文章编号:1671-4598(2018)09-0108-05 DOI:10.16526/j.cnki.11—4762/tp.2018.09.023 中图分类号:TP391 文献标识码:A

总线网络取证信息自动检索风险控制系统设计

彭英杰
(青海民族大学 计算机学院 青海 西宁 810007)

摘要:当前取证信息自动检索系统未对取证信息潜在检索风险进行过滤,导致对取证信息的检索风险控制效果差、检索效率低、误差高的问题,为此设计一种总线网络取证信息自动检索风险控制系统;将输入的取证信息在采集模块中进行收集,检索风险控制模块对采集的取证信息信息进行风险过滤和风险控制后,发送给DSP进行自动检索,采用STM32F407设计接口电路连接采集模块和检索风险控制模块,完成硬件部分的改进;选择高检索相关度节点,利用节点内置文档实现取证信息检索风险的控制,完成软件部分设计;实验结果表明,该系统的检索风险控制效果好,控制精度可达到80%以上,能够为用户提供更有效、更安全的权证信息检索结果。

关键词:总线网络;取证信息;自动检索;风险控制

Design of Risk Control System for Automatic Retrieval of Bus Network Forensics Information

Peng Yingjie
(School of Computer Science, Qinghai University for Nationalities, Chengdong District, Xining 810007, China)

Abstract: At present, the automatic retrieval system of forensics information does not filter the potential retrieval risk of forensic information, which leads to the problem of poor control effect, low retrieval efficiency and high error. Therefore, a system of automatic retrieval of risk control system of bus network forensics information is designed. After collecting the input information in the acquisition module, the risk control module is sent to the DSP for automatic retrieval after the risk filtering and risk control of the collection information information. The STM32F407 design interface circuit is used to connect the acquisition module and the risk control module to complete the improvement of the hardware part. It selects high degree of retrieval relevance node, uses node built-in documents to achieve the control of risk of forensic information retrieval, and completes the software part design. The experimental results show that the system has a good control effect and a control precision of more than 80%. It can provide more effective and safer information retrieval results for the users.

Keywords: bus network; forensics information; automatic retrieval; risk control

0 引言

当前总线网络取证信息的安全检索成为了未来人们获取取证信息与知识的主要手段[1]。由于信息站的建立,信息发布是大量、自由和无顺序的,且在取证信息的传输存储过程中,常会发生异变风险造成取证信息残缺、被篡改等现象,如果没有有效的风险控制措施,在总线网络中检索有用且安全的取证信息较为困难[2]。总线网络取证信息自动检索风险控制技术主要是搜索引擎风险控制技术,搜索引擎风险控制的实质就是一个专用控制器,该控制器可将总线网络中网站的取证信息组成庞大的取证信息数据库,用户使用关键词就可以在取证信息数据库中进行取证信息的检索,找出匹配的取证信息,同时该控制在检索过程中会对取证信息潜在的检索风险进行判断,将判断出的具有检索风险的取证信息进行过滤,并对其风险进行控制,从而降低取证信息的风险,使其能够实现后续的安全检索。要实现总线网络取证信息自动检索时的风险控制,已经有大部分的相关专家和学者对其进行研究,但至今仍未找到比较有效的风险控制途径[3-4]。现有的总线网络取证信息自动检索风险控制系统采用众包的方法进行设计。基于众包模式,利用架构设计总线网络取证信息自动检索风险控制系统,该系统主要包括服务端、客户端以及存储系统和主题系统4个模块。通过主题系统的分布式信息节点向服务器请求上传取证信息数据,利用分布式系统对取证信息数据进行快速处理并将处理的结果进行存储,然后进行检索风险控制软件设计。实验结果表明,该总线网络取证信息自动检索风险控制系统的配置较为简单,支持功能扩展,虽然具有较高的信息检索效率,但无法准确过滤出具有潜在检索风险的取证信息,致使风险控制效果差[5]。

针对上述问题,提出设计一种总线网络取证信息自动检索风险控制系统。实验结果证明,所提系统能有效地对

收稿日期:2018-06-04; **修回日期**:2018-07-09。

作者简介:彭英杰(1979-),男,甘肃省正宁人,硕士,讲师,主要从事计算机应用、电子商务与信息管理等方向的研究。

总线网络取证信息自动检索的风险进行高精度控制，从而实现取证信息的安全检索。

1　检索风险控制系统整体构造设计

要对总线网络中取证信息的自动检索风险进行控制，需对总线网络取证信息的自动检索系统进行改进，在改进的总线网络取证信息自动检索系统的基础上，设计总线网络取证信息自动检索风险控制系统。对总线网络取证信息自动检索风险控制系统进行设计，需以系统的整体结构为基础。

在设计总线网络取证信息自动检索风险控制系统的过程中，依据取证信息自动检索的功能以及自动检索的要求，建立检索风险控制系统的整体结构，该总线网络取证信息自动检索风险控制系统由取证信息采集模块、检索风险控制模块、电源电路、接口电路和取证信息自动检索模块组成，图 1 表示总线网络取证信息自动检索风险控制系统的整体结构示意图。

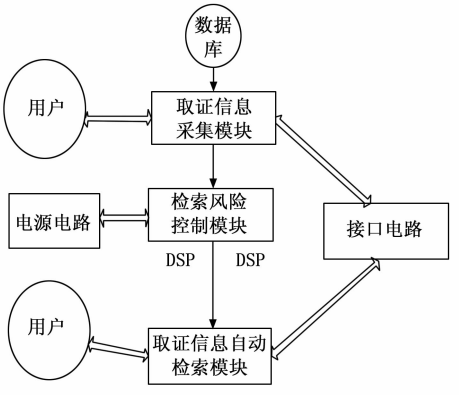

图 1　总线网络取证信息自动检索风险控制系统整体结构示意图

图 1 中，首先从取证信息数据库中输入取整信息数据，将输入的取证信息数据通过取证信息采集模块进行采集，取证信息采集后经过检索风险控制模块对采集后取证信息数据潜在的检索风险进行预测并加以控制，同时检索风险控制模块与电源电路相连接，使得受检索控制后的取证信息数据发送给 DSP[6]，经过 DSP 进行取证信息自动检索，接口电路与取证信息采集模块和取证信息自动检索模块相连接。最终组成了总线网络取证信息自动检索风险控制系统的整体结构。

2　硬件设计

以总线网络取证信息检索风险控制系统的整体结构为依据，对控制系统的硬件部分进行划分设计。总线网络取证信息自动检索系统的硬件部分是由取证信息采集模块、检索风险控制模块、应用 STM32F407[7] 完成的接口电路、电源电路和取证信息自动检索模块组成。各模块的具体设计过程如下：

2.1　取证信息采集模块

总线网络取证信息采集模块主要是完成取证信息的采集任务，采集的取证信息直接影响后续对这些取证信息数据进行处理、检索效率以及检索风险控制的效果，因此取证信息采集模块很重要。该取证信息采集模块首先经过数据库，进行取证信息的剥离、取证信息的隔离和取证信息的转换，而采集模块由通信接口和总线接口连接总控制区域，由总控制中心对其采集过程进行有效控制。

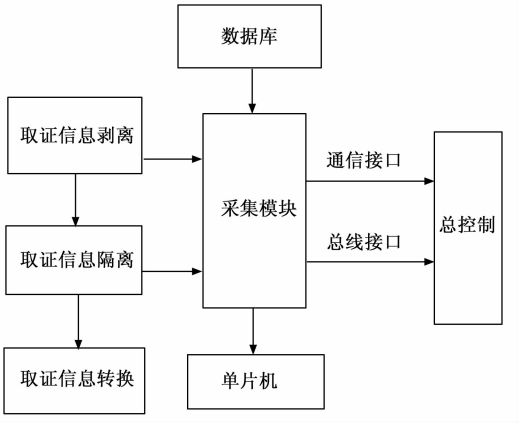

图 2　取证信息采集模块

2.2　检索风险控制模块

经由取证信息采集模块采集到的取证信息，在进行自动检索前，需对其潜在风险进行过滤和控制，即设计了检索风险控制模块。通过进行全局总线网络中取证信息的搜索，对取证信息是否具有潜在风险进行判断，将有风险的取证信息进行过滤，进而对过滤后的取证信息的检索风险进行控制处理。借鉴过滤规则组织模式，结合该过滤规则具有一定的扩展性对检索风险控制模块进行设计。检索风险控制模块的示意图由图 3 所示。

检索风险控制模块主要采用控制器风险控制技术，结合过滤规则与基础的关键字过滤技术，经由取证信息过滤子模块对取证信息进行过滤，提高了信息检索风险控制的准确率。控制器链接的取证信息过滤模块在提交过滤后的取证信息前，先对取证信息是否具有潜在检索风险进行判断，不带有潜在检索风险的取证信息则被过滤出来，进行搜索总线网络的检索，亦或是访问总线网络检索，最后到达内部用户。

2.3　接口电路设计

图 4 表示网络信息自动过滤检索的电源电路图。

电源电路的设计主要选用的型号为 ENC28J60，具有 PBI 接口，符合电路要求。该系统采用 STM32F407 结合 ENC28J60 完成总线网络取证信息的传输，STM32F407 透过对芯片控制实现取证信息的收发通信。芯片连接 PB11 接口，分别连接 PBI 进行中断输出，连接 PB12、PB13、PB14、PB15 进行信息输入引脚。

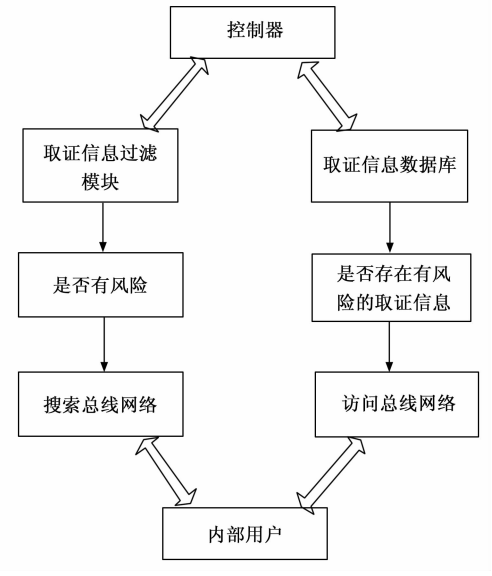

图 3 检索风险控制模块

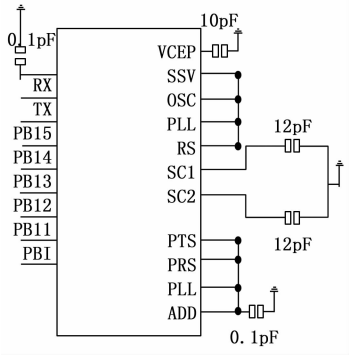

图 4 电源电路图

2.4 接口电路

接口电路主要用于当总线网络产生故障时，将采集的取证信息数据进行存储，待总线网络恢复正常时，将总线信息传输至 STM32F407 芯片[8]。连接通过 ASS 接口进行实现，STM32F407 依据服务器工作的情况进行读写，从芯片的 CSLK 的引脚输入到 ID 引脚，进行读取取证信息时，由 OD 引脚输入至 STM32F407 的 SC 引脚。取证信息的信号进行同步输入与输出。

2.5 取证信息自动检索模块

取证信息自动检索模块主要为实现对总线网络取证信息基本数据与模糊多条件的自动检索。同时还为实现对总线网络取证信息的数据代码以及取证信息的名称进行自动检索，为检索过程提供自动二次模糊检索功能，以提高检索结果的准确性。取证信息自动检索模块针对总线网络中取证信息间的隶属关系进行自主式查询，能查询出取证信息间的层次关系[9]。此外取证信息自动检索模块还可提供用户检索日志与检索信息量统计，实现动态的统计功能。

根据以上各模块的功能结构设计，整个总线网络取证信息检索风险控制系统主要包括取证信息的采集、取证信息的风险控制、无风险取证信息的过滤以及取证信息的自动检索，信息量检索统计，取证信息的新增以及取证信息的校验等主要功能。由此完成了总线网络取证信息检索风险控制系统硬件部分的设计，为系统软件部分的设计提供了优质的硬件环境。

3 软件设计

总线网络取证信息自动检索风险控制系统软件部分的主要核心问题是对检索过程的风险进行控制，其本质则是怎样有效地选择与检索相关度高的节点，找到检索相关度高的节点即可对检索风险进行有效控制。与检索相关度高的节点主要是指具有较多的与检索相关的文档，节点内的文档与检索相关度高的节点[10]。综合对量和质两个因素的考虑，信息检索的节点公式表示为：

$$rel(q,p_j)=\alpha\cdot\exp(rel(q,C_j))+\beta\cdot 1g(|C_j|_d) \quad (1)$$

公式（1）中，$rel(q,p_j)$ 表示考虑质与量得出的查询 q 与节点 p_j 的相关度，$rel(q,C_j)$ 表示质，$1g(|C_j|_d)$ 表示量。α 与 β 表示可调的系数，取值决定 $rel(q,C_j)$ 计算信息的准确性。

总线网络取证信息自动检索系统中的节点构建并维护节点资源描述（PRD），PRD 包含节点内的词条。对于取证信息词条 t_n，运用语言模型（$p(t_n|M_{dk})$）可计算出 t_n 在总线网络取证信息 C_j 中权重 w_n：

$$w_n=\frac{\sum_{d_i\in C_j}p(t_n|M_{d_i})}{|C_j|_d} \quad (2)$$

公式（2）中，$|C_j|_d$ 表示 p_j 节点的总线网络取证信息文档集 C_j 的大小，RAD 表示总线网络取证信息文档的索引信息。利用散度计算检索 q 与总线网络取证信息文档 C_j 的相关度（$rel(q,C_j)$）可表示为：

$$rel(q,C_i)=(q,C_i)=\sum_{t_n\in q}p(t_n|q)\left|\log\frac{p(t_n|q)}{p(t_n|C_j)}\right| \quad (3)$$

由公式（3）能看出，$rel(q,C_i)$ 的值越大，p_j 与 q 就越相关。

对于拥有取证信息文档集 C，在取证信息文档集 C 中存在与 q 相关的信息概率为：

$$p\{R(q,d_i)\mid d_i\in C\}=\left[1-\left(1-\frac{m}{M}\right)\right] \quad (4)$$

公式（4）中，$R(q,d_i)$ 表示取证信息文档 d_i 与检索的 q 有关，C 包含的取证信息较多，C 存在的和检索相关的取证信息概率就越大。

p_i 依据所在节点与检索的相关度 $rel(q,p_i)$，选择部分与 q 相关度高的节点，令这些取证信息节点进行检索任务，再返回查询的结果。将总线网络取证信息节点按与检索 q 的相关度进行从大到小排列，选择一部分的总线网络取证信息节点作为真正执行检索任务的节点。

所有和 p_i 在同一总线网络取证信息节点都需要利用

$rel(q,p_j)$ 进行计算。由公式（1）可知，$rel(q,p_j)$ 需要计算的量较小，其他计算的开销可以忽略不计。由于 p_i 管理的取证信息节点较少，进行排序的操作量也就较少，同时发送检索与返回结果占用的网络带宽也就较少。总线网络取证信息节点的优点是将检索的任务限定在与检索相关的节点，节省了总线网络取证信息节点的计算资源，还同时提高了总线网络取证信息自动检索结果的查准率，从而降低了取证信息的检索风险，实现对取证信息自动检索过程中检索风险的有效控制。

综合以上步骤，增加检索风险控制模块使得系统硬件结构充分对取证信息的潜在检索风险进行控制，并利用风险控制软件应用于检索风险控制模块，对取证信息的潜在检索风险进行精准控制，以确保后续检索的结果具有较高的准确性。

4　实验结果与分析

为证明总线网络取证信息自动检索风险控制系统的性能，需要进行一次实验。在 DSP 环境下搭建总线网络取证信息自动检索风险控制实验平台。实验数据来自 KDD-cup2016 取证信息数据集，该数据集中包括 150 万条取证信息数据。利用改进系统进行实验，观察改进系统的有效性。

4.1　实验参数由来

硬件配置实验平台的 CPU 为 Inter（R）CPUE5－26700，具有 16 个节点，内存为 64G×11 节点，存储为 8T 经过 NFS 进行共享，网络为千兆的以太网，操作系统为 RedHatEnterpriseLinux6.3，Kernel2.6.32。JDKWEI1.7.0－79。总线网络取证信息自动检索风险控制系统具有取证信息过滤转换等设备，可完成取证信息风险过滤等功能。实验的软件平台为 BBS2.1，能有效的进行取证信息的检索。利用上述实验数据对总线网络取证信息自动检索风险控制完成实验。

4.2　实验结果对比

表 1 表示改进系统与文献［8］系统、文献［9］系统的总线网络取证信息风险过滤效准确率对比。

表 1　不同系统取证信息风险过滤准确率对比

采样点数/个	文献[8]系统风险过滤准确率/%	文献[9]系统风险过滤准确率/%	所提系统风险过滤准确率/%
500	88.23	80.24	96.14
1000	84.15	78.26	95.26
1500	81.36	73.24	94.01
2000	76.26	67.14	93.38
2500	71.01	62.23	92.87

分析表 1 可知，文献［8］系统的对总线网络取证信息潜在检索风险的过滤准确率要高于文献［9］系统对总线网络取证信息潜在检索风险的过滤准确率，是因为文献［8］系统是利用众包的模式对取证信息潜在风险过滤的过程进行分析，利用众包的方法是通过主题系统的分布式节点向服务器请求上传数据，利用分布式系统对取证信息进行快速处理并将处理的结果进行存储，文献［8］系统有效地提高了取证信息潜在检索风险的过滤准确率。文献［9］系统是利用分词算法进行取证信息潜在检索风险的过滤，采用结合主索引与增量索引的方案，该系统包含取证信息采集模块、取证信息风险过滤模块以及存储模块，系统执行的主机主要采用多进程方式，进行总线网络取证信息风险的过滤，并将索引进行合并。但是文献［9］系统的风险顾虑精度较差。改进系统对总线网络取证信息自动检索风险控制系统的硬件方面进行全面的改进设计，以取证信息风险控制为前提，实现取证信息自动检索，其中对取证信息的风险过滤过程提高了取证信息潜在风险过滤的准确率。由此证明该方法具有可行性。

网络带宽利用率的大小同样对风险控制的有效性有直接的影响。为此分别对文献［8］系统、文献［9］系统改进系统的网络带宽利用率进行测试，图 5 表示改进系统与文献［8］系统、文献［9］系统的网络带宽利用率（%）对比结果。

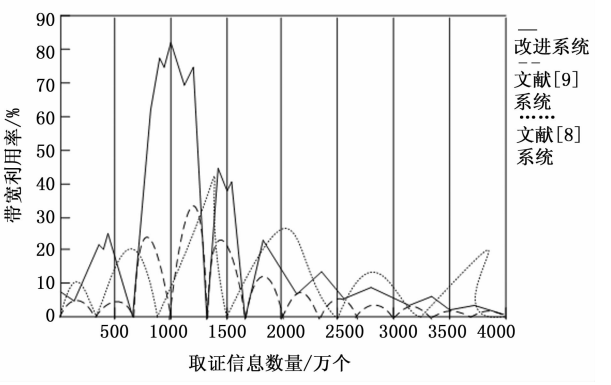

图 5　表示不同系统网络带宽利用率对比

分析图 5 可知，改进的总线网络取证信息自动检索风险控制系统的网络带宽利用率明显低于文献［8］系统基于众包方法的网络带宽利用率，且明显低于文献［9］系统基于分词算法的网络带宽利用率。是因为改进系统的取证信息自动检索风险控制是通过选择与检索相关度高的节点，与检索相关度高的节点主要是指具有较多的与检索相关的文档，节点内的文档与检索相关度高的节点。该风险控制过程能有效地降低网络带宽的利用率。而文献［8］系统是通过众包对总线网络的取证信息进行检索风险控制，文献［9］系统是通过分词算法对总线网络取证信息进行检索风险控制，文献［8］系统的带宽利用率相比文献［9］系统的带宽利用率还更低一些，利用率的波动也较明显，由此说明文献［8］系统与文献［9］系统可行性较低，由此说明改进系统的取证信息自动检索风险控制具

有可行性。

CPU 空间占用率同样会对风险控制的效果产生直接影响。为此测试不同系统的 CPU 空间占用率大小。图 6 表示改进系统与文献［9］系统、文献［10］系统的 CPU 空间占用率（%）对比结果。

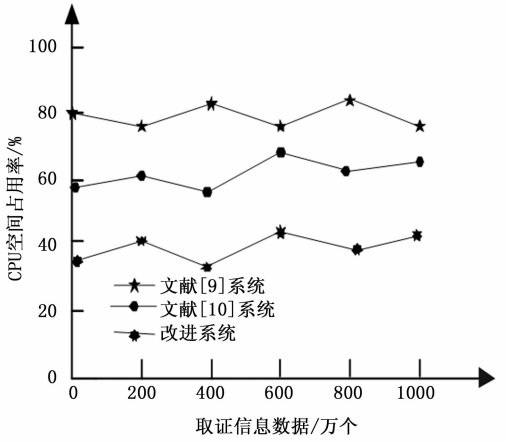

图 6 不同系统的 CPU 空间占用率对比

对图 6 进行分析可知，文献［9］系统的取证信息自动检索风险控制的 CPU 空间占用率明显高于文献［10］系统的 CPU 空间占用率，文献［10］系统采用拓扑特征对总线网络取证信息自动检索风险控制系统进行设计，主要是对各模块的功能进行设计与实现，利用这些功能进行总线网络取证信息的自动检索，然后进行系统的风险控制软件设计完成对取证信息自动检索时的风险进行控制。虽然文献［10］系统相对于文献［9］系统 CPU 占用空间低一些，但和改进系统相对比 CPU 空间占用率还是高一些，由此说明改进系统对取证信息自动检索风险的控制有效性较强。

对比不同系统对取证信息自动检索风险的控制效果，图 7 表示改进系统与文献［8］系统、文献［10］系统的风险控制精度（%）对比。

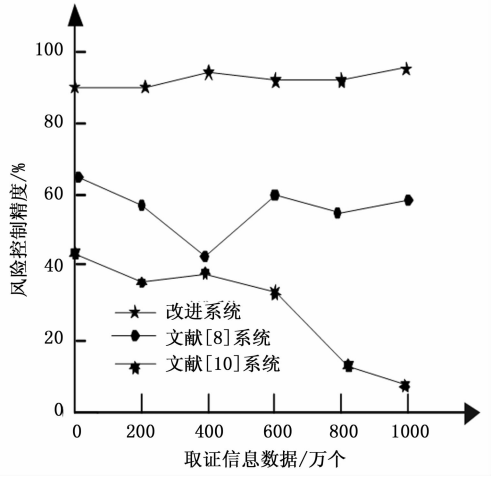

图 7 不同系统取证信息自动检索风险控制精度对比

对图 7 进行分析可知，改进系统的取证信息自动检索风险控制精度明显高于文献［8］系统与文献［10］系统。文献［8］系统的取证信息自动检索风险的控制精度曲线虽然波动不明显，但和改进系统的控制精度曲线相对还是波动较大一些。而文献［10］系统的取证信息自动检索风险的控制精度从信息少时就较低，随着取证信息数据的增加并没有改善。由此说明改进系统能有效得对取证信息自动检索的风险进行控制。

5 结束语

采用当前系统对总线网络取证信息自动检索系统进行检索风险控制时，忽略了对取证细腻潜在检索风险的过滤，致使风险控制效果差，检索效率低和检索误差较高的问题。为此，提出一种总线网络取证信息自动检索风险控制系统。并通过实验进行验证，所提系统能有效地对总线网络取证信息自动检索过程中的潜在检索风险进行控制，满足取证信息的大批量安全检索的需求，提高了风险控制效果，检索效率，降低了检索的误差。随着网络信息检索风险控制的广泛应用和更多的研究者参与到检索风险控制理论与研究中，能够在为用户检索出更有效、更准确、更安全的取证信息方面，发挥巨大的作用。

参考文献：

[1] 张学辉. 基于以太网和现场总线的工业控制网络实训系统设计［J］. 自动化仪表，2017，38（3）：41－43.

[2] 赵小刚. 基于 CAN 总线技术的机械臂自动控制系统设计［J］. 自动化与仪器仪表，2016（10）：56－58.

[3] 魏祥健. 云计算环境下的云审计系统设计与风险控制［J］. 会计之友，2015（1）：101－105.

[4] 许学添，邹同浩. 基于弱关联挖掘的网络取证数据采集系统设计与实现［J］. 计算机测量与控制，2017，25（1）：123－126.

[5] 伍世云，罗 江，王益艳，等. 基于单片机的高校教室照明节能智能控制系统的设计［J］. 电子设计工程，2016，24（23）：180－182.

[6] 汪超群，韦 化，吴思缘. 基于风险控制的 PSCOPF 改进模型及交替迭代算法［J］. 电力自动化设备，2017，37（4）：114－121.

[7] 张 璇，杜 强. 基于静态和动态分析的 Android 短信拦截马自动分析取证方法研究［J］. 计算机科学，2016，43（b12）：30－34.

[8] 张钰婷，邵 勇，顾桂鹏. 基于 CAN 总线的网络门禁控制系统设计［J］. 工业控制计算机，2017，30（2）：29－30.

[9] 王明俭，宋仁平. 基于用电信息采集系统的反窃电研究［J］. 工程技术：全文版，2016（9）：00287－00287.

[10] 曹树金，李洁娜，王志红. 面向网络信息资源聚合搜索的细粒度聚合单元元数据研究［J］. 中国图书馆学报，2017，43（4）：74－92.