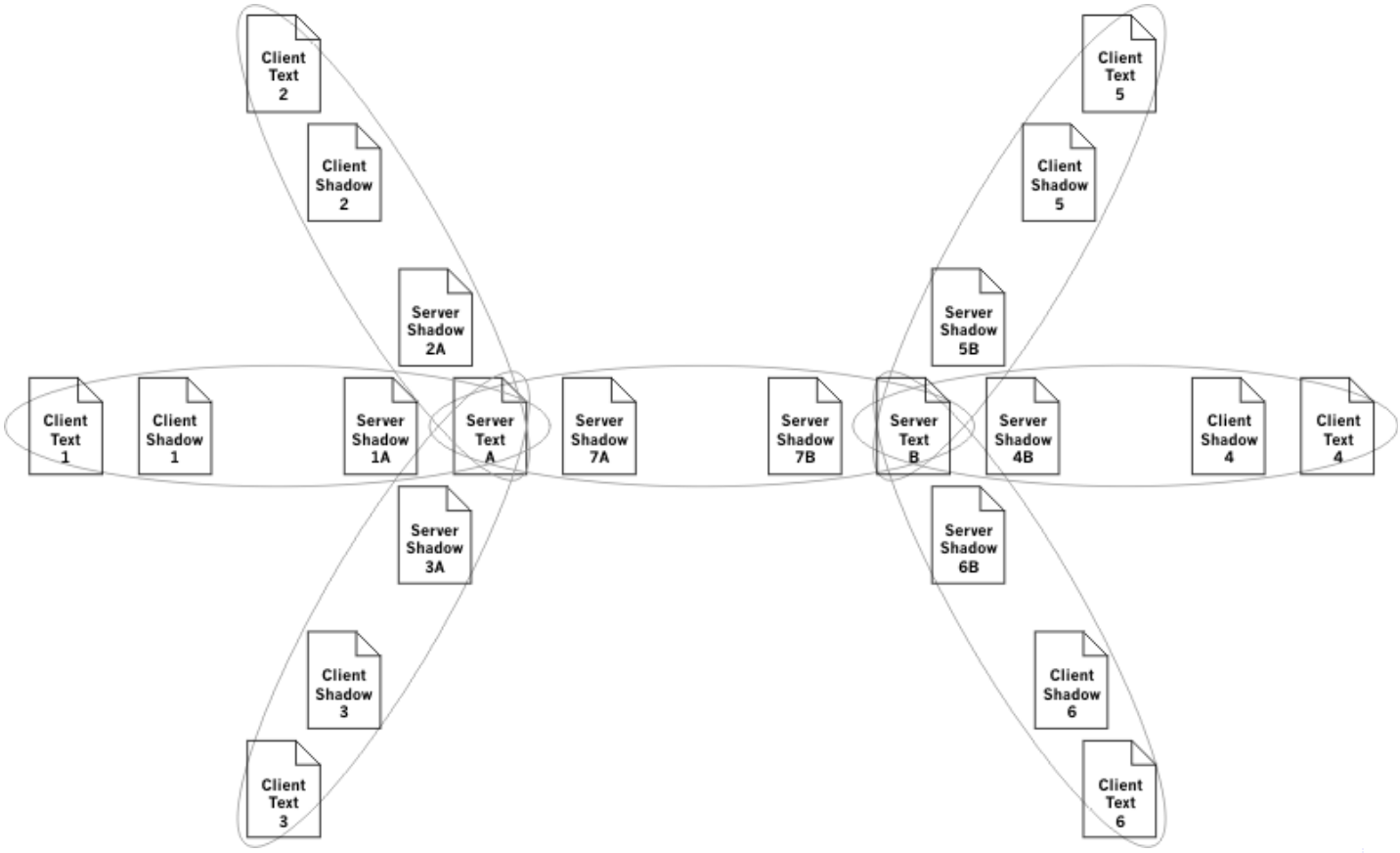


Scalability may become an issue as the number of clients increase. Diff and patch can be expensive operations, thus a server may become overloaded. There are two simple methods of distributing the system onto multiple servers.

One method is to separate the database from the algorithm. Thus one database would service any number of load-balanced servers. A client could hit any server, and as long as the view of the shared database is identical across all servers, the system remains consistent.



Another method is to introduce a server-to-server topology. In the diagram below, the clients are divided equally between two servers and the two servers are linked to each other with exactly the same type of connection as between the servers and the clients. Additional servers may be added seamlessly whenever capacity is exceeded. Servers may only be removed when all their clients depart and they only have a single connection to another server.



As the network expands, a potential problem is latency. Each link might synchronize every five seconds (see section 7). Thus it would take a change from Client 1 up to fifteen seconds to appear for Client 4. As latency increases, so does the potential for non-trivial collisions. Accordingly it is important to avoid a long chain of servers; a balanced tree offers the shortest path between clients, and thus the least latency. Latency may also be reduced by significantly increasing the synchronization frequency between servers. If the servers are located next to each other, then there is no bandwidth cost in synchronizing several times a second.

---

6 Diff and Patch

All the examples in this paper have shown synchronization of plain text. Differential synchronization can handle any content (plain text, rich text, bitmaps, vector graphics, etc) as long as a difference algorithm and a fuzzy patch algorithm is available for the content.

As the most computationally expensive components, improving the efficiency of these algorithms dramatically improves the responsiveness of the system. Likewise, improving the accuracy of these algorithms greatly reduces the number and severity of collisions.

The diff operation is fulfilling two very different roles within the synchronization cycle. The first is to update Server Shadow with the current content of Client Text and Client Shadow. The result should make all three texts identical. This is a simple task which could use any form of synchronization; diff, delta edits or even transmission of the full text. The second operation is more of a challenge: updating Server Text with the changes made to Client Text. Server Text may have changed in the mean time, which means that the diff must be *semantically* meaningful.

For instance, if the word "cat" was deleted and replaced with "hag", then technically one could think of it as the replacement of the first and third letters, with the second letter being preserved. This would be the minimal diff.

```
Client Text:   The cat is here.
Client Shadow: The hag is here.
Minimal Diff:  The ~~c~~h~~at~~g is here.
Semantic Diff: The ~~cat~~hag is here.
```

But this was not the semantic intent of the user. The user changed the word, not two letters. The fact that 'a' was the same in both words was completely coincidental. This distinction matters because if in the mean time another user changed the server's text from "cat" to "cut", the result when applying the first user's patch should be either "hag" (client wins) or "cut" (server wins), but certainly not "hug" (merged differences). An algorithm must be used to expand minimal diffs into semantically meaningful diffs. One such algorithm for plain-text is described in Diff Strategies, along with a set of optimizations to make diff significantly faster.

The patch operation is just as critical to the operation of the system. This system requires that patches be applied in and around text which may not exactly match the expected text. Furthermore, patches must be applied 'delicately', taking care not to overwrite changes which do not need to be overwritten. One such algorithm for plain-text is described in Fuzzy Patch, along with a set of optimizations to make patch significantly faster.

---

7 Adaptive Timing

The frequency of each client's update cycle is a key factor in determining the responsiveness of the system. Insufficiently frequent updates result in more computationally expensive diff and patch operations, major edit collisions, merge failures, and frustration when attempting to interact with other users. Overly frequent updates result in higher network traffic and increased system load.

The most computationally expensive operation is extracting the difference between what was last synced and the current document contents (thus obtaining the changes). This algorithm is $O(n^2)$ where n is the length of the changes. Thus if synchronizations can generally occur one change apart (where one change can be an insertion or a deletion of arbitrary length), then the most expensive operation becomes O(1). An advantage of the Guaranteed Delivery Method described above is that it decouples the differencing operation from the network transmission. Diffs can be taken at frequent intervals (to conserve CPU resources), added to the edit stack, then transmitted in batches at a slower rate (to conserve network resources).

An adaptive system can continuously modify the network synchronization frequency for each client based on current activity. Hard-coded upper and lower limits would be defined to keep the cycle within a reasonable range (e.g. 1 second and 10 seconds respectively). User activity and remote activity would both decrease the time between updates (e.g. halving the period). Sending and receiving an empty update would increase the time between updates (e.g. increasing the period by one second). This adaptive timing automatically tunes the update frequency so that each client gradually backs off when activity is low, and quickly reengages when activity is high.

---

8 Further Issues

Despite the inherent complexity, this synchronization system works extremely well. It is robust, self-healing and (with the proper diff and patch algorithms) impressively accommodating of users who are working on the same text. Try the demonstration of MobWrite, a web-based multi-user editor which uses this differential synchronization.

One limitation of differential synchronization as described here is that only one synchronization packet may be in flight at any given time. This would be a problem if there was very significant latency in the connection. An example would be a client on Mars and a server on Earth. A half hour for the round trip at the speed of light is unavoidable, however it would be better to send a continuous stream of updates in each direction, not waiting for the reply to arrive. The algorithm does not currently support this feature.

Another avenue for exploration would be to keep track of which user was responsible for which edits. Currently the edits from all users are blended together on the server, making attribution difficult. Untangling this blend would allow incoming edits to be visually attributed to specific users, as well as potentially allowing rollbacks of individual contributions and other CVS-like features.