网络空间安全系列教材

ELSEVIER
爱思唯尔

电子数据取证与Python方法

Python Forensics

A Workbench for Inventing and Sharing Digital Forensic Technology

[美] Chet Hosmer 著
张 俊 译 / 邹锦沛 审校

非外借

中国工信出版集团
電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

网络空间安全系列教材

电子数据取证与 Python 方法

Python Forensics

A Workbench for Inventing and Sharing Digital Forensic Technology

[美] Chet Hosmer 著

张 俊 译

邹锦沛 审校

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内容简介

本书是一本电子数据取证的入门书籍，系统介绍如何应用Python编程语言进行电子数据取证软件开发。第1章和第2章介绍Python基本知识和如何建立一个取证开发环境。第3章到第11章针对电子数字取证的各种需求，详细阐述指导性的解决方法，涵盖哈希计算、关键字搜索、元数据提取、网络分析、自然语言处理以及利用云的多进程等专题，并提供大量的源代码实例供读者学习、改进并应用到实际案例。第12章回顾了全书内容，并就未来的发展进行了探讨。

本书适合网络空间安全、网络安全与执法、信息安全、法学、司法鉴定及相关专业的本科和专科学生作为教材，对于从事数字犯罪调查、计算机司法鉴定、内部调查、软件研发等工作的执法人员、调查分析人员、审计人员，以及取证软件和工具的研发人员，也是提升技能、丰富手段的参考书。

Python Forensics: A Workbench for Inventing and Sharing Digital Forensic Technology
Chet Hosmer
ISBN: 9780124186767

版权贸易合同登记号 图字：01-2016-9444

图书在版编目（CIP）数据

电子数据取证与Python方法 /（美）切特·霍斯默（Chet Hosmer）著；张俊译. —北京：电子工业出版社，2017.9
书名原文：Python Forensics: A Workbench for Inventing and Sharing Digital Forensic Technology
网络空间安全系列教材
ISBN 978-7-121-32131-3

Ⅰ. ①电… Ⅱ. ①切… ②张… Ⅲ. ①计算机犯罪—证据—数据收集—高等学校—教材 Ⅳ. ①D918

中国版本图书馆CIP数据核字（2017）第161153号

策划编辑：马 岚
责任编辑：李秦华
印 刷：三河市双峰印刷装订有限公司
装 订：三河市双峰印刷装订有限公司
出版发行：电子工业出版社
北京市海淀区万寿路173信箱 邮编：100036
开 本：787×1092 1/16 印张：15.75 字数：403千字
版 次：2017年9月第1版
印 次：2017年9月第1次印刷
定 价：59.00元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。
质量投诉请发邮件至zlts@phei.com.cn，盗版侵权举报请发邮件至dbqq@phei.com.cn。
本书咨询联系方式：classic-series-info@phei.com.cn。

中译本序

有幸与湖北警官学院的张俊老师相识，我们经常在学术会议间隙，或者相互访问的交流过程中探讨电子数据取证的问题。张俊老师从事电子数据取证的教学和科研工作多年，培养了很多优秀的学生，他还积极参与电子数据取证的司法鉴定工作，办理了许多重大和复杂的案件。张俊老师一直活跃在电子数据取证领域，并不断关注和跟踪国内外最新的技术发展，所以他有心翻译和出版这样一本书，我觉得一切都是顺理成章的事情。

本书作者Chet Hosmer和技术编辑Gary C. Kessler也是信息网络安全业界的专业人士，他们都是知名的专家，有着长期的专业经历和丰富学识。他们将宝贵的经验和专业知识毫无保留地在书中传授，这将有利于专业人才的成长，有助于立志从事电子数据取证的技术人员和法律人士快速进入这一领域。

电子数据取证是一个发展时间相对较短的领域，还有很多的技术问题、法律问题尚待解决。香港大学计算机科学学系的课程也应用了Python语言编程，我们也鼓励从事信息安全专业的学生学习这门语言，并充分利用其简单易学、第三方库功能强大的特点，与专业知识相结合，针对电子数据取证的不同问题或挑战，开展创造性的工作，提出切实可行的解决方案。

这本书提供了广泛的例子，便于不了解编程或只拥有初级技术的开发人员使用。我们盼望着更多充满智慧和理想的新人的加入，期待你们所有人分享思想、知识和经验，一起推进这一事业的发展。

邹锦沛博士（Dr. CHOW KAM PUI）
香港大学信息安全和密码学研究中心（CISC）
香港大学计算机科学学系
2016年11月18日于香港大学

译 者 序

电子数据取证技术的研究和应用在国内得到了越来越多的关注。全世界的专家和学者在提及信息网络安全时，关注的重点多集中于算法理论的研究、技术方法的对抗以及规范策略的制定等，毫无疑问这些问题极为重要。然而，信息网络安全的最后一道防线必然诉诸法律。最近几年电子数据作为一种独立证据类型，逐渐得到各类法规的认可，例如在深圳快播公司涉嫌传播淫秽物品牟利案的庭审中，关于电子证据的控辩，就引发了司法界对电子证据的广泛关注。作为向法庭呈贡证据的侦查人员、电子数据取证司法鉴定人，他们的知识、水平和能力成为关键。但现实的问题是，由于计算机网络技术的飞速发展，以及案件的复杂和规模化，已有的电子数据取证工具在很多情况下并不能替侦查人员或司法鉴定人自动完成他们所需的全部工作，或者完成更高级的取证调查任务。例如，用现有的工具软件提取一部手机的通话、短信和即时通信记录，通常可以得到数吉赫兹的数据，往往一个案件涉及十几部或更多的手机，因此数据量更大。如果侦查人员或司法鉴定人能够快速地编写代码，进行定制的搜索和索引，将极大地提高从这些海量的半结构化数据中高效地得到关键证据的能力。

译者长期从事电子数据取证的科研和教学工作，并作为司法鉴定人，参与了大量重、特大案件的电子数据取证调查工作，在这些过程中一直反复思考上述这些问题。一个偶然的机会，看到了*Python Forensics: A Workbench for Inventing and Sharing Digital Forensic Technology*一书，立刻被它的内容所吸引，感觉找到了问题的答案。本书的特点在于，使用Python语言上手快，第三方库丰富，调查员可以方便地编写代码来完成复杂的、特定的取证任务，而且无须过多关注语言细节，从而能将主要精力放到取证任务本身。另外，本书体现了开发过程满足多伯特（Daubert）证据标准的重要性，也就是设计、开发、测试过程符合特定的证据标准。这些特点也是开发普通程序与取证程序的最大区别。

作者Chet Hosmer不仅是信息网络安全、电子数据取证领域的从业人士，也是一名教育工作者，作为Utica学院网络安全研究生课程的客座教授，他擅长从调查员视角讲解如何用Python语言进行电子数据取证。本书在介绍如何建立一个Python取证环境的基础上，首先详细讲解开发一个取证应用的基本框架，然后每一章都针对网络犯罪取证的一个不同问题，讨论能够自由使用、分享和扩展的Python指导性解决方案，包括哈希、关键字搜索、元数据、自然语言处理、网络分析以及利用云的多进程等。最后，对Python应用于网络犯罪调查，以及更广泛领域的网络安全应用，高性能硬件加速和嵌入式解决方案等在未来的机遇进行了展望。作者认为，能否建造自己的取证工具库，是区分初级取证调查员与专业取证调查员的关键。全书贯穿着一个资深调查员通过长期实践得出的理念。作为前辈，他谆谆告诫后来者：只有当我们（调查人员）理解工具如何工作时，它们才真正是我们的工具。他正是以这一核心指导原则展开了本书全部内容。

电子数据取证是计算机科学、法学等的交叉领域，技术开发人员和法律工作者以不同的知识背景进入该领域，他们从各自角度去理解电子数据证据的提取、分析和呈贡，并且在专业词汇、思维过程、解决问题的方式等方面有很大区别，形成了介于一个自然科学和社会科学之间的“沟壑”。所以，本书的另一个目的是试图建立起工程学（如计算机科学、信息科学）与社会科学（如法学）之间的桥梁。技术开发人员和法律工作者通过阅读本书，提升对取证环境和工具的理解，可以轻松交流和平等参与，从而创造出一个协作而不互损的环境，计算机科学与社会科学都能各尽其力。我在阅读和翻译的过程中，深刻体会到作者毫无保留地把自己积累的宝贵经验传授给读者。本书对读者的编程知识（也许根本没有）不做任何预设。只要肯用功，对书中的例子感兴趣，读者就不用担心自己读不懂，甚至可以将其扩展，开发出适应特定情形和问题的进化版本。

以我有限的水平，要翻译好这样一本同时适合技术开发人员和法律工作者阅读的著作，内心难免不安。感谢作者Chet Hosmer对我每次发邮件向他求教或确认问题的耐心解答。感谢我的领导黄凤林教授和张天长教授，他们给予我很多指导和帮助。感谢武汉天宇宁达公司的CEO郭永建先生，他给出了若干非常专业的修改意见。翻译本书时，我还向徐比超和胡壮求证过有关的示例是否有误，他们是湖北警官学院的毕业生，现在战斗在电子数据取证的最前线。感谢湖北警官学院的学生吴沛沛、欧阳桂申、彭洪飞、沈阳、朱俊妍参与部分翻译校对和代码测试工作。最后感谢我的妻子，她承担了所有的家务重任，让我全心投入工作，并以我翻译本书为骄傲。

由于本人学识有限，且时间仓促，书中翻译错误、不当和疏漏之处在所难免，望读者批评指正。

专家荐语

Hosmer不仅为各种层次的取证分析提供了一个出色的Python取证指南，还眼光独到地阐述了如何建立一个意义非凡的协作环境，这种环境将极大地提升个人、组织以及取证社区的取证能力。对于分析人员、调查人员、管理人员、研究人员，以及其他任何对数字取证感兴趣的人来说，这是一本必须读的书！

—— Michael Duren（CISSP）
Cyber Moxie公司创始人

随着当今技术的快速变化，数字取证工具和实践也不得不快速更新，才能保持某种程度的实用性；调查人员昨天还依赖的技术能力，今天就迅速地过时。然而，随着新技术一起到来的也有新的工具和方法，Python语言就是其中最有可能被调查人员利用的事物之一。本书就是走在这一时代前列的一本书。正因为如此，无论对于初学者还是有经验的调查人员，它都是一本绝好的书。Chet Hosmer做了一项伟大的工作，通过循序渐进的指导，帮助读者更新旧的方法，掌握新的技能；通过合理的组织架构，最大限度地促进内容理解和前后贯通。从本书学到的技能，将有助于读者开发灵活而新颖的工具，并在若干年内发挥作用。

—— Greg Kipper
Verizon公司高级安全架构师和战略官

本书展现了Python应用于现代数字取证的崭新和务实的视野，提出了关于这种语言的强项和劣势的有价值的深刻见解。每一个有见识的取证调查员都值得花时间和精力来了解本书。

—— Russ Rogers
Peak Security公司董事长

本书对于Python取证程序员、很少或者没有Python编程经验的人都非常有用，对一个有经验的程序员来说也是一本很棒的参考书。这本书考虑到了与多伯特规则有关的问题，包括测试和验证，这些对于取证案件鉴定是至关重要的。

—— Zeno Geradts
荷兰法政研究所高级取证科学家和研发协同人

一如既往，Chet Hosmer提供了一个适用于数字取证的，具有全面性和突破性的解决方案和现代平台。这本书写得非常棒，很好用，为所有水平的Python取证程序员提供了一个坚实的基础，还包括关于实证检验的十分必要的讨论。这本书确实很简洁，对于所有想拥有一个数字取证库的人来说，本书值得拥有。

—— Marjie T. Britz博士
Clemson大学

序

2008年6月16日，在2岁的Caylee Anthony家里，有人用谷歌搜索了“防误操作的窒息”的术语。随后还是这位用户，使用Casey Anthony的名字登录了MySpace网站。几个月后悲剧发生，警方发现了这个小女孩腐烂的尸体。检察官以一级谋杀罪指控Casey Anthony，并在3年后对她进行审判。审讯历时6个月，涉及400多份独立证据。遗憾的是，对计算机搜查得到的详细资料却始终没能被提上审讯。检方的计算机取证检查人员使用了一个工具来提取浏览器历史记录。但使用那个工具时，取证检查人员却只搜索了Internet Explorer的历史记录，而没有搜索Firefox浏览器的。这个故事的教训在于，当我们理解工具如何工作时，它们才真正是我们的工具。

与这个取证检查人员的失败相对比，让我们想想最可怕的武装力量——斯巴达军队。斯巴达军队的强大之处在于其士兵的职业化。优秀士兵从年少时起就只学习一种职业——打仗。在这个行业，士兵严重依赖武器和盔甲的品质。每一个斯巴达士兵的责任就是携带他自己的武器和盔甲上战场，而不是配发的武器。儿辈走上战场前，父辈会把武器传递给他。在接下来的文章里，Chet Hosmer会将现代工具和武器传递下去。但作为取证调查员，你的战场也许是硬盘的未分配空间，但一定不是Thermopylae（希腊东部一个多岩石平原）的道路，无论如何，Chet会像斯巴达长老一样，教你打造自己的工具。能否建造你自己的武器库，是将犯下遗漏浏览器痕迹这种粗劣错误的取证调查员，与专业取证调查员区分开来的关键。

余下的章节涵盖了广泛的主题，包括哈希、关键字搜索、元数据、自然语言处理、网络分析以及利用云的多进程等。Chet会在利用Python编程语言教你打造自己武器的过程中，涉及这些精彩的话题。作为Utica学院网络安全研究生课程的客座教授，Chet既是一位教育者，也是一位实践者。作为40多项涉及网络安全、数字取证和信息保障研究计划的主要调查人员，他由于出色的工作而获得国际公认和奖项。就像斯巴达长老一样，他的知识会促进专业人才的成长。所以，请尽情享用本书的内容。就像斯巴达人的妻子们曾在战前告诉她们的丈夫：“带着你的盾牌回来，不然就战死沙场。”

TJ Oconnor

SANS 红与蓝团队网络卫士（SANS Red & Blue Team Cyber Guardian）

前　言

在过去的20年中，我有幸能与一些世界上最优秀、最聪明和最专注的取证调查员一起工作。这些女士和先生们为查明真相而不知疲倦——他们通常都工作在条件不理想的环境和严格的最后期限压力下。无论是追查儿童劫犯、有犯罪组织、恐怖分子，或者仅仅是用老旧手段偷取你钱财的犯罪分子，这些调查人员都面临巨大的压力，必须竭尽全力将手边的工作做到最好。

我时常与开发最新取证产品的业界领袖们交流，同时不断提升其当前软件的基线，以满足最广泛的潜在用户的需求。我也时常与设法解决现实困难的客户们打交道，这些困难需要立刻得到答案，然而包含答案的数据量却在分秒间变得更大。

作为一名科学家和教师，我看到了来自学生、执法人员以及信息技术人士的渴求，他们具备强烈的期待、独特的调查技能以及对问题的理解，更重要的是对于手头问题的创新性的想法。然而在许多情况下，他们缺乏有助于其事业，而且必须具备的核心计算机科学技能。

Python编程语言以及整体环境为创新开辟了一条崭新的道路。最重要的是，这种语言广开大门，兼收并蓄，提供了大量免费工具和技术，能够彻底改革取证证据的收集、处理、分析和推理。这本书提供了广泛的例子，不但让那些几乎不了解编程或者零基础的人群容易理解，对于那些已拥有扎实的技术，想进一步探索、跃升并参与到在取证领域中扩展应用Python的开发人员来说，也非常有帮助。我期待你们的参与，分享你们的知识，带着你们的热情，帮助我们推进这一事业。

适合的读者

本书对于那些渴望学习如何将Python语言应用到取证和数字调查的任何人来说，都很容易理解。我一直认为这本书是跳板和起点，希望能激励读者创建一些伟大的东西，并与世界共享。

预备条件

能使用计算机，熟悉操作系统（Windows、Linux或Mac），能访问互联网，并且具有强烈的学习渴望。

阅读方法

本书组织如下，第1章和第2章侧重于引导性的内容，同时建立起一个免费的Python开发环境。第3章到第11章针对数字取证的不同问题或挑战，提供操作指导性的解决方案，聚

焦于呈现的关键问题并提出实现的参考。我鼓励读者使用、扩展、改进并提升书中提供的解决方案。最后，第12章回顾了全书内容，并就未来的发展进行了探讨。

支持的平台

书中所有例子都是用Python 2.7.x编写，以提供最好的平台兼容性。本书的网站给出了针对Python 2.7.x和Python 3.x的部分代码。当更多的第三方库完全支持Python 3.x时，所有例子都将提供Python 2.7.x和Python 3.x代码。大多数例子都在Windows、Linux和Mac操作系统进行了测试，也将会在完全支持至少Python 2.7.x的其他环境中正常执行。

下载软件

读者可以从python-forensics.org网站获得书中例子的源代码（如果可能，将同时提供Python 2.7.x和Python 3.x的版本）。

评论、提问和贡献

我鼓励读者们主动参与到这一具有首创精神的活动，你们对python-forensics.org网站的源代码库的评论、提问和贡献会被所有人共享。

我鼓励你们所有人分享你们的思想、知识和经验。

致　谢

致以我真诚的感谢：

这本书的技术编辑Gary Kessler博士，你的透彻见解，新鲜观点，深刻的学术理解和指导为这本书增加了极大的价值。你持续不断的鼓励和友情让我十分享受写作的过程。

Elsevier公司的Ben Rearick和Steve Elliot，感谢你们对这个专题的热情以及一直以来的指导和支持。这种精神对我的帮助超出了你们的想象。

我曾经还有很多老师，指导过我多年来的软件开发和取证，正是他们的帮助，我才能构思这本书的内容。这些老师是Ron Stevens，Tom Hurbanek，Mike Duren，Allen Guillen，Rhonda Caracappa，Russ Rogers，Jordon Jacobs，Tony Reyes，Am ber Schroader和Greg Kipper。

Joe Giordano，他在1998年就富于远见地签订了第一个美国空军取证信息战的研究合同。这个合同催生了这个领域中的许多新公司、新发明，并促成了数字取证研究研讨会（DFRWS），以及Utica学院计算机取证研究和发展中心的建立。无庸置疑，你们都是真正的先驱者！

目　录