Network scanning overview

Network scanning for MFP's can be broadly divided into two categories that can be thought of as *pull* and *push.*

Pull scanning: the retrieval of the scanned document is initiated at the user's computer and is "pulled" from the scanner. The common pull methods are Scan to hard drive and Twain.

Push scanning: The scanned document is transmitted, or pushed, to a destination by the MFP using one of a variety of network protocols. Common push methods are FTP, SMB and E-mail scanning.

Pull scanning methods

Scan to hard drive: Also known as scan to box, this method creates a folder or "box" on the mfp's hard disk, that receives the scanned documents, which are then retrieved across the network from the user's workstation. The particular software that can be used to pull these scans depends on the controller in use. Applications include:

- Pagescope box operator.
- Konica Scantrip or Scandirect
- Fiery Remote Scan
- Web browser retrieval via the mfp's web interface.
- Twain driver (covered in more detail below)
- Making the MFP hard drive available as a network share, which can then be mapped to a shortcut on user's computers.

From a network configuration point of view, this is often the simplest to implement, and most reliable method of network scanning. This is primarily because of two reasons:

- MFP's normally have static IP addresses, and it is this address that is used with a pull method, so there are no issues resulting from workstations taking new IP addresses from a server or router.

- Since the scan retrieval is being manually initiated from the workstation, there are very few problems related to network permissions and security, which may block other methods as potential hacking attempts.

"Gotcha's" & "Nice-to-Know's":

- The MFP must have a hard disk on which to store the scans until retrieved. On many models, this is an option. This hard drive can be in the controller or the MFP itself, depending on the model. Consult the product configuration information for details on a specific model.

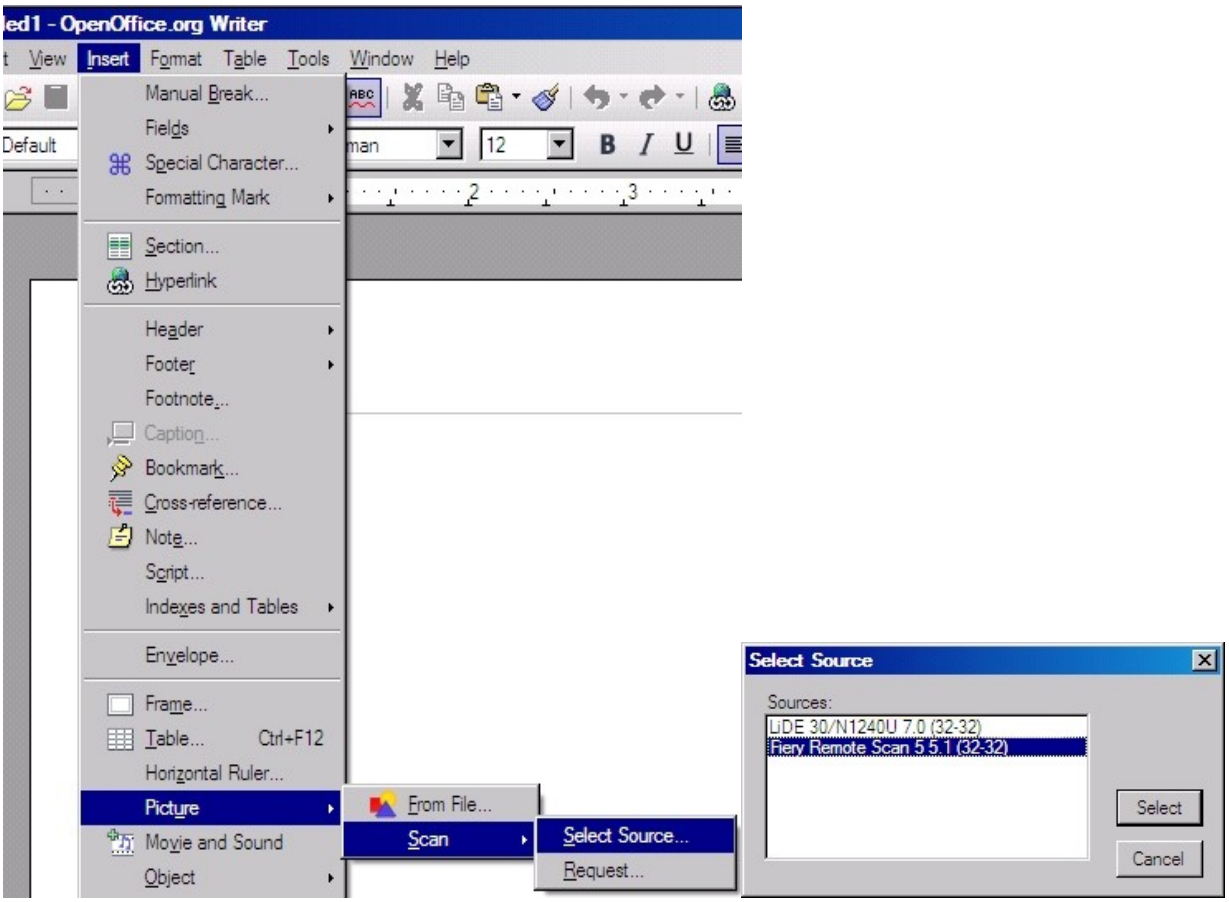
- Some retrieval applications can be set to automatically poll the MFP at set intervals, and save the scans to a predetermined folder. This can make the retrieval of scan fairly transparent, though it increases network traffic slightly.

- On some models, particularly those using Fiery controllers, the web browser method will only allow the document to be opened with the default application, rather than allowing it to be saved directly to the user's hard drive.

Twain: This acronym actually stands for *"Technology Without An Interesting Name".* The Twain standard is used in order to pull images from a scanner or digital camera directly into a software application that supports the Twain standard. Examples of some Twain compliant applications are:

- Adobe Photoshop
- Adobe Acrobat
- Microsoft Word
- Open Office

The exact menu selections for accessing a Twain scanner will vary according to the application in use. As an example, these are the menus within OpenOffice 2.0:



Selecting *"Insert-->Picture-->Scan-->Select source"* will then bring up a dialog box where you can select a scanner from multiple Twain drivers that may be installed on your computer. Selecting the scanner and then clicking *"Select"* will bring up the scanner driver software that the selected device uses.

"Gotcha's" & "Nice-to-Know's":

- Some Twain drivers will import a previously saved scan from a hard drive box on the MFP. Other drivers will initiate the scan from the computer, treating the MFP like a traditional flatbed scanner. Some drivers are able function in both manners.

- Due to its purpose of importing directly into an application, Twain primarily functions with TIFF (Tagged Image File Format) files.

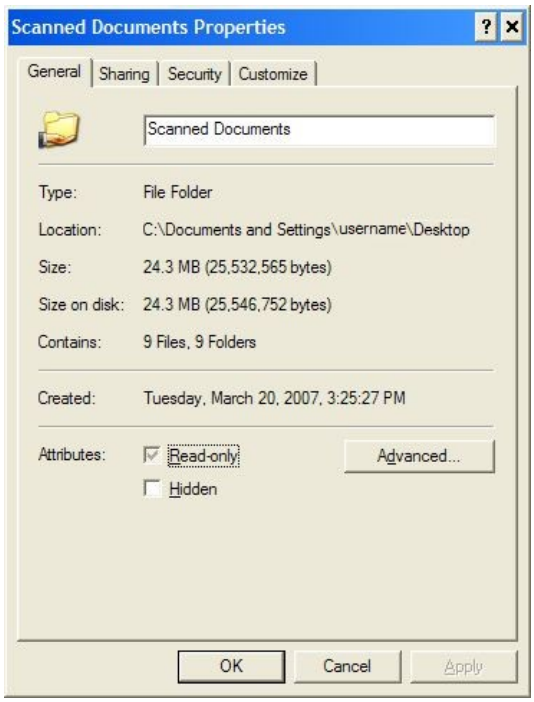
- Twain may be an appropriate choice for a customer doing Optical Character Recognition (OCR) to create text output, or wanting to immediately edit a scanned image within a graphics application such as Photoshop.

Push Scanning methods

SMB Scanning: This method sends the scan to a folder that is shared out on the network. SMB stands for "Server Message Block" and is a protocol used for file and printer sharing by all major operating systems today. SMB shares are normally visible by browsing the network, though actual access to the shares is normally restricted by the use of permissions.

In order to configure SMB scanning, you will generally need to know:

- The IP address or hostname of the computer to which you will be sending scans.
- The name of the network share to which you want the scans to go. As with FTP scanning, it is important to understand the difference between the local folder path and the share name. Some MFP's will ask for the file path in the setup dialog for scan destinations. This is NOT the local folder path on the PC, but rather, the share name. See the following figures:



These are two tabs of the properties dialog for a folder called "scanned documents" that is on my Windows desktop. The "general" tab shows the actual folder location on my local hard disk. The "sharing" pane shows the name with which this folder will be visible on the network. This is the folder path that will be visible by an external device that is accessing it, such as another PC or a network scanner.

- A username and logon password for the PC with rights to access and write files to the desired folder. This is usually the username of the computer user, but can be any account that exists on the computer with the appropriate permissions to the shared folder.

"Gotcha's" and "Nice to Knows":

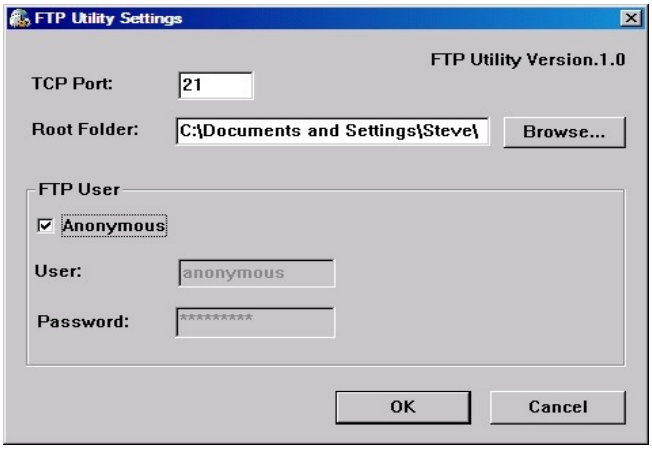
- SMB scanning is simple in theory, however in practice, be prepared to run into issues caused by locked down network permissions, especially when working in a domain environment. Domain environments and especially servers, often require SMB communication signing, which is not yet supported by many MFP's. For additional technical details, see Microsoft knowledgebase article 811497.
- A workaround for the above problem is to create a local user account on the PC, and use these account credentials for scanning, rather than using the normal domain login. A local user can be created with the management console, which is accessed by right clicking on "My computer" and then clicking on "Manage".

FTP scanning: With this method an FTP (File Transfer Protocol) site is created on a computer on the local network. The MFP then uploads the scanned document to the ftp site.

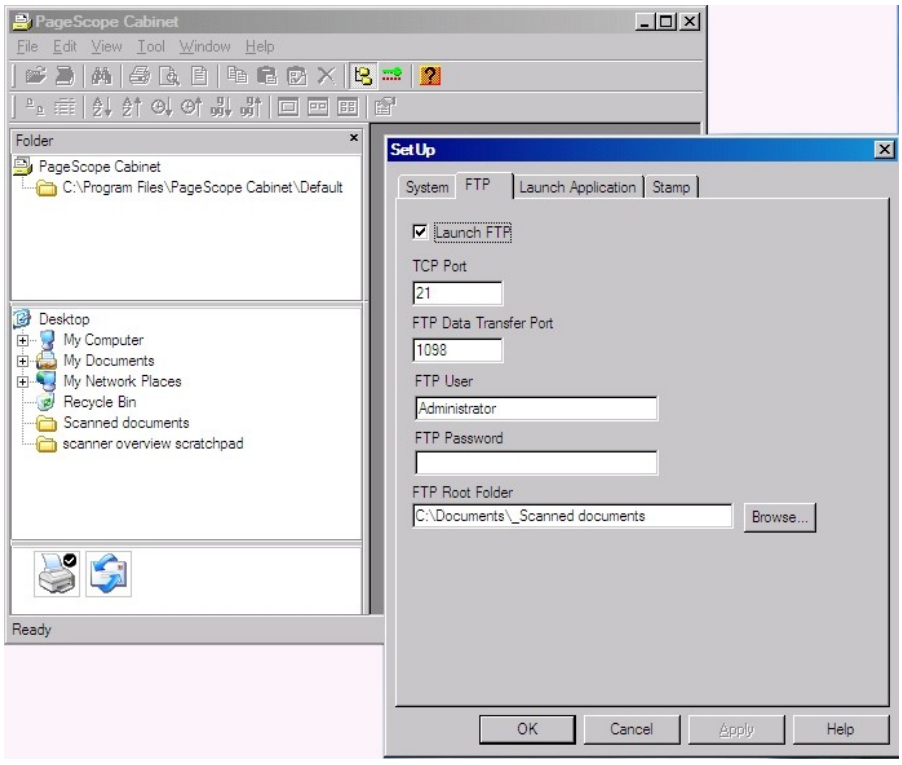
The FTP site can be created on a server and the scans folder then shared out over the network. If the shared folder and shortcuts to access them are created properly, this can be a very easy method for the end user, but requires some network configuration. From the user's perspective, they select the destination on the MFP and scan. Then the user goes to his computer, double clicks a shortcut to the scan folder and the scan is there in the opened folder.

Alternatively, FTP sites can be created on individual workstations, so that the scans are directly sent to the user's workstation. There are a number of programs that can be used to create this ftp server on a computer. Some of them are:

- The Konica-Minolta FTP utility- Included on some user software disks. This program is simple and straightforward. It creates a small FTP server on a client PC.



- Pagescope Cabinet- A document management application included with many Konica-Minolta MFP's. It has the ability to create an FTP server to receive scans and once configured, the FTP service can be run on its own if other functions are not needed.



- Internet Information Services (IIS)- Included with Windows 2000 and XP Professional, as well as with all server editions. Not included with XP home. This program is a little less intuitive to access and configure, but many network administrators like it as it is the FTP server that comes built in to Windows.

There are some concepts related to creating an FTP server that one needs to be aware of. Please refer to the preceding screenshots to help illustrate.

- The FTP root folder: This may be a folder several levels deep within the file structure of the local computer, however, when set as the root of the FTP site, this folder is what is visible to a user logged into the FTP server. It is the top level of the FTP site, regardless of where it is located in relation to the local file system. When setting up scanning on an MFP, the root of the FTP site is indicated by a forward slash: / You can set up subfolders beneath the root if needed.

- The TCP port normally used for FTP is port #21. You can remember this since 21 is the legal drinking age in most states, so FTP can stand for *"Free To Party"*

- The FTP server program must be placed in the Windows startup group, so that it runs when the user logs into their computer. Some FTP server programs can automatically place themselves in this group, others will need to have a shortcut manually placed in the startup program group. An exception to this is the version of IIS built into Windows, it runs as a service without needing a user logged in to execute it.

- Login accounts: The easiest access is to create an "Anonymous" account in the FTP server, which will allow access to the FTP site without any credentials. This is usually not a security problem as long as only the scanned documents folder is what is under the FTP site. If higher security on the folder is required, then a username and password can be created for the FTP site. The username and password will need to be entered in the scan destination setup on the MFP, or the selection made to tell the MFP to use anonymous access.

"Gotcha's" & "Nice-to-Know's":

- Most models are dependent upon the IP address of the computer running the FTP site. This is not a problem if the FTP site is on a server, as servers normally have static IP addresses. If scanning to a workstation, though, most networks use dynamic IP addressing and take their IP addresses from a router. This can result in the workstation periodically acquiring a new address and the scanning destination pointing to this address then will not be able to find it. The time period this address is kept for is called a DHCP lease, and is determined by settings in the network router that is handing out the IP addresses. The lease time can be determined by opening a command prompt and typing the command *"ipconfig /all"* If the lease time is very short (2 days or less) scanning can be expected to fail after the computers are rebooted following a long weekend. If the lease time is 3 days to several weeks, scanning will normally be reliable, but may still occasionally fail if a computer is turned off for an extended period. If a lease time is 3 weeks or more, then scanning can be expected to fail only if something forces a workstation to take a new IP address. The only 100% reliable addressing method with this scanning method is static IP addresses. Please see the glossary for more information on DHCP leases.

- Some newer MFP models can perform FTP scanning based on the hostname of the receiving PC. This solves the previously mentioned problem of computers taking new IP addresses from a router. This function requires DNS to be configured in order to resolve the computers name to it's IP address.

- Sometimes the FTP server will fail to run because another application may be using port 21. The FTP server software will allow the setting of a different port number. (see preceding screenshots) Port 2121 is suggested as an alternative if needed.

Scan to E-mail : In this method, the scan is sent out as an E-mail attachment, usually a .PDF file. In the simplest case scenario, you just need to preconfigure the MFP with the hostname or IP address of an outgoing mail server (SMTP server), enter the destination address for the scan, and the MFP sends it to the SMTP server which then forwards it on to it's destination through the network or internet.

Most mail systems use authentication, however, which adds some additional configuration to the process. Authentication requires that mail be identified as coming from a known and valid mail system user. You can use a username and password that already exists for someone, however this may not be preferred, since all scans done this way will be seen as coming from that individual. It is often beneficial to have a dedicated E-mail account for the scanner. If desired, many modern systems can be configured to receive incoming E-mail using POP. This gives the MFP an ability to print out any incoming E-mails, such as a "bounce" message indicating that an outgoing E-mail could not be delivered.

Some models can also have a number of E-mail addresses preconfigured, so that you can select who the outgoing mail will be seen as coming from. This has an advantage over using a common mail account assigned to the MFP in that any responses will go to that individual's own E-mail.

Some MFP's have the capability of authenticating a user against a network server. When set up this way, an individual logs in at the MFP with the same credentials they use on their own computer, and the scans are sent out as being from them rather than from a generic MFP address. This may be preferred in a network environment where security is more important. Setting this up will require the assistance of the customer's IT department.

Another feature becoming more common is LDAP search. This allows the MFP to search a network directory of information in order to find the E-mail address of someone. Configurations and naming conventions will vary, and the MFP will need to know where it is to search. As with the previous function, it will require the assistance of a network administrator who knows the appropriate settings for his domain.

"Gotcha's" and "Nice to Know's":

- In some environments, you may find a scan can be sent to an internal address within the company, but not to an external address at another domain. Some mail servers are set to behave this way when the outgoing mail does not have credentials for a valid user account. A user will first have to send it to their own E-mail account and then forward the message. Some network administrators and users may prefer this, as it keeps anonymous mail from being sent and the user directly receives any responses. There are two ways around this behavior if it is undesirable:
    - Configure the MFP with valid mail account credentials.
    - The administrator can set the SMTP server to allow unauthenticated outbound mail from the address of the MFP.

- Generally, scan to E-mail is the easiest and most reliable when the SMTP server is located on the local network. Some older controllers (Such as legacy Konica IP series controllers) work best when the mail server is on the same subnet as the MFP. Second best is a server that is remotely located, but still within the company domain. Mail servers that are located outside the customers control (for instance, the mail server run by their ISP) may or may not work easily and are often unreliable, as the customer has no control over the mail server settings, nor access to it.

- If the mail server environment will not permit the scans to go out no matter what is tried, a workaround is to use a standalone mail server application running on a PC on the local network. Two popular application are Postcast Server (http://www.postcastserver.com/)and Argosoft Mail Server (http://www.argosoft.com) . Both are freeware. These work by acting as the SMTP server and they forward the email off to wherever on the internet it needs to go.

- Web based E-mail systems (GMail, for instance) normally use the IMAP protocol to send E-mail. Most MFP's are unable to work with this method and require an E-mail system that supports the SMTP and POP protocols, If a customer is using web based E-mail, they will need to find out from their provider if SMTP/POP E-mail is available.

Proprietary Push methods: These compromise a variety of solutions, with varying capabilities and price points. The MFP sends the scan to a dedicated receiving application on a server or workstation and the scan is routed or saved in the appropriate location. Applications in this category include, but are not limited to:

- IP Scanner/Image Receiver, used with many Konica-Minolta and NEC products
- The Kyocera-Mita scanner file utility
- Ricoh Scanrouter

Due to the variety of proprietary scan methods and applications, it is not possible to go into any real technical detail concerning these.

Glossary

DHCP Dynamic Host Configuration Protocol. This is the protocol that is used for a network device to automatically request an IP address from a network server.

DHCP Lease A DHCP server will "lease" an address to a client for a predetermined amount of time, which is known as the DHCP Lease time. After 50% of the lease time has passed, the client will attempt to renew the lease with the original DHCP server that it obtained the lease from. Any time the client boots and the lease is 50% or more passed, the client will attempt to renew the lease. At 87.5% of the lease completion, the client will attempt to contact any DHCP server for a new lease. If the lease expires, the client will send a request as in the initial boot when the client had no IP address. If this fails, the client TCP/IP stack will cease functioning. The lease time can be found from a command prompt by issuing the command "*ipconfig /all"* and looking for the fields labeled "lease obtained" and "lease expires".

DNS Domain Name System This is the system that is used to resolve a name, such as *www.google.com*, to an IP address. The computer will contact its default DNS server to do this. The default server is normally located on the local network or at the Internet Service Provider. If the default DNS server does not know the address of the requested resource, it passes the request upstream until the name is either resolved or it is determined to not be available.

Domain There are some variations of this definition, depending upon the context it is used in, but for purposes of MFP's it is a group of computers or addresses that belong to a named network entity. For instance, the E-mail addresses of *joeblow@bigcompany.com* and *susieq@bigcompany.com* are within the *bigcompany.com* domain, as are the computer hostnames of *server1.bigcompany.com* and *susiesmac.bigcompany.com*

Dynamic IP A dynamic IP address is one that is taken from a DHCP server. A dynamic IP address is subject to changing either if the lease expires or something happens to force the device to request a new address. Tip: if a device has taken a dynamic IP address starting with 169.254, it indicates that it was not able to contact a DHCP server, and so has picked a random address from the 169.254 range. This could indicate a network problem, or it could be as simple as a patch cable that is not plugged into a jack.

FTP File Transfer Protocol. This is an efficient means of transferring files across a network or the internet. It requires an FTP server application running on a host computer. An FTP site can be accessed with a web browser, or more effectively with a dedicated FTP client application. Most modern MFP's have the ability to upload scans to an FTP server.

Hostname A name by which a network device can be accessed. This is much easier for humans to remember than IP addresses. Hostnames generally do not change on their own, whereas IP addresses in most networks can. An example of a hostname using DNS might be: *bobscomputer.bigcompany.com. bigcompany.com* is the overall network domain, while *bobscomputer* is the individual client on that network.

IP Address A network address assigned to a device and used to address communications to it. An IP address consists of four segments known as "Octets". Even though these Octets are a maximum of three characters long, they are called octets because of being decimal coded binary. In binary, each octet is a string of eight 1's or 0's. An example of a common IP address would be 192.168.1.105

ISP Internet service provider. This is the company that you gain access to the internet through. For home or small office users, this is usually a company such as Earthlink, Charter, AOL, Comcast, Etc.

JPG / JPEG An image file format. JPEG stand for “Joint Photographic Experts Group”, which is the name of the committee that created the standard. The commonly used file extension is .JPG. Jpeg images have (relatively) small file sizes due to the use of a “lossy” compression. What this compression algorithm does is to look at areas of similar color and if it judges them to be the same, it encodes it once and when displaying, repeats that code for blocks that are determined to be the same. The amount of compression can be adjusted when saving an image in most image editing programs. The choices are usually labeled as “size” and “quality”. If an image is saved with a high level of compression, you can see blocky artifacts usually visible around the the edges of solid areas. With a modest level of compression, a Jpeg can be perceived as crisp and clear, but will be a fraction of the size of the same image as a TIFF. Most current MFP’s can produce Jpegs, though some older models may not.

LDAP Lightweight Directory Access Protocol. LDAP is used for accessing a network directory service. For MFP’s, it’s use is to search a directory for a persons E-mail address. LDAP uses port 389

MFP Multi Function Printer (or Peripheral). We once called them “Copiers”.

OCR Optical Character Recognition. The nature of any scanner is that it only produces an image of the page. Even a PDF that is output from a scanner is only an image embedded in the PDF file. If a customer requires a document that can be edited in a word processor, then OCR is needed. OCR recognizes patterns of text and will output a file in a choice of common formats. OCR accuracy is only as good as the original document and it will have great difficulty with skewed and distorted copies of multiple generations. On clean, straight, first generation text, however, modern OCR is very accurate, and becomes more accurate as it is used and learns. One of the more popular OCR packages is Omnipage from Nuance software. Many OEM’s offer their own branded versions of this or similar software, optimized to work with their own products.

Full versions of Adobe Acrobat also have an OCR capability built into them, called “Paper Capture” in some versions and “OCR text recognition” in later versions. This function is more limited than a dedicated OCR package, but will create a text layer within a PDF file so that minor text edits and copying can be done. If Acrobat does not recognize something as text it will simply leave it as an image.

PDF Portable Document Format. A file format created by Adobe. The PDF format is based on the Postscript page description language and is a file type that retains the appearance of the document, regardless of what platform it is viewed on. If a JPG or TIF can be considered the digital equivalent of a photograph, then a PDF can be considered the digital equivalent of a book or magazine. Originally created and manipulated with Adobe’s Acrobat software, the format has now been licensed to many others and has become a defacto standard. PDF files can be created with most recent MFP scanners, as well as numerous software applications from third parties.

POP Post Office Protocol. POP is used by a client retrieve E-mail from a mail server. The commonly used port number for POP is port 110

Protocol A set of accepted communication rules used to enable two devices to work with each other. Some common protocols are TCP/IP, FTP, SMB and DHCP.

Router A router performs forwarding of communications and translation of IP addresses between separate networks, so that communications can pass between them with the networks still remaining independent of each other. In a small office or home environment, the router is what separates your own network from the internet, but still enables you to communicate with other hosts through it. In a small environment such as this, the router will also act as a DHCP server. In large corporate environments these functions may be assigned to dedicated servers.

Service A utility or function that loads and runs upon bootup of the computer, prior to any user logging in.

SMB Server Message Block. This is a file and printer sharing protocol that is supported by all modern operating systems. When you are browsing a network for a shared printer or folder, this is the protocol that is being used. Many MFP's have the ability to send files to a folder that is shared out using SMB.

SMTP Simple Mail Transfer Protocol. This is the protocol that is used to send outgoing E-mail, and to transfer it through the internet to get where it needs to be. The commonly used port number for SMTP is port 25

Static IP An IP address that is permanently assigned to a computer or device. A static IP address can be changed, but it must be done manually, through the setup function for the devices network interface.

Subnet A subnet is a logical, and sometimes physical, division of a large network into smaller, more manageable sub-networks. A single department might be on one subnet, with network bridges or routers in place to allow access to other subnets within the larger network. Many IT departments in large networks will set aside a separate subnet for printers and MFP's.

TCP Port A TCP Port is not a physical port, but rather a logical concept used to access resources. A port might be more accurately thought of as a channel. A network device will listen on these ports or channels for certain types of activity that is expected. For instance, the common port for FTP communication is 21, Outgoing E-mail uses port 25 for the SMTP service, and many printers use port 9100 or 10001.

TIF / TIFF An image file format. TIFF stands for "Tagged Image File Format". The common file extension is .TIF. TIFF files are uncompressed and are far larger than other, compressed file formats. Because of this, graphics professionals prefer them for their higher quality, but they are generally not desired for E-mail transmission. Most MFP scanners can produce TIFF files.

TCP/IP Transmission Control Protocol/Internet Protocol. Although usually referred to together, this is actually two protocols that function together. IP handles the addressing of network devices, while TCP handles the establishment and negotiation of communications sessions.