

**“Highest Performance
Lowest Price”**

Microsoft
GOLD CERTIFIED
Partner

GFI WHITE PAPER

ARCHIVING TECHNOLOGIES

Have you ever considered the impact one, untraceable email can have on an organization or individual's career? With so much corporate information contained within email, it is not surprising that industries and governments worldwide are insisting that all email should be retained. Email archiving is no longer an option for organizations: undoubtedly it solves major problems that systems administrators face daily and it helps an organization fulfill its compliance and regulatory requirements, particularly those stipulating the retention of email for a period of time. The key to a successful implementation is choosing the right email archiving technology for that organization.

Introduction to email archiving

This white paper shall cover the following topics:

- What archiving is and how it integrates with Microsoft Exchange or a messaging solution
- What stubbing is and why Microsoft does not recommend the use of stubbing
- How Journaling in Exchange can allow efficient email archiving
- How does stubbing compare with Journaling?

An email archiving solution addresses the need to retain a copy of all incoming and outgoing email traffic. With a proper email archiving solution it is possible to access email content at any time from a centrally managed location.

Email archiving is the key to the following three needs:

- Manage email server resources efficiently
- Allow virtually unlimited mailbox storage
- Meet legal requirements, compliance and business needs

Email management

The task of managing email is usually split between the system administrators and the end users. System administrators need small mailboxes which contain a limited number of email messages. Mailboxes with large attachments or a large number of email accounts can easily bring down an Exchange or email server. When a server is handling hundreds or thousands of email addresses, system administrators typically put a quota on each mailbox to limit the amount of information stored on server.

On the other hand, setting a hard quota on all mailboxes affects end users in a negative way. They often need to retrieve emails that are weeks, months or even years old. They sometimes need to be able to store emails with large documents. For example, some departments will make extensive use of PDF files for legal reasons or Microsoft Word documents. The file transfer medium of choice for these files is generally email. Additionally, these files need to be retained for a given period of time especially in the case of legal documents or in industries which require email to be retained for a minimum time frame.

Thus system administrators and end users have conflicting requirements. In many cases reaching a compromise is difficult to achieve and at times is simply not an option. To solve this issue one needs to use a proper system which satisfies the needs of both the system administrators and those of the end users. Email archiving can provide an excellent solution by shifting the bulk mail to the archive store. This virtually gives end users an unlimited amount of space while keeping the Exchange or email server clean.

Compliance

Data retention is one of the most important communication issues facing companies worldwide. Many organizations have to comply with one or more regulations that require them to keep email communications available for a period of time.

The following are some of the best known compliance and regulations in the US that require retention of emails:

- Sarbanes-Oxley Act
- SEC Rule 17A-4
- NASD 3110 and 3111
- Gramm-Leach-Bliley Act (Financial Institution Privacy Protection Act of 2001, Financial Institution Privacy Protection Act of 2003)
- Healthcare Insurance Portability and Accountability Act of 1996 (HIPAA)
- Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act of 2001 (USA Patriot Act)

Member states of the European Union are subject to a Directive 2006/24/EC which requires communications providers to retain email data for a period of 6 months.¹

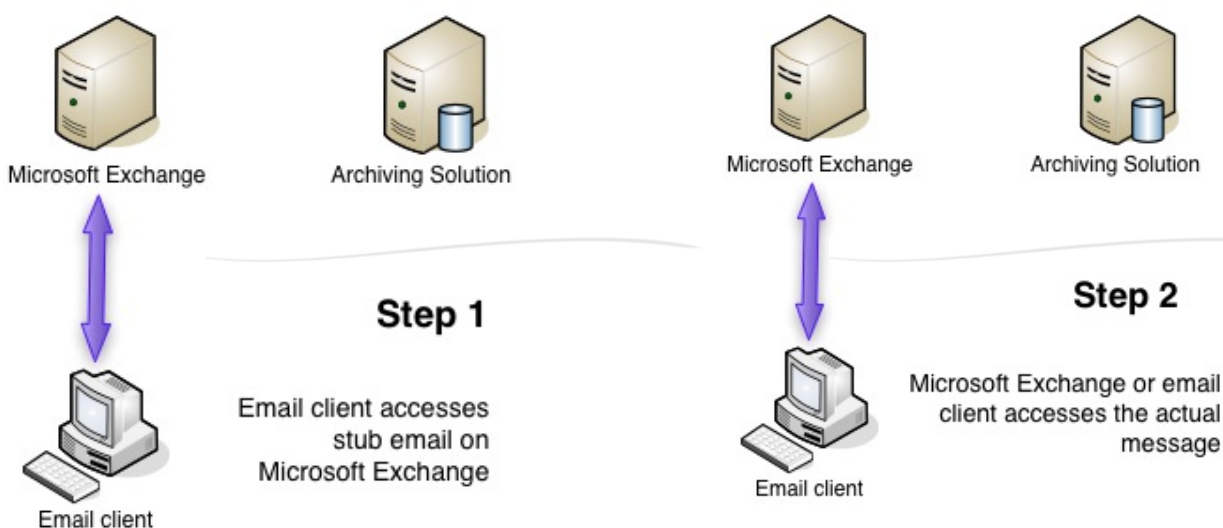
The only way to comply with such regulations to avoid risking legal liability and not create unnecessary burden on the email systems is to make use of an archiving solution. The side benefit is that a proper solution not only helps organizations comply with regulations, but also provides significant advantages when it comes to managing email.

¹ http://en.wikipedia.org/wiki/Telecommunications_data_retention#Data_retention_in_the_European_Union

Archiving technologies

Stubbing

Various archiving solutions make use of a process called stubbing. This involves moving emails from the users' mailboxes to a new location, while replacing the original email in the user mailbox with a small message pointing to the new location of the email. Think of a stub as a shortcut that contains information to point at the actual content. When a user clicks on an email that has been archived, the stub message is read and then the message is retrieved from the area where it has been archived. The idea behind stubbing is that it allows archived emails to be easily accessible to the end users. Rather than containing all the original emails that have been archived, a mailbox will contain stub messages which are much smaller.



However this concept is somewhat problematic. In August 2008 Microsoft published a White Paper on TechNet² entitled "Planning for Large Mailboxes with Exchange 2007". In this paper Microsoft described problems that occur when an email archiving solution makes use of stubbing. Archiving solutions use stubbing to try to solve the storage and performance concerns mentioned previously; however, in this paper Microsoft said that the reduced size of a message does not really avoid the problems that stubbing was meant to prevent.

Microsoft's paper explained how performance issues faced by Outlook users have more to do with large numbers of messages rather than large emails. Over time, an archiving solution working on hundreds of mailboxes will create thousands of small stub messages. Each of these stub messages may be between 2 and 15 kilobytes and still amount to a performance hit since item counts is the primary performance driver for the Exchange store rather than aggregate size.

² [http://technet.microsoft.com/en-us/library/cc671168\(EXCHG.80\).aspx](http://technet.microsoft.com/en-us/library/cc671168(EXCHG.80).aspx)

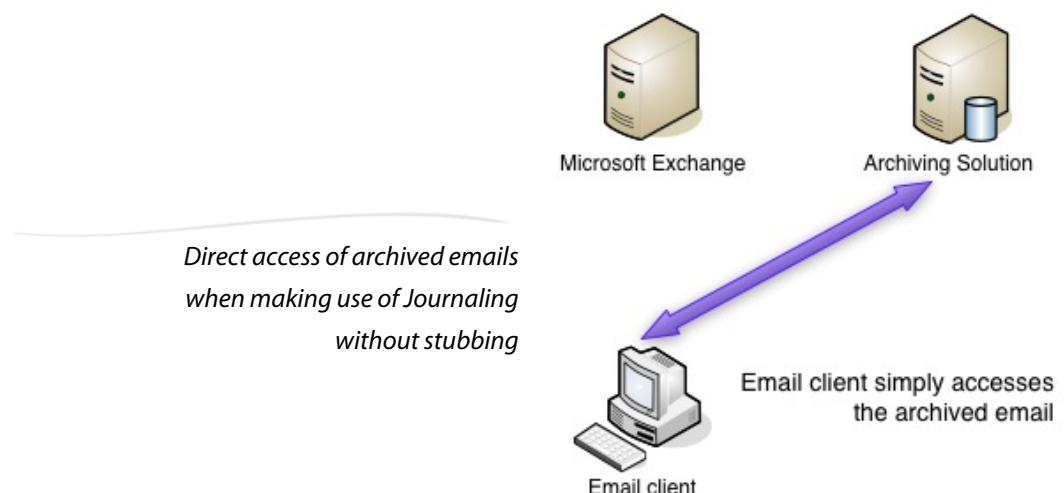
An email archiving solution that makes use of stubbing typically enumerates all emails in a mailbox and replaces the emails with a stub. Enumeration of emails is a processor intensive activity especially when this is done on each mailbox on the Exchange server. Processor intensive operations can have a negative impact on the performance of Microsoft Exchange especially when it is under load serving a large number of Outlook users.

The ability to search archived emails is another important consideration and functionality that makes an archiving solution usable. In the case of a solution making use of stubbing, the content of the original email is not available for searching. This means that users making use of Outlook's search functionality to find old archived emails will not get any useful results. Therefore typical archiving solutions that make use of stubbing render the built-in Outlook searching functionality useless for searching old emails.

Finally, stubbing also changes the way that Microsoft Exchange normally works because third party code has to be installed on the Exchange server itself to enable stubbing functionality. Experienced administrators know that the introduction of any new component to the system can easily affect the availability and reliability of the servers that they administer. This is especially of concern when the new component has to directly affect the way that Microsoft Exchange functions. Therefore system administrators are loathe to install software on their production servers simply to evaluate it, so they may have to create a test server which impinges on the administrators' already valuable time. An evaluation of such an archiving solution therefore will not reflect the load of the live Exchange server.

Journaling

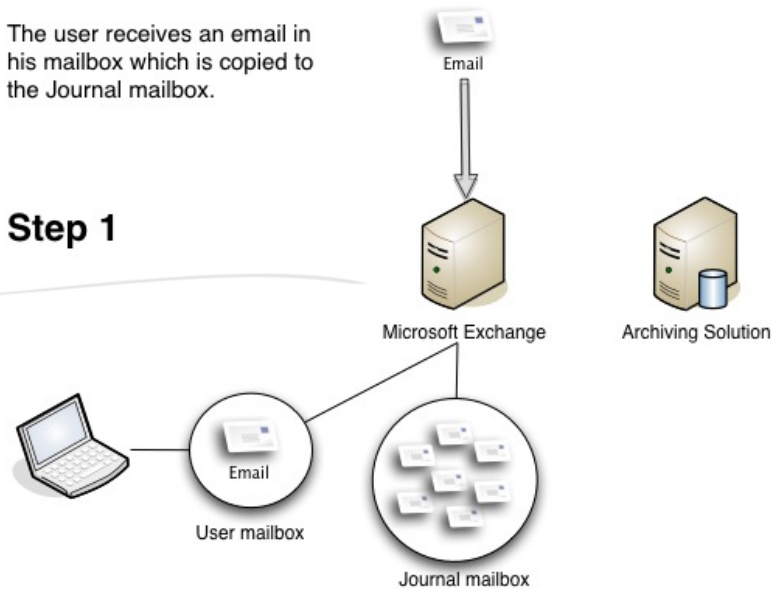
The disadvantages of stubbing can be altogether avoided by making use of a Microsoft Exchange built-in feature called Journaling. This feature provides the ability to record all communications within an organization and works by making available all incoming and outgoing content in a special location on the Exchange server. For archiving, it is particularly useful to make use of Envelope Journaling. Other Journaling options in Exchange only store the message contents and will miss important meta information such as distribution groups. With Envelope Journaling, Microsoft Exchange captures all emails' details that could be required for full compliance, including NDR emails and recipient information such as carbon copy (CC), blind carbon copy (BCC) recipients and members of a distribution group expansion.



Envelope Journaling alone does not provide a manageable solution. The advantage of this feature is that it allows the journaled emails to be fed into an archiving solution while using minimum overhead and requiring no additional code on the Exchange server. An archiving solution can then copy the journaled emails to its own database, clearing the Exchange server from the bulk of emails. Such a system separates email delivery from email archiving. Additionally, the Archive database can be stored on a totally different server. This design allows each component to do what it does best: the Exchange server delivering communications and a proper archiving solution dedicated to efficiently storing emails on a scalable technology such as an SQL database. An archiving solution can then provide additional invaluable functionality such as automated deletion of emails that are older than a specified timeframe and fast searching.

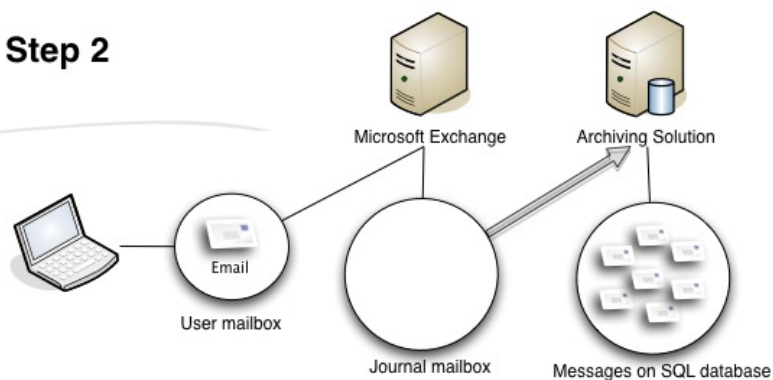
The user receives an email in his mailbox which is copied to the Journal mailbox.

Step 1



Emails are moved from the Journal mailbox to an SQL database, clearing the Exchange from old emails.

Step 2



How does stubbing compare with Journaling?

The following is a table that compares archiving solutions that make use of stubbing and journaling.

	Stubbing	Journaling
Easy to evaluate?	Requires vendor code to be introduced on the Exchange server.	Does not require additional software to be installed on the Exchange Server.
Usage of storage space	Creates thousands of small stub messages which add up over time.	Email messages are safely stored in a scalable database.
CPU usage on Exchange server	The process of enumerating emails and replacing them with stub files is CPU intensive.	All archiving activities are done on a system separate from the Exchange server.
Industry standard	Although used by many vendors, Microsoft does not recommend it.	Supported interface.

Conclusion

Archiving of emails allows organizations to manage their email messages in an efficient manner and comply with the regulations that might apply. What is more important is that different archiving technologies can affect the efficiency of an archiving solution. With reliance on email communications ever-growing, when choosing an archiving solution one would do well to address these concerns!