Large Scale Parallel Print Service

I. Deloose, D. Foster, I. Reguero

CERN, Information Technology Division, Geneva, Switzerland

Abstract

During several years, the CERN PRINT service consisted of a wide variety of printing server platforms handling more than 1300 network printers. Whereas a single UNIX printer server mainly served the 5000 UNIX clients, printing services for the 3500 Windows 9x and 1500 Windows NT PCs went mostly through Novell and NT servers. At the end of 1997, the CERN Printing Project was launched to define a single server model for all clients.

Linux PCs running the RFC1179 (lpd daemon) protocol were adopted as a common solution as queue servers for all CERN network printers. A CERN Printing Package has been designed to integrate the PC-Windows based printing clients into this lpr based environment. This software package is built around a central database, holding the configuration data (printer name, driver, settings,...) of all corporate network printers.

This paper will give an overview of the new architecture including the Linux server environment, the UNIX clients and the CERN Printing. Additionally Java and Perl print clients have been developed in order to provide a GUI for Unix systems and to support other operating systems such as VMS.

keywords Printing,Linux,parallel,cluster,Windows,lpr,NT,CERN

1. Introduction

With the aim of simplifying printing operations, removing single points of failure, offering better stability, and easing integration of new printing technologies, a task force was set and the end of 1997. Most of the software solutions relied on Windows NT Print Servers, but our goal was to obtain a single queue per printer model. The server solution finally implemented relies heavily on Unix Public Domain products under GPL[1]. The high quality of products like the LPRng spooler[2, 3] has allowed us to build a very reliable system that is flexible enough to meet our scalability requirements and to support our heterogeneous environment.

Support of a large number of printer models is offered. This includes various models of HP, QMS, Tektronix, Xerox and Apple printers. The service back-end is implemented as an array of similarly configured Linux PCs running the LPRng spooler. Appletalk protocol support is provided by means of the Columbia Appletalk Package and the Linux DDP module.

A dedicated DNS name hierarchy has been implemented containing aliases that associate the printer name with the specific server serving its queue. Load balancing and fail-over are controlled by the manipulation of the DNS configuration file.

Two sets of print clients are provided, one for Windows and one for the Unix systems. They both use a RFC1179 lpr compliant protocol to communicate with the print server back-end. This solution was preferred over implementing the SMB protocol in the back-end due to the better scalability of RFC1179 because its connectionless nature.

The Unix clients are the LPRng print clients modified at CERN to support our server addressing scheme and to solve backwards compatibility issues.

The new CERN Printing Package for Windows clients was adapted to this RFC1179 based environment. Besides the management tools included for controlling the networked printing infrastructure, it comes with a client package enabling Windows 9x, NT and Windows

2000 users to set-up printer objects on their machine in an automated way. Applications can print via the standard Windows interfaces on NT, Novell (IPX) and LPR (e.g. Linux) print servers.

2. Print Server Environment

2.1. Choice of Printing Protocol

RFC1179 was chosen as the protocol for the print clients to interface to the server back-end. this is the protocol of the BSD lpd printer server. There are several reasons for this choice:

- RFC1179 is the only non-proprietary standard available today.
- The scalability of connection-oriented protocols like SMB is limited to less then hundred queues per server.
- The management of connection oriented protocols can be complex. In the past we often saw that Windows PCs on some network segments were loosing their connections to Novell print servers, thus requiring a manual user intervention and reconnecting to his printer.
- It is available and easy to implement on all our supported platforms.

2.2. Appletalk Support

Although we have tried to migrate from Appletalk to TCP whenever possible, the server back-end has to interface to both TCP and Appletalk printers. As more than 50 % of the printers still are on Appletalk this is an important requirement

2.3. Spooling Software

LPRng is our choice for print spooler software to support RFC1179 printing. It is mostly compatible with the `lpd` BSD print spooler while having a large number of enhancements and new features, some of which are exploited in our project. Features most relevant to us include:

- Lightweight print clients with both SysV and BSD flavors, requiring no databases or spooling setup in the clients.
- Better security, since no SUID root is required in the client software.
- Improved permission and authorization mechanisms: a set of rules determines the type of actions that a given user can perform.
- Improved filter model with support for parameters, and for separated banner page generation, and accounting filters.
- Support for dynamic redirection of queues.
- Excellent debugging facilities.
- Flexible customization model.

2.4. Filters

When printing using the BSD `lpd` model, the server passes the files through a filter that processes them according to a format specified in the control file and the type of device. The processing includes formatting, accounting and banner page generation.

LPRng supports a superset of the BSD printing filter model and allows greater flexibility. In particular, one can specify parameters for filters in the printcap database, and separate functions such as accounting, banner generation, etc. into different filtering programs. This feature was exploited mainly for the banner generation programs.

It should be noted that print jobs coming from Windows clients that use the "CERN Client Package" go through the filters in literal mode, i.e., no formatting is required by the filters since it has already been done by the Windows driver. The filters still have to do other tasks, such as handling communications with the printer, error reporting, accounting, etc.

2.5. Server Clustering

Clustering is required in the server back-end in order to guarantee scalability and resilience to failures. Several alternative architectures were considered, for instance the one in which a "master" front-end machine distributes print jobs to an array of workers. This solution was rejected because it would still present a single point of failure and it would complicate the management of print queues. As a consequence a simpler architecture had to be developed with the following characteristics:

- An array of print servers which are configured with similar queue definition files (`printcap` , etc).
- Printer queues are distributed among the servers to balance the workload.
- A printer is served by a single server at any given time.
- A external naming service directs the print clients to the server assigned to each specific printer.

The advantages of this design are numerous:

- Queue management is easy as only one server sends jobs to the printer. Contention is not possible.
- In case of failure of one of the servers, a reconfiguration of the naming service database suffices to reallocate the queues served by the failed server onto active ones.

2.6. Naming Service Setup

The Domain Name Service (DNS) has been chosen as the printer to server resolver since it is easily accessible from the client software. Moreover a high availability DNS server infrastructure already exists at CERN. A print.cern.ch DNS hierarchy was setup, such that for each printer, an alias of the form printername.print.cern.ch is available. These aliases map printers to the server which serves their specific printer queue.

2.7. Scalability and Fail-over

The architecture implemented can support with two or three PC servers, our current printer population of around 1300 printers, for a client system population of some 2500 Unix workstations, servers and Linux PCs, 5000 Windows PCs and 1700 X-terminals.

The fail-over mechanism currently requires intervention of an operator and results in a slightly degraded mode, but the system is simple, robust and easy to manage.

3. The PC Printing Architecture

From the PC/Windows point of view there is no commercial software available for automated printer and driver installation integrated in a Unix printing architecture. In order to avoid the deployment of platform dependent printing architectures (e.g., Netware and NT for the PC world) the CERN standard Unix printing server architecture has been adopted for the Windows clients. The "CERN Printing Package" architecture[4] for Windows 9x, NT and Windows 2000 consists of the following components (see also figure 1):

- A common database and driver repository defining all CERN network printers (queue name, W95/WNT driver data,...) which is stored on a central application server.
- Administration tools: a set of programs used by the printer administrators to install, modify or delete network printers/drivers.
- Client software: Windows based programs running on the client PCs enabling the user to connect to any CERN network printer in an easy way. The "CERN Printer Wizard" is the key application for printer installation and configuration on the PC client platform.

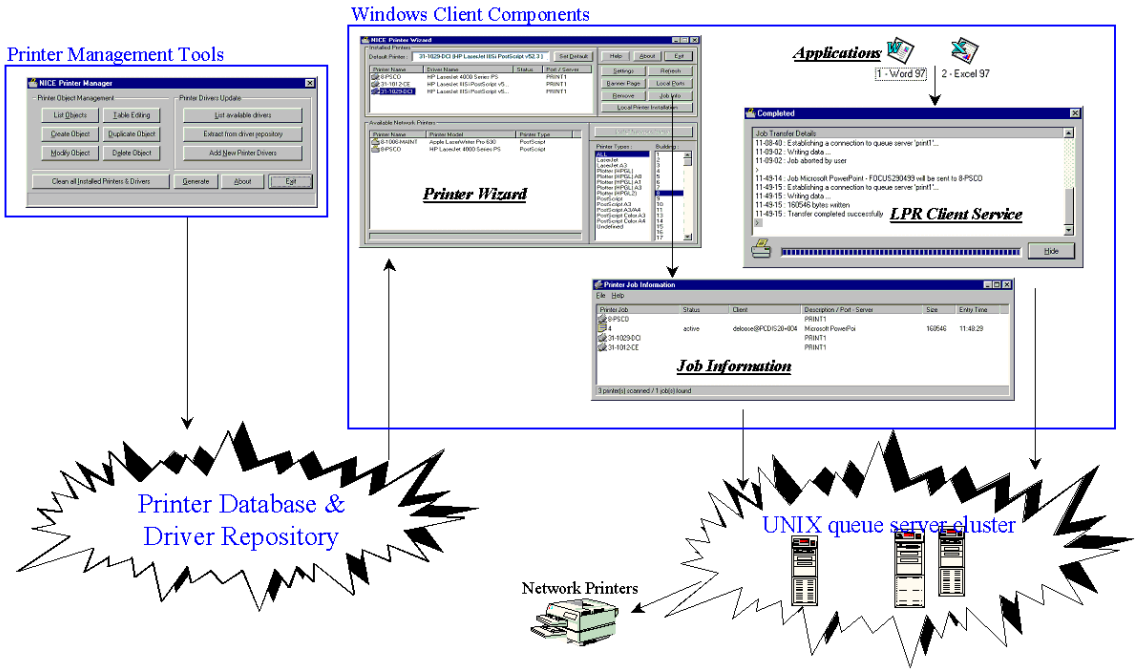

Figure 1: The PC Printing Architecture

3.1. The printer database and repository

MS Access was chosen as the database medium because of its good integration with the PC world and the powerful development environment. The database contains the following information:

- Supported printer drivers: In addition to the standard printer drivers included in the W95 and WNT distribution CDs, the database holds information about third party drivers added by the administrator. For every driver, the database has a reference to the corresponding .inf and

driver files. These files are stored in the printer data repository and used by the client software when a user connects to a printer.

- Available network printers, defined by the name of the printer (identical to its queue name on the Linux server), the printer driver name, as well as the path to the driver files for the W95 and WNT platforms.

3.2. Administration tools

Administration of the printer database and drivers includes the following features:

- Possibility to add third party printer drivers. Any new printer driver which is not part of the standard W95/WNT distribution CDs can be added in an automated way. All information required to install this driver on a client PC is added to the database and the necessarily files are copied to the repository by the "CERN Printer Manager."
- Management of the network printers. For every CERN network printer, the "CERN Printer Manager" allows the administrator to:
  - enter the printer name.
  - select the W95 and WNT driver name from the supported driver list.
  - choose the printer type.
  - define the default settings.

3.3. The Client Package

The client package comes as an InstallShield setup and includes the following printing components:

- The "CERN Printer Wizard": a graphical interface which enables the user to connect to all network printers defined in the database. Connecting to a printer from W95, WNT or W2000 implies installing the driver files on the client PC, importing the default settings and creating the printer port. The driver is downloaded via network share or using embedded FTP. The "CERN Printer Wizard" presents the available printers organized by building number and printer type, so the user knows exactly the capabilities and location of the nearest printers. Other options include the configuration of the printer settings and banner page.
- ”Job Information”: This tool returns information about the current jobs for every connected printer with the possibility to delete them. In addition to job queering from printer queues on Novell and NT servers, job information is queried from the Unix printer servers via embedded **`lpq`** and **`lprm`**. This gives the advantage that a PC user has an overview of all current jobs on the connected printers, even sent by non PC platforms. Another benefit in a heterogeneous client environment gained by having a single server (and hence queue) for a particular printer.
- The “LPR Service” module: This task is permanently running in the background on every Windows PC and is the gateway between the local printer ports and the Unix printer server. It is triggered when a job is being created from an application program. As defined by the "CERN Printer Wizard," the printer port sends the data to a local file on the hard disk, containing the name of the printer. On its turn the LPR Service module transmits the file to the corresponding Linux server after DNS name resolution via embedded LPR.

At CERN, the NICE[5] architecture delivers the client to the desktop but the complete package, including the database and driver repository, can it be installed in any Windows based environment.

4. UNIX print clients

The print clients supported are the LPRng clients (CERN xprint interface) which have been modified to allow the use of DNS to locate the print server. In particular, the modification ensures that printer names that are specified as -Ppname are translated into names like pname@printername.print.cern.ch, thus making the print server back-end scalable in a way that is transparent to the print clients.
In addition, a Java and PERL implementations of xprint was developed for UNIX systems and others such as VMS.

References

[1] GNU. GNU General Public License. Free Software Foundation, Inc., 1991.
[2] Patrick Powell. Managing network printers and print spoolers. In LISA XI, Tutorial T15pm. Astart Technologies Inc., 1997.
[3] Patrick Powell and Julian Mason. "Lprng - an enhanced printer spooler system." In LISA IX Proceedings, pages 13-24. Usenix, 1995.
[4] The CERN Printing Package, http://printpackage.web.cern.ch
[5] The CERN-NICE architecture, http://nicewww.cern.ch