文章编号:1671-4598(2021)06-0142-05 DOI:10.16526/j.cnki.11—4762/tp.2021.06.029 中图分类号:TP391 文献标识码:A

船载系统日志自动分析软件的设计与实现

魏江涛，冯建峰，姜美雷，鲍存军
（中国卫星海上测控部，江苏 江阴 214431）

摘要：船载中心计算机系统的系统日志和测控软件日志记录了大量的系统故障与软件异常信息，如何及时、全面地分析系统日志、软件日志，发现系统运行故障，并及时予以处理、解决，是系统运维人员的一项重要工作；针对当前船载中心计算机系统日志分析现状及存在的问题，系统日志和测控软件日志自动分析软件给出了解决的思路和方法，提出并介绍其实现的技术要点和效果；测试结果表明日志自动分析软件有效地提升了人员工作分析效率，在快速分析故障问题和批量进行系统日志维护方面具有较好应用。

关键词：日志；awk；自动分析；Qt 远程调用

Design and Implementation of System Log Automatic Analysis Software for Space Tracking Ship

Wei Jiangtao, Feng Jianfeng, Jiang Meilei, Bao Cunjun
(China Satellite Maritime Tracking and Control Department, Jiangyin 214431, China)

Abstract: In the computer system of the space tracking ship, there are many system fault and software abnormal information in the system log and measurement and control software log. To the system administration, it is important to analyze the system log and software log timely and comprehensively. In the meantime, the fault must be fixed as soon as possible when the system operation fails. Thus the current situation and existing problems in the system of the computer log analysis are specified, the methods to automatically analyze the system log and software log are improved, and the technical points and effects of its implementation are applied. The test results show that the automatic log analysis software can effectively improve the efficiency of personnel work analysis, and it can be applied in the rapid analysis of fault problems and batch maintenance of system logs.

Keywords: system log; awk; automatic analysis; Qt remote call

0 引言

计算机在运行过程中，可能会产生硬件故障、操作系统故障、数据库故障、中间件故障和应用软件故障。大部分故障信息都会及时记录到系统日志文件和应用软件日志文件中。系统管理员、数据库管理员等维护人员需要经常查看这些日志文件，综合分析，找出故障原因，采取相应措施，保证系统正常运行。

从日志的产生来源上可以将日志分为操作系统日志、数据库系统日志、网络应用服务日志、应用软件日志等。操作系统日志又称为主机日志，由操作系统自动生成，与操作系统行为密切相关。它记录了系统发生的各种事件，此类日志文件一般是纯文本的文件，通常存放在“var/log”目录下。应用软件日志由其运行在服务器或终端微机的应用数据处理软件产生，记录软件运行的状态信息和异常信息，系统运维人员可以通过这些日志解决应用软件某些方面的错误，查看应用软件进程的历史运行信息，排查软件运行过程中出现的问题。因此，对日志信息的检查和分析是保障系统正常运行的重要一环。

awk 是 Linux 系统中一款功能强大的文本分析工具，相对于 grep 查找和 sed 文本编辑，awk 在数据分析、生成报告等方面尤为强大，在 Linux 系统日志处理工作中，发挥重要的作用。通常，awk 以文件的一行为处理单位，每接收文件的一行，然后在这些行上执行相应的命令来进行文本处理。简单来说 awk 就是把文件逐行读入，以空格为默认分隔符将每行切片，切开的部分再进行分析处理。awk 拥有自己的程序设计语言，它允许创建简短的程序，通过这些程序可以读取输入文件、数据排序、处理数据、执行计算以及生成报表等，对于处理文本格式的日志文件，awk 更加轻便高效。

另外，在 Linux Shell 中重定向和管道是 Shell 的一种高级特性，这种特性允许用户人为改变程序获取输入和输出位置。重定向分为输出重定向和输入重定向。Shell 程序在默认的情况下将输出结果输出到标准输出（stdout），通常来说，标准的输出即为显示器，而通过输出重定向后可以把 Shell 程序的输出转移到其他地方。例如如下的命令则将 awk 的输出结果重定向到 home 文件夹的 awk_out 文件中。

收稿日期:2020-10-15； **修回日期**:2020-11-16。

作者简介:魏江涛(1978-),男,陕西人,大学本科,高级工程师,主要从事测量测控方向的研究。

引用格式:魏江涛,冯建峰,姜美雷,等.船载系统日志自动分析软件的设计与实现[J].计算机测量与控制,2021,29(6):142-146,158.

awk ‘ { print 1}’ file> /home/awk _ out

管道将重定向还可以通过“|”将一条命令的输出连接另一条命令的输入。例如：

awk ‘ { print 0}’ file | grep ‘WARN’

awk 命令首先列出 file 文件的每一行，管道“|”接收到这些输出后，并把它们发送到 grep 命令作为其输入，最后 grep 在这行输入中查找包含字符串“WARN”的行，将其在标准输出中显示。通过合理的使用管道和重定向可以很好地提高文本分析效率。

1　船载中心机系统目前日志现状

船载中心计算机系统作为测量船数据处理和传输的神经中枢发挥非常重要的作用。系统由多台服务器和多台操控终端组成，操作系统均部署国产的麒麟操作系统。

系统组成结构如图 1 所示，服务器部署在后台远端，主要有数据处理服务器（SJCL）、汇集分发服务器（HJFF）、轨道计算服务器（GDJS）、数据库服务器（SJK）、仿真服务器（FZ）、数据显示服务器（ZHXS）、电文服务器（DW）等，这些重要的服务器同时进行工作，相互热备份，总共约十几台服务器；操控终端微机部署在前端用于软件操控和数据监视显示，总共有近四十台。船载中心计算机系统的服务器和操控终端微机均安装麒麟操作系统，每台终端微机和服务器在运行过程中，由操作系统记录系统日志，操作系统日志文件由操作系统本身的 syslogd 服务来维护，记录系统运行的重要信息及硬件故障信息，主要是存储在/var/log/messages 和/var/log/dmesg 文件中，系统日志大多数是以一定格式的文本形式存放，系统运维人员通过系统日志文件可以了解系统的运行状况。

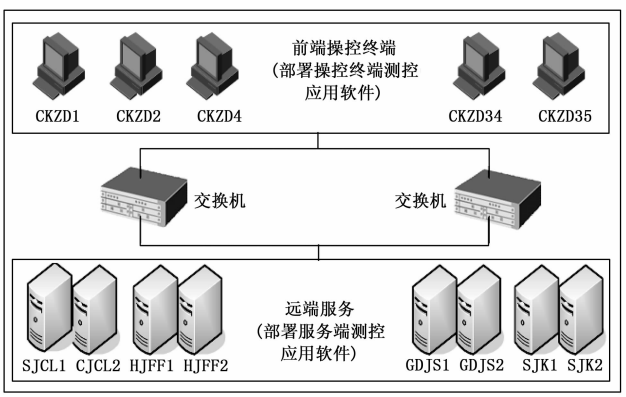

图 1　船载中心计算机系统组成结构图

船载中心计算机系统除了服务器和终端微机外还有另外一个重要组成部分——测控实时数据处理软件（以下简称测控软件），它承担着航天器重要数据的处理工作。

目前随着中心计算机系统硬件结构的改变，测控软件也在不断地变化。由以前运行在单台微机完成所有的数据处理工作逐渐向层体系结构发展，测控软件也分成了两部分：

一部分是运行在后台服务器上的测控业务层数据处理软件，主要有实时数据处理软件、轨道计算软件、数据汇集分发软件等等，这些测控业务层软件安装在不同的后台服务器上完成各自的数据计算处理功能，这些测控业务层数据处理软件部署在后台远端服务器上，没有软件监视界面；

另一部分是运行在前端终端微机上的操控终端应用软件，操作人员通过操控终端应用软件来启动后台服务器的测控业务层数据处理软件，前端操控终端应用软件通过网络和后端服务器上的测控业务层数据处理软件进行数据交互。

后台测控业务层数据处理软件有很多配置项用于完成不同的数据处理。以运行在实时数据处理服务器上的实时数据处理软件为例，实时数据处理软件有近 20 个配置项，主要有数据支持（TSBS）、设备协议转换（DAPC）、数据存储（DADS），设备数字引导计算（TRLF）、设备数据处理（TRTSI）、惯导船姿船位处理（TRPP）、遥测数据处理（TMST）、导航数据处理（NVDA）等等。每一个配置项在运行过程中都会记录一个独立的日志文件（*.log），存放在数据处理服务器的/var/log/amtcass 相应目录下，这样实时数据处理服务器就有 TSBS. log、DAPC. log、TRLF. log、TRPP. log、TMST. log、TRTSI. log 等近 20 个日志文件，记录各个配置项运行信息和出错信息。

同样，运行在其他服务器上的数据处理软件也类似，如汇集分发服务器、轨道计算服务器等，也有相应的独立的多个软件配置项日志文件，这些日志文件中包含了重要的软件运行过程信息，这些信息对分析软件运行情况非常重要。针对目前船载中心机计算机系统的现状，系统运维人员在进行日志分析方面主要面临以下几个问题：

1）日志文件存储比较分散，每台操控终端及服务器的日志存储在本机，系统运维人员逐台分析系统日志和测控软件日志需耗费大量时间，分析效率不高。

2）服务器测控软件日志在软件出现异常时需要快速查看分析，服务器部署在后台远端，测控软件日志存储在各自的数据处理服务器，存储比较分散，系统运维人员在前台操控终端微机操作，因此很难及时查看到后台服务器的测控软件日志信息，这对问题排查和及时应急处置带来一定的困难。

3）船载中心机系统的终端微机和服务器众多，系统日志文件多，操作系统记录的日志信息量大，逐台检查工作量大，在分析系统日志时没有集中收集，系统运维人员面临海量的日志文件分析，容易遗漏和忽略一些重要的系统日志信息和故障信息。

4）测控软件日志和操作系统记录的日志文件多，系统运维人员在人工进行日志清理维护时工作量大，耗费大量时间精力，时效性不高。

为解决上述问题，有必要对日志文件分析模式进行改进，通过日志自动分析软件实现对日志自动分析、故障及时上报、文件定期清理，从而降低人为工作量，提高船载中心计算机系统故障排查的效率和准确率，提升系统运行的整体安全性和稳定性。

2 日志自动分析软件的关键技术及设计

船载中心计算机系统的日志分为系统日志和测控软件日志。

系统日志记录要素如表 1 所示，主要有日期、系统时间和日志描述，记录日志发生的日期和系统时间和系统运行过程中发生的重要事件信息。

表 1 麒麟操作系统记录的系统日志要素

日期	系统时间	日志描述，例如描述系统日志的信息
May 11	17:13:23	kernel: Hardware name: Inspur NF8460M3/NF8460M3, BIOS NF8460M3_1.2.9 06/07/2016

测控软件日志记录要素主要有记录日期、系统时间、测控软件配置项名称、时统时间、日志类型、日志描述等。

测控软件的设备数据处理配置项（TRTSI）如表 2 所示。

表 2 测控软件记录的日志要素

日期	系统时间	子功能名称	时统时间	日志类型	日志描述
May 11	17:55:06	TRTSI	18:23:01.077	ERROR	[THREAD: 9513] PubMonitorData 失败，原因为：发布数据失败！
……	……	……	……	……	……

其中，测控软件日志中记录的日期和系统时间和系统日志中的内容相同，时统时间为日志发生时准确的北京时刻，日志类型主要来标识日志的类别，主要有以下几种：

ERROR：测控软件配置项运行错误信息，该类信息是测控软件日志分析重点关注的信息；

WARN：测控软件配置项运行的警告信息；

INFO：一般信息；

DEBUG：测控软件配置项调试信息；

日志描述为日志发生时的具体内容说明。

2.1 日志自动分析软件的关键技术

2.1.1 日志信息的集中收集方法

日志信息收集是日志自动化分析的关键，由于日志信息都存储在远端服务器和操控终端微机上，具体的日志信息收集如图 2 所示，分为采集层、传输层和分析层，下面具体阐述日志采集的实现方法。

日志信息的采集主要利用 awk 采集脚本来完成，将采集脚本存放于各台服务器和操控终端微机上，awk 采集一般的方法为：

awk '{pattern+action}' {filenames}

其中：pattern 表示 awk 在数据中查找内容的正则表达式，而 action 是在找到匹配内容时所执行的一系列命令。在系统日志信息和测控软件日志信息中根据 action 指令筛选出基于指定规则的日志记录信息。

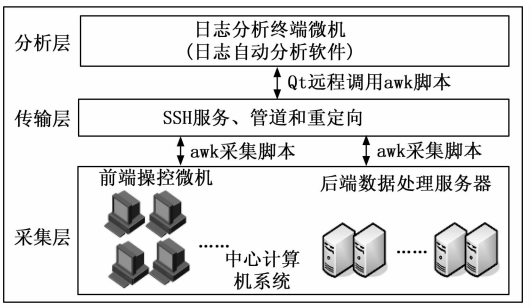

图 2 日志信息采集流程图

日志传输层负责将采集收集到的日志信息传输到日志分析终端微机上，由于日志分析终端微机在远端，日志自动分析软件要获取到各台服务器和操控终端微机的日志信息采集结果，需要利用 Qt 远程调用 awk 采集脚本，同时利用管道和重定向功能将采集结果传输到日志分析终端微机，具体实现步骤如下：

1）日志分析终端微机通过 SSH 服务免密码方式访问服务器和操控终端微机。

SSH 作为 Linux 系统的重要服务，它是 Linux 远程连接重要的方式，通过 SSH 服务可以实现客户端和服务端安全连接，通过使用 ssh－keygen－trsa 命令生成访问服务器的公钥和私钥，将生成的私钥保留，公钥分发到远端服务器和操控终端微机上，这样日志分析终端微机就可以免密码访问远端服务器和操控终端微机。

2）通过 SSH 服务运行存放在远端服务器和操控终端微机上的 awk 采集脚本命令。

通过 SSH 服务可以运行远程的 awk 采集脚本命令，相应运行的命令格式为：

ssh －l name remoteserver'command'

其中，remoteserver 为远程服务器的 IP 地址，command 为存储在远程服务器上的 awk 脚本命令。例如：执行远程服务器 192.168.1.100 脚本的 ssh 调用命令格式如下：

ssh －l root 192.168.1.100'awk.sh'

3）日志采集结果输出重定向。

众所周知，Shell 脚本命令的执行结果在默认的情况下将输出结果输出到标准输出（stdout），通常来说，标准的输出为本机显示器，对于 SSH 服务运行远程的服务器上的 awk 命令也不例外。

在 Linux 系统中，通过输出重定向可以改变输出结果的位置，通过输出重定向后可以把 awk 命令的输出转移到另一个地方去。因此对 SSH 服务运行远程 awk 命令的格式进行轻微改动，修改如下：

ssh －l root 192.168.1.100'awk.sh'＞tmpLogFile

将 awk 命令的执行结果重定向到 tmpLogFile 中，tmpLogFile 为日志自动分析终端中存储执行结果的文件，然后日志自动分析软件就可以对 tmpLogFile 中内容进行解析显示。

4）通过 QProcess 类产生新进程调用 awk 采集脚本。

日志自动分析软件采用 Qt 进行研发，通过 QProcess 类对每一个 awk 采集脚本命令都独立创建一个进程，相互独立运行，互不影响，执行效率比较高。这种方式可以多任务并行运行而不会阻塞相应进程执行。

通过 QProcess 类执行远程调用命令过程如下：

```
QStringprogram="ssh -l root 192.168.1.100 'awk.sh'>tmpLogFile";
QStringList arguments;//创建输入参数存储列表
arguments<<str; //str 为输入的参数
QProcess * myProcess=new QProcess();
myProcess->start(program,arguments); //执行远程调用命令
```

通过以上步骤就实现对日志信息的采集和传输，实现了日志自动分析软件获取远端服务器上日志信息采集结果。

2.1.2　日志的筛选和重复日志的处理方法

日志分析层负责日志文件的解析、分类和显示，按照日志文件内容和日志类型对日志进行分类显示，方便系统运维人员使用。

在日志自动分析时，按照不同的日志类型进行日志信息筛选，利用 awk 命令和 grep 命令进行过滤筛选。测控软件日志自动分析时主要关注 ERROR 和 WARN 日志类型的信息，因此日志自动分析软件在分析测控软件日志时，会按照日志类型分类显示，对 ERROR 类型的日志突出显示，方便系统运维人员分析使用。

测控软件在运行过程中需要实时记录故障信息及运行状态信息，因此测控软件日志可能会出现海量的重复错误日志记录信息，以设备数字引导计算配置项（TRLF）为例，TRLF.log 日志记录的重复日志信息如图 3 所示。

从图 3 中可以看出，在 1 秒钟内软件日志记录了几十条重复的日志记录，而这些海量的重复的日志信息出现会将其他的有用的错误日志信息湮没。

针对这一情况，需要对这些海量的重复的日志信息进行特殊处理，建立过滤条件，通过正则表达式对重复行日志信息进行过滤筛选，只显示部分行的错误日志记录。

在测控软件重复日志自动分析过程中通过建立模式比较规则库对重复的软件日志记录进行筛选。这种筛选的方法提取信息速度较快，对于重复日志的处理，效率比较高，具体的筛选过程如下：

第 1 步：建立模式比较规则库，将重复日志描述的部分建立模式比较规则库；

第 2 步：读取日志文件的日志记录，逐行读取，利用 awk 将日志记录分隔出日志描述部分，和模式比较规则库中规则进行比较，若存在，则转入第 1 步，否则转入下一步；

第 3 步：将该行日志记录进行解析、分类、显示输出，完成重复第 2 步。

这样通过模式比较规则库将海量相同的日志信息进行过滤，只保留有用的日志信息记录输出。

2.1.3　增量日志数据的采集方法

系统日志和测控软件日志在每一次分析时，都应该从上一次检查结束的地方开始，避免重复分析，浪费时间，浪费系统资源，避免事件重复报告，为了解决这个问题，日志的采集时记录了上次采集的结束位置，采用增量日志数据采集的方法，重新采集新生成的日志数据，提高日志自动分析的效率。

2.2　日志自动分析软件设计

日志自动分析软件主要由主线程显示类 MainWindow、系统配置类 ConfigDlgClass、测控软件日志分析类 CKSoftlogClass、系统日志分析类 CSyslogClass、日志清理类组成 ClearLogClass 组成，类之间的关系如图 4 所示。

主线程显示类 MainWindow 主要完成软件日志和系统日志的显示、完成软件初始化工作、系统配置类 ConfigDlgClass 完成软件初始状态的设置，主要设置软件初始启动时扫描测控软件日志和系统日志的时间区间，软件在设计时提供了三个时间段区间，分别是“前 1 小时”、“前 4 小时”、“前 1 天”，默认为“前 1 小时”；测控软件日志类 CKSoftlogClass 主要负责测控软件日志的获取，通过 awk 脚本采集测控软件日志信息，并对日志采集的日志数据进行分析，分析结果送主线程显示类显示；系统日志类 CSyslogClass 主要负责系统日志的获取，同样通过 awk 脚本采集系统日志信息，并对系统日志数据进行分析，分析结果送主线程

```
May 11 17:55:32 SSCL1 TRLF[9988]: [时统时间:18:23:26.701] [ERROR]  [THREAD:9991] 申请[TR_R_PH]失败:回答数据包读取错误, Connection refused
May 11 17:55:32 SSCL1 TRLF[9988]: [时统时间:18:23:26.702] [ERROR]  [THREAD:9991] 申请[TR_R_PH]失败:回答数据包读取错误, Connection refused
May 11 17:55:32 SSCL1 TRLF[9988]: [时统时间:18:23:26.703] [ERROR]  [THREAD:9991] 申请[TR_R_PH]失败:回答数据包读取错误, Connection refused
May 11 17:55:32 SSCL1 TRLF[9988]: [时统时间:18:23:26.704] [ERROR]  [THREAD:9991] 申请[TR_R_PH]失败:回答数据包读取错误, Connection refused
May 11 17:55:32 SSCL1 TRLF[9988]: [时统时间:18:23:26.704] [ERROR]  [THREAD:9991] 申请[TR_R_PH]失败:回答数据包读取错误, Connection refused
May 11 17:55:32 SSCL1 TRLF[9988]: [时统时间:18:23:26.705] [ERROR]  [THREAD:9991] 申请[TR_R_PH]失败:回答数据包读取错误, Connection refused
May 11 17:55:32 SSCL1 TRLF[9988]: [时统时间:18:23:26.706] [ERROR]  [THREAD:9991] 申请[TR_R_PH]失败:回答数据包读取错误, Connection refused
May 11 17:55:32 SSCL1 TRLF[9988]: [时统时间:18:23:26.706] [ERROR]  [THREAD:9991] 申请[TR_R_PH]失败:回答数据包读取错误, Connection refused
May 11 17:55:32 SSCL1 TRLF[9988]: [时统时间:18:23:26.707] [ERROR]  [THREAD:9991] 申请[TR_R_PH]失败:回答数据包读取错误, Connection refused
May 11 17:55:32 SSCL1 TRLF[9988]: [时统时间:18:23:26.708] [ERROR]  [THREAD:9991] 申请[TR_R_PH]失败:回答数据包读取错误, Connection refused
May 11 17:55:32 SSCL1 TRLF[9988]: [时统时间:18:23:26.709] [ERROR]  [THREAD:9991] 申请[TR_R_PH]失败:回答数据包读取错误, Connection refused
May 11 17:55:32 SSCL1 TRLF[9988]: [时统时间:18:23:26.709] [ERROR]  [THREAD:9991] 申请[TR_R_PH]失败:回答数据包读取错误, Connection refused
May 11 17:55:32 SSCL1 TRLF[9988]: [时统时间:18:23:26.710] [ERROR]  [THREAD:9991] 申请[TR_R_PH]失败:回答数据包读取错误, Connection refused
May 11 17:55:32 SSCL1 TRLF[9988]: [时统时间:18:23:26.711] [ERROR]  [THREAD:9991] 申请[TR_R_PH]失败:回答数据包读取错误, Connection refused
May 11 17:55:32 SSCL1 TRLF[9988]: [时统时间:18:23:26.711] [ERROR]  [THREAD:9991] 申请[TR_R_PH]失败:回答数据包读取错误, Connection refused
```

图 3　测控软件 TRLF 配置项日志重复信息

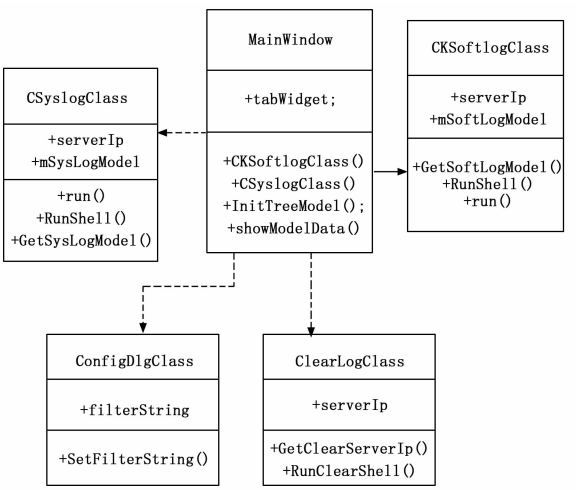

图 4 日志自动分析软件类图

显示类显示；日志清理类 ClearLogClass 主要完成主要服务器的日志清理，通过远程操作，清除服务器的测控软件日志，提高系统运维人员的维护效率。

3 软件测试

日志自动分析软件在麒麟操作系统下研发采用 Qt4.8.5 版本。日志自动分析软件开发完成后，部署在一台终端微机，在实际的应用环境中进行软件测试，通过测试比较输出结果和预期输出结果，从而判断软件功能能否达到预期设计要求。

日志自动分析软件在初始启动时设置日志分析的时段，之后日志再次采集分析时采用增量的方法进行日志分析。以 SJCL1 为例，如图 5 所示，左边树形控件列举出需要分析的服务器和操控终端微机，双击树形控件某服务器或操控终端微机后，日志自动分析软件获取测控软件日志信息和系统日志信息，之后对日志进行分析，给出分析出结果。在图 5 中列出 SJCL1 服务器测控软件各配置项的日志分析情况，按照日志类型错误（ERROR）、警告（WARN）、全部（ALL）进行分类，对于 ERROR 类型的日志进行突出显示。系统运维人员可以点击某个配置项的日志类型，查看具体的日志类型的详细内容，图 5 中下半部分为设备数据处理配置项（TRTSI）的错误（ERROR）日志详细信息。系统运维人员非常方便的查看错误的具体内容信息，为快速分析定位故障提供了强有力的支持。

日志自动分析软件在分析某台服务器测控软件日志的同时，也分析了系统日志文件，并列出分析结果，如图 6 所示，逐行列出了近时间段内系统日志的详细信息。系统运维人员及时了解系统的详细日志信息，在系统日志查看时，还提供了对分析结果再次过滤分析的功能，通过过滤条件分析关注的系统日志信息。

日志清理功能为方便系统运维人员进行日志清理而设计，如图 7 所示，系统运维人员可以通过工具栏选项“日志清理”选择需要清理的服务器，在不需要人员手动逐台

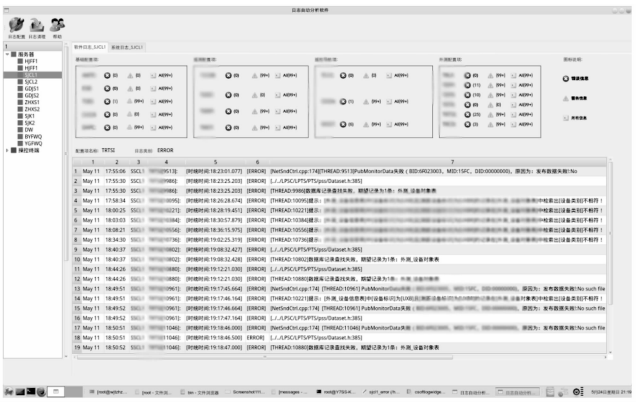

图 5 测控软件日志自动分析结果

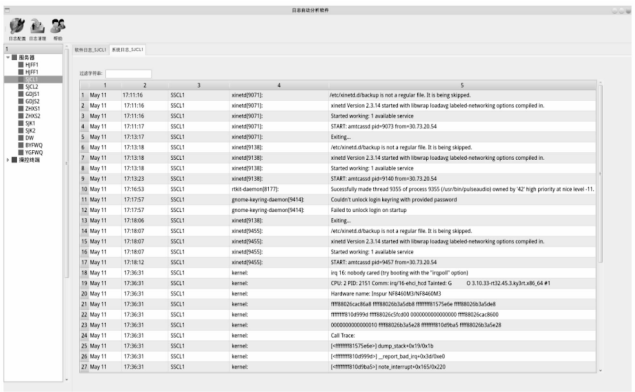

图 6 服务器的系统日志信息

清理众多的测控软件日志文件和系统日志，日志自动分析软件可以很方便的清理日志文件，对服务器测控软件各个配置项日志文件进行批量清理，提升运维人员的工作效率。

图 7 日志清理服务器选择

4 结束语

通过日志自动分析软件，可以有效地实现日志集中化的采集、处理、分析，极大的方便系统运维人员对日志的分析需求，有效防止信息疏漏，提高人员分析问题的效率和准确率。

（下转第 158 页）