Spring, 2006

File Sharing Options: Elements of a Collaborative Infrastructure

- Mason Kortz (PAL/CCE)

When working on a collaborative project, the benefits of sharing files between users quickly becomes apparent. A method has to be found that is supported by the group's collective infrastructure, that works for all members of the collaborative group, and that is (hopefully) efficient, user-friendly, and secure. This article is meant to help scope the file sharing needs of a project and find a method that meets the needs of the project, and all parties involved, without adding undue administrative or technical overhead. This article considers email, FTP, remote disks, and WebDAV as file sharing solutions.

Email, while not a file sharing technology in a strict sense, is so widely used for file transfers that it deserves some mention here. Email is a useful tool because it is ubiquitous, familiar, and in the foreground of many users' work routines. When used to transfer files, email also adds the advantage of adding context to the file, in the form of the email message body. When dealing with small, static files and small groups of recipients, email is a good choice for transferring files. However, email lacks scalability. Large attachments sent to many users can cause clutter to build up quickly. For this reason, many email services limit the size of attached files. Collaboration over email can also be difficult - with many users mass-emailing out revisions to a file, it can be hard to keep track of which, if any, is the authoritative version. Finally, emailed files don't live in a shared space. If a new user wants to start collaborating on a file, the user must first get someone who already has the file to email it out.

Another popular method of sharing files is via FTP, the file transfer protocol. Like email clients, FTP clients run the range from simple text-based clients to slick, stylish graphical interfaces. They all provide the same basic functionality - a way to move files to and from a central server. Because there is a centralized repository, there is one authoritative 'master copy' of the file being shared. A user with access to an FTP server and the proper permissions can actively find any file they are looking for; the user does not need to passively wait for someone else to send the file out. This, combined with the fact that FTP is widely supported on all platforms, makes FTP a good way to move large files between groups of users. Although storage of files is centralized with FTP, editing is done locally. In addition to being time-consuming, the upload/download mechanic means that at any time, many version of a file may exist on individual users' computers. If one user forgets to upload after making an edit, other collaborators may miss an important change.

Remote disks (also known as network drives, network volumes, sharepoints, and many other terms) are extremely useful for file sharing, especially in a collaborative environment. The exact protocols used to access remote disks vary depending on both the server and client platforms - for example, the SMB protocol is used in Windows networks, whereas AFP is more common in Macintosh networks - but they all share certain properties. Remote disks provide a centralized storage location for files, and share

many of the benefits of FTP. The main difference is that remote disks, once accessed, can be treated just like local disks. This means a user can edit and interact with files in place, making the work experience smoother than with FTP, especially when working on projects consisting of multiple files. There is no upload/download or send/receive step to take up time and spread file versions around to many clients. This also means that users need to be more careful about not overwriting each others work - because edits are not done locally, any changes made to a file will overwrite the previous version. Version control and/or backup software are often used with remote disks to mitigate this problem. Another issue is that clients on many different platforms might be connecting to the same remote disk. Many applications and operating systems can leave behind special system files - for example, .DS_Store files on Macintosh OS - that have no meaning on other platforms and show up as useless clutter. This can be mitigated or prevented with the correct client settings, but this requires some extra attention from system administrators. If multiple file sharing protocols are implemented on a single server, as is often the case, system administrator overhead can increase and system security can decrease, as multiple ports must be opened.

Another protocol, WebDAV (Web-based Distributed Authoring and Versioning), bears special mention in the context of collaborative environments. WebDAV is an extension of the HTTP protocol, meaning it works through your web server, but rather than being read like a web page, a WebDAV 'site' is accessed like a file system. This allows WebDAV to use the authentication and encryption provided by the web server, and creates a secure file sharing environment without opening additional ports. As the name implies, WebDAV is designed for collaborative work. Connecting to a WebDAV share is functional very similar to connecting to a remote disk, and provides many of the same benefits. WebDAV provides a built in file-locking system that helps prevent the problem of overwrites, and resource metadata (title, authors, etc.) through XML. Future versions of the protocol will also include native version control support and more refined access control. The main downside of WebDAV is that it is a relatively young protocol, meaning that its feature set is not yet complete. Currently the permissions system provided by WebDAV is not as robust as that provided by a shared file system. Application support is currently slim and even operating system support is inconsistent. Before moving to WebDAV as the mechanism for collaboration of a project, all clients should be checked for compatibility.

So what file sharing mechanism is the best for collaboration? It all depends on the local situation. Remote disks and WebDAV provide a good collaborative environment, but require an amount of administrative overhead that, in many cases, means an over-engineered solution. For small, short-term, one-on-one collaborations, email is a fine solution. This is especially true when the collaboration is between two institutions, and getting the necessary accounts and permissions for high-level file sharing may raise administrative issues. FTP can be very useful when dissemination of files, especially large files, is more the focus than collaboration. FTP can be a good way to move files between institutions without having to create accounts, provided the one of the institutions' infrastructures supports anonymous FTP. For long-term collaboration with a large group of users, setting up a remote disk or a WebDAV share is worth the administrative overhead. Ultimately, it is best to be familiar and comfortable with several methods of file sharing, to best accommodate the infrastructure at hand, the capabilities of the collaborators, and the needs of the project.