

基于多预训练模型的在线隐写盲分析系统研究与实现

袁亚飞¹, 卢伟¹, 冯丙文², 翁健²

(1. 中山大学数据科学与计算机学院, 广东 广州 510006;

2. 暨南大学信息科学技术学院, 广东 广州 510632)

摘要: 在实际应用中, 针对未知隐写算法的盲检测难度非常大, 结合实际应用设计实现了一个在线盲检测系统。在 SRM 算法的基础上, 简化特征提取算法, 提高特征可用性和提取速率; 使用多个预训练检测模型, 并采用加权投票策略判定检测结果; 设计实现一种 3 层系统架构, 分布式后台更加灵活高效; 为了进一步满足实际应用要求, 引入多线程技术, 加快检测速率。实验表明, 单张图片的平均检测时间可达 0.97 s, 并且对多种未知隐写算法均具有良好的检测结果。

关键词: 数字图像隐写; 隐写分析; 多模型; 加权投票; 在线检测

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2017.00164

Online universal steganalysis system based on multiple pre-trained model

YUAN Ya-fei¹, LU Wei¹, FENG Bing-wen², WENG Jian²

(1. School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China;

2. College of Information Science and Technology, Jinan University, Guangzhou 510632, China)

Abstract: In reality, universal blind steganalysis is still a sensitive issue. A universal online steganalysis system that could be used in practical application was proposed. With reducing the dimensions of SRM, it could improve availability and speed up feature extraction. Some effective pre-trained models and weighted voting strategy were used in this system with a B/S architecture, involving a higher speed. In addition, multithread technology was introduced. Experimental results demonstrate that high detection accuracy can be obtained and about 0.97 seconds for single detection with the system.

Key words: digital image steganography, steganalysis, multi-model, weighted voting, online detection

1 引言

信息隐藏是利用现代新兴的数字化技术发展衍生出的隐秘通信方式, 通过利用数字化媒介的

冗余信息, 将有用信息隐藏在普通媒介中, 达到不为人知的目的。根据 Cheddad 等在文献[1]中的总结, 信息隐藏技术主要包括水印 (watermarking) 和隐写术 (steganography) 这 2 种。两者具有一

收稿日期: 2017-02-23; 修回日期: 2017-03-26。通信作者: 袁亚飞, dayinyf@163.com

基金项目: 广东省科技发展专项基金资助项目 (No.2016KZ010103); 广东省自然科学基金资助项目 (No.2016A030313350); 中央高校基本科研业务费基金资助项目 (No.16lgjc83); 中山大学科技成果转化计划基金资助项目

Foundation Items: The Special Funds for Science and Technology Development of Guangdong Province (No. 2016KZ010103), The Natural Science Foundation of Guangdong Province (No. 2016A030313350), The Fundamental Research Funds for the Central Universities (No.16lgjc83), The Scientific and Technological Achievements Transformation Plan of Sun Yat-sen University

定的技术相似性, 水印多用于数字版权保护、物品真伪识别、防止数字作品非法传播、数字作品盗版追踪等, 通过特定方式可以访问嵌入的信息; 而隐写术是将隐秘信息嵌入载体媒介中, 人的感官很难发现隐秘信息存在, 一般方法也很难检测是否存在秘密信息, 提取秘密信息则更加困难。

隐写术载体多种多样, 如图像、音频、视频等, 由于数字图像在网络上随处可见、传播方便、冗余信息较多, 因此成为目前最常用的载体。数字图像隐写术充分利用了图像信息的冗余特性和人眼的视觉特性, 将信息隐写到普通图像中, 且不改变原图像的视觉效果。如 LSBM^[2] (least significant bit matching) 隐写算法, 通过修改图像像素八分位最低有效位的方法嵌入秘密信息, 并且尽可能不改变图像统计信息和视觉效果。目前比较安全的隐写算法, 如 EA^[3] (edge adaptive)、HUGO^[4] (highly undetectable steganography) 等, 倾向于将有效信息嵌入图像纹理复杂区域。这是由于图像纹理复杂区域噪声较多, 像素间依赖关系较弱, 检测难度很大。

随着现代网络技术的快速发展, 由于隐写术的高度隐蔽性, 一旦被非法利用, 在互联网中传递有关政治、军事、经济等重要信息, 有可能会给社会安全带来巨大危害。

2 隐写分析概述

为了对抗隐写术, 检测是否存在隐写信息, 研究人员进行了大量研究和分析。

早期的数字图像隐写分析算法主要从图像质量和统计信息入手, 对一些隐写算法取得了很好的检测结果。但随着隐写算法的不断发展, 信息嵌入的隐秘性大大提高, 特别是 EA^[3]、HUGO^[4]、WOW^[5] (wavelet obtained weights) 等自适应隐写算法的出现, 传统检测算法很难满足实际需求。

随后, 机器学习被引入隐写分析中, 通过提取图像特征, 使用机器学习的方法进行训练和分类, 取得了很好的效果。文献[6]提出了一种基于马尔可夫链的隐写分析方法, 该方法利用自然图像像素之间依赖关系提取图像特征, 取得了较高的准确率。在此基础上, Zou 等^[7]改进马尔可夫模型, 计算图像水平、垂直和对角线方向一阶差

分矩阵, 并进行阈值化, 计算马尔可夫特征; Pevny 等^[8]将马尔可夫特征扩展到二维, 计算图像 8 个方向差分矩阵, 提出了 SPAM (subtractive pixel adjacency matrix) 算法, 特征维度为 686 维, SPAM 算法是低维度空间最经典的通用检测算法之一。

自适应隐写算法引入了信息扰动函数^[9], 最小化隐写带来的图像信息扰动, 大大提高了检测难度。Fridrich 等^[10,11]认为增加特征多样性和维度可以有效提高检测的准确性, 并提出了基于多模型 (rich model) 的特征提取算法。Fridrich 等计算多种像素残余、多阶和多个方向的共生矩阵, 大大提高最终特征维度和多样性, 典型的算法, 如 SRM^[11] (spatial rich model), 维度可达 34 671。

由于基于多模型的特征维度过高, 使用传统的 SVM^[12,13] (support vector machine) 分类器计算开销过大, SVM 已经不适应于多模型特征分类场景, Kodovsky 等^[14]提出了集成分类器 (ensemble classifier) 用于高维度特征分类。集成分类器基于随机森林 (random forests) 思想, 在高维特征中抽取小的特征子空间, 选择多个弱分类器, 如 Fisher 分类器^[15], 对每一种子特征进行训练和测试, 使用多数投票策略得到最终分类结果, 并利用检测错误率优化特征子空间和基础分类器数量, 从而获取最优分类结果。集成分类器能够处理高维度特征, 并且分类速度更快, 随着如今隐写分析特征维度增大, 集成分类器已经取代 SVM 成为了隐写分析领域中主流的分类器。

目前, 多模型特征加集成分类器是最常用的检测手段, 但是这类算法在提升特征多样性的同时, 也增加了计算量, 特征提取时间长, 不利于实际应用。本文选取 SRM 特征中部分子特征组成新的 Sub-SRM 特征, 使用多个预训练模型用于图像检测, 设计了一个高效的在线盲检测系统。

3 Sub-SRM 和多预训练模型

SRM 特征使用多种子特征, 但部分特征有效性不高, 可能会影响分类效果, 很多研究人员对此做了很多改进^[16]。本文结合实际应用, 简化 SRM 特征, 并选用多个预训练模型用于分类, 使用加权投票策略判定检测图片类型。

3.1 特征简化

在 SRM 算法中, 为了提高特征多样性和维度, 选取了不同种类和阶数的特征, 但某些子模型质量不高, 对实际应用来说, 会影响提取速率和检测结果。Sub-SRM 算法从中选取出代表性更强的子特征, 以及 SPAM、WAM^[17] (wavelet absolute moment) 特征进行拼接, 总维度为 6 966, 在降低维度的同时, 尽量保持特征的多样性。子特征选取结果如表 1 所示 (特征命名规则参考文献[11], 其中 S 后面的数字表示计算差分矩阵的阶数, 33 和 55 分别表示计算像素残余时使用了 3×3 和 5×5 大小的滤波核, 35 表示两者的组合)。

表 1 Sub-SRM 特征组成和子特征维度

选取子特征算法	子特征维度
S1MINMAX24	325
S1MINMAX34	325
S1MINMAX54	325
S1SPAM14	338
S2MINMAX21	325
S2MINMAX32	325
S2MINMAX41	325
S2SPAM12	338
S3MINMAX24	325
S3MINMAX34	325
S3MINMAX54	325
S3SPAM14	338
S33MINMAX24	325
S33MINMAX41	325
S33SPAM14	338
S35SPAM11	338
S55MINMAX24	325
S55MINMAX41	325
S55SPAM14	338
SPAM	686
WAM	27
总计	6 966

为验证 Sub-SRM 特征的有效性, 本文进行了 10 次对比实验, 每次实验从 BOSSBase^[18]v1.01 图像库中随机抽取 2 000 张图片, 使用 HUGO 隐

写算法进行隐写, 隐写率为 0.25, 得到载体图片和隐写图片共计 4 000 张。得到的图片平分为 2 个部分, 2 000 张用于训练, 2 000 张用于测试。特征提取算法分别使用 SRM、SRMQ1^[11]和本文提出的 Sub-SRM 特征, 分类器为集成分类器, 记录每次实验的测试错误率。表 2 给出了 10 次实验的最大错误率、最小错误率和平均错误率。

表 2 不同特征间 10 次对比实验结果

特征提取算法	最小错误率	最大错误率	平均错误率	特征维度
SRM	26.90%	28.90%	27.63%	34 671
SRMQ1	27.90%	30.00%	29.13%	12 753
Sub-SRM	27.80%	30.35%	29.26%	6 966

由表 2 的实验结果可以看出, Sub-SRM 特征检测正确率较 SRM 特征稍微有些降低, 但与 SRMQ1 特征结果十分接近, 基本保证了正确率。Sub-SRM 算法的优势在于速度, 通过实验可发现, 在 BOSSBase v1.01 图像库上提取单张图片的 SRM 特征大约需要 63.12 s, 但提取 Sub-SRM 特征只需要 14.30 s, 效率有了很大提升。

3.2 预训练模型选择

在通用检测系统中, 预训练模型对检测结果有很大的影响。通过实验可以发现, 1 000 张图片 (500 张载体图片和 500 张对应的隐写图片) 的预训练模型即可取得较好的检测结果, 较小的模型也有利于节省内存和分类时间。因此, 训练时只随机选取 500 张载体图片及其对应的 500 张隐写图片制作预训练模型。预训练模型命名规则为

$$name = Model_{\{algorithm\}\{bpp\}} \quad (1)$$

其中, algorithm 表示用于制作预训练模型的隐写算法, bpp 表示隐写率。所用特征提取算法为 3.1 节中的 Sub-SRM, 分类器为集成分类器。

实验 1 和实验 2 用于介绍选择制作预训练模型的隐写算法的依据。其中, cover 表示载体图片, stego 表示隐写图片。

实验 1 自适应隐写算法的预训练模型更具有通用性

本文选用 3 种隐写算法 (LSBM、EA、HUGO) 和 2 种隐写率 (0.25、0.5) 制作预训练模型, 分

别进行交叉检测，结果如表 3 所示。实验结果表明，自适应隐写算法的预训练模型更具有通用性。这是由于自适应隐写算法信息嵌入更隐秘，模型检测能力更强；而非自适应隐写算法的信息嵌入机制比较简单，嵌入位置更加明显，检测难度较自适应隐写算法小；因此自适应隐写算法的预训练模型对非自适应算法也有很好的检测效果。非自适应隐写算法模型检测载体图片的正确率虽然比较高，但其对于自适应隐写算法检测错误率过高，不利于实际应用，因此本文不使用这类算法制作预训练模型。

表 3 不同隐写算法模型检测错误率对比

预训练模型	cover	stego		
		隐写算法	bpp=0.25	bpp=0.5
$Model_{LSBM0.25}$	7.6%	LSBM	5.6%	4.8%
		EA	59.6%	16.4%
		HUGO	88.0%	59.4%
$Model_{LSBM0.5}$	9.0%	LSBM	4.8%	4.4%
		EA	57.6%	15.8%
		HUGO	84.2%	59.0%
$Model_{EA0.25}$	17.0%	LSBM	23.6%	20.2%
		EA	17.8%	7.8%
		HUGO	61.0%	25.8%
$Model_{EA0.5}$	7.6%	LSBM	44.4%	41.4%
		EA	45.4%	6.6%
		HUGO	85.2%	55.6%
$Model_{HUGO0.25}$	30.6%	LSBM	4.8%	5.2%
		EA	11.8%	2.8%
		HUGO	28.4%	11.8%
$Model_{HUGO0.5}$	20.8%	LSBM	11.0%	10.4%
		EA	16.6%	6.4%
		HUGO	51.8%	17.0%

实验 2 低隐写率的预训练模型检测隐写图片更有效，高隐写率的预训练型检测载体图片更有效

本文选用 HUGO 隐写算法和多种隐写率，制作预训练模型，分别检测 LSBM、EA、HUGO 算法隐写图片，每次隐写采用 0.25 和 0.5 这 2 种隐写率，检测结果如表 4 所示。

表 4 不同隐写率模型检测错误率对比

预训练模型	cover	stego		
		隐写算法	bpp=0.25	bpp=0.5
$Model_{HUGO0.15}$	38.0%	LSBM	5.8%	6.4%
		EA	10.4%	3.6%
		HUGO	25.4%	10.0%
$Model_{HUGO0.2}$	34.2%	LSBM	4.6%	5.8%
		EA	10.2%	3.0%
		HUGO	26.4%	11.0%
$Model_{HUGO0.25}$	30.6%	LSBM	4.8%	5.2%
		EA	11.8%	2.8%
		HUGO	28.4%	11.8%
$Model_{HUGO0.3}$	29.2%	LSBM	6.0%	6.6%
		EA	13.0%	3.2%
		HUGO	32.4%	12.2%
$Model_{HUGO0.35}$	35.0%	LSBM	5.8%	6.4%
		EA	14.0%	2.4%
		HUGO	36.6%	13.6%
$Model_{HUGO0.4}$	25.4%	LSBM	7.6%	8.0%
		EA	13.6%	4.4%
		HUGO	42.4%	13.2%
$Model_{HUGO0.45}$	22.0%	LSBM	10.0%	8.4%
		EA	16.8%	3.6%
		HUGO	47.8%	17.2%
$Model_{HUGO0.5}$	20.8%	LSBM	11.0%	10.4%
		EA	16.6%	6.4%
		HUGO	51.8%	17.0%

实验结果表明，自适应隐写算法的低隐写率预训练模型检测能力更强，这是由于隐写率较低时，嵌入位置较高隐写率图片更加隐秘，低隐写率图片训练出的模型检测高隐写率图片的能力更强；相反，高隐写率会造成信息嵌入位置更明显，隐写图片和载体图片差异更大，其预训练模型检测低隐写率图片效果会变差，但检测载体图片的正确率会上升。

综合实验 1 和实验 2，以及对检测效率的考虑，本文只选择了 $Model_{HUGO0.15}$ 、 $Model_{HUGO0.2}$ 、 $Model_{HUGO0.45}$ 和 $Model_{HUGO0.5}$ 这 4 个预训练模型用于图片检测。更多的训练模型不一定能提高检测

正确率,反而会增加检测时间,因此本文在系统中只选用了代表性更强的4个预训练模型。

3.3 加权投票策略

由于基于机器学习的分类结果是不确定性结果,单个模型的预测结果十分不可信,在集成分类器中,使用了多个基础分类器,采用多数投票的决策原则判定最终检测结果。在盲检测系统中,笔者认为采用多个模型是提高检测正确率的有效手段。由3.2节可知,不同预训练模型对于隐写图片和载体图片检测能力不同,因此,没有采用集成分类器中的多数投票原则,而是使用了加权投票策略。

在选取的预训练模型中, $Model_{HUGO0.15}$ 、 $Model_{HUGO0.2}$ 由低隐写率图片训练得到,由于自适应隐写算法引入信息扰动函数,低隐写率图片特征十分接近载体图片特征,分类器训练出的模型区分载体和隐写图片的能力比较弱,如果检测结果是载体图片,则可信度比较低,需要适当降低其投票权重;但是高隐写率图片会使用更多嵌入位置,隐秘性减弱,图片特征与载体图片特征差别会比较大,因此 $Model_{HUGO0.15}$ 、 $Model_{HUGO0.2}$ 更容易区分高隐写率图片,如果检测结果是隐写图片,则可信度比较高,可以适当提高其投票权重。同理,对于 $Model_{HUGO0.45}$ 、 $Model_{HUGO0.5}$,如果检测结果是隐写图片,则可信度比较低,适当降低其投票权重;如果检测结果是载体图片,则可信度比较高,适当提高其投票权重。因此,本文采用多模型动态加权投票策略,计算最终检测结果为

$$R = \sum_{i=1}^N w_i r_i \quad (2)$$

其中, $R < 0$ 表示检测结果为载体图片, $R > 0$ 表示检测结果为隐写图片, $R = 0$ 则随机返回-1或+1; $N=4$,表示共有4个预训练模型参与检测; w_i 表示每个模型的投票权重; r_i 表示每个模型分类结果, $r_i \in \{-1, +1\}$, -1表示载体图片, +1表示隐写图片。由于每个模型检测隐写图片和载体图片的能力不同,对其分类结果的可信度也不同,因此, $w_i(i=1,2,3,4)$ 的取值规则为

$$w_i = \begin{cases} w_{i1}, r_i = -1 \\ w_{i2}, r_i = +1 \end{cases} \quad (3)$$

笔者进行了多次实验,最终在本文系统中, w_{i1} 和 w_{i2} 取值如表5所示。

表5 权重分配结果

预训练模型	w_i	
	w_{i1}	w_{i2}
$Model_{HUGO0.15}$	0.6	1.5
$Model_{HUGO0.2}$	0.6	1.3
$Model_{HUGO0.45}$	1.2	0.8
$Model_{HUGO0.5}$	1.3	0.8

这种动态加权投票策略可以充分发挥每个模型的预测能力,尽可能减少单个模型引起的决策误差,提升检测正确率。

4 实验结果及讨论

为了体现本系统的检测准确性,本文检测了多种隐写算法和隐写率图片以及载体图片,所有检测图片均没有参与模型训练,检测错误率如表6所示。

从实验结果可以看出,本文提出的盲检测系统对于非自适应和自适应隐写算法均有较高的检测准确率。特别是对于HUGO隐写算法,在0.25和0.5这2种隐写率下,均超过了其对应预训练模型自检的正确率;此外,对于未知的隐写算法SUNIWARD^[19] (spatial universal wavelet relative distortion),在0.5的隐写率下错误率只有17.0%,对于0.25的隐写率也有较好的检测结果,说明本系统具有良好的通用性。

表6 系统盲检测错误率

cover	stego		
	隐写算法	$bpp=0.25$	$bpp=0.5$
34.2%	LSBM	4.2%	5.6%
	EA	10.2%	3.0%
	HUGO	25.8%	10.2%
	SUNIWARD	38.0%	17.0%

此外,本文目的是实现一个能够实际应用的在线检测系统,因此,检测效率也是一个十分重要的方面。传统的基于多模型的盲检测系统受限于运行效率,难以满足实际应用需求。本文使用了多种技术手段,设计了一种3层式架构,使用分布式的后台提高运行效率和灵活性。在本文的实验环境下,单张图片检测用时约0.97 s,以下

几点有助于加快检测速度。

1) 简化的 Sub-SRM 特征选取了更有效特征, 维度降低, 加快提取速度。

2) 改写原有 Matlab 特征提取算法代码, 改写成 C++ 代码, 速度更快。

3) 引入了多线程技术。

4) 选取代表性更强的预训练模型, 减少分类次数。

5) 预训练模型较小, 可以减少计算开销, 加快分类速度。

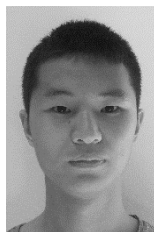
5 结束语

数字图像隐写术由于嵌入信息极为隐秘, 随着网络技术的快速发展而兴起, 而相应的隐写分析盲检测技术仍然有待进一步研究。本文在 SRM 算法基础上, 提出了 Sub-SRM 算法, 使用多个预训练模型, 并依照每个模型分类能力不同, 采用加权投票策略, 设计开发了一个在线检测系统, 在保证检测准确性的前提下, 大大提高了检测速度。下一步将研究更有效的特征提取算法, 设置更合理的决策权重, 提高检测的准确性。

参考文献:

- [1] CHEDDAD A, CONDELL J, CURRAN K, et al. Digital image steganography: survey and analysis of current methods[J]. Signal Processing, 2010, 90(3): 727-752.
- [2] SHARP T. An implementation of key-based digital signal steganography[C]//The International Workshop on Information Hiding. 2001: 13-26.
- [3] LUO W, HUANG F, HUANG J. Edge adaptive image steganography based on LSB matching revisited[J]. IEEE Transactions on Information Forensics and Security, 2010, 5(2): 201-214.
- [4] PEVNÝ T, FILLER T, BAS P. Using high-dimensional image models to perform highly undetectable steganography[C]//The International Workshop on Information Hiding. 2010: 161-177.
- [5] HOLUB V, FRIDRICH J. Designing steganographic distortion using directional filters[C]//Information Forensics and Security(WIFS). 2012: 234-239.
- [6] SULLIVAN K, MADHOW U, CHANDRASEKARAN S, et al. Steganalysis of spread spectrum data hiding exploiting cover memory[C]//The International Society for Optics and Photonics, Electronic Imaging. 2005: 38-46.
- [7] ZOU D, SHI Y Q, SU W, et al. Steganalysis based on Markov model of thresholded prediction-error image[C]//IEEE International Conference on Multimedia and Expo. 2006: 1365-1368.
- [8] PEVNÝ T, BAS P, FRIDRICH J. Steganalysis by subtractive pixel adjacency matrix[J]. IEEE Transactions on Information Forensics and Security, 2010, 5(2): 215-224.
- [9] 凌轶华, 蔡晓霞, 陈红. 应用共生矩阵特征的改进自适应 JPEG 隐写[J]. 计算机应用研究, 2014(11): 3490-3493.
- LING Y H, CAI X X, CHEN H. Improved adaptive steganographic algorithm for JPEG images based on co-occurrence matrix feature[J]. Application Research of Computers, 2014(11): 3490-3493.
- [10] FRIDRICH J, KODOVSKÝ J, HOLUB V, et al. Steganalysis of content-adaptive steganography in spatial domain[C]//The International Workshop on Information Hiding. 2011: 102-117.
- [11] FRIDRICH J, KODOVSKÝ J. Rich models for steganalysis of digital images[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(3): 868-882.
- [12] CORTES C, VAPNIK V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.
- [13] LYU S, FARID H. Detecting hidden messages using higher-order statistics and support vector machines[C]//The International Workshop on Information Hiding. 2002: 340-354.
- [14] KODOVSKÝ J, FRIDRICH J, HOLUB V. Ensemble classifiers for steganalysis of digital media[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(2): 432-444.
- [15] DUDA R O, HART P E, STORK D G. Pattern classification[M]. New York: Wiley, 1973.
- [16] 陈世媛, 汤光明, 高瞻瞻. 基于权值分配的隐写分析算法[J]. 计算机应用研究, 2016, (11): 3468-3471.
- CHEN S Y, TANG G M, GAO Z Z. Steganalysis method based on weight allocation[J]. Application Research of Computers, 2016(11): 3468-3471.
- [17] GOLJAN M, FRIDRICH J, HOLOTYAK T. New blind steganalysis and its implications[C]//The International Society for Optics and Photonics, Electronic Imaging. 2006: 1-13.
- [18] FRIDRICH J, KODOVSKÝ J, HOLUB V, et al. Breaking HUGO—the process discovery[C]//The International Workshop on Information Hiding. 2011: 85-101.
- [19] HOLUB V, FRIDRICH J, DENEMARK T. Universal distortion function for steganography in an arbitrary domain[J]. EURASIP Journal on Information Security, 2014(1): 1.

作者简介:



袁亚飞 (1991-), 男, 河南商丘人, 中山大学硕士生, 主要研究方向为多媒体信息安全与数字取证。

卢伟 (1979-), 男, 河南南阳人, 中山大学副教授、硕士生导师, 主要研究方向为多媒体信息安全与数字取证。

冯丙文 (1985-), 男, 山东东营人, 博士, 暨南大学讲师, 主要研究方向为多媒体安全与数字取证。

翁健 (1976-), 男, 广东茂名, 博士, 暨南大学教授、博士生导师, 主要研究方向为密码学与信息安全。