上海交通大学硕士学位论文

上海交通大学学位论文 LaTeX 模板示例文档

硕 士 研 究 生：某　某

学　　号：0010900990

导　　师：某某教授

申 请 学 位：工学硕士

学　　科：某某专业

所 在 单 位：某某系

答 辩 日 期：2022 年 5 月 9 日

授予学位单位：上海交通大学

Dissertation Submitted to Shanghai Jiao Tong University
for the Degree of Master

A SAMPLE DOCUMENT FOR LaTeX-BASED SJTU THESIS TEMPLATE

Candidate:	Mo Mo
Student ID:	0010900990
Supervisor:	Prof. Mou Mou
Academic Degree Applied for:	Master of Engineering
Speciality:	A Very Important Major
Affiliation:	Depart of XXX
Date of Defence:	May 9, 2022
Degree-Conferring-Institution:	Shanghai Jiao Tong University

上海交通大学

学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：

日期：　　年　　月　　日

上海交通大学

学位论文使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。

本学位论文属于 □**公开论文**

□**内部论文**，□ 1 年/□ 2 年/□ 3 年　解密后适用本授权书。

□**秘密论文**，____ 年（不超过 10 年）解密后适用本授权书。

□**机密论文**，____ 年（不超过 20 年）解密后适用本授权书。

（请在以上方框内打“✓”）

学位论文作者签名：　　指导教师签名：

日期：　　年　　月　　日　　日期：　　年　　月　　日

上海交通大学学位论文 LaTeX 模板示例文档

摘　要

中文摘要应该将学位论文的内容要点简短明了地表达出来，应该包含论文中的基本信息，体现科研工作的核心思想。摘要内容应涉及本项科研工作的目的和意义、研究方法、研究成果、结论及意义。注意突出学位论文中具有创新性的成果和新见解的部分。摘要中不宜使用公式、化学结构式、图表和非公知公用的符号和术语，不标注引用文献编号。硕士学位论文中文摘要字数为 500 字左右，博士学位论文中文摘要字数为 800 字左右。英文摘要内容应与中文摘要内容一致。

摘要页的下方注明本文的关键词（4～6 个）。

关键词：上海交大，饮水思源，爱国荣校

A SAMPLE DOCUMENT FOR LaTeX-BASED SJTU THESIS TEMPLATE

ABSTRACT

Shanghai Jiao Tong University (SJTU) is a key university in China. SJTU was founded in 1896. It is one of the oldest universities in China. The University has nurtured large numbers of outstanding figures include JIANG Zemin, DING Guangen, QIAN Xuesen, Wu Wenjun, WANG An, etc.

SJTU has beautiful campuses, Bao Zhaolong Library, Various laboratories. It has been actively involved in international academic exchange programs. It is the center of CERNet in east China region, through computer networks, SJTU has faster and closer connection with the world.

KEY WORDS: SJTU, master thesis, XeTeX/LaTeX template

目　录

插图索引

表格索引

算法索引

符号对照表

ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数

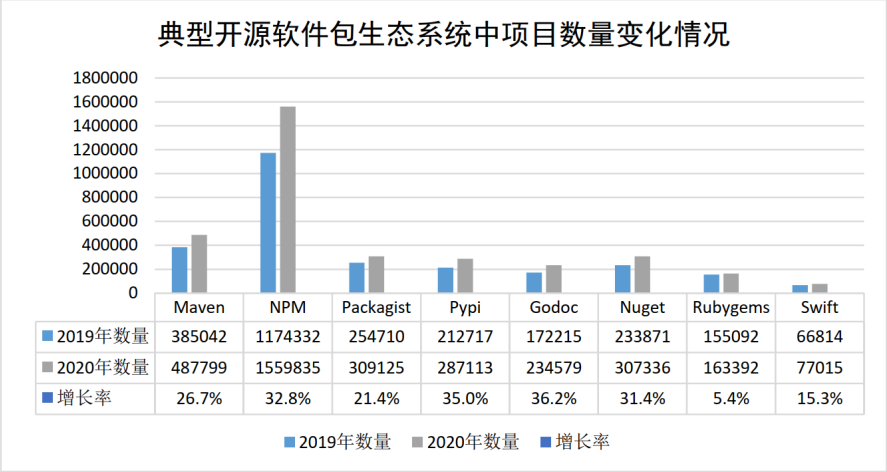
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率
ϵ 介电常数
μ 磁导率

第一章　绪论

1.1　软件供应链

随着容器、微服务等新技术日新月异，开源软件成为业界主流形态，软件行业快速发展。现代软件大多数是被“组装”出来的，不是被“开发”出来的。据 Forrester 统计，软件开发中，80-90% 的代码来自于开源软件。因此，现代软件的源代码绝大多数是混源代码，由企业自主开发的源代码和开源软件代码共同组成。

根据奇安信代码安全实验室的检测与统计[1]，八个典型的开源软件包生态系统发展迅猛，呈现繁荣态势，包括 Maven、NPM、Packagist、Pypi、Godoc、Nuget、Rubygems、Swift。2019 年与 2020 年各开源软件包生态系统增长情况如图1–1：



	Maven	NPM	Packagist	Pypi	Godoc	Nuget	Rubygems	Swift
2019年数量	385042	1174332	254710	212717	172215	233871	155092	66814
2020年数量	487799	1559835	309125	287113	234579	307336	163392	77015
增长率	26.7%	32.8%	21.4%	35.0%	36.2%	31.4%	5.4%	15.3%

图 1–1　2019 和 2020 年八个典型开源软件包生态系统的增长情况

1.2　软件供应链安全

软件供应链的上游软件可能悄无声息地影响着下游产品。开源软件之间的依赖关系错综复杂，在开发过程中，开发者通常借助包管理程序实现自动管理，因此可能意识不到产品中包含里数量巨大的开源软件。一旦某个上游的开源软件被发现安全漏洞，软件开发者无法立即意识到漏洞同时被引入到了产品中，隐含里巨大的软件供应链安全风险。

2020 年 5 月，GitHub 披露了 Octopus Scanner 漏洞[2]，该漏洞是针对 Apache NetBeans IDE 项目的开源软件供应链攻击，影响到了 26 个开源项目。

2020 年 12 月，安全公司 FireEye 发现全球著名的网络安全管理软件供应商 SolarWinds 遭遇国家级 APT 团伙高度复杂的供应链攻击。该攻击在 SolarWinds 的一个数字签名组件 DLL 中插入后门，该后门通过 HTTP 协议与第三方服务器通信。

1.3 安卓软件的供应链安全

Appbrain[3]追踪了 450 个流行的库，统计结果显示它们在安卓生态系统中有着广泛的使用，广告库、社交网络库、以及手机设备分析库尤为受欢迎。如此广泛的第三方库使用在加速开发过程、避免重复造轮子的同时，也吸引着攻击者将目标向软件供应链上游移动，通过利用受欢迎的库的漏洞来达到攻击应用的目的。atvhunter 2-4。来自 Trend Micro 的安全研究团队披露百度提供的 SDK 中的 Moplus 包含的功能可能被恶意使用，以向用户设备植入后门[4]。这一处于软件供应链上游的漏洞已经流入超过 14000 款安卓 APP，可能使得约 1 亿用户处于黑客的攻击风险中。

2022 年 4 月 Google Play 商店内的安卓应用超过 260 万，3 月与 4 月新增应用数量均在 2 万左右，来自其他市场的应用更是不计其数。

如此数量的 APP 包含着不可忽视的供应链风险，但是由于 APP 包含着敏感信息或者具有商业价值的运行逻辑，大部分开发者基于安全和产权的考虑都会将产品进行混淆后再发布。这导致在对 APP 进行安全性检查时更加困难，识别混淆 APP 中引入的上游软件成为了亟待解决的问题。事实上，约 78% 的漏洞都是在间接的依赖中找到，可能带来的安全风险则更加难以发现[1]。

第二章　研究现状

2.1　检测混淆库

随着 APP 混淆技术的成熟，以第三方库能够被容易地区分为前提的方法已不适用，标识符被混淆为无意义的简短的字母组合，比如 *com.google* 可能被混淆为 *a.c*，无法提供关于库的任何信息。图2–1为一个代码混淆的示例，仅从名称无法获得任何关于包的信息。

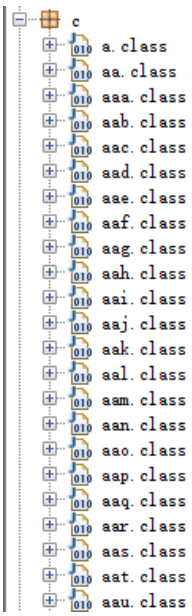


图 2–1　一款 360 软件的 apk 解压后得到的经过重命名混淆的 class 文件

T. Book 等人的工作通过白名单的方法检测 APP 内的第三方库[5]，但这类方法显然无法解决标识符重命名的问题。PEDAL[6]借助机器学习方法，从 SDK 中提取代码特征，并使用包之间的关系信息训练了分类器来识别第三方库。

2.2　检测未知库

一些研究工作提出了在没有已知第三方代码的数据库知识情况下检测 APP 组成成分的方法。此类方法通常首先把开发者代码与第三方代码进行分类，再将第三方代码聚类成不同的组件，组件即一个可能的库的候选，进一步评估候选之间的相似度，当超过相似度阈值的候选的出现次数足够多时就认为找到了一个库。

如 Chen 等人[7]从大量的 APP 中获取库，进行聚类和检测然而这一方法在混淆的情况下表现不佳，因为其假设不同 APP 中包含的库的相同实例拥有相同的包名，混淆打破了这一基本的假设。

为解决包名混淆问题，LibRadar[8]使用特征哈希的方法，不需要基于包名的聚类，而是借助包中的目录结构来识别库的候选，具体来说是将一个候选表示为一个目录树的结构。这引入一个新的假设，即包的结构在混淆过程中不改变。但混淆工具可以将不同的包合并为一个包，很容易打破这一假设。

WuKong[9]和 AnDarwin[10]用控制流图和 API 数量来定义哈希特征，用来计算各候选库的相似度。考虑到混淆工具可能修改一个方法的控制流图，或者移除在 APP 运行中未真正使用的方法，哈希的质量影响着这两类方法的表现性能。

2.3　检测已知库

基于已知库的检测要求关于现存库的知识，如库的基本信息、哈希特征等，在混淆 APP 第三方库识别的场景下，用构建知识数据库的代价换取了更好的表现。

具有代表性的一个工具是 LibScout[11]，用包的结构以及类的哈希作为特征，进行 APP 与数据库中第三方库的匹配，在控制流篡改和包/类/描述符重命名情况下依然有效。但是随着数据库中的标准库代码特征增多，哈希特征的计算也应当考虑更多信息，导致特征生成时间与匹配时间增加。

2.4　检测标准库的版本

现有工作中以版本为目标实现精确检测的并不多，AdDetect[12]仅能够区分广告和非广告的库，基于聚类的方法如 LibRadar[8]，LibD[13]等都没有声明能够检测库的特定版本。

实现版本的检测仍面临着很多问题：

1. 需要处理庞大的数据集。第三方库本身就纷繁复杂，如果再将各个版本考虑进去，将导致需要处理的数据成倍增长。
2. 缺乏精确的表示。一个库的不同版本可能差异微小，如何找到合适的特征来区分这一差别非常关键。
3. 代码混淆的干扰。代码混淆同样会导致库的代码发生改变，这种改变是由不同库引起还是由同一库的不同版本引起，需要被准确的区分。

第三章　研究方法

在参考了多篇文献后，我提出了一种适用于包的结构混淆、包/类/标识符重命名场景的，基于已知标准库的数据库，利用两类信息生成粗粒度/细粒度两级哈希特征的安卓应用第三方库及其特定版本的检测方法。

3.1　方法概述

此方法不依赖于包中的目录结构以及各级名称，因此可以抵抗结构混淆以及重命名混淆，包括了四个步骤：

1. 预处理 jar、aar 和 apk。将来自 Maven 仓库的 jar 包、aar 包以及待检测 apk 处理成便于构建树结构的形式。
2. 构建特征树。根据上一阶段输出，将每个包作为根节点构建特征树，该包内的所有类，不论是根包的类还是子包的类，一律作为树的中间层节点，各类的方法作为叶子节点。特征分为粗粒度、细粒度两级特征。粗粒度特征为方法的描述符的返回值以及参数类型，细粒度特征为该方法的字节码，首先生成叶子节点的两级特征，再利用叶子节点生成中间层节点即类节点的特征。
3. 构建数据库与匹配。根据以上特征生成方法，计算 Maven 仓库中的标准库的特征，并存储到数据库中。对待检测 APP，首先生成粗粒度特征，确定所包含的库，再根据细粒度特征，确定各库的具体版本。

3.2　包的预处理

3.2.1　Dex 与 Class 文件的处理

3.2.1.1　Dex 与 Class 简介

DEX 文件：DEX 文件时 Android 系统中的一种文件，是一种特殊的数据格式，能够被 Dalvik 虚拟机识别并加载执行，类似于 Windows 上的 EXE 可执行文件。将 APK 安装包解压后得到的文件就包含了 DEX 文件，它记载了应用程序的全部操作指令以及运行时数据。当 java 程序编译成 class 文件后，还需要使用 dx 工具将所有的 class 文件整合到一个 DEX 文件里，目的是其中各个类能够共享数据，在一定程度上降低了冗余，同时也使文件结构更加紧凑。DEX 文件大小通常是传统 jar 包的 50% 左右。

CLASS 文件：class 文件是能够被 java 虚拟机识别，加载并执行的文件格式，通过 javac 程序可以从 java 源文件生成 class 文件。class 文件记录了一个类文件的所有信息，不仅包含了 java 源代码中的信息，还包括了 this、super 等关键字的信息。作为一种 8 位字节的二进制六文件，class 中的数据按顺序紧密排列，没有间隙，从而让 JVM 加载更加迅速，每一个类、接口或者枚举都单独占据一个 class 文件。

参考文献

[1] 2021 中国软件供应链安全分析报告[Z]. https://www.qianxin.com/news/detail?news_id=1108. Accessed May 9, 2022.

[2] LAB G S. The Octopus Scanner Malware: Attacking the open source supply chain [Z]. https://securitylab.github.com/research/octopus-scanner-malware-open-source-supply-chain/. Accessed May 9, 2022.

[3] AppBrain. Android library statistics[Z]. https://www.appbrain.com/stats/libraries. Accessed May 9, 2022.

[4] Thehackernews.com. Backdoor in Baidu Android SDK Puts 100 Million Devices at Risk[Z]. https://thehackernews.com/2015/11/android-malware-backdoor.html. Accessed May 9, 2022.

[5] BOOK T, PRIDGEN A, WALLACH D S. Longitudinal analysis of android ad library permissions[J]. arXiv preprint arXiv:1303.0857, 2013.

[6] LIU B, LIU B, JIN H, et al. Efficient privilege de-escalation for ad libraries in mobile apps[C]//Proceedings of the 13th annual international conference on mobile systems, applications, and services. 2015: 89-103.

[7] CHEN K, WANG X, CHEN Y, et al. Following devil's footprints: Cross-platform analysis of potentially harmful libraries on android and ios[C]//2016 IEEE Symposium on Security and Privacy (SP). 2016: 357-376.

[8] MA Z, WANG H, GUO Y, et al. Libradar: fast and accurate detection of third-party libraries in android apps[C]//Proceedings of the 38th international conference on software engineering companion. 2016: 653-656.

[9] WANG H, GUO Y, MA Z, et al. Wukong: A scalable and accurate two-phase approach to android app clone detection[C]//Proceedings of the 2015 International Symposium on Software Testing and Analysis. 2015: 71-82.

[10] CRUSSELL J, GIBLER C, CHEN H. Andarwin: Scalable detection of android application clones based on semantics[J]. IEEE Transactions on Mobile Computing, 2014, 14(10): 2007-2019.

[11] BACKES M, BUGIEL S, DERR E. Reliable third-party library detection in android and its security applications[C]//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. 2016: 356-367.

[12] NARAYANAN A, CHEN L, CHAN C K. Addetect: Automated detection of android ad libraries using semantic analysis[C]//2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP). 2014: 1-6.

[13] LI M, WANG W, WANG P, et al. Libd: Scalable and precise third-party library detection in android markets[C]//2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE). 2017: 335-346.

附录 A Maxwell Equations

选择二维情况，有如下的偏振矢量：

$$\mathbf{E} = E_z(r,\theta)\hat{\mathbf{z}}, \tag{A–1a}$$

$$\mathbf{H} = H_r(r,\theta)\hat{\mathbf{r}} + H_\theta(r,\theta)\hat{\boldsymbol{\theta}}. \tag{A–1b}$$

对上式求旋度：

$$\nabla\times\mathbf{E} = \frac{1}{r}\frac{\partial E_z}{\partial\theta}\hat{\mathbf{r}} - \frac{\partial E_z}{\partial r}\hat{\boldsymbol{\theta}}, \tag{A–2a}$$

$$\nabla\times\mathbf{H} = \left[\frac{1}{r}\frac{\partial}{\partial r}(rH_\theta) - \frac{1}{r}\frac{\partial H_r}{\partial\theta}\right]\hat{\mathbf{z}}. \tag{A–2b}$$

因为在柱坐标系下，$\overline{\overline{\mu}}$ 是对角的，所以 Maxwell 方程组中电场 $\mathbf{E}$ 的旋度：

$$\nabla\times\mathbf{E} = \mathrm{i}\omega\mathbf{B}, \tag{A–3a}$$

$$\frac{1}{r}\frac{\partial E_z}{\partial\theta}\hat{\mathbf{r}} - \frac{\partial E_z}{\partial r}\hat{\boldsymbol{\theta}} = \mathrm{i}\omega\mu_r H_r\hat{\mathbf{r}} + \mathrm{i}\omega\mu_\theta H_\theta\hat{\boldsymbol{\theta}}. \tag{A–3b}$$

所以 $\mathbf{H}$ 的各个分量可以写为：

$$H_r = \frac{1}{\mathrm{i}\omega\mu_r}\frac{1}{r}\frac{\partial E_z}{\partial\theta}, \tag{A–4a}$$

$$H_\theta = -\frac{1}{\mathrm{i}\omega\mu_\theta}\frac{\partial E_z}{\partial r}. \tag{A–4b}$$

同样地，在柱坐标系下，$\overline{\overline{\epsilon}}$ 是对角的，所以 Maxwell 方程组中磁场 $\mathbf{H}$ 的旋度：

$$\nabla\times\mathbf{H} = -\mathrm{i}\omega\mathbf{D}, \tag{A–5a}$$

$$\left[\frac{1}{r}\frac{\partial}{\partial r}(rH_\theta) - \frac{1}{r}\frac{\partial H_r}{\partial\theta}\right]\hat{\mathbf{z}} = -\mathrm{i}\omega\overline{\overline{\epsilon}}\mathbf{E} = -\mathrm{i}\omega\epsilon_z E_z\hat{\mathbf{z}}, \tag{A–5b}$$

$$\frac{1}{r}\frac{\partial}{\partial r}(rH_\theta) - \frac{1}{r}\frac{\partial H_r}{\partial\theta} = -\mathrm{i}\omega\epsilon_z E_z. \tag{A–5c}$$

由此我们可以得到关于 E_z 的波函数方程：

$$\frac{1}{\mu_\theta\epsilon_z}\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial E_z}{\partial r}\right) + \frac{1}{\mu_r\epsilon_z}\frac{1}{r^2}\frac{\partial^2 E_z}{\partial\theta^2} + \omega^2 E_z = 0. \tag{A–6}$$

附录 B　绘制流程图

图 B–1 是一张流程图示意。使用 tikz 环境，搭配四种预定义节点（startstop、process、decision和io），可以容易地绘制出流程图。

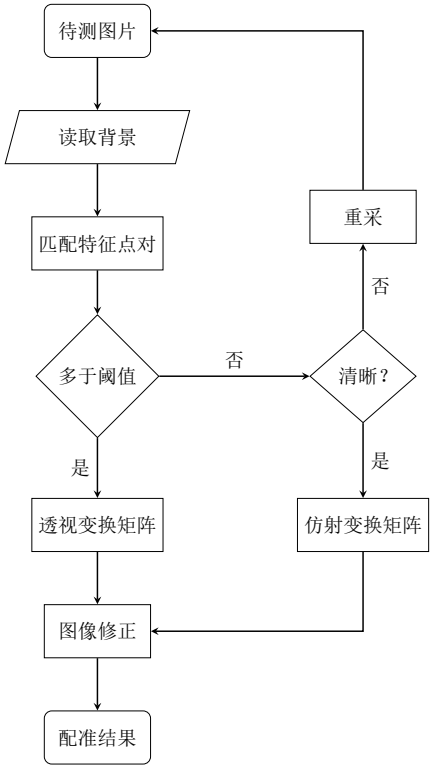


图 B–1 绘制流程图效果

Figure B–1 Flow chart

致　谢

感谢那位最先制作出博士学位论文 LaTeX 模板的交大物理系同学！
感谢 William Wang 同学对模板移植做出的巨大贡献！
感谢 @weijianwen 学长一直以来的开发和维护工作！
感谢 @sjtug 以及 @dyweb 对 0.9.5 之后版本的开发和维护工作！
感谢所有为模板贡献过代码的同学们，以及所有测试和使用模板的各位同学！
感谢 LaTeX 和 SJTUThesis，帮我节省了不少时间。

学术论文和科研成果目录

学术论文

[1] Chen H, Chan C T. Acoustic cloaking in three dimensions using acoustic metamaterials[J]. Applied Physics Letters, 2007, 91:183518.

[2] Chen H, Wu B I, Zhang B, et al. Electromagnetic Wave Interactions with a Metamaterial Cloak[J]. Physical Review Letters, 2007, 99(6):63903.

专利

[3] 第一发明人，“永动机”，专利申请号 202510149890.0

个人简历

基本情况

某某，yyyy 年 mm 月生于 xxxx。

教育背景

- yyyy 年 mm 月至今，上海交通大学，博士研究生，xx 专业
- yyyy 年 mm 月至 yyyy 年 mm 月，上海交通大学，硕士研究生，xx 专业
- yyyy 年 mm 月至 yyyy 年 mm 月，上海交通大学，本科，xx 专业

研究兴趣

LaTeX 排版

联系方式

- 地址：上海市闵行区东川路 800 号，200240
- E-mail: xxx@sjtu.edu.cn