Computing Reviews

Association for Computing Machinery

ThinkLoud®

TODAY'S ISSUE | HOT TOPICS | SEARCH | BROWSE | RECOMMENDED | MY ACCOUNT | LOGIN

REVIEW ?

Search

Review

Hacks, leaks, and revelations: the art of analyzing hacked and leaked data
Lee M., No Starch Press, San Francisco, CA, 2024. 544 pp. Type: Book (9781718503120)

Date Reviewed: May 8 2024

Full Text

Imagine you've come across a trove of files documenting a serious deed and you feel the need to "blow the whistle." Or maybe you are an investigative journalist and this whistleblower trusts you and wants to give you said data. Or maybe you are a technical person, trusted by said journalist to help them do things right--not only to help them avoid being exposed while leaking the information, but also to assist them in analyzing the contents of the dataset. This book will be a great aid for all of the above tasks.

The author, Micah Lee, is both a journalist and a computer security engineer. The book is written entirely from his experience handling important datasets, and is organized in a very logical and sound way. Lee organized the 14 chapters in five parts. The first part--the most vital to transmitting the book's message, in my opinion--begins by talking about the care that must be taken when handling a sensitive dataset: how to store it, how to communicate it to others, sometimes even what to redact (exclude) so the information retains its strength but does not endanger others (or yourself). The first two chapters introduce several tools for encrypting information and keeping communication anonymous, not getting too deep into details and keeping it aimed at a mostly nontechnical audience.

Something that really sets this book apart from others like it is that Lee's aim is not only to tell stories about the "hacks and leaks" he has worked with, or to present the technical details on how he analyzed them, but to teach readers how to do the work. From Part 2 onward the book adopts a tutorial style, teaching the reader numerous tools for obtaining and digging information out of huge and very timely datasets. Lee guides the reader through various data breaches, all of them leaked within the last five years: BlueLeaks, Oath Keepers email dumps, Heritage Foundation, Parler, Epik, and Cadence Health. He guides us through a tutorial on using the command line (mostly targeted at Linux, but considering MacOS and Windows as well), running Docker containers, learning the basics of Python, parsing and filtering structured data, writing small web applications for getting at the right bits of data, and working with structured query language (SQL) databases.

The book does an excellent job of fulfilling its very ambitious aims, and this is even more impressive given the wide range of professional profiles it is written for; that being said, I do have a couple critiques. First, the book is ideologically loaded: the datasets all exhibit the alt-right movement that has gained strength in the last decade. Lee takes the reader through many instances of COVID deniers, rioters for Donald Trump during the January 2021 attempted coup, attacks against Black Lives Matter activists, and other extremism research; thus this book could alienate right-wing researchers, who might also be involved in handling important whistleblowing cases.

Second, given the breadth of the topic and my 30-plus years of programming experience, I was very interested in the first part of each chapter but less so in the tutorial part. I suppose a journalist reading through the same text might find the sections about the importance of data handling and source protection to be similarly introductory. This is unavoidable, of course, given the nature of this work. However, while Micah Lee is an excellent example of a journalist with the appropriate technical know-how to process the types of material he presents as examples, expecting any one person to become a professional in both fields is asking too much.

All in all, this book is excellent. The writing style is informal and easy to read, the examples are engaging, and the analysis is very good. It will certainly teach you something, no matter your background, and it might very well complement your professional skills.

More reviews about this item: Amazon, Goodreads

Reviewer: Gunnar Wolf

Review #: CR147759

Recommendations

Reviewer Selected

Featured Reviewer

Related Topics

Browse	Alerts
Security and Protection (K.6.5)	Add
Unauthorized Access (K.6.5 ...)	Add
General (E.0)	Add

Manage Alerts | More Alerts

SHARE

Would you recommend this review? ○ yes ○ no Enter

Other reviews under "**Security and Protection**":	**Date**
CIRCAL and the representation of communication, concurrency, and time Milne G. ACM Transactions on Programming Languages and Systems 7(2): 270-298, 1985. Type: Article	Oct 1 1985
Computer security risk management Palmer I., Potter G., Van Nostrand Reinhold Co., New York, NY, 1989. Type: Book (9780442302900)	Apr 1 1991
Computers at risk , National Academy Press, Washington, DC, 1991. Type: Book (9780309043885)	Oct 1 1991

more...

E-Mail This Printer-Friendly

REVIEWER'S AREA | MASTHEAD | SUBSCRIBE | NEWS | TIPS | HELP | CONTACT US