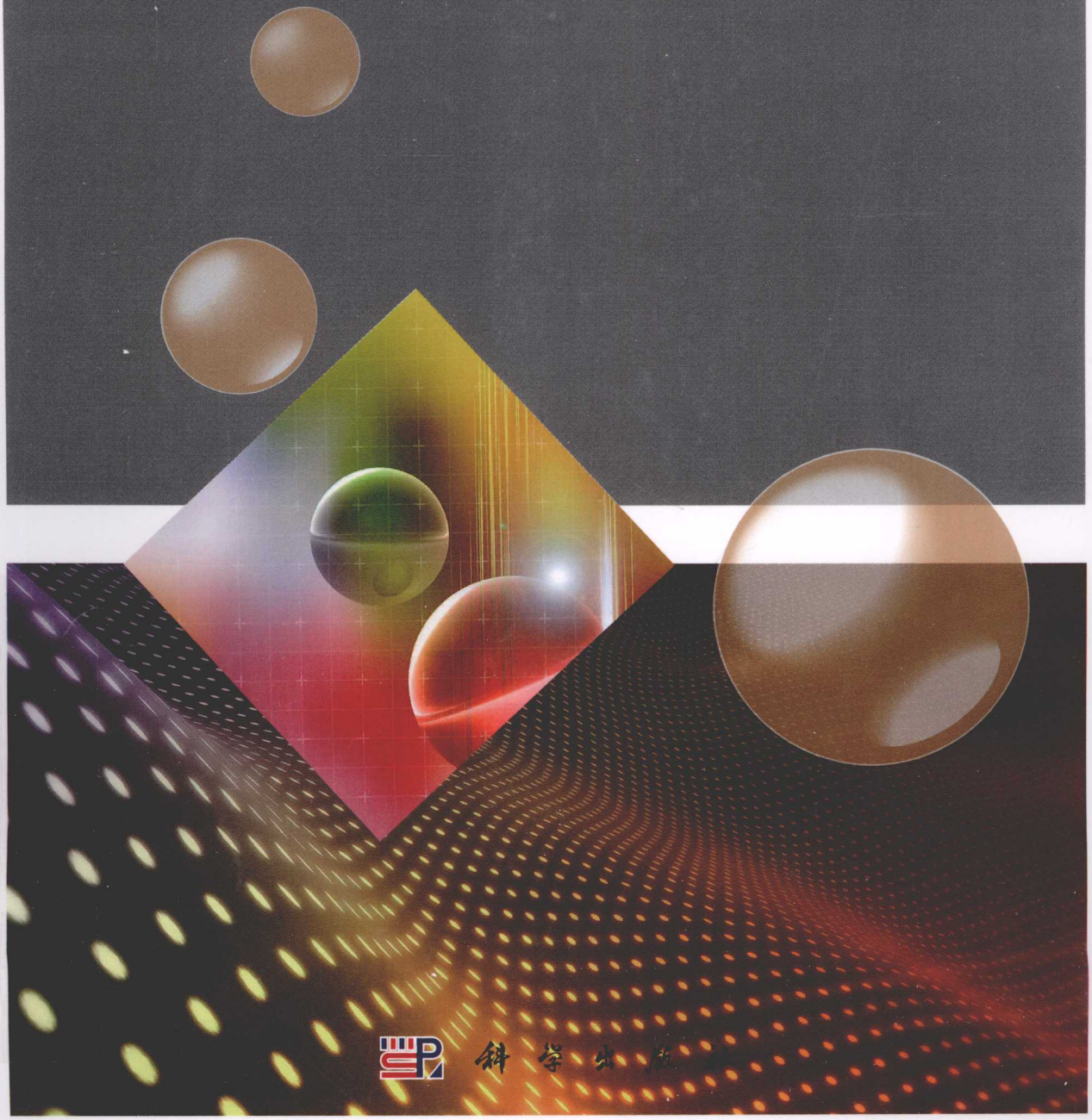
基于水印和特征的软件保护技术研究

罗养霞　王命宇　郭晔　著



科学出版

基于水印和特征的软件保护技术研究

罗养霞　王命宇　郭　晔　著

科学出版社
北京

内 容 简 介

软件水印和软件特征技术的研究涉及密码学、信息学、模式识别等多种学科。该领域的研究有着非常重要的实用价值和广泛的应用领域，在软件识别、版权保护、真伪鉴别、信道保密、篡改提示、恶意代码检测、混淆评估、程序理解和维护等方面有着举足轻重的意义。本书详尽地给出了软件水印和软件特征的各种应用算法及实例，理论基础全面，参考性和可操作性强。

本书可以作为信息安全、软件安全、电子商务安全等领域的技术人员和管理人员的参考书，还可以作为计算机科学与技术专业、软件工程专业的高年级本科生和研究生的入门教材或参考书。

图书在版编目(CIP)数据

基于水印和特征的软件保护技术研究/罗养霞，王命宇，郭晔著. —北京：科学出版社，2015

ISBN 978-7-03-045593-2

Ⅰ. ①基… Ⅱ. ①罗…②王…③郭… Ⅲ. ①电子计算机-密码术-研究 Ⅳ. ①TP309.7

中国版本图书馆 CIP 数据核字（2015）第 209873 号

责任编辑：王杰琼 冯 涛 / 责任校对：王万红
责任印制：吕春珉 / 封面设计：耕者设计工作室

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

北京京华虎彩印刷有限公司印刷
科学出版社发行 各地新华书店经销
*
2015 年 9 月第 一 版 开本：787×1092 1/16
2015 年 9 月第一次印刷 印张：12 1/2
字数：280 000

定价：70.00 元

（如有印装质量问题，我社负责调换〈京华虎彩〉）

销售部电话 010-62136131 编辑部电话 010-62135235（VP04）

前　言

随着网络服务的不断普及，对软件的侵权、盗版、随意篡改和恶意攻击等问题也日趋增多，严重阻碍了软件产业的健康和可持续发展。软件的保护技术受到各国政府、企业家和学者的重视，由此而引发的版权管理、防盗版、软件识别等软件保护技术也成了关键和热点研究问题。

软件下载和在线购买已成为方便管理者统一管理和销售、方便用户下载、升级和使用的主流发展趋势，不过也致使软件的拷贝、编辑、传播或篡改日益增多且逃避限制，这使得软件开发者、软件运营商以及消费者的合法权益受到侵害。对软件的保护有很多，有基于硬件为载体的保护方法，有基于软件为载体的保护方法，也有基于硬件和软件相绑定的方法。软件水印和软件特征识别是两种基于软件为载体的保护方法，前者主要是给软件中嵌入水印信息达到识别的目的；后者主要是提取软件中的不变特征（或关键特征），以达到对该软件或该软件家族的识别。这两种技术都不是用来阻止软件的传播，而是起到跟踪、鉴别的目的，同时这种技术和加密、DRM 等技术相结合，可达到对软件的更好保护。

目前已提出许多软件水印和软件胎记特征算法，但从安全性、鲁棒性等方面还不能满足应用的需求。基于水印和特征的保护体系需要适应网络环境发展和多用户多权限的新需求；软件水印仍面临减裁、扭曲、添加和共谋等攻击；特征检测的准确率不高，不能准确代表和检测软件等。因此，本书展开了基于水印和特征的软件保护技术研究，主要研究内容有基于水印和特征的软件版权协议研究，结合图拓扑、混沌加密等技术的水印算法研究，基于数据流的软件特征分析，基于聚类理论的特征处理和基于提升多属性特征，集成学习的特征识别研究等，为建立安全的软件保护系统提供理论分析和技术支撑。

本书分为十六章：第一章为绪论部分，介绍水印及软件特征相关的研究背景、意义及需要解决的问题；从第二章到第九章，介绍相关软件水印算法和应用；从第十章到第十五章，介绍与软件胎记特征有关的算法和应用研究；第十六章进行总结和展望。

本书作者长期从事水印和特征分析研究，主持和参与国家级、省厅级多个项目成果的总结，同时书稿得到王命宇、郭晔的补充、修改和完善。

本书的出版得到国家自然科学基金项目（No.61170218; No.61272461）、陕西省自然科学基金项目（No.2014JM2-6100）、陕西省教育厅自然科学研究项目（No.15JK1274）等研究的资助，是多年来在水印和特征方面研究成果的凝聚。

本书的出版得到西安财经学院 2014 年学术著作基金资助，得到西安财经学院信息学院的支持和帮助，得到西北大学信息学院安全实验室老师和朋友的大力帮助，本书的出版更是得到科学出版社各位领导和编辑的支持，为该书的成功出版付出了辛苦劳动，特此一并致谢！

由于作者水平有限，书中难免出现疏漏和不当之处，恳请读者批评指正。

罗养霞

2015 年 2 月

目　录

第一章 绪 论

本章首先论述了本书撰写的背景和目的，然后介绍软件的安全及保护技术、水印及特征的应用、主要攻击方法和需要解决的问题，进一步阐述了本书研究的意义。最后介绍本书的主要研究内容。

1.1 研究背景及目的

随着网络服务的不断普及，数据通信，如 WWW、FTP、P2P 对等网络，BT 多通道高速上传下载服务，电子邮件，网络存储，移动存储等，促进了数字内容的共享和交换，数字产品的存储、复制、获得和传播变得越来越容易。与此同时，软件作为数字产品的一种，其复制率高，商业获益大，对软件的侵权、盗版、随意篡改和恶意攻击等问题也日趋严重，严重阻碍了软件产业的健康和可持续发展。2012 年，商业软体联盟（Business Software Alliance，BSA）在全球软件盗版研究报告表明全球软件平均盗版率为 42%，造成的经济损失达 330 亿美元。在软件市场发展较快的国家，即使相对较低的软件盗版率也会带来惊人损失。例如，全球最低的软件盗版率是美国，为 19%，但依然是损失最大的国家，损失达 70 亿美元。我国的软件盗版率在 2012 年比上年有所下降，但盗版率也高达 77%，所造成的损失达 37 亿多美元，是全球排名第二大损失国。在 2012 年另一份 BSA 报告指出[1]，如果中国软件盗版率每降低 10 个百分点，可有利于 IT 产业的利润翻两番，可带动 IT 产业产值从目前现有的 270 亿美元增至 2013 年的 890 亿美元，政府的税收收入也会因此增加 68 亿美元，可用于必要的环境保护或其他公共福利。正因为如此，软件的保护技术受到各国政府、企业家和研究学者的重视，由此而引发的软件保护、软件识别技术、防盗版等问题成为了热点研究问题。

目前，软件保护技术有基于硬件为载体的保护方法，如加密狗[2]、FLEXlm[3]和 SmartCard[4]等；有基于软件的保护方法，如水印[5-7]、特征[6, 8, 9]和加密技术[10]等；也有基于硬件和软件相绑定的方法。基于硬件的保护技术是将额外的专用硬件设备与权限许可证绑定，迫使软件只能在带有该硬件设备的机器上正常运行，可有效地防止软件被非法扩散，但这容易产生硬件适应性问题，增加软件运行开销，不容易实现软件运行环境的迁移，还有可能与其他软件的正常使用发生冲突而带来不便。软件水印属于数字水印的一种，用来携带版权发行者、使用者、开发者和销售者的信息，在软件版权管理中，可用于身份认证、权限约束，防止版权盗用和软件的非法复制等。由于软件水印技术比加密狗等其他技术更具有隐蔽性，比加密更多一层保证，能够提供法律证明，所以基于水印的版权保护研究最近几年日益增多。然而，其主要应用在文本和多媒体方面的版权保护处理，在软件水印和数据库水印方面的研究相对较少，所以这方面研究也面临着许多挑战[11]。软件胎记特征技术[12]（又叫零水印[13]或零知识[14]）是提取软件中的不变特征（或关键特征），以达到对该软件或该软件家族的识别。若两个软件的胎记特征之间

具有较大的相似度，则其中一个软件很可能是另外一个软件的盗版，或者与另外一个软件属于同一家族。基于胎记特征的检验和识别因不会导致软件性能的退化而受欢迎，但目前软件特征的研究才处于一个开始阶段。

近年来，软件下载和在线购买已成为方便管理者统一管理和销售、方便用户下载、升级和使用的主流发展趋势，不过也致使软件的拷贝、编辑、传播或篡改日益增多且逃避限制，这使得软件开发者、软件运营商以及消费者的合法权益受到侵害。使软件版权保护、软件检测和识别等成为迫切需要解决的问题。根据数字版权管理（Digital Rights Management，DRM）的功能将其分为用于控制非法复制和传播的数字版权管理（Copying Proof Based DRM，CP-DRM）和用于版权归属检测的数字版权管理（Rights Ownership Identification DRM，ROI-DRM）[15]。那么，研究如何提高控制非法复制和传播的有效性和版权归属检测的准确性更显得尤为重要。

基于水印和特征的软件版权保护，要求能够借助水印跟踪侵权内容信息、验证身份和权限发放，或能够证明版权归属，水印是作为一种不可见数据隐藏于原始载体中，需要既能达到保护的目的又要减少影响载体的性能；同时当软件被攻击时，水印和特征识别算法要有强的鲁棒性来抵抗各种破坏；当软件出现版权纠纷时，所有者可提取水印或软件特征信息获取对软件的识别和检验，从而保护所有者的权益。所以基于水印和特征的软件保护技术需要满足以下要求。

（1）安全性：软件版权管理在权限发放、验证、传递过程中，既要满足需求又要保证安全。

（2）鲁棒性：水印和特征识别算法本身要具有抗攻击性，不易被破坏或去除。

（3）准确率：软件检测和识别不管是基于水印还是基于软件特征，要求水印和特征都要有高的鲁棒性的同时，还要保证识别的准确率。

（4）性能和效率：尽量保证软件的性能和效率不受影响。

因此，基于以上要求，本书对基于水印和特征的软件保护中的相关技术进行研究，保证软件版权管理中基于安全的水印认证和权限发放协议，提高水印的鲁棒性和隐蔽性，增强软件特征识别的准确率，改进软件特征选择方法和识别方案。从而为建立安全的软件保护系统提供理论分析和技术支撑。

1.2　软件安全及保护技术

1.2.1　软件安全问题

软件安全（Software Security）所涉及的内容较多，问题较复杂，范围较广泛。从用户角度来说，希望软件拥有更高的可靠性，操作性强、功能多、保密性好、性价比高等特点；作为开发商，不但要考虑使用方面的安全问题，还要考虑到软件的防攻击、版权保护和系统安全。

从造成软件安全的原因来划分，可将软件安全分为三个方面的内容，如图 1-1 所示。一是软件自身安全，也是质量安全，软件在开发过程中存在缺陷和漏洞而造成的安全问题。有些软件便于编程或扩展，专门设置“后门”，还有一些专门热衷于寻找漏洞的“高

手”，这些漏洞给计算机软件的安全性带来了严重的威胁。二是意外安全，也属于不当使用安全，由于管理、人为操作不当，缺乏专业的软件防护技术，软件安全意识又不强，而产生的违规操作、人为破坏、数据损坏等安全威胁。三是攻击安全，是由于蓄意攻击和破坏而产生的安全问题，是威胁最多、损失最大，也是目前安全领域最为棘手的问题之一。

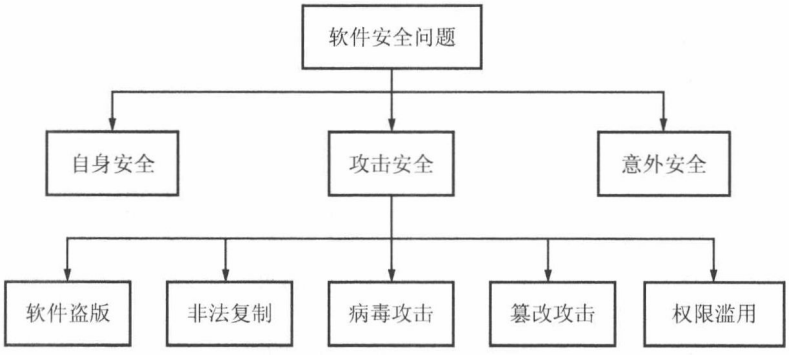


图 1-1 软件面临的安全问题

软件正在遭受多方面的攻击：全球高居不下的软件盗版问题，使不法分子从中牟取暴利，损害开发商的利益。世界各国的政府、企业和学者对软件盗版问题特别关注和重视，并采取措施来遏制盗版行为，但从巨大的盗版市场看，保护技术仍然是杯水车薪。由于计算机软件的易复制性，给软件及代码复制创造了条件，软件及代码剽窃使产品开发和知识产权遭受到严重的威胁；计算机病毒是较为常见的安全问题之一，由于其具有触发性、潜伏性、感染性、破坏性和自我复制性，日益增多的病毒种类，更为隐蔽的传播手段和途径，给计算机软件和软件环境造成危害；由于攻击者有足够多的分析工具和跟踪工具，如调试工具、仿真工具、反编译器、静态和动态分析工具等等，可对软件进行篡改，对机密数据、核心代码、版本标识等进行窃取或破坏；由于版权管理和访问控制中身份认证、版权识别和权限发放存在着安全漏洞，而使软件产生权限滥用、非法使用、数据盗用等问题。

综上所述，加强软件安全、保护软件势在必行。

1.2.2 软件保护和软件识别

为了达到对软件的保护，仅靠立法或制订标准，其收效是有限的。除了加强软件开发设计质量、应用技术和人为意识方面的安全，更重要的是采取适当的技术，从技术角度进行防范。保护技术包括 DRM 中的访问控制、防篡改技术、加密技术、软件水印和胎记特征识别等，如图 1-2 所示。具体包括对网络环境软件的流通和使用保护，加强用户访问和认证的设计，既要满足应用环境需求，又要保证软件关键部分访问路径的安全性，版权检测与识别、用户角色认证等验证手段和程序的安全性和鲁棒性；通过加密技术，对软件中敏感信息、关键代码、嵌入的水印等进行保护，加强软件体系的安全性；软件水印和胎记特征主要是基于软件识别技术，虽然不像加密，具有直接的抗攻击作用，

但由于其不可感知性和隐蔽性较好，且能提供识别依据和法律证明，被广泛研究和应用于软件的认证、真伪鉴别、篡改提示和完整性检验等技术保护中。

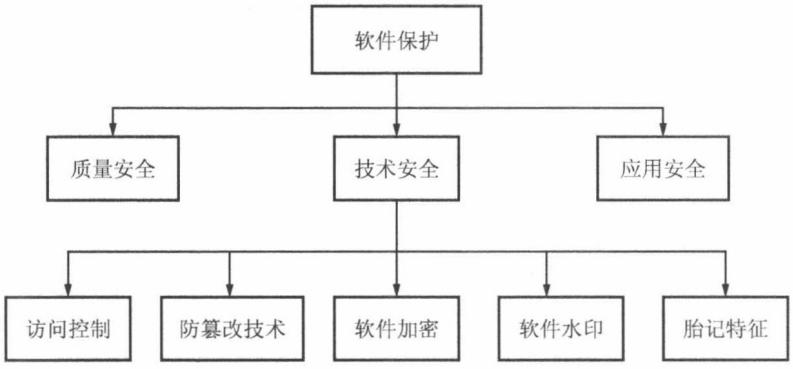


图 1-2　软件软件保护技术

1.3　水印和特征的应用领域及其意义

水印和特征的应用越来越广泛，不但应用于软件的保护和管理，如用于控制非法复制和传播的软件版权管理（CP-SRM）和用于版权归属检测的软件版权管理（ROI-SRM），也广泛应用于恶意代码识别、混淆评估及程序定位与理解分析等研究领域。该技术研究与应用具有十分重要的实际意义，具体可从以下几个方面分析。

1. 版权保护

版权保护中，可将水印作为版权标识，也可将水印作为客户指纹标识。版权标识水印作为加密的一种补充，把版权所有者的身份标识作为水印信息嵌入到软件中，当软件作品被篡改、非法复制或发生盗版等软件纠纷时，版权维护者可从软件中提取并恢复出水印信息，达到维护版权权益和控制非法活动的目的。指纹水印是把用户信息编码成水印，嵌入到软件中，相应地通过验证水印来判断用户的使用权限，跟踪软件发放过程等，也可方便实现对客户的分级控制，对于特定软件，若发现权限越级、未授权使用或无权分发，则可通过提取指纹，查找和跟踪软件根源。这些水印要求具有较好的隐蔽性、鲁棒性和安全性。利用信息隐藏技术，使水印标识不可见，要求既要达到保护的目的，又要不损害软件的可用性和性能。

2. 真伪鉴别

随着计算机及网络技术在各行各业不断普及，不同版本软件在网络环境中也不断涌现，尤其是需求庞大的商业应用软件，如组态软件、金融分析软件、游戏软件、财务软件等，这些软件显示出其丰厚的商业利润，软件的模仿迅速蔓延，基于水印对软件的真伪鉴别的需求也愈来愈迫切。由于水印的类别和用途不同，所具有的隐蔽性、安全性、脆弱性等特性，使得软件真伪鉴别变得更为方便和简单，并日益发挥不可替代的作用。

3. 信道保密

由于水印的嵌入方式是多种多样的，可以是融合通信通道的冗余数据、随机不可感知成分或是纠错码、监督码等载体，增加通信信道传输的可靠性。水印的嵌入可通过加入纠错码或利用监督码的冗余信息进行隐藏。嵌入水印的秘密载体主观上是感觉不到的，又结合通信信道的数据，网络传输也是通过加密，则可有效减少网络信道被攻击的可能，即使攻击者知道秘密水印的存在，但要通过解密来分析信道数据也具有一定困难。

4. 篡改检查

在软件载体或敏感信息保护中，用水印作为检测相应设备是否被攻击者篡改，检查嵌入的水印是否被改变或删除，水印嵌入时可以使用迭代方式加密和解密，或是以相互检验和修复方式对载体敏感信息及水印信息进行篡改检查，或对水印载体允许拷贝的次数进行检查。这类水印通常是使用透明的、脆弱性、盲水印作为水印篡改检测标识。

5. 软件识别

基于特征的软件识别研究是不嵌入信息到软件中，而是提取软件的特征信息，尤其是胎记特征（关键特征），来构造软件的识别标识。有许多研究称为胎记（Birthmark）研究，也有称为软件的“零水印”或“零知识”研究，原因是不嵌入信息到软件中，或称嵌入的信息为零。这比以上提到的水印有两个好处，一是不会有被去除的可能，二是不会影响软件的性能。通过对比两个软件本身的特性信息或特性集合来判别是否是同一版本或同类软件。所以提取的特征是否具有代表性和可信性，对软件的识别非常关键。

6. 恶意代码检测

恶意代码是指以破坏为目的的恶意程序，会造成资源消耗、信息泄露和数据的损坏等。随着恶意代码危害越来越严重，对其识别和检测技术也越来越关注。特征检测主要基于模式匹配思想[16]，从恶意代码的程序分析和提取关键特征，使其代表一个或一类恶意代码，并将该关键特征存入特征库，对未知软件或软件环境进行检测或扫描时，通过该特征的匹配来判断是否为恶意软件。随着软件的多态和变形技术的层出不穷，软件和特征变异的种类繁多，关键特征的提取和过滤是影响识别的关键环节。

7. 混淆评估

代码混淆是指对软件及应用程序进行保持语义的变换，这些变换改变了软件代码的表现形式，但在功能上，使变换后的程序和原来的程序相同或相近。代码混淆提高了软件逆向分析的免疫力，混淆技术分为四类：数据流混淆、布局混淆、控制流混淆和预防性混淆。软件特征的分析与度量研究中，也是对软件操作码特征，API 调用，数据流特征、控制流特征等特征属性进行研究，研究特征的变化和特征间关系有利于用来作为混淆效果的评估和度量。

8. 程序理解和维护

程序的理解和维护需要软件特征研究和特征定位技术支撑，找到具体实现特征的代

码，特征定位技术是实现面向特征程序理解的关键技术。通过特征定位恢复出特征与实体间的跟踪关系，可分为静态特征定位和动态特征定位。对特征的分析和研究有利于进行特征定位，进而促进面向特征的程序理解研究。

综上所述，水印和软件特征的研究有着非常重要的实用价值和广泛的应用领域，尤其对软件保护有着举足轻重的意义。在数字化产品和网络应用环境中，复制方便，盗版严重，基于水印和特征的软件保护技术研究也显得越来越重要。

1.4　对水印和特征的攻击

1.4.1　软件水印攻击

软件水印攻击是指攻击者企图找到水印，移除或破坏水印的行为。在文献 Ginger Myles[17]将攻击类型从破坏方式上可以分为以下四类：减裁攻击（Subtractive Attack），扭曲攻击（Distortive Attack），增添攻击（Additive Attack）和共谋攻击（Aollusive Attack）。

减裁攻击：又称移除攻击，指攻击者去除可能包含水印的部分代码，使水印 w 按提取方法无法正常提取出来。

扭曲攻击：又称变形攻击，指攻击通过混淆或变换，破坏水印 w，使得水印 w 不能被提取出或提出来的水印 w 已被扭曲不能反映正确信息。

增添攻击：原软件已嵌入水印 w，而攻击者在软件中添加自己定义的水印 w′，用两种水印识别函数都能得到不同的水印 w 和 w′，无法裁决或判断究竟那个水印是有效水印。

共谋攻击：从多个相似的软件版本或若干个副本跟踪对比，分析出水印的特征信息和分布信息，从而进行以上三种类型的攻击方式来破坏水印的保护功能。

许多代码混淆技术，因为能够破坏程序代码及结构，所以也能够破坏水印的正常提取，代码混淆可以造成以下影响：①语言的数据结构可以被拆分或被合并，如数组；②数据结构的维度可以被增加或者减少，例如二维数组可以被转化为一维或者三维；③改变程序嵌套循环方式或被重新排列顺序，例如添加循环函数或交换相邻的语句；④重命名类或变量名。那么，如果这些混淆刚好是发生在水印嵌入的代码段中，则会影响水印的正常提取，同样混淆技术对软件特征也是一种攻击。

面对水印的不同攻击，如何提高水印的鲁棒性和隐蔽性尤为重要，特别是针对攻击最多的减裁攻击和扭曲攻击，本书在研究中基于门限方案的部分提取原理和混沌系统的混沌特性来提高水印的抗攻击性。

1.4.2　软件特征攻击

软件具有多态和变形功能，当软件被加密、变形或被多态攻击时，得到的新软件与原来的语义等价，即变换前后，软件功能相同，但程序特征就会发生变化，这些变化体现在改变软件入口，给软件添加指令，修改软件初始数据，改变软件大小，修改运行路径，修改函数调用方式等。把这种使特征发生变化的软件等价变换方式称之为特征攻击方式，主要包括垃圾代码，等价替换，代码混淆等。特征攻击表现不同，可体现在软件