DOI:CNKI:50-1165/N.20121112.1642.014

基于 PXE 网络的远程数据清洗系统*

彭仁明[1],李 岷[2]
(1. 绵阳师范学院 物理与电子工程学院,四川 绵阳 621000;2. 绵阳职业技术学院 信息工程系,四川 绵阳 621000)

摘要:提出了一种基于 PXE 协议的网络互连结构用于数据清除的远程控制系统。数台客户端通过 PXE 协议与主机相连,客户端从主机上下载镜像文件,包括启动文件、配置文件和数据清除主体程序,从而实现远程启动操作系统和自动运行程序进行数据清除。整个操作过程实现实时控制,运行的过程和结果都在主机的屏幕上进行显示。PXE 网络结构互连模式使系统拥有更快的速度和灵活性,实际工程应用表明系统性能优良,数据清除具有规模性,灵活快捷,其性能和效率明显优于传统的单机操作处理方式。

关键词:数据清除;PXE 协议;Linux;Gutmann 算法;PHP
中图分类号:TP3 **文献标志码**:A **文章编号**:1672-6693(2012)06-0059-05

随着现代科技的进步和发展,人们越来越多地使用数码设备存储数据,随之而来的数据安全性和个人隐私保护成了首先需要考虑的问题。一般情况下,人们如果要将个人数据或需要保密的使用记录清除,所采用的方法大都是文件删除、格式化、磁盘分区等基本操作。实际上对于数码存储介质来说,就算是经过这些操作,存储数据依然存在[1],专业人士可以对数据进行恢复处理,重新得到已清除的各种数据。如何保证数据安全,如何将私密数据安全彻底地销毁,已经引起了人们的广泛关注[2]。本文提出了一种网络式数据清洗方案,可以同时处理数十台连接在网络上的数据存储设备,通过 PXE(Pre-boot execute environment)协议将设备与服务器进行连接,利用服务器启动客户机并进行相应的数据清洗工作。所有工作过程和处理结果都可以在服务器端进行实时显示。整个系统在 Linux 平台上进行基本操作,利用 Gutmann 算法进行数据清洗。该系统可以对多台数据终端同时进行数据清洗操作,数据被彻底清除,即使是用专门的数据恢复软件也难以恢复数据。

1 数据清洗方案

一般来讲,常见数据清洗方案有以下几种[3-4]:

1)单机式数据清洗方案。该方案是通过 CDROM,USB 接口等方式直接在单台机器(如 PC、Notebook 和 Server 等)直接启动并运行清洗程序,操作完成后输出所有清洗状态报告。

2)集中数据清洗方案。该方案是将各种存储介质,如 IDE、SATA、SCSI 和 SD/CF 卡集中,同时连接到多硬盘式处理伺服器,并直接运行清洗程序,处理完成后的硬盘可以直接更换,系统可持续进行数据处理,运行完后直接输出所有清洗状态报告。

3)消磁式数据清洗方案。直接将数据设备(IDE、SATA、SCSI、SD/CF 卡)放入消磁设备中,并启动消磁操作,消磁后硬盘无法重复利用,处理后的硬盘只有回收拆散并作破碎处理。

4)销毁式数据清洗方案。直接将需要销毁的存储介质集中起来,并放入销毁设备中进行销毁,最后形成微细颗粒,同时后期还要进行高温处理。

5)网络数据清洗方案。通常情况下,不可能将还可以使用的数据存储设备进行报废式的消磁处理或是销毁处理。另外在工业生产环境或是机关单位进行数据清除处理时,有可能一次性对几十甚至上百台数据终端进行数据清洗工作,如果以人工的方式逐台对设备进行操作,工作效率非常低。网络数据清洗方案可以由服务器同时启动多台数据终端,

* 收稿日期:2012-06-03 网络出版时间:2012-11-12 16:42:01
资助项目:绵阳师范学院基金项目(No. MA2009010)
作者简介:彭仁明,男,副教授,硕士,研究方向为电子信息技术。
网络出版地址:http://www.cnki.net/kcms/detail/50.1165.N.20121112.1642.201206.59_014.html

并同时进行相应的数据清洗工作。所有工作过程和处理结果都可以在服务器端进行实时显示。网络采用 PXE 协议进行终端设备连接,在 Linux 平台上利用 Gutmann 算法对存储介质进行数据清洗。利用该方案可以对多台数据终端同时进行数据清洗,经处理后数据被彻底清除,用专门的数据恢复软件也难以恢复数据。

2 PXE 网络协议

PXE 技术由 Intel 公司开发设计,该协议分为 Client 和 Server 主从两部分[5-6]。Client 通过网络从远端服务器下载系统启动、运行文件。在远程启动过程中,客户端通过 PXE(存放于主板或者网卡上的 BIOS ROM 中)向 BOOTP(Boot protocol)或 DHCP(Dynamic host configuration protocol)服务器发出请求,要求服务器分配 IP 地址,再用 TFTP(Trivial file transfer protocol)或 MTFTP(Multicast trivial file transfer protocol)协议下载启动软件包到本机内存中执行,由启动软件包完成终端基本软件设置,引导预选安装在服务器中的终端操作系统[7-8]。

当客户机启动后,网卡 BIOS 中的 Boot Rom 会发送请求帧,在此帧中有客户机的网卡 MAC 地址,服务器端的远程启动服务收到客户机广播的查询帧后,根据帧中的 MAC 地址,搜寻远程启动数据库中与 MAC 地址对应的配置记录,如无配置记录则引导过程不能继续。

如果配置记录存在,远程启动服务将发送查询帧,帧中包含服务器网卡 MAC 地址,再调用 BOOTP 或 DHCP 以分配客户端名字、IP 地址、服务器 IP 地址及启动映像文件等。工作站收到第一个响应后会发送文件请求帧,要求远程服务器传送启动所需的文件。服务器依据数据库工作站记录查找对应的 Boot block(启动块),并发送给工作站文件数据响应帧,将所需的启动文件传送给客户机当接收完启动文件后便执行启动程序。

由于 PXE 协议将传输的数据划分为很多小的部分,因此数据传输速度与 RPL 启动技术相比运行速度提高很多。PXE 网络协议还可用于引导多种操作系统,如 Windows 以及 Linux 操作系统等。PXE 协议在启动前并不需要客户端的网卡型号以及 MAC 地址,只需在服务器上进行设置便可启动使用本协议的终端,减少在原有方式下分别对网卡进行设置的重复劳动,从而使用的便捷程度和灵活性大大增加。在本系统中,网络的连接建立采用 PXE 协议。

3 远程数据清洗系统设计

3.1 系统概述

远程数据清洗系统由服务器与需进行数据清除的一台或多台客户机构成,系统构成如图 1 所示。

整个远程数据清洗系统网络的基本操作平台由 Linux 系统所构建,由于本项目系统运行于互连的计算机网络,在实际工程应用环境中,存在着多种类型终端设备、多任务处理操作和多 CPU 运行等特点,而 Linux 系统具有良好的系统兼容性、通用性、RAM 保护模式、动态加载程序、多用户和多线程等特点,可以良好地支持系统正常运行。

在 Linux 系统平台基础上,系统还包括 PXE 模块,Busybox 模块,数据清洗模块以及 PHP 显示模块等。PXE 模块作为远程客户机启动和操作系统传输的基本枢纽,将系统各功能模块相互连接构成整体。数据清洗模块运用 Gutmann 算法对数据进行处理,PHP 显示模块将系统的数据处理过程、流程和控制操作等进行实时显示,同时系统还包括 BusyBox 模块、MD5 加密和无盘网络启动服务等。

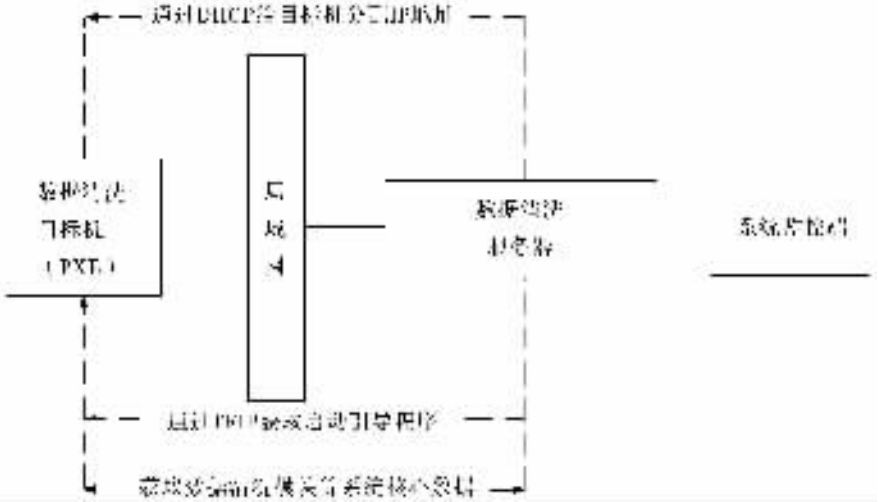

图 1 远程数据清洗系统

3.2 PXE 模块

在本系统中,数据清洗服务器与预清洗的目标机由 PXE 协议进行连接。与通常的 PXE 网络不同,本网络系统是对连接于网络中的目标机进行数据清洗工作,因此系统需要支持目标机的远程启动,并在连接启动完成后,由服务器端下载数据清洗程序、相关的服务程序和状态显示监控程序等,从而进行相应的数据处理,并将处理的过程和结果数据传回系统监控端,从而实时掌握目标机的各项配置数据以及当前工作状态。

3.3 BusyBox 模块

在客户机上进行远程无盘启动,整个操作系统和数据删除模块都只能在其内存上直接启动和运行,而硬盘等存储介质是作为数据清除的对象,并不能参与系统的启动和运行。考虑到客户机可能存在的多种类型,因为系统资源有限,对远程电脑启动和操作运行所需要的软件系统的大小要求就十分苛刻,需要该系统体积小巧并能顺利完成系统启动、远程数据清除以及相关的数据传送工作。

BusyBox 最初是为 Debian GNU/Linux 安装盘编写,其目标是在一张软盘大小的空间里创建一个可引导的 GNU/Linux 系统,用于系统的安装。BusyBox 模块集成了与 Linux 相关的重要工具,包括小型的 Http 服务器和 Telnet 服务器,利用该模块,可以实现最小化的网络 Linux 操作系统及其启动和运行。BusyBox 将很多基本工具合并到一个可执行程序中,让它们可以共享这些相同的元素,这样可以产生更小的可执行程序。实际上,BusyBox 可以将大约 3.5 MB 的工具包装成 200 KB 左右大小,这就为可引导的磁盘和使用 Linux 的嵌入式设备提供了更多功能。

BusyBox 有良好的树型组织结构,基于用途对工具进行了分类并存放于子目录中。Makefile 配置编译和安装所使用的各个文档存放于根目录中;网络及进程工具位于 Networking 目录中;标准的模块工具位于 Modutils 目录中;编辑器存放于 Editors 目录中。BusyBox 提供了编译选项,可以根据需要编译和正确的调试,它的编译选项表见表 1。

在本项目中,考虑到操作系统和清洗程序对运行空间的苛刻要求,选用 BusyBox 作为系统的引导和传送对接模块。

3.4 数据清洗模块

本系统主要目的是在工程环境中,将目标机存储介质中的数据擦除,但一般删除方式和格式化等操作并不能彻底清除数据,专业人士利用数据恢复工具可以再建和重构出数据。

表 1　BusyBox 编译选项表

Make Target	Discription
help	显示 make 选项的完整列表
defconfig	启用通用配置
allnoconfig	禁用所有应用程序
allyesconfig	启用完整配置
allbareconfig	启用所有应用程序(不包括子特性)
config	基于文本的配置工具
menuconfig	基于菜单的配置工具
all	编译 BusyBox 二进制文件和文档
busybox	编译 BusyBox 二进制文件
clean	清除源代码树
distclean	彻底清除源代码树
sizes	显示所启用的应用程序的文本/数据大小

将无规律的随机数据信息覆写入原储存介质进行数据清洗,如有人对处理后的介质进行数据恢复,将无法恢复出原有保密数据,而只能恢复出覆写后的数据,达到数据保密及数据清洗效果。

数据覆写有逐位覆写、随机覆写以及跳位覆写等形式,美国国防部要求必须对所要清洗的数据用不同的数据序列进行 7 次覆写,所采用的标准为 DOD 5220-22M。考虑到本项目实际的工程应用环境,以及数据清除的完备性和有效性,采用 Gutmann 算法进行数据清洗。

Gutmann 算法用于对计算机硬盘等存储介质进行数据清洗,该算法用一系列预定义参数对目标介质进行 35 次数据写入和覆盖,用于清除原有的各种数据,经此方式处理后存储介质将无法恢复原始数据。

对于不同硬件结构的存储介质,为了安全地将数据进行彻底清洗,Gutmann 算法提供了 3 种不同的写入参数,分别对(1,7)RLL(Run-length limited),(2,7)RLL,MFM(Modified frequency modulation)编码格式的存储介质进行数据覆盖写入,如果预先明确了存储介质的类型,可用对应的参数数据写入覆盖即可。数据的写入内容和步骤顺序见表 2。

3.5 用户认证过程

在数据清除中,对有用数据的保护非常重要,因此为防止误操作,在进行正式删除之前,对用户操作权限验证很有必要。常见的数据加密算法有很多,本系统采用 MD5 算法进行验证加密。MD5 算法是用 512 位分组来处理输入的信息,而算法的输出由

4个32位分组组成,将这4个32位分组级联后将生成一个128位散列值,再利用这个数值进行加密方面的确认操作。可有效地进行用户身份的验证,保证系统使用的安全性。

表2 Gutmann算法的数据写入过程

顺序	写入数据	对应介质编码形式		
1~4	随机数据			
5	01010101 01010101 01010101 0x55	(1,7) RLL		MFM
6	10101010 10101010 10101010 0xAA	(1,7) RLL		MFM
7	100100100100100100100100 0x92 0x49 0x24		(2,7) RLL	MFM
8	010010010010010010010010 0x49 0x24 0x92		(2,7) RLL	MFM
9	001001001001001001001001 0x24 0x92 0x49		(2,7) RLL	MFM
10	00000000 00000000 00000000 0x00	(1,7) RLL	(2,7) RLL	
11	00010001 00010001 00010001 0x11	(1,7) RLL		
12	00100010 00100010 00100010 0x22	(1,7) RLL		
13	00110011 00110011 00110011 0x33	(1,7) RLL	(2,7) RLL	
14	01000100 01000100 01000100 0x44	(1,7) RLL		
15	01010101 01010101 01010101 0x55	(1,7) RLL		MFM
16	01100110 01100110 01100110 0x66	(1,7) RLL	(2,7) RLL	
17	01110111 01110111 01110111 0x77	(1,7) RLL		
18	10001000 10001000 10001000 0x88	(1,7) RLL		
19	10011001 10011001 10011001 0x99	(1,7) RLL	(2,7) RLL	
20	10101010 10101010 10101010 0xAA	(1,7) RLL	MFM	
21	10111011 10111011 10111011 0xBB	(1,7) RLL		
22	11001100 11001100 11001100 0xCC	(1,7) RLL	(2,7) RLL	
23	11011101 11011101 11011101 0xDD	(1,7) RLL		
24	11101110 11101110 11101110 0xEE	(1,7) RLL		
25	11111111 11111111 11111111 0xFF	(1,7) RLL	(2,7) RLL	
26	100100100100100100100100 0x92 0x49 0x24		(2,7) RLL	MFM
27	010010010010010010010010 0x49 0x24 0x92		(2,7) RLL	MFM
28	001001001001001001001001 0x24 0x92 0x49		(2,7) RLL	MFM
29	011011011011011011011011 0x6D 0xB6 0xDB		(2,7) RLL	
30	101101101101101101101101 0xB6 0xDB 0x6D		(2,7) RLL	
31	110110110110110110110110 0xDB 0x6D 0xB6		(2,7) RLL	
32~35	随机数据			

3.6 系统监控模块

系统的基础平台为Linux,由于本系统的目的是对目标机进行数据清洗,目标机由服务器远程启动,并将相应的数据清洗等核心程序通过PXE协议传送到目标机,系统可供运用的系统资源非常有限,考虑到PHP与Linux良好的兼容性,而在运行时只消耗很少的系统资源,因此系统监控模块采用PHP编写。

系统需要收集目标机硬件信息及相关数据,并将清洗处理参数、过程与结果实时地显示在服务器端,也可以将所有处理过程和数据结果打印进行分析和保存。其中参数包括数据存储介质接口类型、存储介质尺寸、容量大小、连接存储盘符数、清洗处理进程和异常情况显示等,终端状态显示如图2所示。

图2 远程数据清洗系统监控端状态显示

4 结束语

远程数据清洗系统有效地利用 PXE 无盘网络启动技术,实现了远程启动、数据采集、数据传输和清洗工作。经过实际工程运用,证明该方法可靠、有效、快捷方便,在数据安全、数据保密等工程应用方面有着广阔的应用前景。

参考文献:

[1] Dei C T,Simoes P,Bastos F,et al. Integration of PXE-based desktop solutions into broadband access networks [C]// Network and Service Management (CNSM),2010 International Conference. [S. l.]:IEEE Press,2010:182-189.

[2] Li J H ,Zhang K,Zhang F. Network center's highly-efficient management solutions based on intel PXE-based remote cloning system [C]//Advanced Computer Control (ICACC),2011 3rd International Conference. Harbin:IEEE Press,2011:408-411.

[3] Fang A P,Ma X N,Wang Z. Design and implementation of instrument cluster's management unit based on embedded Linux[C]//Electronic Measurement and Instruments. ICEMI'07. International Conference. Xi'an:IEEE Press,2007:3-254-3-258.

[4] Yuan T,Ren G Q,Wu Q Z. Implementation of real-time network extension on embedded Linux [C] //Communication Software and Networks. ICCSN '09. International Conference. Macau:IEEE Press,2009. 163-167.

[5] Dellinger M,Garyali P,Ravindran B. Chron OS Linux:a best-effort real-time multiprocessor Linux kernel [C]//Design Automation Conference (DAC). 48th ACM/EDAC/IEEE. New York:IEEE Press,2011:474-479.

[6] Hughes G F,Coughlin T,Commins D M. Disposal of disk and tape data by secure sanitization [J]. Security & Privacy,2009,7(4):29-34.

[7] Oliveira S R M,Zai ane O R. A unified framework for protecting sensitive association rules in business collaboration [J]. International Journal of Business Intelligence and Data Mining,2006,1(3):247-287.

[8] Rahm E,Do H H. Data cleaning:Problems and current approaches [J]. IEEE Data Engineering Bulletin,2000,23(4):3-13.

Remote Data Sanitization System Based on PXE Network

PENG Ren-ming[1], LI Min[2]

(1. School of Physics & Electronic Engineering, Mianyang Normal University, Mianyang Sichuan 621000;
2. Dept. of Information Engineering, Mianyang Vocational and Technical College, Mianyang Sichuan 621000, China)

Abstract:A remote control system using in data sanitization field is introduced in this paper. The client downloads image files including starting configuration files and data sanitization files, then supports start-up operation of system and data sanitization. Data sanitization processes don't need the hard disk of client. The remote data sanitization system has network structure and processing flow. System performance including flexibility and efficiency is better than that of traditional system, and PXE protocol network structure means high speed and reliability. It is proved by practical application that the system is steady and reliable.

Key words:data sanitization;PXE;Linux;gutmann method;PHP

(责任编辑 游中胜)