西安交通大學

硕士学位论文

基于 IBF 的对称可搜索加密研究与实现

学位申请人：王鹤宇

指导教师：侯迪 教授

学科名称：计算机科学与技术

XXXX 年 X 月

Research And Implementation of Symmetric Searchable Encryption Based on IBF

A dissertation submitted to

Xi'an Jiaotong University

in partial fulfillment of the requirements

for the degree of

Master of Engineering

By

Heyu Wang

Supervisor: Prof. Di Hou (导师姓名全拼，例如 Anxue Zhang)

Associate Supervisor: Prof. Yong Qi

Computer Science and Technology

XXX XXXX (英文日期，月在前，年在后，例如：September 2017)

硕士学位论文答辩委员会

基于 IBF 的对称可搜索加密研究与实现

答辩人：王鹤宇

答辩委员会委员：

XXXXXXXX 大学 XXX：_____ (主席)

XXXXXXXX 大学 XXX：_____

XXXXXXXX 大学 XXX：_____

XXXXXXXX 大学 XXX：_____

XXXXXXXX 大学 XXX：_____

答辩时间：XXXX 年 XX 月 XX 日

答辩地点：XXXXXXXXXXXXXX

摘 要

随着云计算和大数据的发展，越来越多的用户数据被存放在云端。这些数据中通常包含一些敏感数据，如个人身份证号，健康记录等等。如果数据以明文形式存放在云端，则很容易导致隐私信息的泄露。因此为了保护用户的隐私数据，会在将数据上传到云服务器之前对其进行加密。为了使用云端的密文数据，通常会先用大量的网络带宽将密文下载，在本地完全解密后再检索。然而这种方案有以下两个致命缺点：1. 如果云服务器上含有大量数据，一一下载会占用大量网络带宽。2. 对已下载的文件完全解密会占用大量本地计算资源，效率极低。解决此类问题的技术称为可搜索加密 (Searchable Encryption,SE), 该技术要求只有被授权的用户才具有检索能力。

本文提出并实现了一种新的对称可搜索加密方案，并将其和已有的方案在实验上进行了对比。同数据库使用索引提高搜索效率一样，对称可搜索加密方案通常会针对明文文件集构建相应的安全索引 (不泄露相应明文文件集合信息的索引)。本文基于一种新型的数据结构——不可区分布隆过滤器 (Indistinguishable Bloom Filter,IBF1)[1]，提出了一种新的安全索引构建算法。关键词在被送往云服务器进行搜索前，需要通过某种变换来隐藏关键词中的明文信息，同时使得变换后的词又能够在安全索引上进行匹配搜索。这种变换被称为陷门 (trapdoor) 生成算法，本文也提出了相应的陷门生成算法和陷门在安全索引上的搜索算法。为了验证方案的有效性，分别用 C 语言实现了我们提出的方案以及另一种基于 IBF 的可搜索加密方案。从 1. 安全索引构建时间，2. 安全索引占用的磁盘空间，3. 搜索安全索引花费的时间这三个方面对两种方案进行了对比。在真实数据集上的实验结果表明，本文提出的方案极大的减少了安全索引的构建时间和对磁盘的使用空间情况，同时搜索时间和另一方案处于同一数量级 (ms 量级)。同时本文提出的方案也满足目前提出的最强的安全模型——自适应安全模型。

安全索引占据的磁盘空间以及搜索时间 3 个方面进行了对比。本文第一次提出了使用作为树节点来构造一棵变长不可区分布隆过滤器树 (Variable-length Indistinguishable Bloomfilter Tree,VBTree) 的方式来构建安全索引，同时也给出了与 VBTree 相对应的陷门 (trapdoor) 算法以及针对相应陷门的服务器端的搜索算法。

关 键 词：对称可搜索加密，IBF，VIBtree

论文类型：应用基础

ABSTRACT

英文摘要正文每段开头不缩进，每段之间空一行。

The abstract goes here.

LaTeX is a typesetting system that is very suitable for producing scientific and mathematical documents of high typographical quality.

KEY WORDS: Xi'an Jiaotong University, Doctoral dissertation, LaTeX template

TYPE OF DISSERTATION: Application Fundamentals

目 录

CONTENTS

1 LaTeX 介绍

本章对 LaTeX 排版系统做一个简要介绍，希望没有使用过 LaTeX 的同学对 LaTeX 有一个初步认识。

1.1 LaTeX 是什么

LaTeX 是一款排版软件，和其它排版软件 (例如 Word) 相比，LaTeX 具有非常明显的优势和不足。其最大的优势是高质量、高水准的专业排版效果；最大的缺点是使用门槛高，需要具备一定的编程基础[①]。对于习惯于抽象思维的科技人员而言，与精美的排版效果相比，LaTeX 的确缺点是微不足道的，只要经过短时间 (一周足已) 的学习和实践，就可以编写出高质量的科研论文。

LaTeX 的基础是 TeX，TeX 诞生于 20 世纪 70 年代末到 80 年代初，用来排版高质量的书籍，特别是包含数学公式的书籍。有趣的是，这样一款排版软件并非在排版业界产生，而是由著名计算机科学家 Donald Ervin Knuth (中文名高德纳) 在修订其七卷巨著《计算机程序设计艺术》时设计的。

虽然 TeX 功能非常强大，但是多达 900 多条的排版命令让排版人员使用起来非常不便。因此 20 世纪 80 年代初，Leslie Lamport 博士给 TeX 编写了一组自定义命令宏包，并取名为 LaTeX，其中 La 是其姓名的前两个字母。LaTeX 拥有比原来的 TeX 更为规范的格式命令和一整套预定义的格式，可以让完全不懂排版技术的学者们很容易地将书籍和文稿排版出来。LaTeX 一出，很快风靡全球，在 1994 年 LaTeX 2_ε 完善之后，现在已经成为国际上数学、物理、计算机等科技领域专业排版的事实标准，相关专业的学术期刊也都采用 LaTeX 作为投稿格式。

1.2 为什么用 LaTeX

虽然论文排版是一项基本技能，但是从实际情况看，同学们经常被各种格式整得晕头转向。加之 Word 排版不够美观，版本管理麻烦，排版效率低下，因此开发 LaTeX 论文模板非常重要。国际上许多著名的出版机构和学术期刊都有自己的 LaTeX 模板，国内外许多高效也有自己的硕博论文 LaTeX 模板。事实上，LaTeX 已经成为科技出版行业的国际标准，特别是数学、物理、计算机和电子信息学科。

采用 LaTeX 排版主要有以下优点：

1. 排版质量高：主要体现在对版面尺寸的严格控制，对字距、行距和段距等间距的松紧适度掌握，对数学公式的精细设计，对插图和表格的灵活处理，对代码和算法的优美呈现，等等。

[①] 因为 LaTeX 的资源非常丰富，有许多模板可以使用，这些模板已经为用户定制好了排版格式，所以单纯从使用的角度看，使用 LaTeX 的门槛其实并不算高。

2. 安全稳定：自发布以来 TEX 和 LATEX 没有发现系统漏洞，不会出现死机或者系统崩溃而导致编写的内容来不及保存。
3. 灵活方便：LATEX 的源文件是纯文本文件，文件大小比 Word 小很多，不会因为文容的增加而导致文档打开、编辑、保存和关闭等操作变慢。因为 LATEX 在编译时才将所有源文件和图表汇总，故撰写内容时可以随意增删章节和图表。并且和大部分程序设计语言一样，LATEX 具有注释功能，作者可以在源文件任何地方添加注释，而不会影响最终生成的文档。
4. 格式和内容分离：LATEX 将文档格式和文档内容分开处理，作者只要选择合适的模板，就可专心致志地撰写文档内容，文档的格式细节则由 LATEX 模板统一规划设置。特别是文献管理能力非常强大，这给撰写像博士论文一样需要大量引用参考文献的文档提供了很大便利。
5. 免费开源：LATEX 软件完全免费，源代码也全部公开，并且相应的配套软件也都采用开源的方式。

无论你是因为羡慕 LATEX 漂亮的输出结果，还是因为要给学术期刊投稿而被逼上梁山，都不得不面对这样一个事实：LATEX 是一种并不简单的排版软件，不可能只点点鼠标就弄好一篇漂亮的文章。还得拿出点搞研究的精神，通过不断练习，才能编排出整齐漂亮的论文。一旦你掌握了如何使用 LATEX 撰写出精美漂亮的论文时，你会发现你的决定是明智的，你的投入是值得的。

1.3 怎样用 LATEX

本模板在 Windows + TeXLive2016 + Texsdudio 平台下开发，采用 XeLaTex 编译。虽然之前也开发过一个基于 CTeX 的模板，但是经过多方面比较发现 TeXLive+XeLaTex 处理中文更好，所以基于 CTeX 的模板没有共享。

本模板不能在 CTeX 软件下使用，必须采用 TeXLive，并且编译方式是 XeLaTeX。TeXLive 每年更新一个版本，我用的是 TeXLive2016。文本编辑器可以根据自己的喜好选用，我用的是 Texsdudio，这款开源软件非常不错，推荐大家使用。

本模板的源文件通过主目录下的 main.tex 统一管理，setup 文件夹中存放格式定义和封面、摘要、目录等内容，body 文件夹中存放论文正文章节的源文件，appendix 文件夹中存放附录、致谢和声明等内容。

本模板只提供论文的格式定义，不提供 LATEX 的详细使用方法。因为 LATEX 的资源非常丰富，大家可以在网上查找资料和并参与讨论，这样学习效率更高。我关注的两个网站是：`http://bbs.ctex.org/forum.php` 和 `http://www.latexstudio.net`；参考的两本书是 “The Not So Short Introduction to LATEX 2_ε” 和 “LaTeX2e 完全学习手册”。

1.4 实际显示内容

crap

2 图表公式排版

虽然本模板不讲解 LaTeX 的详细使用方法，但是为了方便大家使用本模板撰写论文，本章对论文写作中经常用到的**图**、**表**、**公式**等内容的排版方法做一个简单介绍。

2.1 图

2.1.1 单幅图

图 2-1 是用 TeXLive 自带的宏包 Tikz 绘制而成，Visio 画不出这么好看的图。

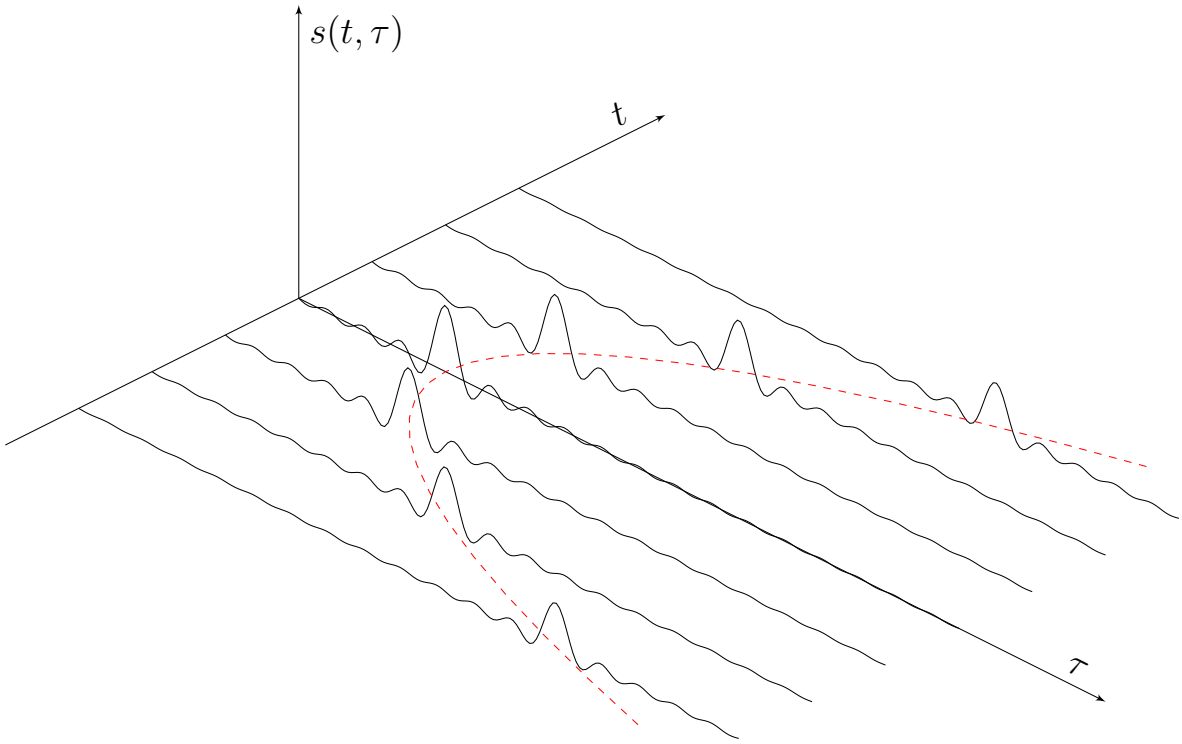

图 2-1 雷达回波信号 (注意：图注是五号字)。

2.1.2 多幅图

如果一幅图中包含多幅子图，每一幅子图都要有图注，并且子图用 (a)、(b)、(c) 等方式编号，如图 2-2 所示。

2.2 表

表格要求采用三线表，与文字齐宽，顶线与底线线粗是 $1\frac{1}{2}$ 磅，中线线粗是 1 磅，如表 2-1 所示[①]。

① 注意：图表中的变量与单位通过斜线 / 隔开。

(a) 灰色的交大校徽

(b) 蓝色的交大校徽

图 2-2 交大校徽

表 2-1 表题也是五号字

Interference	DOA / degree	Bandwidth / MHz	INR / dB
1	−30	20	60
2	20	10	50
3	40	5	40

2.3 公式

2.3.1 单个公式

LaTeX 最强大的地方在于对数学公式的编辑，不仅美观，而且高效。单个公式的编号如式 (2-1) 所示，该式是正态分布的概率密度函数[2]，

$$f_Z(z) = \frac{1}{\pi\sigma^2}\exp\left(-\frac{|z-\mu|^2}{\sigma^2}\right) \tag{2-1}$$

式中：μ 是 Gauss 随机变量 Z 的均值；σ^2 是 Z 的方差。

2.3.2 多个公式

多个公式作为一个整体可以进行二级编号，如式 (2-2) 所示，该式是连续时间 Fourier 变换的正反变换公式[3]，

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\mathrm{d}t \tag{2-2a}$$

$$x(t)=\int_{-\infty}^{\infty}X(f)e^{j2\pi ft}\mathrm{d}f \tag{2-2b}$$

式中：$x(t)$ 是信号的时域波形；$X(f)$ 是 $x(t)$ 的 Fourier 变换。

如果公式中包含推导步骤，可以只对最终的公式进行编号，例如：

$$\begin{aligned}\mathbf{w}_{\mathrm{smi}}&=\alpha\left[\frac{1}{\sigma_n^2}\mathbf{v}(\theta_0)-\frac{1}{\sigma_n^2}\mathbf{v}(\theta_0)+\sum_{i=1}^{N}\frac{\mathbf{u}_i^H\mathbf{v}(\theta_0)}{\lambda_i}\mathbf{u}_i\right]\\&=\frac{\alpha}{\sigma_n^2}\left[\mathbf{v}(\theta_0)-\sum_{i=1}^{N}\mathbf{u}_i^H\mathbf{v}(\theta_0)\mathbf{u}_i+\sum_{i=1}^{N}\frac{\sigma_n^2\mathbf{u}_i^H\mathbf{v}(\theta_0)}{\lambda_i}\mathbf{u}_i\right]\\&=\frac{\alpha}{\sigma_n^2}\left[\mathbf{v}(\theta_0)-\sum_{i=1}^{N}\frac{\lambda_i-\sigma_n^2}{\lambda_i}\mathbf{u}_i^H\mathbf{v}(\theta_0)\mathbf{u}_i\right]\end{aligned} \tag{2-3}$$

3 参考文献格式

参考文献格式应符合国家标准 GB/T-7714-2005《文后参考文献著录规则》。中国国家标准化管理委员会于 2015 年 5 月 15 日发布了新的标准 GB/T 7714-2015《信息与文献参考文献著录规则》。因为二者的差别非常小，所以采用了新的标准。标准的 BiBTeX 格式网上资源非常多，本模板使用了李泽平开发的版本，该版本提供了多种参考文献的排序规则。学校博士论文规范指定了两种排序方法：一是按照文献的引用顺序进行排序，二是按照作者姓氏加出版年份进行排序。本模板采用第一种排序规则，第二种排序规则的使用方法请参考文献 [4]。

致　谢

致谢中主要感谢导师和对论文工作有直接贡献和帮助的人士和单位。

一般致谢的内容有：

1. 对指导或协助指导完成论文的导师；
2. 对国家科学基金、资助研究工作的奖学金基金、合同单位、资助或支持的企业、组织或个人；
3. 对协助完成研究工作和提供便利条件的组织或个人；
4. 对在研究工作中提出建议和提供帮助的人；
5. 对给予转载和引用权的资料、图片、文献、研究思想和设想的所有者；
6. 对其他应感谢的组织和个人。

致谢言语应谦虚诚恳，实事求是。字数不超过 1000 汉字。

用于双盲评审的论文，此页内容全部隐去。。

参考文献

[1] Li R, Liu A X. Adaptively secure conjunctive query processing over encrypted data for cloud computing[C]. 2017 IEEE 33rd International Conference on Data Engineering (ICDE): volume 1, 2017: 697-708. DOI: 10.1109/ICDE.2017.122.

[2] Manolakis D G, Ingle V K, Kogon S M. Statistical and Adaptive Signal Processing[M]. Norwood: Artech House, Inc., 2005.

[3] Vetterli M, Kovacevic J, Goyal V K. Foundations of Signal Processing[M]. Cambridge: Cambridge University Press, 2014.

[4] Lee Z. GB/T 7714-2015 参考文献 BiBTeX 样式 [M/OL]. 2016. `https://github.com/zepinglee/gbt7714-bibtex-style`. DOI: fuck.

附录 A　公式定理证明

附录编号依次编为附录 A，附录 B。附录标题各占一行，按一级标题编排。每一个附录一般应另起一页编排，如果有多个较短的附录，也可接排。附录中的图表公式另行编排序号，与正文分开，编号前加“附录 A-”字样。

本部分内容非强制性要求，如果论文中没有附录，可以省略《附录》。

排版数学定理等环境时最好给环境添加结束符，以明确定理等内容的起止标志，方便阅读。官方模板未对这些内容进行规范，本模板中定义的结束符采用 ◊，例子的结束符采用 ⧫，定理的结束符采用 □，证明的结束符采用 ■。

定义 A.1 (向量空间)： 设 X 是一个非空集合，$\mathbb{F}$ 是一个数域 (实数域 $\mathbb{R}$ 或者复数域 $\mathbb{C}$)。如果在 X 上定义了加法和数乘两种运算，并且满足以下 8 条性质：

1. 加法交换律，$\forall\, x, y \in X$，$x + y = y + x \in X$；
2. 加法结合律，$\forall\, x, y, z \in X$，$(x + y) + z = x + (y + z)$；
3. 加法的零元，$\exists\, 0 \in X$，使得 $\forall\, x \in X$，$0 + x = x$；
4. 加法的负元，$\forall\, x \in X$，$\exists\ -x \in X$，使得 $x + (-x) = x - x = 0$。
5. 数乘结合律，$\forall\, \alpha, \beta \in \mathbb{F}$，$\forall\, x \in X$，$(\alpha\beta)x = \alpha(\beta x) \in X$；
6. 数乘分配律，$\forall\, \alpha \in \mathbb{F}$，$\forall\, x, y \in X$，$\alpha(x + y) = \alpha x + \alpha y$；
7. 数乘分配律，$\forall\, \alpha, \beta \in \mathbb{F}$，$\forall\, x \in X$，$(\alpha + \beta)x = \alpha x + \beta x$；
8. 数乘的幺元，$\exists\, 1 \in \mathbb{F}$，使得 $\forall\, x \in X$，$1x = x$，

那么称 X 是数域 $\mathbb{F}$ 上的一个**向量空间** (linear space)。 ◊

例 A.1 (矩阵空间)： 所有 $m \times n$ 的矩阵在普通矩阵加法和矩阵数乘运算下构成一个向量空间 $\mathbb{C}^{m\times n}$。如果定义内积如下：

$$\langle A, B\rangle = \operatorname{tr}(B^H Q A) = \sum_{i=1}^{n} b_i^H Q a_i \tag{A-1}$$

其中 a_i 和 b_i 分别是 A 和 B 的第 i 列，而 Q 是 Hermite 正定矩阵，那么 $\mathbb{C}^{m\times n}$ 构成一个 Hilbert 空间。 ⧫

定理 A.1 (Riesz 表示定理)： 设 H 是 Hilbert 空间，H^* 是 H 的对偶空间，那么对 $\forall\, f \in H^*$，存在唯一的 $x_f \in H$，使得

$$f(x) = \langle x, x_f\rangle, \qquad \forall\, x \in H \tag{A-2}$$

并且满足 $\|f\| = \|x_f\|$。 □

证明： 先证存在性，再证唯一性，最后正 $\|f\| = \|x_f\|$。 ■

附录 B 算法与代码

对于数学、计算机和电子信息专业，算法和代码也是经常用到的排版技巧。

B.1 算法

算法描述使用 algorithm2e 宏包，效果如算法 B-1 所示。

Input: $\mathbf{x}(k), \quad \mu, \quad \mathbf{w}(0)$
Output: $y(k), \quad \varepsilon(k)$
for $k = 0, 1, \cdots$ **do**
  $y(k) = \mathbf{w}^H(k)\mathbf{x}(k)$ // output signal
  $\varepsilon(k) = d(k) - y(k)$ // error signal
  $\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\varepsilon^*(k)\mathbf{x}(k)$ // weight vector update
end

算法 B-1 LMS 算法详细描述

B.2 代码

源代码使用 listings 宏包，LMS 算法的 Verilog 模块端口声明如代码 B-1 所示。

代码 B-1 空时 LMS 算法 Verilog 模块端口声明

```
module stap_lms
#(
parameter       M               = 4,    // number of antennas
                L               = 5,    // length of FIR filter
                W_IN            = 18,   // wordlength of input data
                W_OUT           = 18,   // wordlength of output data
                W_COEF          = 20    // wordlength of weights
)(
output  signed  [W_OUT-1:0]     y_i,    // in-phase component of STAP output
output  signed  [W_OUT-1:0]     y_q,    // quadrature component of STAP output
output                          vout,   // data valid flag of output (high)
input           [M*W_IN-1:0]    u_i,    // in-phase component of M antennas
input           [M*W_IN-1:0]    u_q,    // quadrature component of M antennas
input                           vin,    // data valid flag for input (high)
input                           clk,    // clock signal
input                           rst     // reset signal (high)
);
```

攻读学位期间取得的研究成果

研究成果包括以下内容：

1. 已发表或已录用的学术论文、已出版的专著/译著、已获授权的专利按参考文献格式列出。
2. 科研获奖，列出格式为：获奖人 (排名情况). 项目名称. 奖项名称及等级, 发奖机构, 获奖时间.
3. 与学位论文相关的其它成果参照参考文献格式列出。
4. 全部研究成果连续编号编排。

样例：

[1] Wei ZY, Tang YP, Zhao WH, et al. Rapid development technique for drip irrigation emitters[J]. RP Journal,UK., 2003, 9(2):104 110 (SCI: 000350930600051; EI: 03187452127).

[2] 魏正英, 唐一平, 卢秉恒. 滴灌管内嵌管状滴头的快速制造方法研究 [J]. 农业工程学报, 2001,17(2):55 58 (EI:01226526279,01416684777).

[3] 姜锡洲. 一种温热外敷药制备方案: 中国,88105607.3[P].1989-07-26.

[4]

用于双盲评审的论文，只列出已发表的学术论文的篇名、发表刊物名称，必须隐去各类论文检索号、期号、卷号、页码；专利号；日期等。

答辩委员会会议决议

论文提出了××××××××××××××××××××××××××××××××××××××。

论文取得的主要创新性成果包括：

1. ××。
2. ××。
3. ××。

论文工作表明作者在××××× 具有 ××××× 知识，具有 ××××× 能力，论文 ××××××××××，答辩 ××××××××××××××。

答辩委员会表决，（× 票/一致）同意通过论文答辩，并建议授予 ××× （姓名）××× （门类）学博士/硕士学位。

1. 填写内容应与学位（毕业）审批材料中答辩委员会决议书一致。
2. 无需签名。
3. 盲审论文仅保留“答辩委员会会议决议”标题。

常规评阅人名单

本学位论文共接受 X 位专家评阅，其中常规评阅人 X 名，名单如下：

王 XX	教授	西安交通大学
李 XX	教授	XXXX 大学
田 XX	教授	XXXX 大学

盲审论文仅保留“常规评阅人名单”标题。

学位论文独创性声明 (1)

本人声明：所呈交的学位论文系在导师指导下本人独立完成的研究成果。文中依法引用他人的成果，均已做出明确标注或得到许可。论文内容未包含法律意义上已属于他人的任何形式的研究成果，也不包含本人已用于其他学位申请的论文或成果。

本人如违反上述声明，愿意承担以下责任和后果：

1. 交回学校授予的学位证书；
2. 学校可在相关媒体上对作者本人的行为进行通报；
3. 本人按照学校规定的方式，对因不当取得学位给学校造成的名誉损害，进行公开道歉；
4. 本人负责因论文成果不实产生的法律纠纷。

论文作者 (签名)：　　日期：　　年　　月　　日

学位论文独创性声明 (2)

本人声明：研究生_____ 所提交的本篇学位论文已经本人审阅，确系在本人指导下由该生独立完成的研究成果。

本人如违反上述声明，愿意承担以下责任和后果：

1. 学校可在相关媒体上对本人的失察行为进行通报；
2. 本人按照学校规定的方式，对因失察给学校造成的名誉损害，进行公开道歉；
3. 本人接受学校按照有关规定做出的任何处理。

指导教师 (签名)：　　日期：　　年　　月　　日

学位论文知识产权权属声明

我们声明，我们提交的学位论文及相关的职务作品，知识产权归属学校。学校享有以任何方式发表、复制、公开阅览、借阅以及申请专利等权利。学位论文作者离校后，或学位论文导师因故离校后，发表或使用学位论文或与该论文直接相关的学术论文或成果时，署名单位仍然为西安交通大学。

论文作者 (签名)：　　日期：　　年　　月　　日

指导教师 (签名)：　　日期：　　年　　月　　日