Architecture of the ScriptServer PAN Printing System

≡ *Introduction* ≡

The ScriptServer PAN Printing System has been designed to provide a high quality production printing capability in a modern distributed computing environment. The computer programs that provide the printing services run on Windows NT systems, Unix systems, or a combination of both. The printing "clients" that use the services can be running Windows, Unix, Mac OS, or any operating system that supports a web browser or the LPR/LPD protocol.

The system has been designed to appear to users as a service available on their local area networks. It has a high quality, high speed graphical user interface, accessed via a Web browser, through which users can submit and monitor the progress of their print jobs and system administrators can configure and control the printing system. In addition the system includes command shell interfaces for both Unix and Windows NT systems, support for print job submission using LPR compatible clients, and an interface for Windows NT applications through the "print" command on the "file menu."

The system will print preformatted PostScript or PCL files on appropriately capable TCP/IP connected printers. The TCP/IP connections may be either direct or use the LPD protocol.

Printing options include the number of copies to be printed, simplex or duplex printing, media selection and other printer specific processing options, optional flag and trailer pages, and a job scheduling priority.

For files which are not preformatted, the system includes an ASCII to PostScript conversion facility. In addition to the printing options listed above, this facility implements landscape or portrait orientation, control of margins, fonts, line spacing, page headings, page numbering, line wrapping, and tab emulation.

Finally, the system includes a complete PostScript Soft Form capability which can eliminate the need for preprinted forms.

It is the purpose of this document to provide an overview of the architecture the ScriptServer PAN Printing System. The document discusses the essential design ideas, explains the concepts on which the system is based, and provides a rationale for the design. Other documents are available which discuss the specific features of the system in detail.

≡ *Why "PAN"?* ≡

"PAN" is an acronym for "Print Area Network." Similar to the concept of Storage Area Networks (SANs), where the location of data is abstracted from the host on which it resides, ScriptServer PAN abstracts the printer/host association and moves the queues associated with printers into a centrally managed network of print queues.

Several trends in printing technologies lend themselves toward the need for a centrally managed network of print queues such as that provided by ScriptServer PAN:

- printers are more complex – with features such as multiple input trays, duplexing, and stapling becoming commonplace and available at lower cost

- printers are faster – users get their output much faster than just a few years ago; however, when things go wrong, it can be difficult to determine that a problem has occurred before an incomplete job has been deleted from a queue or a ream of paper has been wasted

- printers are networked – since they are directly accessible from any computer in the network, management and control of the printers is decentralized; in many cases, each computer accessing a printer will have its own queue, operating independently of other queues on the same printer. These queues connect on a "catch-as-catch-can" basis; there is no arbitration between queues.

- printers have full graphics capabilities – this enables the use of electronic or soft forms which can be printed concurrently with the information they contain and eliminates the requirement for more expensive pre-printed forms.

It is interesting to note that in this printing system, there are no "printer" objects. To PAN, printers are merely the physical hardware that queues perform their work on.

≡ *Architecture* ≡

The design goal of the ScriptServer PAN Printing System is to significantly improve the performance and lower the cost of administering medium to large scale distributed printing systems.

This goal has been achieved through an advanced product architecture. This architecture, which is specialized for supporting high volume, high quality production printing, has been developed over a number of years and is based on our experience and the experience of our customers with both batch and desktop printing systems.

Although the system consists of a number of software modules that can be distributed across several host computers, it is designed for centralized management and control. The comprehensive Web browser based graphical user interface provides a uniform means for interacting with the system for both end users and system administrators. The printing model used by the system embodies a powerful new set of abstractions for specifying both the capabilities of printing equipment and the processing requirements of print jobs.

For example, the heart of the architectural model employed by the ScriptServer PAN Printing System is the concept of print "areas." A print area is logical grouping of co-scheduled print queues. A collection of queues may be grouped into a print area based on similarity in printer capabilities, on physical location of the printers, or on similar operational of functional requirements. Any number of individual print queues and print areas may be configured concurrently on the same host computer. Also, any number of hosts may be participating in a ScriptServer PAN network. This design results in a highly scalable implementation that can,given sufficient network capacity, support thousands of disparate printers across dozens of hosts while retaining a central point of management.

The combination of these architectural elements has resulted in a design which attains a substantial reduction in the complexity of applying and managing a distributed production printing system while at the same time providing access to the complete range of capabilities of modern, full-featured, high performance printers.

≡ *Centralization* ≡

Traditional distributed printing systems have been based on a printing model which consists of the following basic elements:

1. "Physical printers."

2. "Printer driver software" specific to each type of printer.

3. "Queues" — often more than one — for each of the printers.

4. "Jobs" — which are really "files" generated in one way or another in one format or another.

5. "Hosts" on which queues reside and from which jobs are submitted and printers are driven.

6. A set of "commands" or some other kind of "user interface" for submitting jobs and monitoring their progress. These commands often specify the settings of device dependent parameters.

7. Another set of "commands" and associated "configuration files" for administrators of the system.

8. A "network" of some kind connecting hosts to hosts and hosts to printers.

   Note that this model does not identify any central elements other than the network. Therefore, it is not so much a "distributed printing system" as a "distributed set" of printing system components. This is a perhaps subtle but nevertheless important distinction. As systems begin to scale up in size, this lack of central organization can become a significant cause of administrative difficulty and loss of reliability.

   Therefore, the PAN architecture adds the following two new basic elements to the general distributed printing system model:

9. A centralized database which specifies the static structure of the printing system. This database contains all system configuration information. This configuration is served on request to any distributed component requiring it.

10. A centralized database that contains the complete real time dynamic state of the system. Information in this database is also served on request to any system component that requires it.

These two databases form the structural nucleus of the PAN printing system. Their presence provides the system with a pronounced sense of "identity" or "one-ness" so that the PAN system appears to both administrators and users as a single entity. No matter how many hosts are participating, how many queues have been configured, or how many printers are connected, there is just one PAN system and "it" is what both users and administrators see.

≡ *User Interface* ≡

Viewed most straight forwardly, the PAN system is a Web site. It is a Web site that provides access to and "controls" a network of distributed printing facilities. The primary access to this Web site is through a standard Web browser.

The implementation of the browser based PAN user interface is greatly facilitated by the centralized data bases mentioned above which collect all of the information about both the structure and state of the printing system in one place. The user interface provides immediate access to this information through one uniformly designed and related set of screens. From both the user and administrator point of view, the design of the interface effectively centralizes control of the distributed set of printing facilities and provides a single method for accessing all of the capabilities provided by the system.

The browser interface is state of the art dynamic HTML. Extensive use of JavaScript provides multi-frame displays, rollovers, tabbed dialogs, interactive prompts, instantaneous parameter range and consistency checking, a method for saving and recalling sets of parameter settings, and a database of user preferences. Current status information is available for all printing jobs that includes the page currently being printed, the bytes remaining to be printed, and an estimated time for job completion. These rapidly varying components of state information can be viewed in real time on the browser screens using a variation of server push technology.

Note that a comprehensive graphical user interface is the hallmark of the so-called "vertically integrated application" which has been the key to the explosion in productivity through the use of LAN based workstations. This design approach, which provides access to the full set of an application's features in a single interactive control regimen, is usually considered to be best approach for building products which are easy to learn and "user friendly." However, applications that must rely exclusively on such an interface can sometimes suffer from a lack of sufficient interoperability and flexibility to handle special situations.

To avoid this problem, the PAN system augments its browser interface with several alternatives for both job submission and system administration. For example, the system can accept jobs from any Windows NT application, or any LPR client. Also, it has a complete "command line" interface so that it can be fully controlled from shell scripts or perl programs.

Finally, PAN does not use proprietary "internal" file formats but has been designed to use only standard "interchange formats." For example, in addition to being able to print jobs consisting of ASCII, PostScript, or other standard file formats, all of its work files, data bases, configuration files and event logs are plain ASCII text which can be readily processed by other applications, edited, or searched directly using standard text processing programs such as vi and grep.

≡ *The PAN Print Job* ≡

Job submission in the PAN system involves combining three essential entities in a "print" operation. These three entities are the data that is to be printed, the queue to which the job is to be submitted, and the processing specification that controls aspects of the printing process.

The physical representation of a print job in the PAN system is a pair of files. One of these, which we call the "data file," contains the data that is to be printed. The other, which we call the "work file," contains the specifications that control the printing process. The processing specifications are based on the options selected when the job is submitted. These specifications are "abstract" meaning that they specify only "properties" of the printing process; they do not specify particulars of device control.

In production printing applications it is often the case that a job requiring a number of special option selections is submitted periodically as part of a company's normal operation. In these cases, encapsulation of this set of option selections is the key to ensuring dependable results. The PAN system provides a very general mechanism for submission option encapsulation called the PAN form. Conceptually, a PAN form may be thought of as a "set of instructions" which can be used to insure that particular data files are always printed in exactly the same way.

≡ *Queues and Area Queues* ≡

In the PAN system the only job processing abstraction that the user sees is the "queue." One always submits a job "to a queue." One never prints a job "on a printer." Therefore, all print queues in a PAN installation are global and have a unique name. While every queue must reside on some host, specific hosts are not referenced during job submission. As in the case of particular printers, there is no way to refer to a particular host during job submission.

Of course, a physical printer is associated with every queue and it is possible for more than one queue to share a printer. However, it is never required that more than one queue share a printer. Since queues are global, all participating hosts may submit jobs over the network directly to the particular host that is providing a particular queue. This differs from systems in which each submitting host has its own queue for each printer on the network.

One consequence of this difference is that it is easy to see at any point in time exactly how much work has been queued for each printer. A second consequence is that scheduling of the jobs in the

queue is uniform and unpredictable scheduling caused by "races" between hosts for access to printers is eliminated. A third consequence is that queues may be readily moved from one host to another without affecting anything in the user interface.

The PAN system regards the queue itself as active. That is, it is "the queue" which actually does the work — not the printer. The printer that is associated with a queue is "utilized" by the dequeing process for each queue but aspects of the printing process itself are controlled by properties of the queue.

A realistic practical example of how this capability is used must wait until more of the system has been explained. For now, as a few very simple examples, it is possible to configure a queue which will always shift ASCII files submitted to it five spaces to the right. It is possible to configure a queue that will always print on canary paper. Of course, it is possible to configure a queue which will do both, that is print the job right shifted five spaces on canary paper.

It is also possible to associate many other much more complex operations with queues. For example, the PAN system implements Soft Forms which are described below. A Soft Form such as company letterhead can be associated with a queue and then anything printed on that queue will come out on company letterhead.

≡ *Area Queues* ≡

Most simply stated a PAN Area Queue, which we call an Area, is a co-scheduled set of regular queues. Instead of submitting a job to a particular queue in this set, one can submit a job to the "area." When this happens, the job will be selected for printing by the first queue in the set to become free.

As an example of how concept this might be applied, note that many organizations tend to locate the printers used by particular departments at a single physical location. For example, suppose a department has two printers and they are both in a room called, "the East Wing IO Lobby." A PAN administrator can configure an area called EastLobby that contains two queues, one for each of the printers.

From the point of view of job submission, EastLobby just becomes, in effect, another "queue." When you submit a job to EastLobby it will print on the first of the printers in the area to become available.

The PAN system implements two new concepts that are the foundation of the PAN printing model. Used in combination, these concepts allow the creation and specification of abstract job processing specifications. The PAN "resource" is a global identifier established by the system administrator. This identifier is used to designate either an attribute of a print job or a capability of a queue. For example, a resource may be a media selection choice, a finisher option, an encapsulated set of submission options, or the background image for a Soft Form.

Resources are used by system administrators to compose forms and a PAN "form" is a "set of PAN resources." A job which through its form specification expresses a "requirement" for resource X should be submitted to queues for which a definition for resource X has been provided.

Much more will be said about PAN forms and PAN resources in the following section. For now, the important point is that the key to understanding the concepts of forms and resources is to realize that the simple and intuitive description of the area queue as a co-scheduled set of regular queues over looks the following critical consideration:

A general implementation of area queues defined in this way requires that job processing options be specified abstractly.

Here is an example that shows why this is. Suppose that the two printers belonging to the department mentioned above each have multiple trays. Printer A has canary paper in tray 2. Printer B has canary paper in tray 4. If one submits a job to printer A, in order to get canary paper one must "specify" tray 2. Similarly, if one submits a job to printer B, in order to get canary paper one must "specify" tray 4.

But if the user has to specify the tray with some printer configuration dependent job submission option, for example, "tray=2" or "tray=4," what happens when these two printers are grouped into a PAN area like EastLobby? It is not known at the time the job is submitted to the area queue which of the printers A or B will be used to print it.

The implementation of PAN areas thus requires that media selection be expressed abstractly at job submission time. The same is true for the whole list of printer dependent submission options.

Therefore, in the PAN system, a command something like

```
print /area=EastLobby/form=Canary job.data
```

can be used to provide the solution. When the job is "submitted" it is copied into the possibly remote area queue EastLobby. Part of the submission is a record of all submission parameters which includes, in particular, the option /form=Canary.

The system administrator will have defined this form to include a "media selection resource" for selecting a tray containing canary paper. At the time the job is dequeued, which is to say actually printed, the definitions of any resources specified by the job's form parameter are looked up in a data base of resources which is rooted at the queue chosen for processing, that is, in this case, either the queue for printer A or the queue for printer B.

These definitions of the resource, "select_canary" say, are different in these two cases. In particular, the one for printer A contains a tray selection command which selects tray 2 and the one for printer B contains a tray selection command which selects tray 4. Providing these "specializations" of the definition of "select_canary" which correspond to the particular queue being used to print the job permits the media to be specified abstractly as simply "Canary" when the job is submitted, while the determination of the concrete printer dependent tray selection command is deferred until the job is printed.

In summary, implementation of the PAN area queue which allows uniform parallel processing of jobs by a group of printers of possibly different types and capabilities entails the implementation of abstract job processing specifications.

≡ *PAN Resources and PAN Forms* ≡

Considering the example from the previous section in more detail, in order to solve the general problem of selecting "canary paper" across a range of different printers at least three things have to be done:

1. The print job itself has to be able to specify that "canary paper" is desired.

2. The "system" has to have a record of which tray the canary paper is in on every printer that has it.

3. The specific "tray select" command string for selecting this tray has to be known for each of these different and possibly even different types of printers.

Note that the above list is in a particular order — it goes from the submission command "inward" so to speak toward a higher degree of implementation detail. This is the natural way in which the features of a system are presented to an end user. However, from the point of view of

understanding the mechanism, a clearer discussion results by considering this list in the reverse order, that is, from the implementation "outward." This is the point of view that must be adopted by a system administrator.

Taking this latter approach then, to solve the second and third problems the PAN system administrator has to "define" the abstract concept of "canary media selection." This takes two steps: 1) give this concept a name, and 2) associate this name with an appropriate meaning.

It is very important to realize that this definition is not simply a single "attribute-value" pair of the type usually found in a standard manufacturer or device specific PPD. Instead, the PAN resource is a more general cross product and site configuration dependent concept. The definition of a PAN resource will typically have several specializations, one corresponding to each printer or type of printer expected to implement "printing on canary."

In general resources are defined in a three level hierarchy. The root level of resource definition is the "queue" level. In this case the resource definition is associated with a specific print queue.

The next level is called "shared" in our terminology. This level is used to define resources that are expected to be shared by more than one print queue. The typical application of shared resources is to define device specific resources, for example finisher option selection commands, for all printers of a certain type.

Finally, the most general level is "global." This level is used to obtain the definitions of resources not defined at any other level.

In the PAN system every "print job" is associated with a "PAN form." If no specific form is specified when the job is submitted it will be associated with the "default form" defined for the queue from which it is printed. We say in general that a "PAN form" is a collection of "resources." (More precisely a PAN form "specifies" a collection of resources which are to be included with the job when it is printed.)

In general, a "resource" is a set of information which can be one of five types —

- Layout
- Finisher
- Image
- Media
- Other

Layout resources generally specify page layout information to be used by the ASCII to PostScript translator such as margins, number of lines per page, fonts, font sizes, and so on. They may also be used to encapsulate a number of other job submission options.

Image resources specify background images to be used with Soft Forms (not to be confused with the more general PAN forms) which are discussed in their own section below.

Media resources specify command strings for media selection.

Finisher resources specify command strings that set finisher options.

Other resources are command strings of unspecified purpose.

The PAN system administrator is responsible for defining all the resources used in each installation.

To continue the example, suppose printer A is a Konica 4060 and printer B is a Tek Phaser 740. Suppose the 4060 is associated with a queue called Speedy and the 740 is associated with a queue called Pretty. These two queues are both members of the print area EastLobby (which is itself a queue). The Konica has canary paper in tray 2, the 740 has canary paper in its "middle tray."

The system administrator wants to define a form called "Canary" which users can use to get their jobs printed on canary paper.

The idea is to have the command line command

```
print /area=EastLobby/form=Canary job.data
```

(or equivalent browser interface submission command) print job.data on either the Konica 4060 or the Tek 740 and have the output on canary paper.

The exact details of operating the browser interface to the resource editor are given in the PAN Reference Manual. Using this interface it is a routine process to define select_canary so that it looks something like this —

Queue = Pretty

```
Resource select_canary =

<</MediaPosition 1
   /MediaClass null
   /TraySwitch false
   /ManualFeed false
>> setpagedevice
```

Queue = Speedy

```
Resource select_canary =

   statusdict begin 2 setpapertray end
```

With these definitions in place, the first problem, that is, making it possible for the user to specify canary paper, can be solved by defining the form Canary as follows:

```
Form Canary = (select_canary)
```

That is, the form Canary consists of the single resource select_canary.

As an example of a form which contains more than one resource, suppose a layout resource, right_shift_5, is defined which will shift an ASCII file five places to the right. To make this capability available to the user community, the system administrator might define a form called Right5 as follows:

```
Form Right5 = (right_shift_5)
```

Users could then get this effect by

```
print /area=EastLobby/form=Right5 job.data
```

To make the effect of both the select_canary and the right_shift_5 resources available the system administrator would define

```
Form CanaryRight5 = (select_canary,right_shift_5)
```

Here the form is a set of two resources and

```
print /area=EastLobby/form=CanaryRight5 job.data
```

will put the output on canary paper, shifted five places to the right.

Note that forms and resources are separate tiers of the hierarchy of abstraction. Form names are "published" by the user interface of the PAN system. That is, the set of forms that an installation has defined determine the characteristics of the printing facility seen by the end users sees. This is an inherently much simpler description of an installation's capabilities than a list of all of the device dependent features of all of its printers.

The resource names, on the other hand, are not published. These abstractions are intended to be created and used primarily by system administrators for the purpose of composing and "publishing" forms to be used by their user community.

To assist in meeting special requirements, the command line interface supports job submission with unnamed forms. For example, it is possible to say

```
print /area=EastLobby/form=(select_canary,right_shift_5) job.data
```

given that the indicated resources have been defined.

In summary, a PAN resource is an installation defined feature of the printing process that can be referred to by name. Resources are general because they can be tailored to an installation's particular needs and powerful because they can have multiple definitions.

Individual resources are not intended to be used as job submission parameters. Instead, particular combinations of resources which are to be included with a job are collected in installation defined sets called PAN forms.

PAN Forms allow system administrators to present an installation's printing facilities to their user communities in a convenient, device independent manner that can be made highly compatible with an organization's operational requirements.

≡ *Soft Forms* ≡

One of the most important and also one of the most complex production printing applications is the so-called "soft form." Many companies use pre-printed forms for quotations, invoices, letterhead, and so on. Since modern laser printers have complete graphics capabilities it is possible to eliminate the pre-printed form and just print the form background image and the data on the form at the same time. A form produced in this way is called a soft form and successful deployment of soft forms can save customers using large numbers of forms a considerable amount of money.

The ScriptServer PAN Printing System has been specifically designed to support the printing of high quality soft forms and it is in the printing of soft forms that the full generality of the design of the system is utilized.

Note, however, that our system is a "printing system" which facilitates the use of soft forms; it is not a turnkey soft form system. In particular, we do not provide software to do the artwork for the soft form images. These images, which we expect to be supplied in encapsulated PostScript format, may be produced by a range of high quality graphics programs such as Adobe Illustrator, Photoshop, CorelDraw, etc. Once produced, soft form images are stored in our resource database as "image resources" which may then be associated with one or several PAN forms.

Usually, the text that is printed on a pre-printed form is ASCII and it is generated by one of the customer's report generators. It is the job of the report generator to make sure the text "lines up" on the form correctly. However, in many cases in practice, when the change is made from the pre-printed form to the soft form image, it is found to be desirable to make certain modifications to the form image and it therefore becomes necessary to make corresponding adjustments in the positioning of the text.

The PAN printing system includes an ASCII to PostScript translator which has a full range of positioning, scaling, rotating, and font selection capabilities. This allows adjustments to be made to the text without having to modify the report generator. The parameter settings needed to control the translator for each soft form can be encapsulated and saved in a "layout resource." This layout resource, along with the image resource specifying the background image can be combined in a PAN form.

Finally, often special paper, specified by a media resource, and finishing options, specified by a finisher resource, are also needed by a soft form. All of these resources can be combined in a single PAN form.

When this form is specified on a job submission, given that the definitions of the constituent resources provided by the system administrator are sufficiently general, the proper versions of the background image, the corresponding layout adjustment, media selection, and finisher options will be used no matter which physical printer is used to print the job.

In particular, if the resources contained in a form have been given definitions for each of the queues comprising a PAN area, then jobs requesting that form may be submitted to the area.

This has two important advantages for larger organizations. First, the fact that print areas contain multiple printers means that high volume is possible when printing large numbers of forms. Second, reliability is greatly improved because the failure of a single printer in the area does not prevent continued production of the form.

Finally, note that a particular PAN form which results in printing on a Soft Form may be established as the "default form" for a particular queue. In this case, any job sent to that queue will automatically come out on the form. One use for this might be to dedicate a particular queue to printing on company letterhead.

☰ *Conclusion* ☰

The ScriptServer PAN Printing System has been designed to solve a general problem that is known for its many diverse and often obscure complexities. The product is therefore itself a fairly complex system with many aspects and features that cannot be easily or completely described from a single perspective.

The document has covered the main ideas used in the design of the system and, in summary, the single most important idea embodied in the PAN system is the idea of abstraction.

This idea has been applied to create two distinct views of the printing system — one appropriate for system administrators, the other appropriate for end users. The system administrator's view includes specific hosts and printers. The system administrator's view also includes "resources" which are site dependent abstract printing system capabilities whose definitions may include strings of PostScript or PJL code.

However, in general terms, the "system" as seen by end users consists mainly of simply a list of queues and a list of forms.

In conventional printing systems the "print queue" is simply a list of files waiting to be printed. In the PAN system, the PAN queue itself is a very powerful abstraction: a PAN queue is really an abstract "print plant." This plant takes on "jobs" consisting of a data file to be printed and a "form" saying what is to be done. This form is expressed in abstract, device independent terms. Some of these terms are defined by the printing system and others are the resources defined by installations.

The range of power of this "abstract print plant" known to end users as the PAN queue goes all the way from simply directly copying the data file to the printer associated with the queue using the default medium to performing a number of transformations on the data, combining it with a Soft

Form image, printing it on specifically selected media and providing particular finishing. Complete control of this entire process is provided through a single encapsulated set of parameters called the PAN form.

Therefore, when one submits a job "to a queue" "using a form" a considerable amount of interpretation of the request is done automatically by the system. This interpretation is based on definitions provided by the installation and tailored to their requirements. The result is a substantial reduction in complexity and a related significant increase in reliability for a broad range of high volume, high speed, high quality production printing applications.