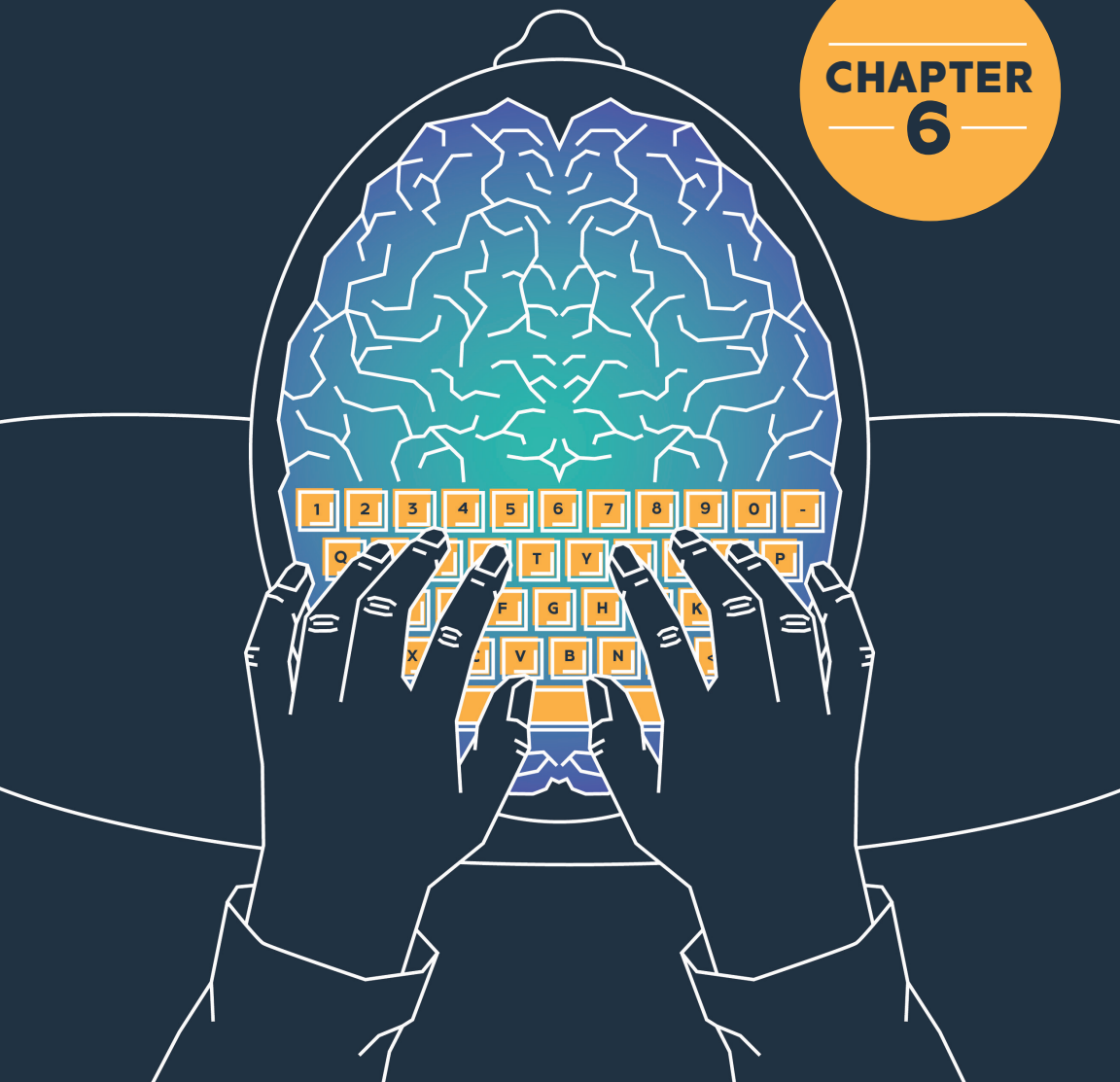
DEEP THOUGHT

A CYBERSECURITY STORY

Alex Blau, Alexandra Alhadeff, Michael Stern, Scott Stinson, Josh Wright

ideas42.org/cyber

CHAPTER 6



CHAPTER 6

D33PTH0UGH1 EXITED THE SUBWAY PLATFORM at 59th St. and Lexington Ave. with the crowd of morning commuters. Her stomach ached with hunger, and her head throbbed from caffeine withdrawal. She was desperate to find something to eat but was painfully aware that there was little time to scour her surroundings for food. She started walking south down Lexington; on the first block were no places to grab a quick bite, but just down 57th St., she spotted a small coffee shop and headed toward it.

She wasn't close enough to smell the roasting coffee or freshly baked pastries, but her imagination was working overtime, and she was certain that her sense of smell was just powerful enough to take in the wafting aromas of the cafe. But, before she halved the distance between the end of the block and the shop, her burner phone rang. Coffee and pastries would have to wait.

D33pTh0ugh1 didn't know who the caller was, but she knew why they were calling. She answered the phone with the bubbliest voice she could muster, "Hello, this is Julia Short, corporate communications. How may I help you?"

The man on the other line introduced himself as a reporter for a regional newspaper and asked to confirm the validity of a press release that had just crossed his desk. Continuing to play the part, D33pTh0ugh1 confirmed that the press release had in fact been issued by CapitalCorp that morning, but that the company was not going to be able to comment.

"We're publishing the story in the next twenty minutes, are you sure there's nothing you want to offer up?" asked the reporter.

D33pTh0ught1 stood outside the coffee shop, watching with longing as people entered. "We appreciate your call, but as I said, we're not making any additional public statements at this time, which includes Mr. James Robinson. We're currently working through the details and will be making a public statement shortly."

"We'll run this story without comment from CapitalCorp then."

"We understand. Thank you," she said and hung up the phone.

Seconds later another call came through, this time from a major newspaper, and after that another call from a TV station in Boston. D33pTh0ugh1 continued to hold court outside the coffee shop for another 30 minutes, answering calls and giving the same canned response over and over again. By the time she finished, she had spoken with 23 different reporters from all over the country representing a constellation of publications, TV networks, news blogs, and radio programs.

The phone stopped ringing at 10 AM, just as planned, and by that point, the hunger pangs were nearly unbearable. But, before venturing into the cafe, D33pTh0ugh1 promptly dismantled the burner phone by removing the battery, snapping the SIM card in half, and throwing the whole mess down a nearby sewer grate.

She walked into the crowded coffee shop, and, standing in line, eyed what looked like the most delicious cheese Danish in all of New York City. She mused to herself about how smoothly the whole operation had gone. A day earlier she had logged into the CapitalCorp VoIP[i] admin portal without a hitch, and quietly set up call forwarding across all of the corporate communications lines so all incoming calls would be sent directly to her burner phone from 9:30 AM to 10 AM. The CapitalCorp communications team would only now be noticing that they were having an unusually light morning, but it was too late for them—it was a done deed.

"Thank you, David," she muttered under her breath as she stepped up to the cash register.

////

DAVID NOTICED THAT HIS FACE was beginning to feel leathery. He had been sitting in the same beach chair for over an hour and hadn't reapplied sunblock. His

[i] VoIP, or Voice over Internet Protocol is a technology that allows for the delivery of voice communication and phone services via the Internet and other Internet Protocol networks.

children were still in the water, playing "King of the Mountain" on an inflatable trampoline anchored fifty feet off the beach. Their splashes punctuated the rhythmic sound of the lapping sea as they successively tumbled from the inner tube into the water. David reached down beside the chair and blindly fumbled through the beach bag for his phone.

"They won't be this age forever," he mumbled as he snapped a photo of his children mid-flight, thrown from the trampoline back into the sea. "What do you think?" he asked his wife, holding the phone out across the sandy threshold between their two beach chairs.

She lazily turned her head to look and returned to sunbathing without comment.

"What did you think?" David asked again, determined to get a response.

"I couldn't see it. The sun is too bright," she said, motionless.

David sat up and brought the phone as close to his eyes as he could, shielding the sun with his hand like a visor. He inspected the shot for a second. "I think it looks good. I'm going to post it."

"Awesome," she said dryly.

David opened his photo-sharing app, slapped on a nostalgic filter, and posted the picture with the hashtags *#bahamas2016* and *#collinsfamilyvacation*. He then added image number eighty-five to social media in the album "Collins Family Vacation 2016." He put the phone back into the bag and exchanged it for some sunblock, which he began reapplying to his face.

"You know, it's already too late," his wife said.

"What's too late?"

"You're sunburned. Putting sunblock on isn't going to protect you at this point. You're better off getting out of the sun."

David's phone chirped in the bag, and, soon after, chirped again.

"Can you take a look at that for me?" David asked.

"You want to see who liked your post?" she asked, looking over her sunglasses at David with a prodding smirk.

"No, those were emails. Can you take a look for me? I have sunblock on my hands."

She pulled out the phone out from the beach bag. Two new email notifications showed on the screen, one from something called "Voicenet" and another from someone named Theresa. "Who's Theresa?" she asked, tossing the phone into his lap.

"It's my boss's assistant," David said with a hint of stress in his voice. He quickly finished applying the sunblock to his face and wiped his hands clean with his towel. He swiped open the phone, leaving a greasy smudge on the glass. The subject of Theresa's email read, "[Urgent] Phone help." David read the email.

> Hello, David, I hope I'm not interrupting your vacation! Dana asked that I set up some complicated phone forwarding for her while she's attending some meetings out of town next week. I went into her Voicenet account and tried to put everything in place, but I got a message telling me that you needed to approve the changes. Did you receive an email from Voicenet about it? If yes, can you approve the changes and let me know? Thanks in advance, you're a lifesaver!

David went back to his inbox and saw the message from Voicenet unopened under Theresa's email. The subject read "[noreply] action requested." David opened the email to find a short note from Voicenet and a link to the admin portal. The notification read:

> The user Dana Mattrick has attempted to make changes to settings on their account that require administrator permissions. To review and approve these changes, please sign into the Voicenet administrator portal.

David clicked on the link and ended up on the portal login page. He entered his username and password into the respective textboxes and clicked "Login." The page took a minute to load and eventually sent him to an error page. David frowned. "The session must have timed out," he said to himself.

David held his phone up higher and hit the back button to return to the login page, hoping to catch a stronger Wi-Fi signal from the resort this time. The login page loaded again, and after re-entering his credentials, and holding the phone back up again, he pressed the "login" button. This time, the admin portal home page loaded. But, scrolling through the page, it wasn't immediately clear what he needed to do. There were no messages in his message center, and no popups providing him any information about the permissions. But, before he could make much of it, his phone chirped again, and another email arrived. It was from Theresa, and all it said was: It looks like it's working now! Thanks! David replied with No problem. He closed his phone and tossed it back into the beach bag.

His wife turned to him. "What was that about?"

PHISHING FROM AUTHORITY

What would you do if your boss, or boss's assistant, asked you to complete a task ASAP? David did what we all do: he complied. Unfortunately, in this case, Theresa did not send the email, and David became the unwitting victim of a phishing attack.

Phishing is a type of attack where a bad actor tries to extract personal information from someone through email, chat, and even over the phone by posing as a trusted person or entity. Phishing remains one of the most frequently used attack techniques by bad actors, and there are many different strategies for extracting information effectively. In this scenario, D33pTh0ugh1 chose to masquerade as one of an authority figure. In fact, emails sent from authority figures, and especially those that include urgent requests, tend to work for the attacker.[1,2,3] But why do people quickly, if not automatically, comply with requests from authority figures?

In his seminal book *Influence,* Robert Cialdini discusses how people can be influenced by those they perceive to have authority. Our deference to authority is likely conditioned, as we're all brought up to obey and defer to people who are in authority positions, (e.g. parents, teachers, etc.). The mechanism through which this **COMMAND AUTHORITY** functions is the perception of consequence – that if a request from someone in an authority position is disobeyed there might be a cost.[4] Authority, however, is not necessarily an objective characteristic. People tend to associate cues like role or job title, appearance, and assertiveness with authority. Additionally, people may overweight information conforms to their mental model of authority. Because we utilize these cues to approximate authority, those same cues can be used maliciously to provide the appearance of authority in a phishing attack.

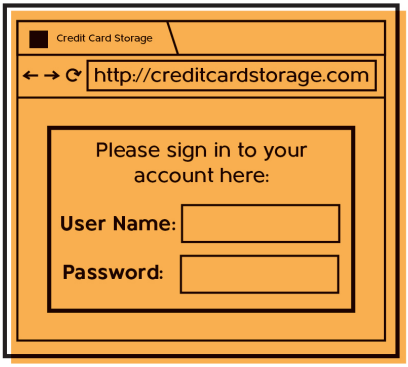
Phishing emails use corporate names and logos to build a façade of legitimacy. Information from a recognized authority can provide a valuable shortcut for deciding how to act in a given situation. One way organizations and service providers can help reduce the effectiveness of phishing attacks that use authority is to provide users with clear, up-front channels for how specific types of information will be collected or how notifications will be disseminated. These channels should not be easily spoofed by bad actors (e.g. take these communications offline, or only allow them within proprietary interfaces), but are still standard and accessible channels for end users.

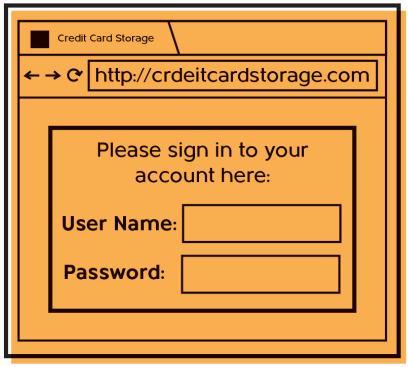
"Nothing," he said, "the boss needed a little help with her phone." He pulled the sunblock back from the bag and finished reapplying it to the rest of his body before lying back down on the chair.

"I told you, you already have a burn," said his wife.

"I know, I know," he said, rolling over in the opposite direction.

Can you spot the difference?





////

AN EMAIL FROM DAVID POPPED UP in the theresa42@mailserv.com account that D33pTh0ugh1 had created, which read No problem. D33pTh0ugh1's little phishing expedition had been successful, and she now had the information she needed to complete the attack.

A small bit of scanning through David's social media accounts showed that his vacation travel had been quite regimented for the past several years. August trips to the Bahamas had been a family staple ever since his youngest had entered elementary school as evidenced by a post in 2010 showing a picture of David on the beach holding the child up in the air. The caption on the photo that read: *First family trip to the Caribbean and Jessica can't stop talking about her upcoming first day of school! Is this my daughter?!!* After that year he had posted photo albums for each consecutive trip that they had taken over the six-year interim. Phishing David when he was

PRIMED TO SEE WHAT THEY WANT YOU TO SEE

Even though David clicked on a link that contained a typo, why wouldn't he recognize a spoof of a frequently visited page like his login screen? Surely, David would quickly notice a difference if D33pTh0ugh1 couldn't replicate the browser experience accurately, right? Not necessarily.

Research into site spoofing has shown that people often fall for well-spoofed pages because they tend to evaluate the legitimacy of websites based on the site's and the professionalism of the design, and not necessarily the page's URL.[5] What people look for when evaluating a product or an experience are **SALIENT CUES** (e.g. familiar visual interface, professional design, etc.), which may or may not provide valid information about the actual characteristics the user is trying to assess (e.g. security).[6] Moreover, the salient cues users do look for may not be the ones that would provide them with insights about the relative security or insecurity of a web page.

Additionally, D33pTh0ugh1 told David that he needed to sign into a portal, which ensured that David would direct his attention to the details of the login interface, as opposed to other visual cues. This phenomenon is an extension of visual **PRIMING**—the idea that "what we have recently seen and attended to strongly influences how we allocate visual attention."[7] In this case, David was primed to expect a familiar process (e.g. the login screen), which in turn made him less likely to pay attention to other details and to notice that he was handing his username and password to D33pTh0ugh1 on a silver platter.

To design around this problem, web developers and UX designers might build processes into browsers or email interfaces that redirect users' attention toward the "right" salient cues. For instance, before loading a link embedded in an email, the email client might prompt the user to confirm that the URL that they are traveling to is valid. An additional level of support for users who are less familiar with URLs would be to provide rules of thumb to help users better evaluate whether the URL is, in fact, safe.

least likely to be paying much attention to work seemed like a prudent strategy for D33pTh0ugh1, and what better time than during a family vacation on the beach?

D33pTh0ugh1 had to build a trap and lure David into it, which was not a simple task. Creating a convincing spoof website to capture login credentials required keen attention to detail. To be convincing, the user needed to see what they anticipated seeing, which meant ensuring the admin portal looked and felt exactly like the real one. The user interface, links, and other page attributes needed to be exact replicas, and the URL had to be familiar too. Because she couldn't use the exact URL, D33pTh0ugh1 decided that typosquatting[ii] on the admin portal URL might work. The actual portal URL was portal.voicenet.com, but by taking the original URL and switching around the placement of the 'o' and 'r' in 'portal,' she could register a new website at protal.voicenet.com, a small enough change that David was unlikely to notice. But once he entered his credentials, where would he go? It would be nearly impossible to build a fully functioning spoof of the admin portal itself with all the essential details, so she needed to figure out some other diversion that wouldn't draw suspicion. After thinking about it for a bit, she decided that she could build an error page to make it look like the connection didn't go through, and embed a link back to the real login URL so David could try to log in again and do so successfully.

Sending the emails out to prompt David to log in was a little more complicated. Masquerading as someone else is not terribly difficult over email, but it often requires finding an open outgoing mail server, which, nowadays, were few and far between. Open SMTP servers were mostly a thing of the past, as contemporary mail server software closed the SMTPs by default.[iii] However, it was still possible to sniff out occasional open SMTPs, and D33pTh0ugh1 knew a professional spammer in China through personal connections in the deep web who might be able to help. She got in contact with him, and they worked out a deal that he would let her know if one opened up during the period that David was on vacation, but that he couldn't make any promises about how long it would be open.

[ii] Typosquatting is technique designed to direct users to fake websites by intentionally inserting a typographical error that often goes unnoticed, or is likely to be typed by accident. Here, D33pTh0ugh1 leveraged the insight that humans can genreally raed wodrs taht jubmle the cotnents betewen the frist and lsat lettres.

[iii] SMTP (Simple Mail Transfer Protocol) is the method used by email servers to send our emails. By Open SMTP server, the hacker is referring to open mail relay servers which are configured to allow anyone to send emails through them. In contrast, closed mail relay servers only send and receive emails from known users. In the early days of the Internet, SMTP servers were open relays, but today, most are closed to avoid exploitation from spammers and worms.

INSECURITY BY DEFAULT

David unwittingly made the attack a little bit easier with his social media habits. Posts and even entire photo albums of his family were visible to the public. Why didn't David switch his privacy settings? One reason is that users sometimes have incorrect mental models about the default level of security and privacy they may have when using a service like a social networking site[8] or an Internet-connected product. When incorrect, mental models about the security defaults can be especially problematic because defaults are very sticky.

To illustrate how defaults work, consider retirement savings. Policymakers and employers observed that they could increase retirement savings by changing the default. Originally, employees had to opt-in to their company's 401(k) plans, but relatively few people did so. By changing the default from opt-in to an opt-out, not only did enrollment rates in 401(k) increase significantly, but the default contribution rates had a strong impact on savings.[9]

Defaults can be a powerful tool for swaying behavior both positively and negatively, and this is no less true when it comes to cybersecurity. One example of this is a recent distributed denial-of-service (DDoS) attack on the DNS provider Dyn, which caused massive outages and network congestion for many websites. The Dyn attack was executed using Mirai malware, which turned millions of Internet of things (IoT) devices (many of which were WI-FI enabled cameras) into a botnet that spanned the globe. Attackers recognized that many of the various IoT devices were still password protected with the default passwords that had been set by the manufacturer—they had never been reset by the users—making them easy to compromise.[10] Had the manufacturer automatically required users to reset the passwords as soon as the device was turned on or provided a random password for each separate device instead of a standardized default, this kind of event may have been avoided.

Default security settings are powerful because people are unlikely to change them. Organizations need to determine whether opt-in policies are reasonable when it comes to security, fully taking into account how people actually act. Instead, service providers and device manufacturers could make lower levels of security and privacy an opt-out decision from the beginning. Or, if opting out isn't feasible, service providers and device manufacturers could force consumers to make a considered decision about their security preferences during their first-time experiences through thoughtful UX design.

"These things close up almost as soon as they open," he said, "so you're going to have to take the opportunity when it comes."

She was all right with that and set to work crafting the emails. The one from Voicenet needed to look automatic but also provide a sense of urgency. To focus David's attention on following through on the Voicenet email instructions, and reduce the likelihood that he'd scrutinize the email too much, she decided to craft another email to contextualize the request. D33pTh0ugh1 decided that sending it from CapitalCorp's CISO herself might seem a bit odd – what executives make those sorts of requests for themselves? – so instead, she pretended to be the CISO's executive assistant, Theresa, which turned out to be an effective decision.

In the middle of August, when David was on vacation, the message came through from her Chinese contact that there was an open SMTP that she'd be able to co-op for her attack, and she immediately set to work.

"Have at it," he said, and so she did.

////

THE BARISTA RETURNED TO the cash register with a cup of coffee in a to-go cup and a cheese Danish in a small wax paper bag.

"Best cheese Danish in all the city," said the barista, "I promise."

D33pTh0ugh1 rummaged through her purse and happily handed over the eight dollars and change she owed, leaving some extra in the tip cup on her way out. Once back on the street she took a small sip of coffee, which burned as it hit her tongue, and then a bite of the Danish to try to soothe the already numb taste buds. Despite the coffee being too hot, both it and the Danish were delicious, even more so than she had anticipated – *the barista was right,* she thought.

For a moment, looking out at the taxi cabs and morning commuters, she felt calm. All of her well-laid plans had unfolded, and now all she had to do was to wait and watch to see where the whole thing would land. In a sense, despite the work that she put in, it wasn't that hard. The systems that she compromised, the passwords she collected, the trickery she had played on the reporters, all of her successes came down to the fact that people are predictable. But, there was a poignancy in it. All of these people, going about their happy little lives, had no idea how close they were to making a misstep and becoming a victim; they had no idea how, for

a person like D33pTh0ugh1, they were all like wounded animals on the savannah, completely unaware of their limp.

"If they were more like Spock," she thought to herself, "then I'd be out of the job." But part of her wished that she could be out of the job.

She took another bite of her cheese Danish, and her personal phone rang in her purse. She took the phone out, looked at the number, and answered.

"We're going to be starting soon," said the voice on the other line.

"I'm heading in now," D33pTh0ugh1 replied. "I should be there in 15 minutes." She hung up, put the phone back in her bag, and began her walk downtown.

---

[1] Ferguson, A. J. (2005). Fostering e-mail security awareness: The West Point carronade. *Educase Quarterly, 28*(1), 54-57.

[2] Workman, M. (2008). Wisecrackers: A theory-grounded investigation of phishing and pretext social engineering threats to information security. *Journal of the American Society for Information Science and Technology, 59*(4), 662-674.

[3] Butavicius, M., Parsons, K., Pattinson, M., & McCormac, A. (2016). Breaching the Human Firewall: Social engineering in Phishing and Spear-Phishing Emails. *arXiv preprint arXiv:1606.00887.*

[4] Cialdini, R. B. (2007). *Influence: The psychology of persuasion.* New York: Collins.

[5] Dhamija, R., Tygar, J. D., & Hearst, M. (2006, April). Why phishing works. *In Proceedings of the SIGCHI Conference on Human Factors in computing systems* (pp. 581-590). ACM.

[6] Tom, G., Barnett, T., Lew, W., Selmants, J., (1987) "Cueing the consumer: The role of salient cues in consumer perception," Journal of consumer marketing, vol. 4 iss: 2, pp.23 - 27

[7] Kristjánsson, Á., & Campana, G. (2010). Where perception meets memory: A review of repetition priming in visual search tasks. *Attention, Perception, & Psychophysics, 72*(1), 5-18.

[8] Lin, J., Amini, S., Hong, J. I., Sadeh, N., Lindqvist, J., & Zhang, J. (2012, September). Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 501-510). ACM.

[9] Madrian, B. C., & Shea, D. F. (2001). The power of suggestion: Inertia in 401 (k) participation and savings behavior. *The Quarterly Journal of Economics, 116*(4), 1149-1187.

[10] Brian Krebs (October 21, 2016) Hacked Cameras, DVRs Powered Today's Massive Internet Outage.

[11] Krebs on Security. Accessed from https://krebsonsecurity.com/2016/10/hacked-cameras-dvrs-powered-todays-massive-internet-outage/ on October 31, 2016

// APPENDIX

BEHAVIORAL CHALLENGES IN CYBERSECURITY

END USER SECURITY SETTINGS

The Problem

Operating systems, popular web-based services, including social media sites like Facebook, and some IoT hardware offer users the opportunity to set and modify settings that can impact users' privacy and security. The design of these settings interfaces (by software or hardware companies) and their management by end users have profound implications for the security of the user's personal information, and in turn, any enterprise of which the user is a member.

Hackers can use information on social media to take a better guess at passwords or to set a personalized spear-phishing email. Thieves can see vacation dates and know when to rob a home or business. Bad actors can leverage insecurities in Wi-Fi enabled devices to construct intricate botnets or spy on unsuspecting individuals.

To keep users safe, end users need to understand, maintain and periodically update security and privacy settings. While the availability of security and privacy settings may well give each user the freedom to 'control' her personal information or access to hardware, in practice, users may adopt (either consciously or unconsciously) settings that are less secure than they intend or would prefer. For instance, in one study of Facebook users, researchers reported a gap between privacy intentions and actual behaviors (settings and sharing) for every study participant.[1]

Simple statement of the behavior we want to change

Users do not change default settings, and rarely review or alter settings that affect the privacy and security of their personal information and the devices they use. We want users to be aware of and attend to their personal security settings across all devices and services they use.

Behavioral Insights

>>USERS TEND TO KEEP CURRENT OR DEFAULT SETTINGS. All people tend to have an emotional preference for the way things are, known as *status quo bias*. Status quo bias is one reason users are not likely to shift away from default settings. This reluctance to change the setting is agnostic to the settings themselves, whether designers and developers design those settings for security or to facilitate openness. This user behavior highlights the importance of *defaults.* If the settings are less-than-secure by default when users first begin using a product, users are not likely to change the privacy or security configuration. How software providers set the defaults has a powerful influence on the overall security of user data and the hardware they use. The idea of designing a default is at the core of one of the most famous applications of behavioral science: defaulting employees into retirement plans and pre-setting contribution escalations.

>>THE DESIGN OF SETTINGS OPTIONS MATTERS. Users need to be able to find, navigate to, and then understand settings options. When users can't find the settings or when users don't clearly understand the settings options, they are less likely to make changes. Security researchers have shown that while users have a general, high-level understanding of the importance of privacy settings, they view adjusting the privacy settings not as an essential step to protect against hacks, but as a type of cost for using the 'free' service. Through this lens, users have a limited amount of 'goodwill' for trying to figure what the appropriate privacy settings should be, and there is a drain on that goodwill when the settings options are difficult to understand.[2] One explanation offered by behavioral science for the evaporation of user goodwill is the idea of *choice overload.*[3] Studies on shoppers in the market for consumer products have shown that even if a customer states a preference for maximum freedom of choice when that customer has myriad options presented to them, she becomes demotivated and is less likely to purchase a product at all.[4] In the context of modifying settings, if choosing the right settings is hard, users may avoid taking action.

Additionally, a user's engagement (or lack thereof) with the security settings during their first user experience and the visual layout of those options influences whether users will be attentive to settings in the first place, and which op-

tions users are more likely to select. The manner in which options are presented and arranged is sometimes referred to as *choice architecture.* Limiting the number of settings and options a user must choose from is one way a designer may alter the choice architecture in order to avoid choice overload.

It's safe to assume that popular, well-capitalized platforms, especially social media, have invested heavily in defining both the security settings, the options available within each setting, and studied how changes in choice architecture influence users' settings choices. This type of data and continued research is a significant step in learning how best to design small nudges that help users make secure settings choices.

>>INSECURE DEFAULTS MAKE UNSAFE BEHAVIOR TOO EASY. An additional aspect of settings design is the way insecure defaults can unwittingly promote insecure user behavior. The primary example is an operating system default to join an open Wi-Fi network, a behavior that is known to risk personal data. The default to an open network is what psychologists call a *channel factor* or something that makes it easier for someone to maintain their current intentions. A user wants to get online; the default to connect to an open Wi-Fi network makes following through on that intention easy. As a general rule of thumb, settings should be designed to make secure behavior easier and insecure behavior harder, not the other way around.

>>USERS UNDERESTIMATE THE RISKS OF SHARING PERSONAL INFORMATION. Having a mental model or feeling of *'who am I? No one wants my data'* is reflective of a tendency towards overconfidence. Users are assuming the probability of not being hacked is considerably greater that it actually is. Two key contextual features likely lead to overconfidence in these instances. First, users often underestimate how much information they intentionally share online and are often unaware of how much data is available to be captured through their internet-enabled devices.[5] Second, users don't realize how the data they share online over time and on different software platforms can be aggregated by hackers to gather a fairly robust portrait of who the user is offline.

Design Concepts

1 **Force choice around secure defaults.** Flipping the default from insecure to totally secure could have a significant positive effect on user security, but may not be feasible, or even preferable for the user in all circumstances (e.g. it could hinder the usability of a service or a device to an unnecessary or unfavorable degree). Instead, online service providers, software developers, and device manufacturers should start with stringent default settings, and then force users upon the first interaction with the product or service to set the security settings to their own preferences. By doing so, product and service providers can avoid users' status quo bias, and provide a moment of action for the user to think critically about their security preferences.

2 **Standardize privacy measures and settings across services.** In the world of food, whether it's cereal or chips, the nutritional information on a package label is formatted identically, making key measures – calories, saturated fat – easy to find. Like foods, software services vary widely, but there are likely a few measures, such as the total number of people who can see a post shared on social media, which could be easily developed and standardized across services.

3 **Provide users feedback on overall privacy and security.** While software or services may provide users with many settings, each of which is modifiable in different ways, the overall security of a user's account may not be salient to them. A single salient metric could take into account information from privacy settings, password strength, and the user's behavior (e.g. logging on from an open Wi-Fi network) and give users meaningful feedback on the security of their account. Additionally, service providers could also give users actionable pointers about how they can remediate insecurities when they arise, giving users an opportunity to improve their overall security score.

4 **Leverage social norms.** Social media platforms that store significant amounts of personal information and facilitate social networks for users can take advantage of the important signaling function of social norms by showing a user the relative security of their peers, or information about their most secure peers. For instance, when sending users notifications about updating their security preferences, service providers could include information

about the number of close contacts who have recently updated their security preferences.[6] Additionally, once users look at their security preferences, service providers could give users information about the number of close contacts that utilized specific security features. This intervention could also include the concept of providing clear feedback, as described above, by using a standardized metric of comparison across users.

PHISHING

The Problem

In 2016, users experienced the highest number of phishing attacks ever recorded. Over 1.2 million attacks were registered in 2016 by the Anti-Phishing Working Group, a global coalition of law enforcement agencies, representing a 65 percent increase in registered attacks over the previous year.[7] Awareness does not appear to be the deciding factor, as users still click on malicious links and downloads despite knowing the risks.[8] In fact, many of the most sophisticated and damaging cyber attacks begin with a well-executed spear-phishing attack. Nearly two-thirds of IT decision makers say the that spear-phishing is their top concerns,[9] and in testimony to Congress, the CEO of Fire Eye stated that, since 2015, 95% of the breaches they've remediated began with a spear phishing email.[10]

Simple statement of the behavior we want to change

Users click on malicious links in emails that spoof or mimic banks, technology companies, coworkers, or any social/professional affiliation of the user. The link itself may initiate installation of malware, may lead the user to a fake (but familiar looking web page) to capture the user's credentials, or the user may unwittingly reveal information by corresponding with the sender. We want users to avoid clicking on malicious links sent via phishing attacks.

Behavioral Insights

>>PEOPLE COMPLY WITH REQUESTS FROM AUTHORITY FIGURES. When individuals with authority make requests, be it in person, via email, over the phone, or through any other medium, people have a tendency to comply.[11] Bad actors perpetrating phishing attacks use this insight to get their unwitting victims to disclose information or download malware onto their computer by masquerading as a person of authority, such as a supervisor, professor, doctor, or another figure with perceived influence. By just using an authoritative title, phishing attacks can trigger a quick-acting *heuristic*, or mental shortcut for deciding how to act, which causes people to equate a request from a person of authority as something with which they should comply.[12]

>>PHISHING PLAYS ON FAMILIARITY. Familiar people, experiences, and institutions can engender feelings of trust in individuals.[13] However, in the virtual world, it is very easy for bad actors to copy the visual and experiential cues that individuals find familiar, such as corporate logos, web pages, and the names of friends and family. By presenting familiar cues to the user, bad actors can build a façade of legitimacy, and lead users to do things they shouldn't do such as download malware or disclose personal information.

>>PHISHING EMAILS PRESSURE USERS TO ACT QUICKLY. Phishing emails are often crafted to create a sense of urgency for the targeted user. By using trigger words such as "alert," "urgent," or requesting that the user responds or completes a task "ASAP," attackers can prompt users to think and act too quickly, making it less likely that they'll notice that they're falling into a trap.[14] Part of the reason creating a sense of urgency might be effective is because people are loss averse, and will do what they can to avoid losses where possible.[15] If people perceive that they'll lose something if they don't act quickly, they may be more prone to act without thinking.

>>PHISHING EMAILS AND SPOOFS TAKE ADVANTAGE OF OUR LIMITED ATTENTION. Phishing emails and spoofed web pages almost always contain information that can indicate to the user that the email, attachment or web page is malicious (e.g. pixelated images, slightly different URLs, etc.). However, users may not always be attentive to those details because of *limited attention.*[16] Attention, much like a limited resource, gets depleted when in use. For instance, if a user directs their attention to some aspect of a user interface, they will have less attention to direct to other details. Additionally, bad actors executing phishing attacks can *prime* their victims to be attentive to specific details, while simultaneously directing their attention away from cues that would signal that the email, website or attachment may be malicious. For instance, bad actors might send an email asking someone to log into their account, priming the victim to be more focused on the login interface than other cues of malicious intent such as URLs, or pixelated graphics.

>>PHISHING EMAILS EXPLOIT OUR CURIOSITY. Pictures from a party, a sample contract from a competitor, celebrity gossip – sometimes the desire to look obscures a user's ability to weigh the likelihood that the email may be a phishing attack. In at least one study, researchers triggered a person's curi-

osity by using traditional 'reverse psychology,' suggesting that the recipients received a message in error and should not click on a link to an external photo-hosting website. In another example, researchers drafted phishing emails that appeared to present recipients with a personalized opportunity, such as a journalist wishing to write about the recipient's work with a convenient link to the reporter's previous writing.[17] By exploiting peoples' desire to close the *curiosity gap,* bad actors can manipulate users into clicking on links and downloading files that they shouldn't.

Design Concepts

1 **Provide real-time feedback.** 'Just-in-time teaching' can help users connect actions to consequences, eventually pausing the 'fast thinking' that characterizes so much email behavior. Researchers have already shown how real-time feedback and just-in-time training can be effective at teaching users how to identify and avoid phishing attacks and website spoofs in real-world environments.[18,19] Organizations interested in reducing phishing rates should consider adopting these sorts of tools across their enterprises.

2 **Slow user reactions.** To make users more attentive to the little cues and details that characterize phishing attacks, UX designers could build interfaces to help users 'slow' their thinking. While slowing down users may be in conflict with the productivity goals of organizations, but it may be a necessary step in improving enterprise security. One way to accomplish this might be to embed small hassles into the email user experience. For instance, when clicking on a link or file within or attached to an email, the user could be prompted, via a pop-up, to consider whether the link or attachment is from a trusted source. If the user is unsure, an available call to action could be used to quickly and easily send a confirmatory email back to the sender. Slowing down the user in such a way could improve their identification of malicious emails.

3 **Reward savvy behavior.** Recognize employees who pass sophisticated phishing tests or catch an actual spear-phish with public recognition or financial incentives. While pure incentives are not inherently behavioral, well-constructed incentive programs can have the added effect of getting users to be more attentive to the details of emails, making it more likely users would catch potential phishing attacks before they occur.

4 **Adjust cultural norms through rules of thumb.** Develop organizational policies that disallow sharing of links or attachments through email to avoid any ambiguity when a potentially malicious link or attachment shows up. Instead, provide employees with new platforms and rules of thumb about how to send links and attachments to colleagues through other enterprise services. Additionally, if and when a link or attachment appears in an email in-box sent from a fellow employee, establish heuristics that guide employees to ask the sender about whether they had sent the link or email intentionally. Simply asking, "Hey, did you send me this?" can be the difference between a successful attack and one avoided.

5 **Add more information to URL address bars.** By mixing up colors and mixing in words and padlocks, a web browser can purposely recapture a user's attention and focus, thus increasing the likelihood that the user will spot a spoofed URL.

[1] Madejski, M., Johson, M. and Bellovin, S. A Study of Privacy Settings Errors in an Online Social Network. Retrieved from https://www.cs.columbia.edu/~smb/papers/fb-violations-sesoc.pdf

[2] Kirlappos, I., Sasse, M. A. (2012). Security Education against Phishing: A Modest Proposal for a Major Rethink. *IEEE Security and Privacy Magazine 10*(2), 24-32

[3] Iyengar, S.S.; Lepper, M.R. (2000). "When choice is demotivating: can one desire too much of a good thing?". *Journal of Personality and Social Psychology. 79:* 995–1006

[4] Schwartz, B. (2004, January). *The paradox of choice: Why more is less.* New York: Ecco.

[5] Debatin, B., Lovejoy, J., Horn, A., Hughes, B. (2009). Facebook and Online Privacy: Attitudes, Behaviors and Unintended Consequences. *Journal of Computer-Mediated Communication. 15*(1), 83-108.

[6] Das, S., Kramer, A. D., Dabbish, L. A., & Hong, J. I. (2014, November). Increasing security sensitivity with social proof: A large-scale experimental confirmation. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security* (pp. 739-749). ACM.

[7] APWG (Feb, 2017). Phishing Activity Trends Report 4th QR 2016. APWG. Accessed from: http://www.antiphishing.org/resources/apwg-reports/apwg_trends_report_q4_2016.pdf on March 3, 2017

[8] Benenson, Z., Gassmann, F., & Landwirth, R. (2016) Exploiting curiosity and context: How to make people click on a dangerous link despite their security awareness. Accessed from: http://paper.seebug.org/papers/Security%20Conf/Blackhat/2016/us-16-Benenson-Exploiting-Curiosity-And-Context-How-To-Make-People-Click-On-A-Dangerous-Link-Despite-Their-Security-Awareness-wp.pdf

[9] Cloudmark (2016) Survey Revelas Spear Phishing as Top Security Concern to Enterprises. *Cloudmark Security Blog*. Accessed from: https://blog.cloudmark.com/2016/01/13/survey-spear-phishing-a-top-security-concern-to-enterprises/ on October 16, 2016

[10] Mandia, K. Testimony before the U.S. Senate Select Committee on Intelligence. 30 March 2017. Available at: https://www.intelligence.senate.gov/hearings/open-hearing-intelligence-matters-1. Accessed 3/30/2017.

[11] Milgram, S. (1963). Behavioral Study of obedience. *The Journal of abnormal and social psychology, 67*(4), 371.

[12] Cialdini, R. B. (2007). *Influence: The psychology of persuasion.* New York: Collins.

[13] Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of personality and social psychology, 9*(2p2), 1.

[14] Vishwanath, A., Herath, T., Chen, R., Wang, J., & Rao, H. R. (2011). Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems, 51*(3), 576-586.

[15] Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The journal of economic perspectives, 5*(1), 193-206.

[16] Dhamija, R., Tygar, J. D., & Hearst, M. (2006, April). Why phishing works. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 581-590). ACM.

[17] Benenson, Z. Exploiting Curiosity and Context: How to make people click on a dangerous link despite their security awareness. Retrieved from https://www.blackhat.com/docs/us-16/materials/us-16-Benenson-Exploiting-Curiosity-And-Context-How-To-Make-People-Click-On-A-Dangerous-Link-Despite-Their-Security-Awareness.pdf

[18] Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2010). Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology (TOIT), 10*(2), 7.

[19] Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., Blair, M. A., & Pham, T. (2009, July). School of phish: a real-world evaluation of anti-phishing training. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (p. 3). ACM.