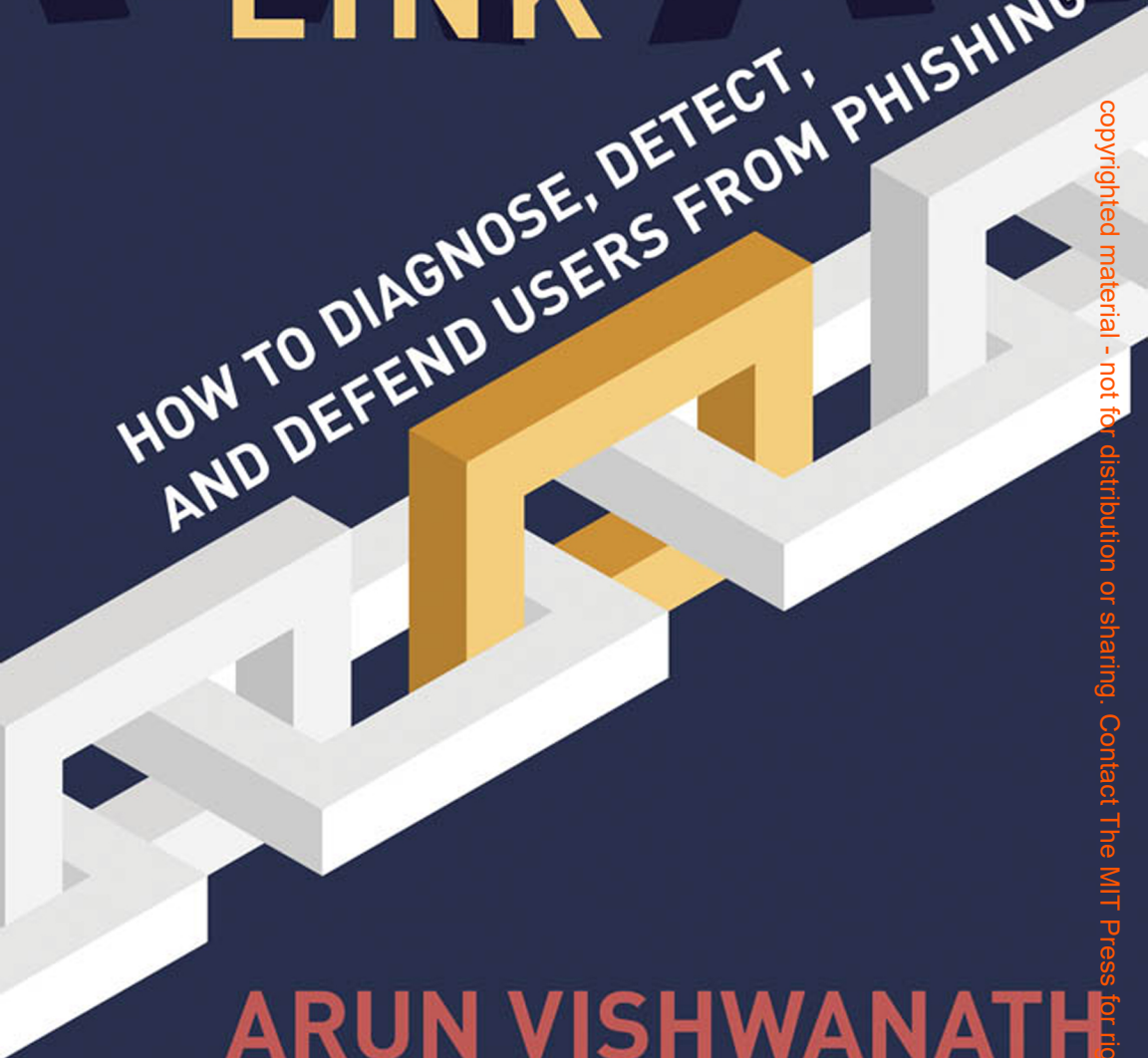
THE WEAKEST LINK

HOW TO DIAGNOSE, DETECT, AND DEFEND USERS FROM PHISHING



ARUN VISHWANATH

THE WEAKEST LINK

How to Diagnose, Detect, and Defend Users from Phishing

ARUN VISHWANATH

The MIT Press
Cambridge, Massachusetts
London, England

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Adobe Garamond Pro by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Vishwanath, Arun, author.
Title: The weakest link : how to diagnose, detect, and defend users from phishing / Arun Vishwanath.
Description: Cambridge, Massachusetts : The MIT Press, [2022] | Includes bibliographical references.
Identifiers: LCCN 2021060548 (print) | LCCN 2021060549 (ebook) | ISBN 9780262047494 (hardcover) | ISBN 9780262371964 (pdf) | ISBN 9780262371971 (epub)
Subjects: LCSH: Phishing. | Computer security. | Computer networks—Security measures. | Computer crimes—Prevention.
Classification: LCC HV6773.15.P45 V56 2022 (print) | LCC HV6773.15.P45 (ebook) | DDC 364.16/8—dc23/eng/20220307
LC record available at https://lccn.loc.gov/2021060548
LC ebook record available at https://lccn.loc.gov/2021060549

Contents

INTRODUCTION

It was Monday, November 24, 2014. Employees logging into their computers were greeted by a locked screen, across which flashed a menacing image of a fiery red skull with long tentacles with the message "Hacked by #GOP." Accompanying it were sounds of gunfire, a poorly worded warning alluding to the theft of all the company's internal data, and a deadline of 16 hours to comply with a demand. This was the beginning of a hostage situation—one that would rewrite the rules of cyber warfare forever.

The company under siege was Sony Pictures Entertainment (SPE). A hacker group named GOP, short for Guardians of Peace, demanded the stoppage of *The Interview*, a Seth Rogen movie slated for release on Christmas Day that featured a comical plot to assassinate North Korean leader Kim Jong-un.

As harried SPE employees restarted their computers, the malware kept finding newer hosts, quickly leaping from computer to computer, jumping through networks and then through servers. Within an hour, the attack had infected all SPE computers in Los Angeles, then New York, and soon across all continents. Within a few more hours, everything digital—files, data, emails, messages, scripts, storyboards had been irretrievably lost.

Writing for *Fortune* magazine, Peter Elkind detailed the scale of the destruction: "It erased everything stored on 3262 of the company's 6797 personal computers and 837 of its 1555 servers. To make sure nothing could be recovered, the attackers had even added a little extra poison: a special deleting algorithm that overwrote the data seven different ways. When

this was done, the code capped each computer's startup software, rendering the machines brain-dead."[1]

Over the next few weeks, the hackers dumped batches of confidential files on publicly accessible file-sharing hubs. These included emails among SPE's leadership team, the salaries and social security numbers of 47,000 employees, passports and visas of various cast and crew members, unfinished and finished scripts of yet-to-be-released movies, and even information about SPE's corporate vendors, such as the salary data of over 30,000 employees of Deloitte, its accounting firm. In all, hackers stole and released over 100 terabytes of data.

By mid-December, the Federal Bureau of Investigation (FBI) had officially attributed the breach to North Korea. The overall cost for system cleanup and recovery would be a staggering $45 million. That's without accounting for the firing of SPE's studio chief, Amy Pascal, and others in top management; the loss of revenue from the leaked movies and scripts; the class-action lawsuits from employees and vendors; and the months of embarrassment from the trove of confidential emails that revealed not just the insides of the movie business but also SPE executives' antipathy toward President Barack Obama and various Hollywood stars and starlets.

But while the media were busy covering the salacious gossip, there was a critical question no one asked: how could a country like North Korea pull off such a major cyber breach? To put this in context, all of North Korea's 24 million inhabitants have access to just about 28 websites, and only 0.3 percent of its entire population—7,200 people—have unrestricted web access.[2] In contrast, SPE's parent company, the Sony Group, employs 114,000 people around the world, all of whom enjoy unrestricted web access. Even by itself, SPE's revenue of $8 billion is more than the combined import and export revenue of North Korea.[3] So how did this technologically unsophisticated nation push one of the world's foremost technological corporations back to the precomputer age, where employees were now resorting to Post-It notes and bulletin boards for communication? The answer is spear phishing—a virulent, internet-based social engineering attack that I had been tracking, researching, and warning about for almost a decade.

Spear phishing uses old-school confidence tricks to get unsuspecting users to provide their online account passwords on fake websites or click on malicious hyperlinks and attachments that provide direct access to their device. Attacks are deployed via email but can just as easily be sent using social media, text messages, USB storage devices, Wi-Fi networks, or Bluetooth connections.

Using this access, a hacker can lurk on a victim's computer, virtually move within an organization's computer networks, and install destructive programs to siphon off data or prevent access to it. Basic programming skills and a free email account for deployment are all it takes to craft one of these attacks, and, since the attacker takes on the virtual identity of the system's approved users, there is no way to identify and block them once they enter the system.

Sony, like just about every other company in the world, provided all its employees with email and internet access. This made spear phishing the perfect attack vector for a less technologically developed adversary such as North Korea. They needed a free email account and needed malware concealable in a hyperlink. One outgoing internet connection could procure this. Using it, they could craft the attack and send it to everyone at Sony—and they could do this repeatedly. All it would take to start the breach was for just one of the thousands of SPE's email users targeted to click on the malware. Email users were the conduit for the breach—they were the weakest link.

The SPE attack was a game changer because it showed how spear phishing could easily co-opt users and wreak havoc on organizations. Thanks to the ubiquity of emails, this virtual surface area for exploit within any organization's network infrastructure was vast and easy to access. Now that the hackers who launched the SPE attack had exposed this potential, hackers everywhere began to copy them. Weak links everywhere were being targeted, and no organization was safe anymore.

SPEAR PHISHING ATTACKS EVERYWHERE

Not long after the SPE attack, in February 2015, came news of another major breach. Hackers had stolen the personal health data of 80 million customers of Anthem, Inc., the nation's largest health insurance provider.

Close on its heels, in June, news media reported an even bigger breach. This time, Chinese military hackers had exfiltrated 21.5 million user records from the Office of Personnel Management (OPM)—the agency that manages human resources and conducts security clearances for all US government employees. Many considered it the biggest breach of sensitive federal government data in modern American history. It was.[4]

In July 2015 came news of an attack on adult dating website Ashley Madison. This time, a hacker group calling itself The Impact Team had stolen 25 gigabytes of data from the site's servers, including the personal details of 37 million users, 13 gigabytes of corporate emails, and all company financial records. The hackers held the website's owner, Avid Life Media (ALM), hostage, demanding they shut down Ashley Madison and its sister website Established Men. When this demand wasn't met, the hackers released all the stolen data on a public file-sharing site.

As with Sony, the fallout caused months of public embarrassment for ALM, resignations of many key personnel, millions of dollars in cleanup costs, and the payout for a major class-action lawsuit. Sadly, a few embarrassed users committed suicide. The virtual attack had now taken real lives.[5]

News of attacks soon started pouring in from all over the world. By 2017, spear phishing had become the leading threat to cybersecurity globally, responsible for 93 percent of all breaches. It was the attack vector for everything from gathering information to installing malware, key loggers, and backdoors to exfiltrate data,[6] and it was being used by activist hackers ("hacktivists"), disgruntled employees, terrorist groups, organized crime networks, and state-sponsored espionage rings from Russia, China, Iran, and other nations.

Within a few years after the SPE attack, a spear phishing attack had targeted every person in the world with an email account.[7] Many had already become a victim of some breach—most just weren't aware of it yet.

The situation was even direr for organizations. Many were constantly under attack, and the sheer ferocity of the attacks was astonishing. In a meeting I attended, the US Senate's sergeant at arms reported blocking 23 million emails in 2017; the US House of Representatives reported blocking millions more. That year, the Department of Defense received 36 million malware-laden emails per day, or 13 billion emails.[8] The responsibility for

stopping this deluge fell on information technology (IT) managers—on whose shoulders it has remained.

THE "PEOPLE PROBLEM"

IT managers all over the world have taken one of two approaches. The first, rooted in the engineering tradition, focuses on using technology to thwart attacks. The goal here is to reduce the quantity of spear phishing emails that reach users and, if the emails do reach anyone, restrict the damage.

The sheer volume of email-based exchanges across all the many devices people use limits this approach. It is difficult for IT to read all online communication, monitor all devices, and effectively limit all spear phishing emails from getting to users without invading their privacy. The other limitation comes from the speed of current computing innovations, which changes so rapidly that proactively finding the weaknesses in each new technology—let alone keeping the virtual hatches on all of them tightly latched to forestall all breaches—is impossible.

This realization has led to the second approach: "hardening" people. Organizations harden their employees, or make them more resistant to spear phishing, by using different awareness training approaches that aim to educate users and improve their ability to detect the deceptive clues in the emails.

Most cybersecurity training takes one of two forms. One is a type of didactic training that involves in-depth, focused educational sessions conducted online or in classroom-type settings. Most such training uses an exam or test after the training to certify the end user's readiness. The second form of training involves subjecting employees to simulated spear phishing attacks (called pen tests) followed by a short educational module embedded in the attack. In most such "embedded training programs," the educational module, provided only to the users who fail the test, reveals the deception clues they missed in the email. The assumption here is that being deceived shocks users and makes them more receptive to the training that follows.

Today, driven by the fear of getting hacked, almost every organization in the US—and increasingly all over the world—deploys one or both forms of training. For many federal and state government employees and

people working in specific sectors, completing cybersecurity training is mandatory, so organizations expend significant resources not just in actual expenses but also in time spent on training.

Objective data on training effects paints a bleak picture, however, showing training as either having no effect or having merely a short-term influence on user behavior. In academic research experiments, within days and at times a mere few hours after being trained, users revert to their established patterns of email use and fall victim to the very same phishing attacks they were trained to detect.[9] Other evidence comes from my own interactions with organizational leaders who have privately shared their frustrations with the limited efficacy of training on their users. Perhaps the best evidence comes from continued news stories of breaches in technology companies (e.g., Adobe in 2013, Deloitte in 2017, the Financial Services Information Sharing and Analysis Center in 2018, and Microsoft in every year since at least 2019), where users are not just highly trained but also technologically adept. So, even users in companies that people turn to for advice on spear phishing fall victim to it.

Faced with such data, the usual reaction from IT leaders in organizations is to train their users even more. Whenever training fails, organizations resort to even more training. This is because everyone in IT believes that training works. Most think it is so effective that they are okay with punishing users who fail training tests. Admiral Michael Rogers, former head of the US National Security Agency (NSA), famously stated the need to court-martial users who fail spear phishing tests.[10] While most organizations don't go that far, many use punishments ranging from mild warnings to users being reprimanded, singled out, or humiliated for their failure. This strategy of shaming and flaming usually leads to even more training being inflicted on the user, and if none of this improves spear phishing awareness among employees—which happens more often than admitted—IT managers usually end up blaming their users for being untrainable. They call it the "people problem," a catch-all term for everything users do that can lead to security breaches in organizations.

But it gets worse. Not only is training today blindly prescribed to treat the people problem of security, IT managers also use it as a metric of the overall cyber readiness of an organization. The pressure for this comes from regulators and policymakers, as well as from cybersecurity insurers, who have been

asking for quantitative cyber risk metrics for their actuarial computations. To fuel more sales, many training companies also advocate this approach. As a result, organizations increasingly use their user training data—pass or fail on didactic training or failure rates on embedded phishing pen tests—as a metric of user cyber risk in their organization. This compromises an already flawed approach in at least five important ways.

For one thing, the current paradigm of using pen-test success or failure rates as a diagnostic for cyber risk conflates outcomes with causes. Failing a phishing test is an outcome, no different from having an automobile accident. Just as an automobile accident can happen because of factors within the control of the driver (e.g., speed, attentiveness, quality of car) and outside factors (e.g., quality of road signage, other drivers), failing a spear phishing test is also merely an outcome caused by many factors.

Second, if our goal was to reduce accident rates in a group of accident-prone individuals, the last thing we would do is try to teach them by inflicting an accident on them and punishing them for it. But that is precisely what all the many phishing tests do: they find crafty ways to make users fail a test and then accord punitive treatments for failing. If this were our approach to driver's ed, it would lead to more paranoid drivers rather than safer ones.

Third, if our goal was to improve how accident-prone drivers could drive better, our immediate assumption wouldn't be that they all simply lack safe-driving awareness. After all, some in our group might have had accidents because of poor eyesight, others may be driving impaired from some substance in their system, and still others might be inattentive because talking or texting while driving distracted them. Simply imparting more-rigorous knowledge wouldn't help those for whom knowledge or awareness wasn't a reason for their poor driving. Likewise, repeatedly conducting awareness training on everyone by using the same approach does little to protect the organization if the underlying reasons for phishing susceptibility vary by user.

Fourth, if our goal is to make users capable of safely driving a single car on a stretch of road under the most predictable circumstances, then perhaps having them comply with the same limited set of instructions is all we need. But driving is far more complex and requires adaptations based on the type of car, traffic, road surface, and weather. Auto safety requires some

compliance, such as always wearing seatbelts, but also requires a good degree of "internalization"—a social science term that means incorporating the information into one's consciousness and using it as a principle to guide decision-making. It also needs good driving habits such as routinely checking rearview mirrors, looking over your shoulder when merging, and signaling before turning. Instilling all this requires more than training. It requires imparting knowledge, fostering motivation, and nurturing good driving habits.

Compared to roadways, the information superhighway is vaster and more unstructured. Cyberattacks are not just more frequent but also more unpredictable, because they mutate with new software. Some vulnerabilities (known as "zero days") aren't even known until a hacker exploits them. Yet our approach to cybersafety is less developed. We train everyone under the singular assumption that they lack awareness. Everyone gets the same awareness knowledge, regardless of its relevance or fit to their needs. Everyone is then subjected to even more testing—which, when many users invariably fail, leads to more of the same training. Nowhere in this process is there any measurement of what users needed, what they lacked, or why they failed. The outcome would be just as poor with automobile safety if we failed to consider a driver's vision problems and instead focused on repeatedly teaching them how to read road signage.

Finally, the goal of auto safety is to make not just one driver safe but to make the entire system safe. A big impetus for this has been the auto insurance system, which rewards and punishes drivers for safe driving. But the system isn't arbitrary. There are insurance adjusters and law enforcement personnel who investigate the cause of accidents, interview drivers, and get feedback. The system allows access to this information, creating accountability. It also gives feedback, allowing users to reduce premiums, giving them agency over their points. In turn, this has driven demand for automobiles with more safety features, for manufacturers to innovate, and for users to be more careful while driving—which together have made automobile safety the norm rather than the exception.

A similar approach is missing from our present regimen for training users in cybersecurity. While we use different phishing simulations and score users who pass or fail, we punish those who fail, usually with even more

training, and ignore those who pass. Today, insurance companies indemnify organizations from cyber losses, but they, too, rely on the training data provided by the organization's IT department to decide on premiums.

The problem is that no one really knows, or cares to know, why the user passed or failed the test—not the person in IT crafting it or even the user, who is neither asked about nor understands their performance. So, it's unclear whether the test was flawed, whether passing or failing was a fluke, or whether the outcome resulted from some conscious or unconscious thought or habitual action of the user. Because of this, it's hard to trust the scores from such training, which does little to build accountability or give users a sense of agency regarding their security posture.

Building a culture of cybersafety requires that we make users accountable for security—their own and that of the entire organization. This means more than simply scoring pen-test failures. It means understanding the thoughts and actions of users that led to their failure or success, the likelihood of them doing so again, and the likelihood that others would have fared similarly. It means receiving feedback both about the test and about the user. It also means giving users feedback, along with a road map for improving their security posture. If done consistently, this can help build user trust in the scoring process, create accountability, and foster a culture of cybersafety. Doing so will help make not just one user safe but everyone, from others in the organization to those the user takes this learning to—vendors, customers, and even people in their household. This is how cybersafety can change from being a dictate enforced by IT managers to a cultural and societal reality.

Achieving this requires an approach that can accurately capture the thoughts, actions, and habits that lead to user risk. Such an approach needs to be agnostic to the operating system and other software they are using, and the work they are doing. This way, we can assess and track the cyber risk of anyone, anywhere, and prepare them for present, potential, and even zero-day attacks. The approach also requires a risk classification system that is simple for all users to not just use but also comprehend. This way, we can tell them about the cyber risk they pose and help them reduce it.

For organizations, knowing the causes of user risk can lead to more-effective policies. Knowing why some users among them weren't at risk,

meaning what some did right—a factor ignored in all our current approaches—can lead to novel solutions or interventions that can be learned, and this knowledge can be built on. Because such solutions emanate organically from within the user population, they can address the intraorganizational causes for user risk in ways no vendor-based training approach developed from outside ever can. This, by default, ensures that policies and security postures align with how users perform their work-related tasks. It further ensures that users won't find workarounds for policies that stand in the way of work, so IT staff can police less. Organizations can then reallocate these security resources and, for a fraction of the effort and cost, achieve cyber resilience.

Implementing such a risk assessment approach requires a method that is reliable, transparent, and compatible with what's already being done presently. This way, any organization, whether on Main Street or Wall Street, can use it. This risk classification system and scoring method is what this book provides.

A DIAGNOSTIC FOR IDENTIFYING THE WEAKEST LINK

Conducting a cyber risk assessment that meets all the above criteria may appear daunting, but it doesn't have to be. Over the course of a decade, my research on technology users has uncovered the reasons some people fall victim to spear phishing and some don't. It has identified the metric and developed the method for measuring these reasons.

My research used simulated social engineering attacks on different users and figured out *how* and *what* people think that makes them susceptible or resilient. It developed a method for capturing these thoughts and behaviors reliably.

This work led to an important realization: there were less than a handful of user-level factors that led to deception or its detection. By measuring just these, any organization could pinpoint who among their users posed a risk to the enterprise and understand why.

But this approach faced an important limitation in that it required 40 or more survey questions. The questions were also complex, some involving terms that only users with a command of English would understand. Employees in many organizational settings couldn't or wouldn't be able to answer the questions.

My focus shifted to the development of a practical, globally usable measure. The need was for an efficient measure that was easy to understand and answer by anyone, in any organization, anywhere, within a few seconds. Through another series of research studies, this quest culminated in the identification of two simple questions, the responses to which were just as valid. These two questions and the approach for asking them is what I call the Cyber Risk Survey (CRS).

The CRS follows a phishing pen test, which many organizations already do as part of their training protocols. But, unlike the tests that are presently being done, it uses a defined framework for crafting the email and deploying it. I developed this framework, called the V-Triad, in 2017 and presented it at Black Hat USA. (The full presentation is available on YouTube.)[11] In 2020, researchers from the National Institute of Standards and Technology (NIST) used the same approach I'd used and came to similar conclusions about phishing pen tests—independently underscoring the validity of the V-Triad.[12]

But the CRS achieves more than simply scoring pen-test quality. It helps design the optimal phishing pen test, collect feedback about it, deploy it on users, and capture users' susceptibility to it. The measures help in understanding who is at risk now as well as who is likely to be at risk in the future. This enables IT managers to pinpoint how much risk each user poses relative to others and locate the weak links among users.

The CRS is also underpinned by another framework: the Suspicion, Cognition, Automaticity Model (SCAM). Developed by my research team in 2016, this model comprehensively explains the user thoughts and actions that lead to social engineering deception or its detection. The model is built on years of testing on users who were subjected to various types of simulated social engineering attacks.

The SCAM provides the explanatory framework for the CRS. Using it to guide analysis, IT managers can go beyond simply locating the weak links; they can now explain why the links are weak. They can also identify the lower-risk users and understand the specific thoughts and actions that attenuated their risk. They can use this information to craft best practices that are particularly suited for the organization and face little resistance from users.

Explanations of the causes of user risk can be communicated with users. This can create a shared understanding of security issues, an internalization of solutions, and a culture of cyber safety. Last but not least, the CRS provides a net resilience score (NRS) for the overall organization. This score serves as a metric of organizational preparedness against future social engineering attacks, which can be used to compare the performance of organizational groups and for tracking improvements in their resilience over time.

Thus, the CRS is a comprehensive approach to the threat of social engineering. IT managers no longer need to worry about the adequacy of their user-focused approaches or wonder which user in the organization poses the biggest risk to the organization's cybersecurity. They can use the CRS along with its associated metrics to identify risky users, create accountability, craft policies and protections based on these results, and develop a culture of cybersafety. Using the information in this book, IT managers can solve their people problem of social engineering.

THE GOAL OF THIS BOOK

This book arms the reader with the knowledge necessary for detecting, diagnosing, and defending against cyberattack by using the Cyber Risk Survey approach. It details the steps for implementing the approach, from how IT should conduct the ideal phishing pen test to how they should administer the CRS and generate the net resilience score.

Generating a valid NRS requires that managers understand social and cognitive science, empirical research, and approaches to analyzing behavioral data. It also requires understanding the fundamentals of social engineering and pen-testing development. This book covers all these topics. It provides a working knowledge of cognitive and behavioral science theories that explain how and why people get deceived via social engineering. It presents the CRS questions, the SCAM framework for analyzing responses to the questions, and the mechanism for converting user responses to the pen test into an NRS. It also covers the V-Triad and teaches how to craft the ideal phishing pen test and assess its quality.

This book further delves into cyber hygiene—a term that is widely used in security circles but is poorly understood—explicating what cyber hygiene

means, what it entails, and what it doesn't. It presents the Cyber Hygiene Inventory (CHI), a scientifically developed tool for quantitatively measuring a user's cyber hygiene. It provides the inventory and discusses how the CHI and CRS can work together to quantify user risk. Finally, the book presents case studies of how different organizations used the CRS and CHI to achieve cyber resilience. In short, the book equips the reader with all tools for solving the problem of social engineering.

WHY I WROTE THIS BOOK

My single most important driver for writing this book was frustration. You see, the November 2014 ransomware attack on SPE was not just a watershed moment in cybersecurity history. It also marked a turning point in my career. Until then, I had been a social-behavioral scientist studying users and how social engineering attacks deceived them. This followed years exploring the psychology of users and what drove them to adopt and use technology.

I'd been drawn to the study of social engineering by a series of spear phishing attacks that targeted my university's email users in 2009. The attacker asked users to update their passwords or risk losing access to their emails—a common attack today but less so back then. It was different from the typical poorly worded "Nigerian Prince" emails. The attacker focused on what mattered: a relevant subject line, a bold warning, a short deadline, and a highlighted hyperlink for changing a password.

The subject line is what most people viewed first; its relevance hooked them. The warning and deadline did the rest. Grammar and spelling errors were there, but the threat of losing access to their emails caused everyone to ignore such flaws and reeled them in. They hurriedly clicked on the centrally placed blue hyperlink, which took them to another official-looking website that captured their passwords. The attacker had crafted the email to take advantage of how people's minds worked online—it was designed by someone who understood user psychology and how to take advantage of it.

No one really noticed much of this or cared: not the users, most of whom complied, or even those in the IT department, who simply asked users to change their password—with another email, of course—and moved on. The problem was being solved using the same approach that caused it, a "solution"

that many organizations (such as retail chain Target and credit-monitoring agency Equifax) implemented years later, with dire consequences.

Back in 2009, I was testing different content-based persuasive approaches on users, looking for ways to make them optimally use email, social media, and mobile devices to harness their fullest potential. The spear phishing attacks appeared to have a similar focus but with a darker agenda. They, too, were using content and other persuasive techniques to make online users do something—open, click, enable, accept—except their goal was to co-opt the user and misutilize the access.

I began by surveying the victims. I wanted to understand how they thought, the actions they took, and the impact their actions had. The attack's overwhelming success soon stood clear. Not only had most users complied, but none were even aware that they had been deceived. Many had acted without thinking, and those who had paused to think didn't have the know-how to correct their decision. Everyone had either thoughtfully or inadvertently complied. Most were victimized. This new form of attack was the ultimate vector of online deception. Anything could be concealed behind a hyperlink, even within an email, in an attachment. From there, the only limiting factor would be the motivations of the hacker.

My research focus completely shifted to examining this new vector. Much of my early work focused on replicating attacks reported in the media. I read about attackers using social media to impersonate people, using different appeals to elicit a response, using different devices (e.g., USB sticks, texting apps) to entice usage, and spoofing or mimicking different web pages to deceive users. After procuring the requisite approvals for studying human subjects, I experimentally re-created many of them. The goal was not just to examine how each attack succeeded but also to identify the specific facets that caused deception.

These tests initially used students in the US and abroad, then employees in organizations that had volunteered. With each study, I improved my approach, iteratively refined my measurement techniques, and better understood the problem. Each study gave me a better understanding of what hackers were doing, how their attacks were evolving, and how vulnerable organizations really were to this new form of attack.

By 2014, I was among the few researchers who had extensively studied the phenomenon of spear phishing from the user deception perspective. Thus, when the infamous Sony attack occurred in November of that year, I not only knew what the North Koreans had done but also exactly how they had accomplished it. I also knew what was coming—now that the proverbial genie was out of the bottle.

But the media completely ignored this. They focused on the salacious, inside Hollywood gossip: on what the emails said, not what the attack signified. I wanted to bring attention to what really mattered—the spear phishing attack, why it had succeeded, and the looming threat of more like it.

To this end, in December 2014, I wrote my first CNN opinion piece discussing how the North Koreans had accomplished the SPE breach by compromising users and what organizations and policy makers must do to stop such attacks. It would be the first of many more such attacks, because, as I'd expected, hackers all over the world started launching even bigger, bolder, and more consequential attacks.

So, I followed up with even more media pieces. Whereas once I had focused solely on addressing the research community, I now focused on communicating directly with cybersecurity professionals, technology users, IT managers, and policy makers. I became a technologist writing in the public interest. My goal was to bring attention to the core user-focused problems in cybersecurity and offer solutions to them. My hope was that by drawing attention to them, organizations and policy makers would attend to them before they mushroomed and became intractable.

I was among the first to call for user cyber hygiene—a term that has since become common in the field—and the first to develop an approach for measuring it. I advocated the implementation of two-factor authentication as a default setting, and called out the dangers of single sign-ons and fake social media profiles. I called for a nationwide system to forewarn others of cyberattacks, for privacy apps to have better "warning labels" and five-star rating systems, and for us to be concerned about the rise of mobile-based attacks, internet-of-things (IoT), and artificial intelligence. These ideas were featured on CNN and in *Wired* magazine, the *Washington Post*, *Politico*, *USA Today*, the *Chicago Tribune*, *Scientific American*, and other leading outlets.

I became one of the foremost experts on human cyber vulnerability. I was invited to present my research at the Coalition for National Science Funding (C-NSF) on Capitol Hill, multiple times at the US Senate and House by the Senate sergeant at arms, and at leading security conferences such as Black Hat. I presented at the US Army Cyber Institute at West Point, New York; at Johns Hopkins University; and to the cybersecurity thought leaders at the NSA, FBI, Department of Homeland Security (DHS), and the Obama White House's Office of Science Technology and Policy (OSTP). But despite the coverage, the publications, and the presentations, the problem of social engineering at its core remained. Nothing really changed. Attacks kept happening, becoming more frequent and significant.

Many of the problems I'd exposed were ignored—only to become even bigger in the years to come. For instance, after a spate of attacks on US hospitals, in 2016 I wrote about the rise of ransomware, calling it the year of the ransomware. I was wrong. It wasn't just in 2016; it's been every year since then. In 2021, we witnessed some of the biggest ransomware heists: Colonial Pipeline, which supplied half of the US East Coast's gasoline; JBS, the world's biggest meat supplier; and Kaseya, a supplier of remote monitoring software to IT companies that serves over one thousand small- and medium-sized businesses all over the world. What had been ransom demands of a few hundred dollars in 2016 are now up to $70 million in the case of Kaseya. What could have been stopped then is now going to cost millions more to clean up.

All the more frustrating was that the solutions I'd offered all along could have prevented much of this. I had presented a diagnostic for assessing users' cyber risk beliefs, their cognitive-behavioral schemas, and their habits; I'd developed a mechanism for using this to train users; and I'd even designed a tool for measuring cyber hygiene and developed an approach for conducting cyber hygiene assessments, all of which were publicly available.

Most security companies chose to ignore this. They doubled down on their own products, which they had already spent millions developing. Security researchers did the opposite. They took to rebottling my suggestions. Some relabeled risk belief as overconfidence, others called it user curiosity, while still others termed it the users' lack of technical efficacy. It's a phenomenon

best described by Mark Twain as the old habit of calling everything new. This habit is endemic in academia, where scholars are rated on originality, not substance or value, so every academic is motivated to reinvent, rename, and reframe a problem rather than test or improve any existing one. Adding to this is the low stake in their pursuit. Neither a breach nor a data loss or its fallout matters to them personally or professionally.

But it does to IT managers and organizations the world over. They are at the mercy of vendors at one end, regulators at the other, and their customers throughout. They are caught between technologies that they cannot operate without and attackers who are ceaselessly finding vulnerabilities in them. For them, social engineering isn't a product problem or a dialectic debate. It is a clear and persistent problem—one that threatens the very survival of their organizations.

As I noted earlier, this frustration with the process of social science and the security vendor space led me to write this book. If you are reading this, you likely share many of these frustrations. You have realized that whatever has been done so far didn't help you pinpoint the weak links among your users, understand why they posed a risk, craft evidence-based risk mitigation solutions, or define your entire organization's social engineering vulnerability.

This book is the product of my principled frustration and the answer to yours. It distills years of research in cyber science, both my own and that of others, into actionable solutions in one comprehensive presentation. It is written such that these solutions can be directly applied without needing a vendor, a product license, or a subscription fee. For nothing more than your attention and willingness to try, you can use the knowledge and techniques from this book to solve the problem of social engineering.

WHO IS THIS BOOK FOR?

If you are working to stop social engineering, this book is for you. If you are the IT manager creating policies to improve cyber resilience, the security staff member working to reduce the organization's exposure to social engineering, or the consultant helping them accomplish this, then this book is squarely

for you. It is also of value to you if you work for the insurer indemnifying organizations, the law enforcement agencies working to protect them, and the regional and national-level policy makers enabling them.

The Cyber Risk Survey approach complements existing approaches for user training while significantly improving what's being done. The Cyber Hygiene Inventory provides the missing toolkit for quantifying and tracking users' thoughts and actions related to digital safety. Both the CRS and CHI fit readily with many organizations that already have a training pipeline that includes pen testing.

For such organizations, this book provides the missing framework for understanding and quantifying user risk. For organizations that don't already have a training protocol in place, it provides the framework for creating one, including the information necessary for developing a pen-testing approach from the ground up and for tracking cyber hygiene levels.

The book's contents are also of value to consultants and training organizations hoping to distinguish themselves in the marketplace. Using its approach, consultants can identify the users who are at risk and diagnose the reasons for it. Rather than prescribe more user training like every other consultant does, they can help organizations develop novel, evidence-based solutions that meet the needs of users.

This book is of value to cybersecurity insurers as well. By implementing the CRS across different organizations, insurers can assess cyber risk across a sector. They can do more than just audit the organizations presently at risk from social engineering. Now they can explain why their client organizations are at risk and provide them with ways to improve it. This helps insurers make more-reliable actuarial estimations and build better relationships with organizations they insure.

Finally, the contents of this book are also of value to national security and public policy makers looking to improve cyber resilience. By measuring cyber risk and cyber hygiene across a region, they can create applicable legislative policies. They can identify and target law enforcement and investigative resources to regions where they are most needed. These are more effective than the present vendor-provided estimates that have failed to deliver.

With the breadth of audience in mind, this book focuses on providing actionable knowledge. In place of lengthy commentary on social science theory and the intricacies of various malware, it uses case studies and vignettes. Instead of lengthy references and reams of empirical data, the book emphasizes brevity and explanation. The goal throughout is to move IT managers toward implementation of the CRS—so they can diagnose, detect, and defend their users before the next social engineering attack.

HOW THIS BOOK IS ORGANIZED

This book will take you on a journey that leads to cyber resilience. The discovery begins in chapter 1 with an overview of how social engineering evolved into what it is today. We'll trace the root of this method of deception—the vector—to surprising places. We find it metastasizing in the highways of precolonial India. We'll discuss why this was the case and how, by applying the then budding science of criminology, a young British East India Company colonel ultimately defeated it. The lessons of that era, while lost to time, are even more relevant today—and we'll see why.

Chapter 2 examines the proliferation of email-based social engineering. Most people attribute the success of this vector of attack to email's ease of use and to the internet's ubiquity, but these are proximate explanations. They don't explain why other forms of online attacks aren't similarly successful. The chapter traces the ultimate reasons for the popularity of email-based social engineering to the supply and demand of user data, the fundamental flaws in the internet's system of authentication, the ease of finding users online, and the waning power of IT departments. These are system level causes that aren't going away any time soon. The chapter then explains how these have advantaged the social engineer.

Chapter 3 details the various approaches that organizations have taken to combat social engineering. It covers the different technical approaches as well as the more prevalent phishing penetration testing (or pen-test) approach to improving security awareness. If you are in IT, you are familiar with most of these, especially the technical approaches. Keeping with the

spirit of the book, the chapter delves less into each technique of protection and focuses more broadly on each approach and its inherent pitfalls and strengths. The chapter culminates with an in-depth look at security awareness training and what we know about its effectiveness.

Chapter 4 provides the missing scientific framework—one that should have been the starting point to combat social engineering all along: an understanding of why people fall victim to it. The chapter explains how people think and act online and how elements in an attack are crafted to deceive them. The explanation of how people think centers around their cognitive processing styles and their beliefs about online risk. The explanation for how people act focuses on reaction, nonconscious habits, and their precursors. We'll discuss how these interact and explain *why* people get victimized.

Chapter 5 goes a step further. It presents the Suspicion, Cognition, Automaticity Model (SCAM) framework and discusses how the patterns of thinking and behaving online causally determine phishing susceptibility. The chapter explains *how* victimization occurs. Just as the depth and intensity of coughing serves as a marker of potential human illness, the chapter discusses a key symptom in the SCAM that acts as a marker of user susceptibility to phishing. This marker serves as a single convenient, quantifiable measure of deception detection.

Chapter 6 focuses on another missing part of the cyber risk assessment: a reliable pen test. The chapter explains how present-day pen testing follows no theory, guiding framework, or logic. Other than trying to trick people into clicking, there isn't much of a governing rationale for any test development. Because of this, much of the data from pen tests is not comparable and should not be used to diagnose risk. This chapter presents a solution to this problem: the Vishwas Triad (V-Triad)—an empirically derived framework that helps in designing a reliable and valid pen test. You'll also be introduced to the Cyber Risk Survey(CRS), an approach for measuring the validity of pen tests and user cyber risk.

Chapter 7 pulls the preceding chapters together. It shows how the CRS can be implemented by IT departments to conduct a comprehensive user cyber risk assessment. It details the steps to be taken by IT managers, from their assumptions about risk assessment to how they should create a suitable

pen test, establish its baseline validity, deploy it, assess its outcomes, and measure its impact. The chapter explains how the data from the test can then be used to generate a net resilience score for the overall organization.

Chapter 8 tackles cyber hygiene—often seen in policy making and IT circles as a solution for reducing cyber risk from users. But cyber hygiene is an idea without guiding principles. There is no clarity on what it entails, what it doesn't, and when we know that someone has enough of it. We'll discuss the pitfalls of this. The chapter then delves into organizational design theory and presents five governing principles that help define cyber hygiene and shape our expectations from it. We'll then discuss a measurement tool, the 25-question Cyber Hygiene Inventory (CHI), which can be used to track user hygiene. The chapter will provide the entire tool and culminate with a discussion on the links between cyber hygiene and cyber risk, and how the CRS and CHI can be used in tandem to create cyber resilience.

Chapter 9 provides real-world examples of the cyber risk assessment process being implemented. It demonstrates how organizations of different sizes faced with a variety of challenges have implemented the CRS and CHI to assess cyber risk from social engineering, create a culture of security, and achieve user resilience against cyberattacks. This chapter provides working examples that elucidate how organizations have approached cyber risk from users and achieved resilience. At the end of this chapter, you will recognize the possibilities that lay ahead of you.

The book concludes in chapter 10 with a look into the future. It examines the changing IT landscape and discusses how our present security approach to users will likely make things worse for enterprise security. We'll look at how the advantages in favor of the hacker will tilt all the more so. We'll then assess how the CRS and CHI help redress this imbalance and reduce these advantages. The book's journey ends with a discussion of how doing what's advocated in the book will lead to a change in the role of IT from that of a first responder, showing up to put out fires, to that of an agent of change.

This book will change the way you diagnose, detect, and defend your users. It will revolutionize the way you combat social engineering. Let's begin the journey.