



Do You Know Where Your Messages Are?

By Jason Sherry

The need for message archiving

In most organizations, an estimated 83 percent of all communications are electronic, with the vast majority of those communications going through an email system. In a 2004 survey of 840 U.S. companies, 21 percent of respondents had their email and instant message data subpoenaed, up from 14 percent in 2003. The cost of providing this data can easily run into the hundreds of thousands of dollars, for which the organization may be solely responsible. Besides the legal concern, regulatory compliance requirements, like those covered in HIPPA, Sarbanes-Oxley (SOX), and the Gramm-Leach-Bliley Act (GLBA), increase the need for a message archiving and retention solution. The last factor in supporting the need for such a solution is the IT overhead involved with the ever increasing size of mailboxes and messaging databases.

→ Contents

The cost of storage	1
Retaining messages.....	2
Finding the needle in the haystack.....	2
Losing messages and knowledge	2
How Exchange helps	3
Reducing mailbox size	3
PST management.....	4
Filling in the gaps with Sherpa Software	4
Electronic polices are not enough	6



Do You Know Where Your Messages Are?

The cost of storage

The cost of storage per megabyte has been on a continuous drop since the first electronic storage solution. At the same time the storage requirements have continued to grow at an almost inverse rate. Ten years ago, 100GB of disk space on your Exchange server would have seemed outrageous for most organizations. Today, that has become nearly a minimum for an Exchange server. With the release of Exchange 2003 SP2, Microsoft has increased the maximum database size in the Standard Edition of Exchange from 16GB to 75GB. Microsoft has also increased the scalability of Exchange so that a single server, or cluster node, can host more than 4,000 mailboxes. In addition, the number of messages users send each day continues to grow—as does the size of attachments in messages. So while the overall cost per megabyte has continued to fall, the management cost of the servers that host the data has stayed about the same, or increased on a per server basis due to server consolidation and increased server complexity.

The increase in storage utilization has also increased the time it takes to back up, restore, and carry out off-line database maintenance tasks. While backup and restore technologies have continued to increase their throughput rate, the increase in throughput has been outstripped by the growth of storage utilization. Even at a backup rate of 50GB/hour, which is the theoretical maximum of mid- to high-end DLT tape drives, it would

take four hours to restore 100GB of data; restore typically takes twice as long as backups. This time doesn't take into account the additional troubleshooting and recovery steps required to restore normal operations on the server being restored. In the case where an Exchange server had died completely due to a catastrophic failure, a restore time longer than an hour might not be that critical. However, restore time does become a concern when you're trying to restore individual messages from a handful of mailboxes. In this case, the restore time is only a portion of the time it will take to obtain the data in question. First, the tapes must be found and then restored. Once restored, the individual mailboxes must be mounted or extracted so they can be searched, and the messages in question extracted. This same process must then be repeated for each set of tapes that contain data needing to be restored.

Even though the cost of storage has dropped, restoring an individual server is still a time consuming and costly task and the larger the database is, the greater the cost. In the situation where a selected number of items need to be restored, the per item restore cost can be significant. Thus, a best practice to keep disaster recovery, maintenance windows, server down time, and storage cost down is to decrease the size of Exchange database files. However, this doesn't help reduce the time to find and restore messages and mailboxes that may exist across backup tapes that span months or years.

Retaining messages

For those organizations required to meet certain business, legal, or regulatory requirements, just keeping the last X days of mail in Exchange for all users isn't an option. Those organizations may need to keep months or years worth of email for certain users. In such cases, messages must be retained in a way that doesn't significantly impact the size of the organization's Exchange databases. In addition, to meet such requirements an email message can never be permanently deleted, except for the case of system, spam, or other messages that would never be included in a legal or regulatory information gathering event. To be able to address these needs, the standard configuration of Exchange will not work. In this case, an organization can choose to implement Exchange Message Journaling, which copies all messages received and sent to mailboxes in a particular database to another mailbox. But message journaling comes with the additional cost of more storage, because every message sent to the selected database is being copied, not to mention the administrative overhead of managing the journaling mailbox. This mailbox must be

managed because it can grow in size very rapidly as all messages are copied. To keep its size down, messages must be removed from the mailbox and stored in some other fashion. This could be accomplished with ExMerge or Outlook by saving messages to Personal Storage Files (PSTs), but these files also require space and must be managed. The other quandary with message journaling is that the journaling mailboxes will contain only messages starting from the time journaling was first enabled.

Finding the needle in the haystack

Even if the data needed was available in some fashion, the next problem is finding the relevant data. If all messages are left in individual user mailboxes, searching those mailboxes for specific items becomes a complex undertaking. If message journaling is used and all messages are left in the journaling mailboxes, which isn't likely, this task becomes easier. However, the journaling mailbox will contain only messages from the time journaling was enabled. In most organizations, the amount of mail flow will require that older messages in the journaling mailboxes are saved to PSTs or some other medium. Let's not forget about all those user or Outlook created PST files spread across servers, desktops, and laptops. Again, the issue with searching multiple locations arises. Exchange does not provide a search front-end. All searches must be done in Outlook when logged onto an individual mailbox. But not being able to find the relevant data can be very costly in the way of legal, regulatory, and other financial penalties. These penalties can easily run into tens or hundreds of thousands of dollars.

Losing messages and knowledge

A final reason to archive messages is to protect corporate intelligence. In most organizations that use mailbox limits, and even in those that do not, users will create PSTs. Outlook will also create a PST if its auto archive functionality is enabled. In most cases, messages are normally moved from the user's mailbox into local PST files. Because the files are normally stored locally on the user's computer, there is no easy way to centrally discover, search, and collect data from them. This exposes an organization to data loss and theft. Take the case of a CxO in an organization. Like most company officials they have a laptop they use when traveling, which is a large percent of their time. If they have PST files on that laptop, there is a good chance they contain some key company information, maybe even some intellectual property that an organization would not

want to lose, or worse, see in the hands of competitors. Without using drive-level encryption, which can be bypassed with enough effort, the data in those PSTs can easily be accessed by anyone with physical access to the system or drive. In the case of a hard drive crash or system loss or theft, chances are the data stored locally on that system will be lost forever. With the appropriate archiving solution in place, users should never use PSTs, and the data in any existing PSTs should be imported into the archiving solution and removed from all systems.

How Exchange helps

Exchange 2003 does provide out-of-the-box capabilities to address some of the issues cited above. The first feature is Journaling, which will copy all messages sent and received to mailboxes on specified Exchange databases to a single mailbox. Message journaling is crucial to ensure that all messages going in or out of an organization have been copied. Without message journaling in place, once a user receives a message he could choose to permanently delete it, which removes it from his mailbox. If that deleted message is needed later, after deleted item retention has expired, the recovery of the message requires a database restore. Because most organizations won't be able to leave all messages ever sent in a journaling mailbox, at least from the point journaling was enabled, journaling must be managed. The process to manage the journaling mailbox and the data it once contained can get quite involved as time goes on.

Reducing mailbox size

The other feature of Exchange that can help is Mailbox Manager's policies. You can use these policies to reduce mailbox sizes (Figure 1) by aging or deleting older items from a mailbox. You can process items in any folder in a user's mailbox based on the age or size of the item. The actions taken by mailbox manager are limited though (Figure 2). The first option generates only an administrator report of the mailboxes that exceed the policy, which leaves it up to the administrator to ask the user to clean up her mailbox. The second option moves the items to the user's Deleted Items folder, which doesn't reduce the mailbox size until the Deleted Items contents have been cleared out. The third option moves items to a new root level folder in Outlook called "System Cleanup," but it doesn't remove the items from the user mailbox. To reduce the size of the user's mailbox, another mailbox manager policy must be created to clean up the "System Cleanup" folder. The last option deletes the items from the user's mailbox. Whenever the

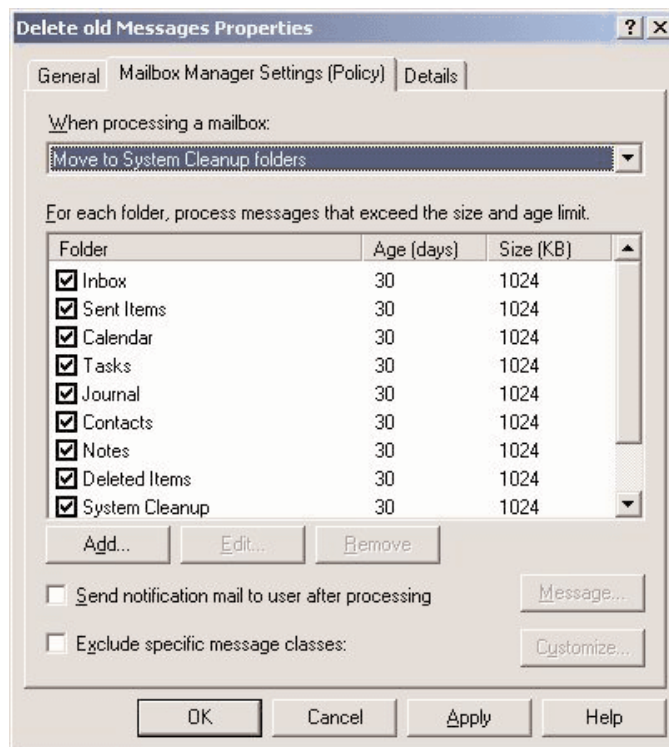


Figure 1: Exchange Mailbox Manager

"Delete Immediately" option is used or the user deletes all items in her Deleted Items folders, the storage overhead for those items is actually not released until the delete item retention limit has been reached. The problem with mailbox manager is that once items have been deleted from a user's mailbox, the user can no longer access or recover those items. Those items are removed from the messaging system completely, unless a copy exists in a journaling mailbox. Therefore, using mailbox manager can greatly impact end users and the support staff, because users will realize that messages are disappearing from their mailbox. This in turn will cause additional work for the Exchange administrators who will need to recover items the users say they need. Users will also start using PST files more aggressively, which further increases administrative overhead for IT. This is compounded by the fact that mailbox manager is limited to just filtering items based on their size, age, and

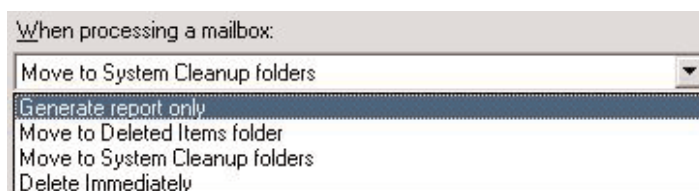


Figure 2: Mailbox Manager

type. There is no way to filter or exclude items based on their contents or keywords, message flags, read or unread status, or other item attributes.

PST management

The next major area of concern and administrative overhead is the management of PSTs. PSTs are very common in almost all organizations because users have learned to use them to keep their mailbox size down, to “clean up” their mailbox, and bypass mailbox manager policies. Outlook also includes an Auto-Archive feature that by default runs every two weeks and will move items from the user’s mailbox to a PST created by Outlook. It is not uncommon for users to have multiple PST files spread across their laptop and desktop. All of this makes locating and managing PST files a very difficult task due to the many different locations where PSTs might exist. This is further compounded by the fact that some clients may only connect to the corporate network occasionally. Unfortunately, no tools are included with Exchange to discover PST files or easily import them into a centralized database for indexing, searching, and item recovery. PSTs must be managed because they can contain confidential corporate information or copies of messages that may be needed to meet the request for a legal or regulatory investigation or audit. Another reason to eliminate PSTs is that they can easily be lost, due to hard drive failure or system loss, or data corruption. PST corruption was very common in larger PSTs before Outlook 2003’s new PST format. Before Outlook 2003, PSTs were limited to 2GB, so once the user hit this limit they would need to create another PST file. The larger the PST file got, the greater the chance for corruption. Even if all PSTs could be consolidated to one server, Exchange and Outlook don’t provide any way to search for data across multiple PSTs. So while finding data in multiple remote PSTs is nearly impossible, having all PSTs stored on a single server doesn’t provide much of a benefit. A further limitation is that users must be connected via

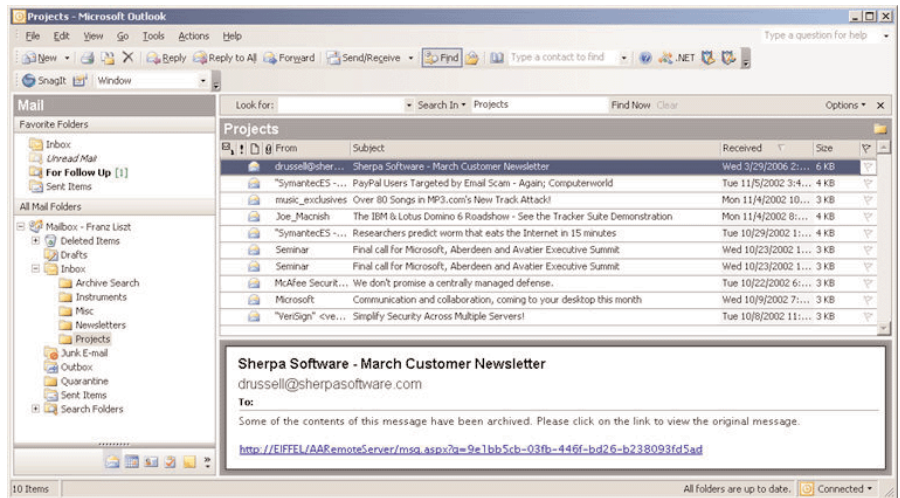


Figure 3: Stub Message

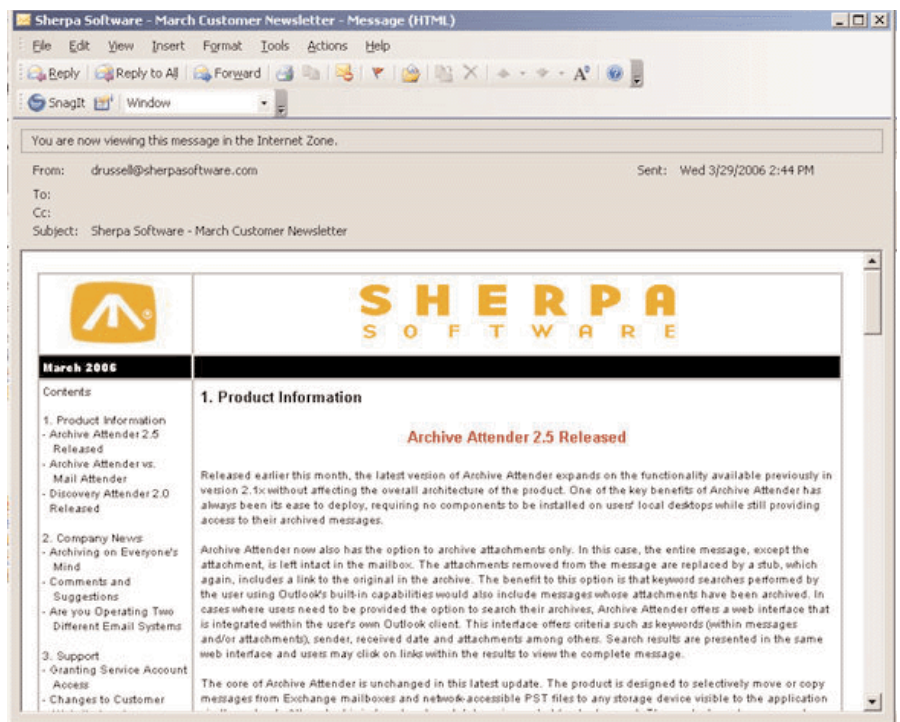


Figure 4: A stub message opened in Outlook

LAN to their PSTs to prevent serious performance issues when accessing them.

Filling in the gaps with Sherpa Software

To meet business, legal, regulatory, user, and IT requirements, a better solution than what Exchange provides is required for most organizations. Sherpa Software provides two key products to address the shortcomings of mailbox

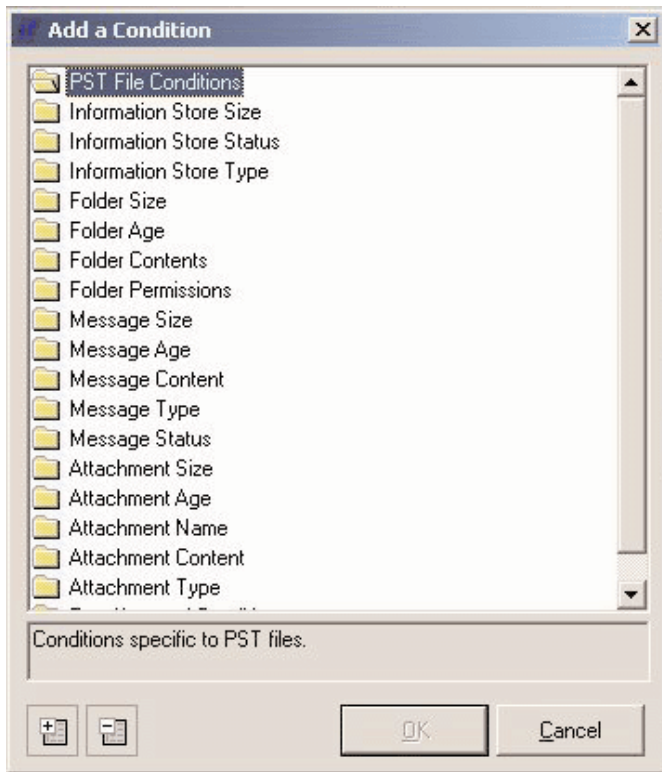


Figure 5: Mail Attender

manager, assist in the management of journaling mailboxes, manage PSTs, and reduce the end-user impact of message archiving. Sherpa's Archive Attender was designed to help organizations manage their storage utilization of Exchange while at the same time provide end user's with the ability to easily access messages that have been archived. You can use Archive Attender to copy or move all messages matching certain criteria in a centralized archive. To reduce user impact, when Archive Attender removes a message from a user's mailbox a stub message can be left in the place of the original message (Figure 3). This "stub message" replaces the original body of the item with one that tells the user the item has been archived and it includes a URL to the item in the archive. When users click on the stub the message is opened in Outlook for reviewing, replying, or forwarding (Figure 4). Once the data is in the archive it is indexed and available for searching and restoration. Sherpa's second product, Mail Attender, provides extensive reporting and management capabilities for Exchange. Mail Attender provides the flexibility to define what items it will process and what actions it will carry out on those items. The product supports about 50 different types of rule conditions (Figure 5). You can use these rule conditions to granularly control and report on data in Exchange (Figure 6). Once the rule condition has been set, you can select any one of the 60 dif-

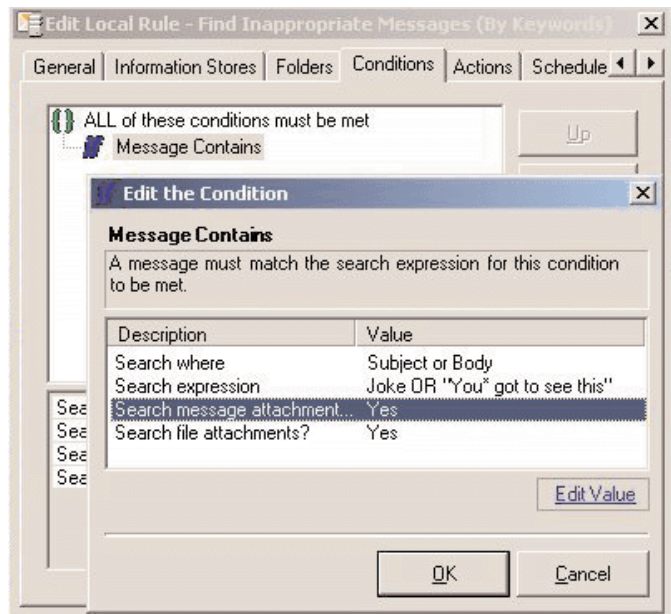


Figure 6: Condition

ferent actions for execution. Examples of actions include reporting on matching items, replacing an attachment with a ZIP file that contains the original attachment, deleting messages and/or attachments, and copying messages to another mailbox or PST file. Mail Attender can also inform the end-user when certain actions are carried out. For example, adding the text "The file attached to this message was ZIPed." With Archive Attender, you can use message archiving rules to set up standard archiving policies for an organization (Figure 7). With the addition of Mail Attender, both products can be integrated together. This allows for the use of the extensive rules supported by Mail Attender to control message archiving with even greater granularity. You can use this level of flexibility, for example, to define rules to copy all business-critical messages to the archive while ignoring or deleting non-business related messages.

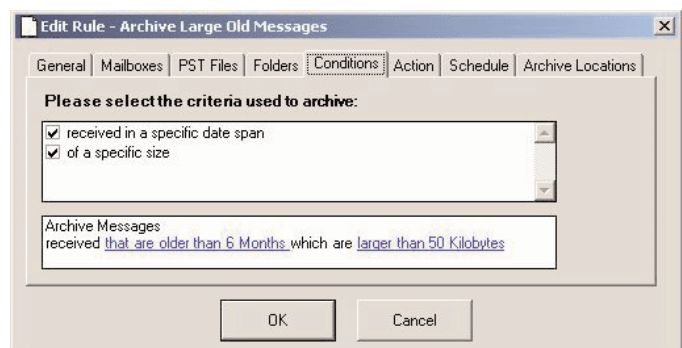


Figure 7: Archive Attender

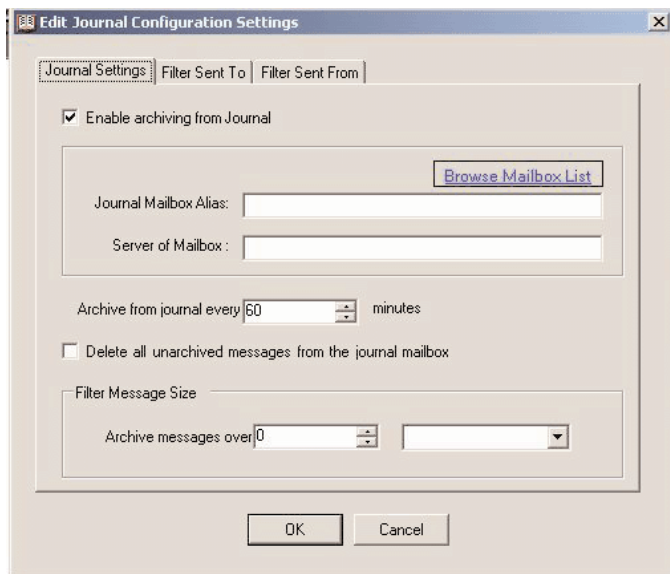


Figure 8: Journal Mailbox

PST management, another area where organizations need help, also is addressed by Archive Attender and Mail Attender. You can use Archive Attender to move items from PST files into a centralized archive. You can carry out the same archiving actions on PSTs that you can on items in a mailbox. This functionality lets organizations archive items from a PST and automatically create stub messages in the same folder structure as they existed in the user's PST file. The folder structure that previously existed in the PST file is placed under a folder in the user's mailbox called "Imported from PST file." This support provides users the ability to easily find messages that used to be in their PSTs. In addition, Mail Attender provides essential assistance in the discovery of PSTs on the network. Once Mail Attender knows the location of PSTs, the same rule conditions and actions can be carried out to report on the contents, archive individual items, copy items, or any of the other actions supported in Mail Attender.

Archive Attender also includes support to help manage a journaling mailbox. To ensure that all messages sent in or out of your messaging system are monitored, message journaling must be enabled for all mailboxes, or at least on those that require this level of accountability. You can configure Archive Attender to archive and delete all messages in the journaling mailbox on a set schedule (Figure 8). In addition, Archive Attender can filter out messages to or from certain email addresses (e.g., system mailboxes or

Internet mailing lists). This functionality will prevent the journaling mailbox from growing too large and increasing maintenance windows.

With the combination of Mail Attender's support for flexible rules and actions and Archive Attender's ability to archive messages while providing end-users the ability to search, review, and restore messages, these products can help organizations meet their business, legal, and regulatory requirements. They also can help reduce storage utilization, decrease downtime, reduce backup and restore windows, and lead to an overall reduction in the management cost of Exchange. With the additional functionality provided by Mail Attender, organizations can get proactive in the management of mailbox size limits, removing business inappropriate emails, and reporting on key messages being sent or received by their users.

Electronic polices are not enough

Mail archiving and management products like those provided by Sherpa Software provide only part of the solution needed to meet legal and regulatory requirements. Organizations must set "paper" policies and ensure that all users affected by those policies know and agree to comply with them. Once the official company's policies are in place, software should be used to enforce those policies to provide "best effort" enforcement. Any organization required to meet regulatory compliance guidelines should consult a professional in that field to make sure both their paper and electronic policies meet the minimum requirements. All organizations should also consult legal council to find out what policies make the most sense to ensure minimum legal liability and exposure.

Jason Sherry is an Infrastructure Architect with Pro Exchange, where he specializes in Active Directory, Exchange, and SharePoint implementations and migrations. Before joining Pro Exchange, Jason worked at NetIQ and Configuresoft as a product manager for a total of seven years. At both companies he oversaw the Active Directory and Exchange management product lines. Jason started working with enterprise-level corporations at Digital Equipment Corp, now HP, in 1994. He has spoken at various industry trade shows, such as Microsoft TechEd, Microsoft Manageability Summit (MMS), and Gartner IT Expo. He has also written for various electronic and print mediums on Active Directory, Exchange, and scripting.