Dropping the Ball on Dropbox

by M. E. Kabay, PhD, CISSP-ISSMP
Professor of Computer Information Systems
School of Business & Management
Norwich University, Northfield VT

Last week I nearly caused myself a heart attack when I thought I was backing up my Dropbox shared folders: all the files disappeared!

I use Dropbox< https://www.dropbox.com/ > to share files with students on student research projects and with colleagues to share course materials when we collaborate in teaching. For other projects, we use Google Drive< https://drive.google.com/ >. Similar functions are available on our university network through shared disk drives and SharePoint< http://office.microsoft.com/en-us/microsoft-sharepoint-collaboration-software-FX103479517.aspx?fromAR=3 >.

Because any change in a shared folder affects everyone depending on the shared resource, accidents can have serious consequences. For example, one of the folders I wiped out of our Dropbox was where my colleagues and I share all the current versions of chapters destined for a new edition of a reference book that has over 80 contributors. Another was where I store all the homework and exam materials for my statistics course – hundreds of hours in development time on the textbook, the homework, and the exams.

Luckily, the disappearance of all the Dropbox files was minor: it turns out that I accidentally backed up everything to the same hard drive as the Dropbox folder – so I actually *moved* instead of *copying* the files. It took less than a minute to move the 2.32 GB back into the right folder; however, it took significantly longer to update the remote drives at Dropbox with the files – and then update the other users' hard-disk versions – because of bandwidth limitations.

The incident has reinforced my belief that taking independent backups of such shared resources is essential. My daily routine includes launching backup processes that generate dated differential backups (backing everything up since the last backup), including backups of the Dropbox.

Differential backups are fine, but recovering from a real disaster would require restoring all the data from the latest full backup and then updating from each differential backup. Unfortunately, that process puts deleted files back in place and one must delete unwanted files manually.

Lately I have switched to taking full dated backups of the Dropbox every day. Recovering from an accident consists simply in restoring everything from that full backup.

When there are multiple users, the likelihood of a problem increases exponentially – that is, as a function of the number of users, *n*.

For any shared resources where all participants have the ability to add, change or delete content, we should remember how to calculate the likelihood of a problem when *n* different users are involved:

- Let the probability of a failure caused by user (i) be p(i)
- Then the probability of not having a failure caused by user (i) is [1 - p(i)]

- And the probability that all the n users will cause no failure is
  [1 - p(1)]*[1 - p(2)]*[1 -p(3)]* … [1 - p(n)]
- For simplicity in this example, let p(i) = p for all n users
- Thus the probability that there will be no failure is simply
  P{no errors} = (1 – p)^n
- So the probability that at least 1 user will cause a failure is
  P{at least one error} = 1 - (1 – p)^n

If the chance that a user will make a mistake in the shared files is, say 0.01 per year and there are 10 users, then

P{at least one error} = 1 – (1 - 0.01)^10 = 1 – 0.99^10 = 1 – 0.904 = 0.096.

That is, there is a roughly 1 in 10 chance that someone in the group will cause an error in the next year.

Change the number of users to 100 and the picture is much more alarming:

P{at least one error} = 1 - 0.99^100 = 1 – 0.366 = 0.634 or roughly 2 in 3.

Up the number of participants another order of magnitude to 1,000 users and we have

P{at least one error} = 1 - 0.99^1000 = 1 – 4.3e-05 ≈ 1.

In conclusion, I think it is essential that any group sharing files determine who shall take *daily* backups of the materials. As long as there is a system for backing up and the backups are safely stored, it should be possible to recover from accidental (or deliberate) damage.

On a related note, agreeing on file-naming conventions is essential for effective collaboration. My practice is to save files with incremented version numbers (e.g., text_v08.docx is saved as text_v09.docx) when I finish working on a document. For convenience, I also save the file immediately without the version number (text.docx) so that it represents the latest version and I can pin that filename to the list of recent files for quick access. However, if there are several people editing a file, I strongly recommend that the group agree on using TRACK CHANGES so that it's clear who did what.

If there are many people working simultaneously, problems will arise due to the same problems that occurred in early versions of databases. The *lost update* problem occurs when uncontrolled access to a file allows Albert to open text.docx and then moments later, Bob to do the same. Each merrily updates the file; if Albert saves it first, then when Bob saves to the same filename, Bob's changes will destroy the record of Albert's changes.

Databases control such problems by providing locking strategies. If our shared folder doesn't allow that, one easy solution (although it requires everyone to use it reliably) is to define a folder such as "in use" and move the original file there so no one else uses it. In the shared folders for the new edition of the book I mentioned above, our team uses suffixes for the name of the chapter; e.g., "ch45" is not currently being updated; "ch45_au" is being modified by the author(s); "ch45_ed" is being edited; "ch45_pub" has been sent to the publisher.

Happy collaboration!

* * *

M. E. Kabay,< mailto:mekabay@gmail.com > PhD, CISSP-ISSMP, specializes in security and operations management consulting services and teaching. He Professor of Computer Information Systems in the School of Business and Management at Norwich University. Visit his Website for white papers and course materials.< http://www.mekabay.com/ >

* * *