

基于二维码的内外网物理隔离环境下的数据交换

韩 林^{1,2} 张春海¹ 徐建良¹

(中国海洋大学信息科学与工程学院 青岛 266100)¹ (国家海洋局第一海洋研究所 青岛 266061)²

摘 要 针对保密性较高的工作数据或者其他原因导致的内外网物理隔离环境下数据交换困难的问题,通过对二维码生成和解析过程的研究,并且利用二维码可以携带数据、成本低和可随载体移动的特性,提出了使用二维码来解决一些特殊情况下的数据交换问题。由于单个二维码可携带数据有限,提出利用 Protocol Buffer 格式和 LZMA 压缩算法来简化和压缩需要通过二维码传输的数据,对于大型数据则采取多个二维码的组合方式传输。另外还简述了基于二维码的数据交换的应用前景。

关键词 二维码,物理隔离,数据交换,ZXing,Protocol Buffer,LZMA

中图分类号 TP391 文献标识码 A

Data Exchange Based on QR Code in Physically Isolated Internal and External Network Environment

HAN Lin^{1,2} ZHANG Chun-hai¹ XU Jian-liang¹

(College of Information Science and Engineering, Ocean University of China, Qingdao 266100, China)¹

(The First Institute of Oceanography State Oceanic Administration, Qingdao 266061, China)²

Abstract To solve the problem of high confidentiality work data or data exchange problems in internal and external network physical isolation environment caused by other reasons, and through the research of QR code generation and the parsing process, we found that QR code has several characteristics, such as it can be used to carry data, low cost and it can follow the characteristics of mobile carrier. Then we raised the point of view that the QR code can be used to solve the problem of data exchange of some special cases. Because a single QR code can carry limited data, we put forward using the Protocol Buffer format and LZMA compression algorithm to simplify and compress the data that QR code transmits. And we used the combination of multiple QR code to solve the problem of large amounts of data. This paper also expounded the application prospect of data exchange based on QR code.

Keywords QR code, Physical isolation, Data exchange, ZXing, Protocol Buffer, LZMA

1 前言

中国大洋样品馆的主要职能为样品的收集、整理、保存、分配和属性数据的提取^[1]。样品馆内只有少数设备可以连接外网,工作中站位和样品等数据是需要保密的,只允许在其内部网络中访问,并且只允许通过光盘进行内外网之间的数据交换。

现行的大洋样品管理系统是对样品相关数据进行管理的系统,它在内网环境下运行且不允许连接到外网。而即将上线的大洋样品申请书系统则是一种方便海洋专家提交样品申请的网站应用,是连接在外网上使用的。在目前实际工作中,这两个系统交换数据时需要光盘作为介质,而光盘并不方便,写入和读取速度较慢,且易损耗。分析上述两个系统中可能的交换数据,申请样品的申请书信息较为特殊,它们有固定的格式并且一个申请书的数据量并不大。

随着互联网时代的来临,二维码也因其其在微信、淘宝等项目上的广泛应用而流行起来,二维码以它生成简单、读取快速

等特点受到众多应用开发者和使用者的青睐。目前国内关于二维码的研究也较为热门,但大多数研究的重点是二维码的生成解析和其简单的应用,利用二维码传输数据的研究还比较少。虽然二维码所能携带的数据量是有限的,但在其可容纳的数据量下,发挥二维码的优点以进行应用还是可以带来一定的方便。

针对内外网物理隔离环境下的数据交换问题,目前较多的研究是通过设计专业的硬件和软件来解决。王济意在《基于物理隔离技术的安全信息交换系统》中设计了具有物理隔离功能的硬件板卡机构和通信机制^[2]。唐利峰在《一种物理隔离数据交换系统的设计》中也提出了一种基于物理隔离卡的系统结构^[3]。张海建和孙丹在《二维码扫描技术在数据传输中的应用》中提出使用二维码作为传递项目短消息的介质,将消息转化为 Unicode 字符然后通过生成和解析二维码进行传输^[4]。汪松、李晓华和王延红在《一种在物理隔离的网络间实现数据安全交换的方法》中提出一种基于二维码传递原材料需求和回执数据的方法^[5]。

本文受国际海域资源调查与开发“十二五”课题资助项目(DY125-25-01)资助。

韩 林(1992—),男,硕士生,主要研究方向为数据库理论及应用,E-mail:hanlin824@sina.com;张春海(1963—),男,硕士,教授,主要研究方向为数据库理论及应用、软件工程,E-mail:chai@ouc.edu.cn;徐建良(1969—),男,博士,教授,主要研究方向为计算复杂性理论、计算机软件与理论,E-mail:xjl9898@126.com。

鉴于大洋样品申请书信息的特点,将二维码作为大洋样品管理系统和大洋样品申请书系统之间传输大洋样品申请书信息的介质是可行的,并且利用相关技术对数据进行预处理并优化二维码传输流程是值得研究的。

2 QR code 码和 ZXing

2.1 QR code 简介

二维码的种类很多,常用的二维码的码制是 QR code。QR code 又称为快速响应矩阵码,与其他二维条码相比,它具有数据量大、超高速识读、全方位识读、应用方便、纠错能力强和能够有效地表示汉字等优点^[6]。本文采用了 QR code 码作为传输数据的二维码载体。

QR code 由名义上的正方形模块构成,组成一个正方形阵列,它由编码区域和包括寻像图形、分隔符、定位图形、校正图形在内的功能图形组成。编码区是由文本信息转化的二进制码组成,它包括格式信息、版本信息、数据和纠错码字^[7]。其结构如图 1 所示。

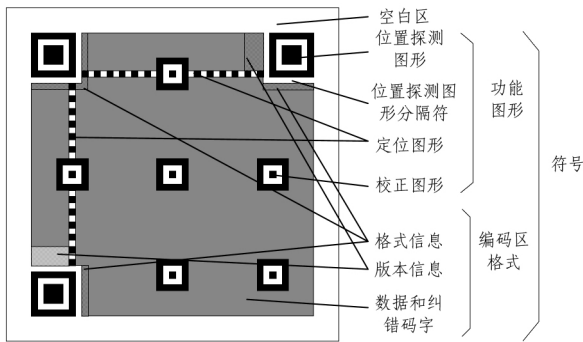


图 1 QR code 符号的结构

2.2 ZXing 简介

ZXing 是由 Google 公司主导开发的一个开放源代码库,它用 Java 语言实现了对多种格式的一维和二维条码的图像处理解析的支持,并且提供了其他多种语言的程序接口,代码托管于 Github^[8]。本文利用 ZXing 库实现对 QR 码的编码和解码,目前最新版本的 ZXing 为 2015 年 8 月 19 日发布的 ZXing 3.2.1。

2.3 程序初步设计与实现 ZXing 源代码的修改和使用

现有申请书采用 XML 格式传输,部分信息如图 2 所示,由于二维码支持以混合字母、数字和汉字组成的数据生成二维码,所以尝试直接将申请书生成二维码。

```
<?xml version="1.0" encoding="UTF-8" ?>
<overview version="0.1" count="" date="" voyage="">
<application type="B1">
  <number>B1001C001</number>
  <applyTime>2014/12/12</applyTime>
  <applicant>
    <name>张三</name>
    <departmentName>中国海洋大学</departmentName>
    <cellPhone>13333333333</cellPhone>
    <email>zhangsan3@126.com</email>
  </applicant>
  <leg>K</leg>
</application>
```

图 2 申请书部分信息

根据用户需求分析所设计的使用流程:与外网连接的申请书系统将申请书信息直接传给生成端程序,生成端程序利用 ZXing 库把数据生成二维码,并显示在电脑屏幕;内网环境下的室内系统利用接收端的摄像头拍摄生成端屏幕上的二维码,并传给接收端程序进行解析得到可用的申请书信息,此流程如图 3 所示。程序采用 WinForm 实现,对摄像头控制使用 TouchlessLib 库。

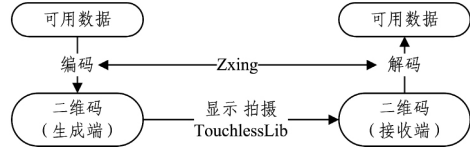


图 3 程序初步设计流程

直接将申请书信息生成二维码,其数据传输效率较低,因为一个二维码所能携带的真正的申请书信息的数据相对较少,并且可能出现一个二维码无法承载一个申请书中所有的信息数据的情况。尝试改进程序,只传递真正的申请书数据并且压缩数据量。

3 Protocol Buffer 格式和 LZMA 压缩算法的应用

3.1 Protocol Buffer 简介及使用

Protocol Buffer 是 Google 公司开发的一种数据描述语言,广泛地应用于 Google 内部,用于结构化数据的描述、传输和存储,并且它是和平台无关的语言,可以支持 Java、C++、python 等多种语言环境,其第三方扩展可以支持更多语言^[9,10]。其最大的特点是基于二进制进行传输,相比于 XML 和 JSON 这样的数据描述语言,虽然它的可读性比较低,但其数据传输效率是最高的。

Protocol Buffer 的使用方法如下:首先按照其官方文档规则及申请书具体属性,编写数据信息格式文件(SiteApplicationBase.proto 和 SiteApplication.proto);然后使用 Protocol Buffer 编译器编译 SiteApplication.proto 文件;其次将生成的自动存取类文件引入生成端和接收端程序项目中,即可使用该文件中的 Build 函数和 ParseFrom 函数进行序列化和反序列化。

3.2 LZMA 压缩算法简介及使用

LZMA 是 7zip 这一开源的压缩软件中默认的一种压缩算法,是运用“字典模型”的经典的 LZ 系列算法经改良和优化后的最新版本^[11]。LZMA 压缩算法的主要特征表现为高压缩比、压缩速率高等。

LZMA 的使用也较为方便,将 7zip 官方的 LZMA SDK 下载并引入到生成端和接收端程序项目即可,利用其中的 Encoder 类和 Decoder 类来实现压缩与解压缩。

3.3 改进程序的设计与实现

将上述两项技术应用到程序中,以改进二维码的传输效率,改进后的程序流程如图 4 所示。

使用 Protocol Buffer 将申请书信息二进制化,相比 XML 或 JSON 格式的数据,Protocol Buffer 格式的数据少了大量的额外数据。此处所谓的额外数据指的是 XML 中的标签和 JSON 中的键,因为 Protocol Buffer 传输的数据中主要为有用的数据部分,而用于标记数据的标签值和类型信息已被简化,每个标记信息最少仅需一个 byte。文件大小对比见表 1。

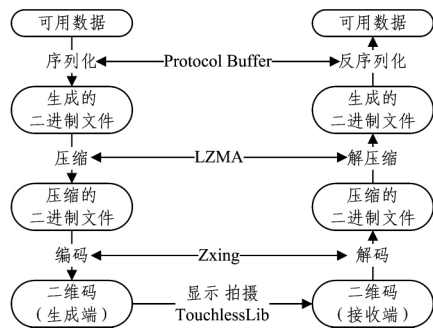


图4 改进后的程序流程

应用 LZMA 压缩算法后,将二进制的申请书信息数据压缩,数据量再次减少,基本可以实现一个二维码携带一个完整的申请书信息数据,单个二维码在最大规格下可承载 2953 个 8 位字节型数据。一个测试用的申请书在不同格式下的文件大小对比见表 1。可以看出在使用了 Protocol Buffer 和 LZMA 压缩后,文件大小已经大大地减少,一个二维码已经可以容纳。

表 1 不同数据格式的文件大小比较

数据格式	文件大小(byte)
XML	11645
JSON	8075
Protocol Buffer	6872
Protocol Buffer+LZMA	1201

由于被 Protocol Buffer 序列化后的数据为二进制数据,所以需要修改 ZXing 源代码,以适应二进制数据的二维码生成和解析二维码后对二进制数据的提取。首先修改 ZXing 源代码中二维码生成部分的代码,使二进制数据绕过 ZXing 对传入数据的格式判断和数据编码操作,将格式设置为 8 位字节型数据模式,再将其传入下一个步骤,如图 5 所示。

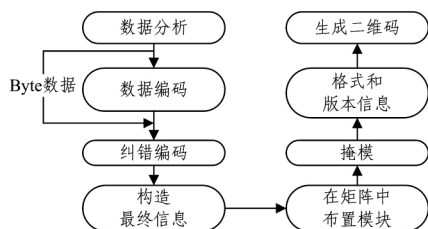


图5 ZXing生成二维码代码修改

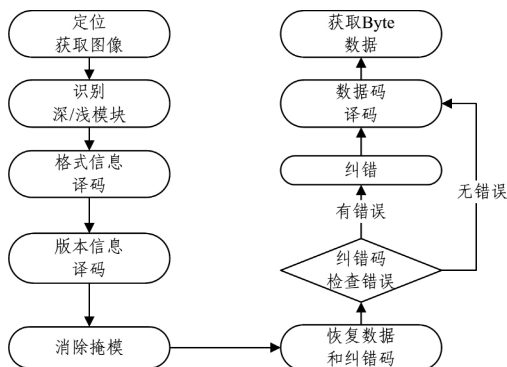


图6 ZXing解析二维码流程

而对于生成端的修改,初步想法是绕过 ZXing 对解析二维码后得到的二进制数据的译码,但通过对 ZXing 源代码的

研究,发现 ZXing 返回的解析二维码的结果对象中含有未译码的原始二进制数据,故可以直接获得。流程如图 6 所示。

4 申请书数据量变化的解决方法

随着我国大洋领域研究的发展和业务量的增加,申请书所包含的内容可能会调整或增加,届时可能出现即使采用本文提出的方法进行压缩后,一个二维码也无法承载单个申请书数据量的情况。针对这种情况,本文给出以下解决方法。

首先,二维码生成端程序将申请书信息经过 Protocol Buffer 和 LZMA 压缩算法的加工,再将加工好的二进制数据分割成几段单个二维码可以承载的数据量,然后于每段数据的开头添加顺序标志位,最后分别生成二维码,并同时显示在屏幕中。

其次,二维码接收端程序捕捉到多张二维码后,自动将捕捉的照片利用图像边界识别并切割成单张的二维码再解析,并且根据解析出的每段数据的开头的顺序标志位将二进制数据重新组合在一起,继而还原出申请书的信息。

结束语 本文研究了将二维码作为传输大洋样品申请书信息数据的介质的方法。针对一个二维码所能携带的数据量有限而无法承载一个完整的申请书信息的情况,又提出利用 Protocol Buffer 和 LZMA 技术来压缩传输中的数据,效果比较理想。

这种方法也适用于其他数据量不大的结构化数据的传递。例如针对目前的快递单可能出现的个人信息泄露问题,可以将快递单信息的二进制数据加密后再生成二维码,而快递公司用特制的能解密的二维码扫描设备可以读取快递单信息,以保证快递信息的安全性。

参考文献

- [1] 杜林涛. 大洋样品仓库信息化管理的研究[D]. 青岛: 中国海洋大学, 2013
- [2] 王济意. 基于物理隔离技术的安全信息交换系统[D]. 西安: 西安电子科技大学, 2015
- [3] 唐利峰. 一种物理隔离数据交换系统的设计[J]. 信息技术, 2009(10): 30-32
- [4] 张海建, 孙丹. 二维码扫描技术在数据传输中的应用[J]. 计算机光盘软件与应用, 2012, 20: 65-66
- [5] 汪松, 李晓华, 王延红. 一种在物理隔离的网络间实现数据安全交换的方法[J]. 电脑编程技巧与维护, 2015(7): 95-96
- [6] 郭琳. QRCode 二维码编译码及自动识别技术的研究[D]. 昆明: 昆明理工大学, 2008
- [7] 中国物品编码中心. GB/T 18284-2000 快速响应矩阵码(QR code)[S]. 北京: 中国标准出版社, 2004
- [8] 杨前华, 邓韦. 基于物联网技术的手机条码识别系统的研究[J]. 微计算机信息, 2012, 10: 335-336
- [9] 殷昊, 沈奇威, 王纯. ProtocolBuffer 在 Android 企业云通讯录中的应用[J]. 电信科学, 2012, 9: 148-151
- [10] Kaur G, Fuad M M. An evaluation of Protocol Buffer[C]//IEEE Southeastcon. 2010: 459-462
- [11] 刘坚, 李胜乐, 王子影. 基于 LZMA 的数据库压缩存储应用研究[J]. 大地测量与地球动力学, 2009, 29(6): 144-147