Compliance Cautions: Investigating Security Issues Associated with U.S. Digital-Security Standards

Rock Stevens*, Josiah Dykstra†, Wendy Knox Everette‡, James Chapman†,
Garrett Bladow§, Alexander Farmer†, Kevin Halliday*, and Michelle L. Mazurek*
*University of Maryland, College Park; †Independent Security Researcher; ‡Leviathan Security Group, Inc.; §Dragos, Inc.
{rstevens,khalliday,mmazurek}@cs.umd.edu, josiahdykstra@acm.org,
jchapman39@gatech.edu, wknox@wellesley.edu, gbladow@dragos.com

***Abstract*—Digital security compliance programs and policies serve as powerful tools for protecting organizations' intellectual property, sensitive resources, customers, and employees through mandated security controls. Organizations place a significant emphasis on compliance and often conflate high compliance audit scores with strong security; however, no compliance standard has been systemically evaluated for security concerns that may exist even within fully-compliant organizations. In this study, we describe our approach for auditing three exemplar compliance standards that affect nearly every person within the United States: standards for federal tax information, credit card transactions, and the electric grid. We partner with organizations that use these standards to validate our findings within enterprise environments and provide first-hand narratives describing impact.**

We find that when compliance standards are used literally as checklists — a common occurrence, as confirmed by compliance experts — their technical controls and processes are not always sufficient. Security concerns can exist even with perfect compliance. We identified 148 issues of varying severity across three standards; our expert partners assessed 49 of these issues and validated that 36 were present in their own environments and 10 could plausibly occur elsewhere. We also discovered that no clearly-defined process exists for reporting security concerns associated with compliance standards; we report on our varying levels of success in responsibly disclosing our findings and influencing revisions to the affected standards. Overall, our results suggest that auditing compliance standards can provide valuable benefits to the security posture of compliant organizations.

I. Introduction

Many digital-security guidelines, such as those provided by the National Institute of Standards and Technology (NIST), present best practices for system owners and digital-security technicians to improve their overall security posture [43]. These guidelines are designed to protect intellectual property, sensitive resources, customers, and employees from security risks. Exemplar protection mechanisms include installing anti-virus applications on all systems and conducting background checks on employees before providing privileged access.

Over the years, governments and organizations have elected to adopt these guidelines as compliance controls: mandatory policy and technical controls that must be enforced across applicable organizations. Non-compliance with these controls is typically followed by significant fines, revocation of access, or employment termination [6]. To illustrate this fact, one energy company was recently fined $10 million for non-compliance [40].

Because of these sometimes-hefty punishments, organizations often commit significant personnel, time, and other resources to maintaining compliance with standards and preparing for compliance audits. For example, one organization we partnered with for this study has allocated 10% of their total workforce to focus solely on compliance. A cursory search in July 2019 showed job openings for compliance auditors across many Fortune 500 companies, with salaries ranging from $46,000 to $96,000 annually based on experience [18]. This indicates the emphasis that many companies place on adherence to compliance standards.

Further, compliance standards are often presented as a proven metric for improving security. The International Organization for Standardization routinely provides metrics on how compliance standards keep users and businesses safe online [29]. Some federal-level programs and businesses develop and deploy systems that are fully compliant with established standards as an implicit seal of security [19], [3], and some organizations actively use digital compliance standards to shape their defensive strategies [31], [52]. Because compliance itself is treated as a first-class security property (with potential financial penalties), standards are often used as checklists, even if they were not written with word-literal interpretation in mind.

Despite the significant emphasis placed on compliance with these standards, their actual efficacy is not well understood. While they may provide important security benefits, it is also possible that they lull security practitioners into a false sense of security, conflating high compliance audit scores with strong security. It is also possible that standards which are useful as general guidelines can become problematic when interpreted legalistically as checklist requirements. In this paper, we report on a two-part study investigating these questions.

First, we conducted a line-by-line audit of three publicly available, widely-adopted compliance standards that affect anyone in the United States who pays federal taxes, conducts credit card transactions, or uses electricity: Internal Revenue Service (IRS) Publication 1075 (P1075), the Payment Card Industry Data Security Standard (PCI DSS), and the North American Electric Reliability Corporation (NERC) Critical Infrastructure Protection for system security management (CIP 007-6). We applied rigorous content-analysis techniques adapted from social science to identify security concerns and categorize them based on root cause and estimated risk levels. In the context of this paper, we define security concerns as *any*

Network and Distributed Systems Security (NDSS) Symposium 2020
23-26 February 2020, San Diego, CA, USA
ISBN 1-891562-61-4
https://dx.doi.org/10.14722/ndss.2020.24003
www.ndss-symposium.org

security control or policy within a compliance standard that can lead to sub-optimal security conditions when implemented as written. We then partnered with experts who confirmed (or rejected) a subset of our results based on their past experience: two experts from New York City Cyber Command (for P1075), a CIP framework co-author, and a PCI standards council member.

Despite organizations such as NIST insisting that compliance programs were never intended to be used as audit checklists, all four of our compliance experts reported first-hand experience with auditors using compliance documents as line-by-line checklists, supporting our decision to treat them as such for analysis purposes [38]. We identified 148 security concerns across the three standards that our researchers assessed would exist when organizations follow compliance programs "by-the-letter."

These security concerns range in risk (assessed based on probability of occurrence and associated severity) from low to extremely high and include issues relating to vague requirements, outdated technology, and improperly protecting sensitive information. Some security concerns could potentially be addressed with straightforward rewrites of the standards and minor changes at compliant organizations, while others likely cannot be remediated without significant, potentially impractical, investment by affected organizations.

The compliance experts validated our findings, confirming 36 of 49 as definite security concerns and 10 as plausible, while rejecting only three. Further, compliance experts confirmed that problems like poorly defined time windows and unclear division of responsibility — trends we observed across the three standards we examined — can manifest in real-world ways that increase risks.

In Section VII, we discuss our efforts to responsibly disclose our findings. Our experience reveals that no viable process for reporting exists. Despite this, our findings have already resulted in one change in PCI DSS standards and are being included in discussion for further updates to both PCI DSS and CIP.

To our knowledge, we are the first to systematically assess multiple compliance standards for insecure practices and identify a range of associated security concerns that may occur within compliant enterprise organizations. Our results highlight the difficulty of establishing standards that are responsive to the fast-moving security space, general enough to apply in multiple contexts, and robust enough to use as line-by-line checklists for compliance auditing. Accordingly, we provide recommendations for improving compliance standards and the overall compliance ecosystem.

II. Background

Digital security compliance programs within the United States date back to the Computer Security Act of 1987, which required agencies to protect sensitive systems and conduct security training [39]. Many programs implement a "carrot-and-stick" approach to compliance, in that organizations are rewarded for successful programs and levied with sanctions for compliance deviations. In this section, we briefly review past studies involving digital security compliance and its impact on organizations.

Compliance audits force organizations to balance being "inspection ready" and sustaining daily operations, such as providing essential services or selling goods. Because of this careful balance, many organizations choose to perform compliance actions only before a pending audit, and then neglect further security maintenance until another audit requires them to repeat the process [46]. This behavior meets the security minimums for compliance standards, but fails to adhere to the spirit of secure practices. Moreover, evidence shows that fully-compliant organizations can still suffer data breaches. Auditors certified Target as PCI-compliant in September 2013, just before it suffered a massive data breach in November 2013 [46]. We highlight sections of compliance standards that may permit similar incidents to occur again and provide recommendations for mitigation.

Previous studies highlight cultural disconnects between developers, engineers, and compliance officials that create issues when digital security measures are "bolted on" after software development is complete [12], [7]. To combat these issues, entities must find ways to overcome organizational behaviors and factors that affect secure software development [56]. Some organizations have embedded compliance experts within development teams to encourage grass-roots-style compliance integration [12]. Other organizations found that threat modeling could proactively identify security gaps that may exist in compliant solutions [12], [4]. Some organizations have even overhauled their physical network topology to meet federally-mandated requirements, restructuring their teams and network architecture to limit the scope of auditable systems within their environment [26]. This, too, meets the letter of compliance requirements while seeming to contradict the intended goals. In this study, we identify several unintended security implications within technical controls and implementation processes that could affect organizations as they alter their normal business practices for compliance adherence.

Numerous studies focus on how humans perceive compliance standards and modify their behaviors based on those perceptions. Julisch highlighted numerous factors that shape organizational decision-making when investing in compliance measures, often seeking new security technologies that are out-of-the-box compliance ready [31]. Beautement et al. describe the "compliance budget," the human factors behind the implementation of compliance controls; their research illuminated ways to improve security and compliance readiness through resource allocation optimization [8]. Building upon previous works, Puhakainen and Siponen found that training employees to better understand compliance standards can improve organizational behaviors and shift employees toward implementing more secure practices [48]. Additionally, Hu et al. found that managers who "lead by example" and implement top-down management initiatives encourage employees' compliant security behaviors [24]. Our study is a significant departure from previous studies, as we do not focus on improving adoption rates within organizations. Instead, in this study we assume organizations are 100% compliant with the letter of the standard and focus on the insecure practices and security concerns that may exist anyway.

ID	Employment[1]	Role[2]	Org Size	IT Exp (yrs)	Edu[3]	Docs[4]
R1	A, G	M, R	500	18	MS	I,P,N
R2	G	M, R	10k+	16	PhD	I,P
R3	A, G*, I	M, R	100	20	BS	I,N
R4	I	M, R	35	15	JD	I,P
R5	A, G*, I	M, D	100	8	BS	I,N
R6	G	M, D	100	5	BS	I,N
E1	G, I	M	150	10	BS	I
E2	G	M	150	15	MS	I
E3	G*, I	M, D	1k	18	MS	P
E4	A, G*, I	R	5k	20	MS	N

[1] A: Academia, G: Government, I: Industry, *: Previous experience
[2] M: Management, R: Research, D: Development
[3] BS: Bachelor's, MS: Master's, PhD: Doctorate, JD: Juris Doctorate
[4] I: IRS P1075, P: PCI DSS, N: NERC CIP

TABLE I: Researcher and expert demographics

III. Method

In the first step of this study, our researchers comprehensively audited three compliance standards to identify potential security concerns. To validate these concerns, we then recruited four experts to provide their assessment of our findings. We performed quantitative and qualitative analysis on expert responses to identify discrepancies and also derive additional context for applicability within enterprise environments.

This study occurred from October 2017 through September 2018 and was ruled not human subjects research by our ethics-compliance office, as we communicated with experts in their professional capacity and did not collect personally identifiable information. Due to the sensitive nature of unmitigated data vulnerabilities within real environments, we generalize many of our findings to protect networks and systems.

A. Compliance-standard audit

Our team of six researchers designed the audit to systematically evaluate three unrelated compliance standards in a repeatable manner. Each researcher audited a subset of the standards, with at least three researchers per standard (as shown in Table I). Our objective was to identify issues that might negatively affect digital security, including policies that expose sensitive information and processes that create issues due to ambiguous implementation guidance.

First, all six researchers conducted a complete audit of IRS Publication 1075, following a content-analysis process drawn from social-science research. Each researcher independently examined each line of the standard. At each of several predetermined milestones within the document (e.g., the end of a section), the researcher would log their findings, including the section title where the issue was found, the exact phrase deemed problematic, a short description of the perceived issue, and references to related, publicly known issues. If a researcher found multiple issues within one phrase or section, they logged each separately. For every logged issue, all other researchers would indicate (1) if they found the same issue independently and (2) whether they concurred with the finding. If there was not unanimous consensus on an issue, we discarded it but maintained a record of the disagreement.

We then calculated the inter-coder reliability — a measure of consistency among independent auditors — for independently discovering issues in IRS P1075. We calculated our Krippendorff's Alpha (α), which accounts for chance agreements [21]. We obtained reliability $\alpha = 0.815$ for P1075; an α value above 0.8 indicates high reliability [33], [34]. Having developed a reliable auditing process, we divided into subgroups to parallelize the remaining effort. Four researchers audited NERC CIP 007-6 and three researchers audited PCI DSS. One researcher (R1) audited all three guidelines. The subgroups attained $\alpha = 0.801$ and 0.797 respectively.

We further analyzed the identified issues using iterative open coding, a process for creating and applying category labels (known as a *codebook*) to data [53]. In particular, the researchers who audited each standard coded each identified issue in that standard for perceived root cause, probability of occurrence, and severity. We resolved all disagreements among coders and developed a stable codebook by establishing a unanimously agreed-upon definition for coded terms, adapting many terms from the Composite Risk Management (CRM) framework [61] and the Information System Risk-based Assessment framework [15].

Our final codebook described four root causes for security concerns. A *data vulnerability* is an issue that will result in a data breach or compromise of sensitive information. An *unenforceable security control* cannot be enforced as written; these controls should be reworded or removed from the compliance standard. An *under-defined process* is an issue explicitly missing instructions or details that are required for a secure implementation, resulting in security gaps. An *ambiguous specification*, in contrast, is vague or ambiguous about some implementation details, such that different readers could interpret it differently. Some interpretations could potentially result in either an inappropriate action or inaction. Throughout Sections IV-B, V-B, and VI-B, we describe our audit findings using these root causes.

We used the following terms and definitions for probability: *frequent* occurs often and is continuously experienced; *likely* occurs several times; *occasional* occurs sporadically; *seldom* is unlikely, but could occur at some time; and *unlikely* we assume it will not occur. We used the following terms for severity: *catastrophic* results in complete system loss, full data breach, or the corruption of all data; *critical* results in major system damage, significant data breach, or corruption of sensitive data; *moderate* results in minor system damage or partial data breach; and *negligible* results in minor system impairment. Using a risk assessment matrix adopted from the CRM framework (Figure 1), we then calculated each issue's risk level — a function of probability and severity — as extremely high, high, moderate, or low [61].

Best practices suggest that empirical research should be conducted by personnel with extensive domain knowledge [47]. Accordingly, the auditing researchers possess an average of 14.3 years of digital security experience within academia, the federal government, and industry. Each researcher grounded their audit findings in their past digital security experiences. Additional information about the data set is in Appendix A.

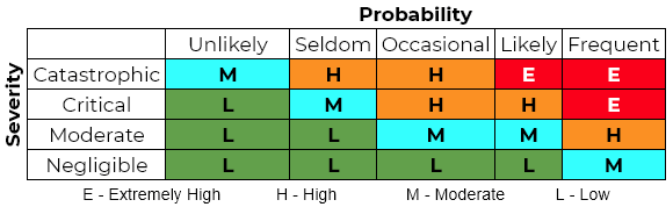

Fig. 1: Security concern risk levels. Levels were assigned based on a Composite Risk Management risk-assessment matrix that includes both probability of occurrence and impact severity.

B. Expert validation process

To obtain external validation of our findings, we established partnerships with real-world organizations and compliance subject-matter experts to confirm or reject our findings. We asked the experts to classify our identified issues in one of three categories: confirmed, plausible, or rejected. A confirmed issue indicates that the expert has previously observed security concerns associated with the issue or that observable consequences from the issue actively exist within an enterprise environment. A plausible issue occurs when the expert has not personally observed security concerns related to the issue but agrees such security concerns could manifest within other organizations. A rejected finding indicates that there is no observable evidence of security concerns related to the issue within a live environment, or that there are related security factors that we had not considered.

We used a series of closed- and open-ended survey questions to elicit information from each expert (detailed in Appendix B). In addition to directly validating or rejecting each issue, the experts were asked to provide additional context from their personal experience. We presented the issues to the experts in a randomized order, providing the referenced section title, exact text from the section, and a short narrative describing the perceived issue.

After collecting data from each expert and removing rejected findings, we used the Wilcoxon signed-rank test to compare researchers' assessment of probability and severity with our experts' responses for PCI DSS and NERC CIP 007-6; we used the Friedman test (omnibus) with planned pairwise Wilcoxon signed-rank tests for comparing IRS P1075 responses, for which we had two expert validators [63], [14]. We also conducted open-ended discussions with the experts to discuss similarities and differences in assessments.

Partner criteria. We established the following criteria for partnering with organizations: (1) the organization must regularly be subjected to audits, must regularly audit other organizations, or must contribute to the content of the relevant compliance standard, (2) the provided validators must have at least two years of experience with the relevant standard, and (3) the organization must be able to mediate responsible disclosure of our findings.

After months of negotiation, we established memorandums of understanding with three organizations that met our criteria. Leaders within each organization nominated several compliance experts; we sent each candidate an email outlining the voluntary nature of the study as well as our motivation and goals. Table I shows the qualifications of our four volunteer experts. Experts completed their surveys during regularly scheduled work hours and did not receive any additional monetary incentives for participating.

Issue selection. We note that an essential tenet for partnering with experts is minimizing disruption to their daily responsibilities. Research suggests that the quality of survey responses decreases over time, and excessive time away from work may result in an expert terminating their participation in the study [25]. To this end, we designed our surveys for experts to complete within 60-90 minutes of focused effort; actual completion time averaged 84.8 minutes. Given our limited pool of experts, this required us to select only a subset of our findings to validate; as described in detail below, we selected the issues to validate semi-randomly, while prioritizing the extremely-high-risk and high-risk issues.

Pilot. Prior to deploying our protocol with partnering organizations, we piloted surveys to pre-test relevance and clarity with security practitioners familiar with auditing and compliance standards. We updated the study protocol based on pilot feedback. After two iterations, we arrived at the finalized questionnaire in Appendix B. Our two pilot experts currently conduct digital-security penetration testing against organizations, providing technical remediation recommendations for discovered security concerns.

C. Limitations

Our recruitment letter and consent waiver explained the purpose of the study. Thus, there may be self-selection bias in which personnel most interested in the study were more likely to anonymously participate. However, this may also suggest that our experts were prepared to think more critically about reported issues.

All of our experts work directly in compliance and their intimate working knowledge with compliance standards reduces the possibility of demand characteristics — a condition in which participants unconsciously change their behavior to perform well within a study [44]. Our study questions the validity of the compliance standards that serve as the basis for the experts' employment. This suggests that the experts would be in many cases predisposed to underestimate problems within these standards. Additionally, our validation method does not elicit expert feedback for false negatives – issues that our original analysis may not have detected. As such, we consider expert responses to provide a lower bound for validity.

The organizations we partnered with for this study have similar structures, missions, and technologies to other organizations that adhere to our selected compliance standards; however, there may exist specific organizational characteristics that affect their specific implementations or inhibit generalizability. As such, validating the presence of our discovered security concerns within partnered organizations' environments does not mean that all organizations adhering to similar compliance standards have security concerns, and the rejection of one of our findings does not imply that another organization elsewhere does not have security concerns. Nonetheless, our results can indicate systemic issues that organizations need to account for when assessing their levels of digital security risk and provide

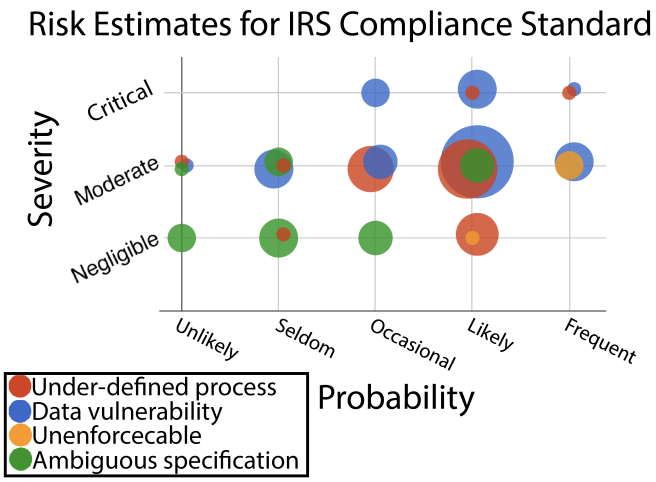

Fig. 2: Distribution of security concerns identified for IRS P1075. Color indicates the type of security concern; each dot indicates by size how many security concerns were identified with a given type, severity, and probability. Data vulnerabilities were most common (n=37).

novel insights into the impact of compliance standards on digital security in enterprise environments.

Lastly, we audited each compliance standard without considering other security controls in complementary documents. For this study, we assume that organizations implement compliance standards perfectly and limit our scope to finding security concerns in the documents as written.

IV. Results: IRS P1075

A. Overview

IRS Publication 1075 provides mechanisms for protecting and controlling access to federal tax information. IRS P1075 applies to all U.S. federal, state, and local agencies that receive Federal Tax Information (FTI) from the IRS or from secondary sources like the Social Security Administration [28]. Of the three standards we assessed, IRS P1075 is the longest standing, dating back to 1996 [27]. We audited the 2016 revision, which was the most current version available at the time of this study.

FTI security potentially affects every federal taxpayer. Organizations such as the Office of Child Support Enforcement from the U.S. Department of Health and Human Services rely upon IRS P1075 for securing the networked infrastructure of child support financial records [60]. Companies such as Amazon offer cloud infrastructure services that are fully compliant with P1075, marketing their virtual private server services to customers who need a "turn-key" solution for systems that transmit or receive FTI [3].

P1075 is written for information technology security professionals responsible for securing FTI. Key provisions include definitions for terms, parties authorized to access FTI, record-keeping requirements, physical controls for securing data, technical controls for secure storage/transmission, inspection protocols, and sanctions for non-compliance. The IRS Office of Safeguards coordinates and conducts compliance audits of entities possessing FTI. Of the three standards we assessed, P1075 has the weakest sanctions. There are no provisions for the issuance of fines for insecure practices, and the strictest sanction available to inspectors is data revocation after failure to adhere to a prescribed corrective action plan. However, non-compliant organizations can apply for data revocation waivers that extend their access to FTI for six months; according to the standard as written, this process can continue indefinitely despite continued non-compliance. This process has the potential to minimize the impact of sanctions while allowing insecure practices to persist. Overall, IRS P1075 was qualitatively and quantitatively the weakest of three documents we assessed during this study.

B. Findings

Our audit of P1075 identified a total of 81 independent issues across 309 individual security controls (Figure 2). Of these, we agreed that two issues presented an "Extremely High" risk, whereas 13 were "High," 32 were "Moderate" and 34 were "Low" risk according to the Risk Assessment Matrix (Figure 1). In addition, we discarded 15 initially identified issues, including 11 discarded when researchers found implementation details that were clarified in later sections of the standard and four resulting from researcher disagreements. All four issue disagreements related to nuanced interpretations of ambiguous portions of the standard.

Security concern trends. We identified five issues involving portable devices (e.g., mobile phones and laptops) and seven involving cloud-based data storage solutions. We associate the prevalence of these issues with shifts toward bring-your-own-device regimes and an increased reliance on cloud-storage solutions over on-premises servers [17]. These emerging solutions require specialized security measures and create inconsistencies with the best security practices that professionals have developed over the past few decades [54].

Of the 81 issues we identified within P1075, Section 9 had 40 technical controls with security concerns. Of note, Section 9 has several obsolete controls such as password expiration period requirements (which is shown to encourage insecure practices such as writing newly rotated passwords near user workstations) [20], [55]. In this particular instance, the *IRS mandated organizations to make a worse security decision than the decision they might have made in the absence of P1075.* Below we present detailed examples of findings based on their associated root cause.

Data vulnerability. We identified 37 issues that would establish conditions for a data breach if controls and processes are implemented as described in the publication. One example in Section 9.3.6.8 outlines processes for restoring backups once a compromise in a live system has occurred. As written, P1075 does not require technicians to verify the integrity of backups before restoration, meaning that technicians could revert to a state that an attacker has already infected (giving them persistent access) or revert to a vulnerable state that an attacker could re-exploit, reestablishing access to sensitive data [49]. Real-world trends stemming from ransomware support the urgency of backup integrity checks [50]. We assessed this high-risk issue to have a likely probability and moderate severity.

Section 9.3.5.11 includes provisions for user-installed applications. Environments that store or transmit FTI should be highly secure and should only be used for FTI — other functions and services should occur outside the FTI environment. As such, there should be little to no need for user-installed software, especially given that users are one of the primary vectors for introducing malware into environments [2]. Section 9.3 should instead mandate application whitelisting for installation attempts, limiting the subset of authorized applications that anyone can install on the system. A more stringent recommendation would include revoking user-installation privileges altogether, forcing trusted system administrators to establish a safe baseline of applications allowed to interact with FTI. We assessed this high-risk issue to have a likely probability and critical severity that can place FTI at risk.

We identified an extremely-high-risk issue within Section 1.4.4 "Information Received From Taxpayers or Third Parties," which limits the responsibility for securing FTI. According to this section, the IRS is only responsible for securing data originating from the IRS as FTI, excluding data received from customers like federal tax returns. Additionally, this section includes provisions for removing FTI protections on data if an entity replaces IRS-sourced FTI with the exact same information sourced from another party. This is analogous to eliminating protections for top-secret government data simply because the same information can be bought on the black market. This mandated behavior allows organizations to bypass security measures and remove protections on the data P1075 is meant to safeguard. We recommend that P1075 enforce protection for all FTI, regardless of source.

Section 1.4.3 defines certain data as personally identifiable information (PII) but does not protect the names of individuals associated with the person filing the return – such as a spouse or dependent. This high-risk issue may allow an attacker, for example, to develop a targeted spearphishing campaign against an individual. We recommend expanding the definition of PII to include sensitive information about all persons listed.

Unenforceable controls. We identified three controls that are unenforceable. For example, Section 4.7 provides several measures for secure telework access to FTI. P1075 provides many requirements for physical data protections, such as badge-based control and on-premises guards; these are infeasible in the case of telework, as most personnel with FTI access at their private homes cannot abide by these types of controls. Additionally, IRS inspections of private residences for physical security compliance seems fraught with complications. We recommend that either the IRS ban residential-based telework programs until it can verify that all locations with FTI access are compliant with physical security requirements, or that the standard acknowledge that these physical controls are not actually required. We assessed this high-risk issue to have a frequent probability and moderate severity.

Under-defined process. We identified 27 issues that reflect processes that are not sufficiently detailed for a secure implementation. One such issue within Section 8.3 states that "every third piece of physical electronic media must be checked to ensure appropriate destruction of FTI." Given the disparate possible sizes of electronic media, this particular section should recommend accounting for logical storage size of the media instead of its physical instantiation. This would ensure that media with larger storage volumes are highly prioritized for destruction validation. We assessed this as a moderate-severity, moderate-risk issue with a likely probability.

One low-risk issue occurs in Section 1.4.7, which limits human access to FTI based on "need to know" but does not consider machines or systems with a "need to access" data. Administrators must limit system access to FTI to prevent unauthorized access or manipulation of data, especially for systems performing aggregate analysis that may inadvertently disclose sensitive information.

Section 9.3.13.3 covers background checks for personnel with access to FTI. Our researchers assessed that this section could create information gaps at the federal, state, and local levels. For example, information about an individual who mishandled sensitive data at a previous job may never have entered federal databases. These extremely-high-risk information gaps increase likelihood for insider threats and risks to data, and highlight the need for aggregating multiple sources of data for thorough background checks.

We identified another issue in Section 9.3.5.8, which outlines a procedure for establishing an Information System component inventory (i.e., a listing of authorized devices that operate within an organization). As written, this procedure does not require the inventory process to be tied to a "ground truth," meaning there is no comparison of which devices should be operating within an organization with which devices actually are. This is dangerous, as it could permit a rogue system to persist on a network or even be inventoried as a legitimate system. Providing a rogue system with legitimate access within a sensitive environment obviates the need for an attacker to deploy malware within the environment and reduces the likelihood that any defensive sensors would ever detect anomalous activity from the attacker. We assessed this moderate-risk issue to have an occasional probability and moderate severity. Industry recommendations integrate asset inventory with supply acquisition, ensuring that only company-purchased, legitimate systems are on the network [9].

Ambiguous specification. We found 14 issues involving insufficient details that create ambiguity or uncertainty throughout P1075. P1075 uses vague terms such as "significant change" throughout, without ever defining thresholds that auditors deem to be significant. As an example, Section 9.3 outlines "Access Control Policy and Procedures" that must be reviewed (by whom?) every three years or whenever a significant change occurs. This subjectivity allows reviewers to deem any or all changes insignificant to circumvent a change review. Additionally, the document's use of passive voice clouds the responsibility for conducting the review — ambiguous controls can create security gaps through inaction. We believe each mandate should use active voice and assign a responsible individual (e.g., an office manager or system administrator) for each requirement. As presently written, an individual who works in an organization's talent recruiting department with no security training would be a sufficient reviewer for access-control policy. We assessed these moderate-risk issues with a likely probability and moderate severity.

C. Expert validation

For assessing the validity of our IRS P1075 audit, we partnered with New York City Cyber Command (NYC3). NYC3 is a city government organization that oversees the cybersecurity of 143 separate departments and agencies as well as more than 300,000 people. In addition to defending NYC against cybersecurity threats, NYC3 is responsible for ensuring compliance with government-mandated policies. In particular, the NYC3 team includes five full-time employees and three consultants who focus solely on security compliance. Each of the 143 departments within the city government also has an internal, full-time compliance teams.

IRS P1075 applies to the vast majority of these 143 entities. NYC3 advises other NYC entities on P1075 compliance and is also subjected to IRS audits. We coordinated directly with two NYC3 governance and compliance officials to assess the validity of our findings with respect to a particular subdomain under NYC3's purview that must comply with P1075 standards. This subdomain consists of a controlled internal network that contains FTI and supports approximately 150 users. NYC3's last formal P1075 audit was in January and February 2018, where three on-site auditors used the standard as a line-by-line checklist to assess NYC3's compliance. Of note, preparation for this inspection consumed the compliance team as well as several technicians for approximately four months prior to their inspection date.

Because of their limited time availability, we asked our two NYC3 compliance officials (hereafter referenced as Experts E1 and E2) to assess 20 issues (25% of our total 81 issues). In order to cover issues at all risk levels but prioritize significant concerns, we included both extremely-high-risk issues, and then randomly sampled 10 of the 13 high-risk issues, four of the 32 moderate issues, and four of the 34 low-risk issues.

When validating P1075 issues, E1 and E2 were able to directly examine their network for the presence of security concerns caused by issues identified by the researchers, in a kind of white-box penetration test [16]. This was possible because, unlike E3 and E4, E1 and E2 are officials with administrator privileges within the audited subdomain. The two experts analyzed our findings independently and did not discuss their findings with one another during the study. Overall, these experts confirmed 17 of our findings, rejected two issues, and indicated one issue could be plausible within their own or another environment.

When comparing our risk estimates to those of E1 and E2, we found no statistical difference between severity estimates (omnibus $p = 0.54$), but our researchers assessed issues to be statistically more likely with medium effect ($p = 0.0001$, 0.034, < 0.0001; $r = 0.485$, 0.336, 0.533 for omnibus and then pairwise researchers vs. E1, E2 respectively). E1 indicated that his knowledge of current and on-going initiatives most likely biased his responses, making it hard for him to follow instructions to consider each issue only "if standard is followed as written and nothing else" (as written in Appendix B). This response supports our notion in Section III-B that participant-validated responses represent a lower-bound for this study.

The issue that E1 and E2 classified as plausible rather than confirmed comes from Section 1.4.7 "Need to Know." E2 indicated that NYC3 data scientists incorporate the principle of least privilege for systems, service accounts, and user accounts, which would prevent unauthorized access and manipulation of FTI. E1 added that NYC3's PKI infrastructure assists with controlling access to "need to access" data. Both participants indicated they were unsure if this security concern was ever present within NYC3, but that it could be present within other organizations.

E1 and E2 rejected our finding for Section 1.4.3 PII, indicating NYC3 always encrypts entire tax records while in transit and rest, and that this is standard practice for organizations with access to FTI. Thus, associated individuals' PII are always protected, invalidating our finding. However, because this is not codified within P1075, there is no certainty that other organizations adhere to this "standard practice."

NYC3 also rejected our finding associated with Section 9.3 background checks. E1 indicated that it is standard practice to aggregate personnel records from the locations an individual has lived or worked to determine if the individual should have access to sensitive information, thus rejecting our finding. Because P1075 does not mandate data sources or how far back in history to consider, there is no certainty that other organizations conduct this practice.

Additional defenses. E1 and E2 identified several controls pervasive throughout NYC3 that help reduce or eliminate the impact of many of our researcher findings. Of note, NYC3 requires a Change Control Board (CCB) that E2 believes "is an essential risk-mitigating factor" for addressing many of the confirmed P1075 security concerns, such as Section 9.3.5.11 "User-Installed Software." The CCB evaluates all user requests for system modifications and holistically considers the change's impact to security. If the CCB approves the change, it authorizes a trusted administrator to conduct the software installation and adds the change to the system's secure baseline. Additionally, NYC3 incorporates a real-time, automatic asset manager which alerts their Security Operations Center any time a new device is added to their networks. This defensive strategy eliminates the security concern we identified in Section 9.3.5.8 "Information System Component Inventory."

It is important to note that these defenses employed at NYC3 exceed the baseline security standards required by P1075 and mitigate issues that P1075 either fails to account for even causes. We cannot assume that all organizations will recognize the need for these additional mitigations and be willing to invest in them.

V. Results: PCI DSS

A. Overview

The Payment Card Industry Data Security Standard (PCI DSS) applies to all entities that store, process, and transmit credit-card-holder data for major branded credit cards [45]. Guidance in this standard includes building and maintaining a secure network and systems, protecting cardholder data, and monitoring/testing networks. PCI DSS v1.0 dates back to 2004 as a program led by Visa; the PCI Security Standards Council (SSC) was formed in 2006 by American Express, Discover, JCB International, MasterCard and Visa to enhance PCI DSS [45]. We audited the 2016 v3.2; v4.0 was in development during this study.

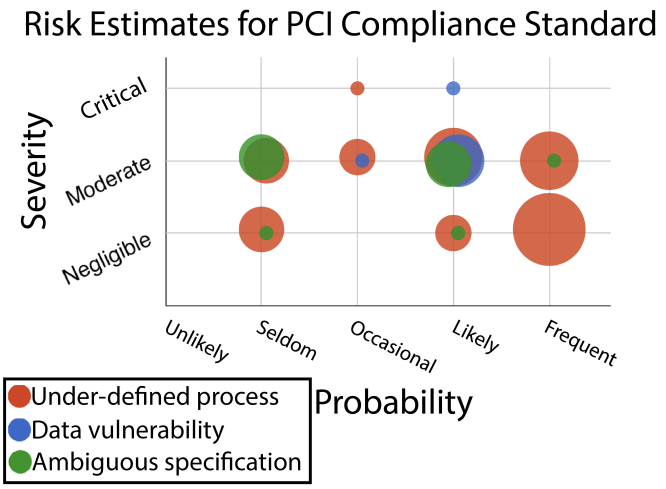

Fig. 3: Distribution of security concerns identified for PCI DSS. Color indicates the type of security concern; each dot indicates by size how many security concerns were identified with a given type, severity, and probability. Under-defined processes were most common (n=29).

PCI DSS affects every person within the United States who makes credit card purchases and every organization that accepts credit card payments. A U.S. Federal Reserve study showed that consumers spent $5.98 trillion with credit cards in 2016, highlighting the importance of securing the systems that support those financial transactions [59]. PCI DSS authors designed the document to be accessible to assessors and the technicians charged with implementing the technical controls.

Qualified Security Assessors perform PCI DSS audits after attaining the appropriate inspection certifications. According to one such assessor (not an author or an expert validator), audit frequency varies for merchants and service providers depending on their number of supported annual transactions [37]. On-site audit teams vary from one to three personnel per inspection; these personnel focus full-time on auditing the PCI DSS compliance of other organizations. The assessor indicates that penalties for non-compliance are common, but vary in size based on the severity of infraction and size of customer base. Monthly fines that can range from $5,000 to $100,000 and continue until compliance issues are resolved. If a data breach occurs as a result of non-compliance, companies may be responsible for consumer services (e.g., credit monitoring) or may have payment-processing privileges revoked.

B. Findings

Within the 851 independent controls specified by PCI DSS, we identified 46 security concerns: eight high-risk, 22 moderate-risk, and 16 low-risk (as shown in Figure 3). We discarded six other potential issues, all of which were under-defined processes that did not result in any insecure practices or conditions.

Security concern trends. We identified four issues related to improperly identifying sensitive information. PCI DSS focuses heavily on protecting primary account numbers (PANs) that are tied to credit cards but fails to protect other information that could lead to PAN access, such as passwords or password-recovery information. Additionally, we identified 10 issues involving technical controls that lack timelines for required action. For each required action, the standard should specify either a fixed interval for repetition or for a triggering event with an ensuing deadline. Below we present discovered PCI DSS issues, organized according to perceived root cause.

Data vulnerability. We identified seven security concerns that could establish conditions for a data breach. One example of a high-risk vulnerability stems from Section 1.4, which includes mandates for securing personal computing systems within the cardholder data environment (CDE). We recommend disallowing any personal electronics within the CDE network segment; more broadly, all services and systems should be limited by "need to access" cardholder data. Personal devices and activities increase the likelihood of malware or other unauthorized access and generally are not necessary within CDE network segments [2]. We assess this security concern to have a likely probability and critical severity.

A tangentially-related moderate-risk security concern stems from the "Network Segmentation" section of PCI DSS, which scopes the standard's safeguards to only the network segment that contains cardholder data. Effectively, this provision would allow an organization with no security controls outside of the CDE to pass an audit as long as the CDE itself is protected in accordance with PCI DSS specifications. Allowing vulnerable servers and systems within the same network as the CDE could provide attackers with a landing point into internal portions of the network and establish conditions for lateral movement into the CDE from adjacent network segments (through well-known attacks such as VLAN hopping). Due to the series of security holes that must be present for such an attack to occur, we assessed that exploitation of this vulnerability would be seldom but critically severe for affected systems.

Another data vulnerability is present within the "PCI DSS Applicability Information" section, where PCI DSS defines sensitive authentication data. PCI DSS does not consider passwords to be sensitive authentication data and does not protect information an attacker could use to reset service passwords (e.g., email addresses, Social Security Numbers, and dates of birth). The social engineering attack against Naoki Hiroshima's `@N` Twitter account leveraged similar pieces of information to access protected accounts [23]. Given that publicly-available articles detail how unprotected information can lead to unauthorized access, we assess this security concern to have a moderate severity and likely probability.

Under-defined process. We identified 29 issues with process specifications that are insufficient for a secure implementation. Section 3.2.1 calls for assessors to select a sample of system components and examine data sources to detect cardholder data that is improperly stored. Sampling is an insufficient process, considering the simplicity of searching for cardholder data that adheres to a well-known format. We recommend improving this section to mandate assessors use automated tools on all CDE systems to detect improperly stored cardholder data. Based on the moderate severity of exposed cardholder data and the frequent likelihood insufficient checks occurring, we assess this issue to be a high-level risk.

PCI DSS features two high-risk under-defined processes

in "Requirement 5: Protect all systems against malware and regularly update anti-virus software or programs" and Section 5.1. These sections rely solely on antivirus to prevent malware infections. Numerous data breaches have shown that antivirus alone cannot protect against all malware [11]. These limited-scope requirements leave organizations exposed to multiple attack vectors that will most likely occur frequently and have moderate severity. These sections should mandate additional controls such as application whitelisting, blocking access to areas that permit persistence (e.g., Windows Registry Keys), and enforcing least-privilege access.

Section 1.3.7 focuses on limiting the disclosure of private, internal IP addresses from firewalls and routers, but fails to discuss any other services that could leak the same information, such as a domain name server or internal files (e.g., Word documents) improperly exposed to search engines. Attackers have leveraged common techniques such as "Google Hacking" to discover internal network configurations and sensitive systems like a domain controller [35]; expanding the scope of this moderate-risk issue to limit external enumeration would improve its security.

Sections 11.1.c and 11.1.d actually incentivize less-secure practices. Each subsection defines additional audit checks that an assessor must conduct only if a particular security control is in place (wireless scanning and automated monitoring, respectively). Under this policy, financial sanctions associated with non-compliance could lead a security professional not to implement a security control at all rather than risk having it assessed as non-compliant — if it is not present, the organization is automatically compliant. These two particular controls would have a negligible overall impact if they were not in place; therefore, we assess this to be a low-risk issue. We recommend that if the PCI SSC believes these security controls are important, they should be mandatory rather than optional; otherwise, these sections should be eliminated entirely.

Ambiguous specification. We identified 10 issues within PCI DSS in which insufficient details create ambiguity or uncertainty. An example of a high-risk security concern with a frequent probability and moderate severity stems from Section A1.1 and limits the usage of Common Gateway Interface (CGI) scripts to control privileged access to cardholder data. This control is sound but is overly narrow; in modern systems, there are a variety of applications that could access or manipulate cardholder data in ways similar to CGI scripts. We recommend simply replacing "CGI scripts" with "applications" to improve the clarity of this control.

Section 11.3.3 discusses corrective action plans for vulnerabilities discovered during penetration tests. The section does not specify how soon after a penetration test vulnerabilities must be addressed, nor the party responsible for fixing the vulnerabilities. Based on the researchers' past experiences with organizations delaying remediation, we assess this security concern to have a high risk level with a frequent probability of occurring and a moderate severity. Moreover, the non-validator assessor we spoke to confirmed that in his experience, organizations often delay remediation, and typically dedicate one to two full-time employees for 30-40 days prior to an inspection to ensure remediation is complete just in time [37]. We recommend this section specify a time limit (based on vulnerability severity) for addressing issues discovered during a penetration test and clarify the party responsible for fixing the vulnerable condition.

C. Expert validation

To assess our PCI DSS findings, we partnered with an organization that is a PCI SSC member. Expert E3 represented this organization, possessing 18 years of experience advising the security practices of large financial organizations, assessing organizations' adherence to PCI DSS security controls, and conducting digital security assessments against networked environments. E3 confirmed past utilization of PCI DSS as a line-by-line checklist as they audited organizations in the past.

We asked E3 to assess all eight high-risk issues and a randomly-sampled subset of seven moderate issues and five low-risk issues; this accounts for 43% (20 of 46) of the issues from our audit. E3 confirmed 18 of the issues and categorized the remaining two as plausible, although he had not experienced them.

We observed no statistical difference between probability estimates between E3 and our auditors ($p = 0.77$),but E3 assessed issues to be statistically more severe with medium effect ($p = 0.003$, $r = 0.469$). During our post-survey discussion with E3, he stated that the financial impacts of digital security breaches involving cardholder data caused him to increase his assessed impact of each issue — had these issues been present within another business sector, E3 would not have assessed them as severely.

The first issue assessed as plausible rather than confirmed involves Section 1.3.7 and information disclosure. E3 indicated that internal data exposure is "inconsequential if boundary configuration is correct," meaning an administrator is successfully limiting which inbound connections from external entities are allowed to communicate with private IP addresses. However, E3 acknowledged that the security concern would exist if these additional controls are not in place.

The second issue E3 flagged as plausible rather than confirmed involves Section 5.1's reliance on anti-virus software. According to E3, organizations have lessened reliance on anti-virus for protection; he argued that Section 5.1 would have minimal impact on organizations with defensive strategies for protecting network segments, user accounts, and key resources.

Additional defenses. E3 recommended account-protection solutions such as multi-factor authentication to mitigate concerns such as VLAN attacks or insufficient protection of passwords.

As discussed for P1075 above, both the issues E3 assessed as only plausible and his recommended additional defenses hinge on additional security controls beyond the PCI DSS standard; we cannot necessarily assume non-mandated controls will be applied.

VI. Results: NERC CIP 007-6

A. Overview

The North American Electric Reliability Corporation (NERC) Reliability Standards define the requirements for planning and operating North American bulk power systems

(BPSs), defined as large interconnected electrical systems consisting of generation and transmission facilities and their industrial control systems [41]. All BPSs within the continental United States, Canada, and the northern portion of Baja California, Mexico must comply with NERC Reliability Standards, meaning that these security controls affect most people living within these areas. The NERC Critical Infrastructure Protection (CIP) Committee, which oversees the set of standards, comprises representatives from 30 companies and municipalities across North America [42]. Although NERC is an international not-for-profit, its regulatory authority stems from section 215(e) of the Federal Power Act and Title 18 Code of Federal Regulations §39.7. The set of standards that make up CIP date back to 2009; in this study, we audited CIP 007-6, which is the 2014 revision of the Systems Security Management standard. CIP 007-6 includes key sections for securing ports and services, patch management, malicious code prevention, event monitoring, and access control.

NERC Regional Entities are the organizations responsible for conducting audits and monitoring adherence to the compliance standards within their assigned geographic region. On-site audits typically last one week and occur every three years. Accordng to our expert validator E4, a NERC Regional Entity employs four to seven auditors per assessment, drawn from a pool of full-time employees. Auditors typically conduct 7-30 audits per year. E4 also noted that organizations allocate a large portion of their operating budgets toward compliance and often spend one year preparing for their audit.

Of the three standards we assessed, NERC has the strongest sanctions (which can actually create security concerns, as discussed in Section VI-C). The maximum fine for a compliance violation is $1M (U.S.) per day; NERC or the applicable Regional Entity determines the monetary fine [40]. According to our expert participant, fines for NERC non-compliance are common. Recently, NERC levied a $10M fine against Duke Energy for 127 security infractions between 2015 and 2018 [22].

Qualitatively and quantitatively, CIP 007-6 had the strongest security controls of the three documents we assessed (shown in Figure 4), but numerous issues exist that we believe create security gaps within compliant organizations.

B. Findings

NERC CIP has 79 individual controls. Our internal audit identified 21 total issues; we categorized one as extremely-high-risk, four as high-risk, six as moderate-risk, and 10 as low-risk. We discarded one additional issue that we identified as a duplicate entry.

Security concern trends. Seven of the 21 issues we identified deal with overly vague terms such as "when feasible" or "unnecessary" without defining feasibility or necessity. For example, Section 5.7 calls for limiting authentication attempts or generating alerts when feasible. The subjectivity of these statements can lead to misinterpretations of the standard and potentially permit insecure actions. Mandatory compliance standards should be mandatory; either administrators must limit authentication attempts or it is merely a suggestion. Additionally, none of the 21 issues we identified specify which entity is responsible for specific actions, which can

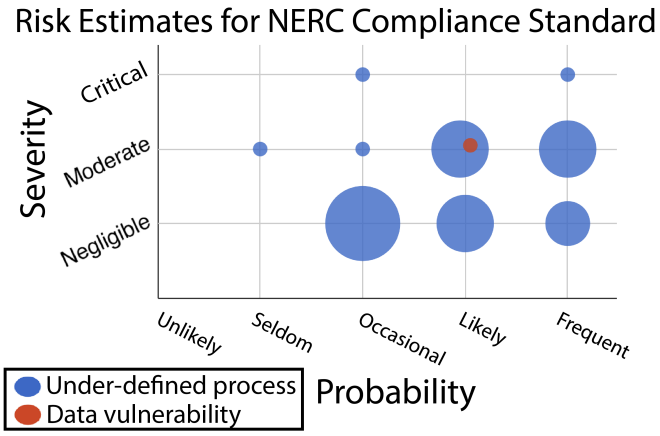

Fig. 4: Distribution of security concerns identified forNERC CIP 007-6. Color indicates the type of security concern; each dot indicates by size how many security concerns were identified with a given type, severity, and probability. Under-defined processes were most common (n=20).

s

lead to inaction. Notably, NERC identified "confusion regarding expectations and ownership of tasks" as a key problem contributing to Duke Energy's non-compliance and eventual fine [22]. Below we present detailed examples of findings, organized by their perceived root cause.

Data vulnerability. Based on our assessment, CIP 007-6 only has one moderate-risk issue pertaining to a data vulnerability. Section 5.1 states that administrators should "[h]ave a method(s) to enforce authentication of interactive user access, where technically feasible." This caveat allows legacy equipment with no provision for authenticating authorized users to endure within a secure environment. It is well-documented that legacy systems often have no password, transmit unencrypted passwords, or never change passwords from their default settings [13]. This permits attackers and insider threats to easily gain control of legacy systems, which could range from sensitive databases to the logical system "off switch." We argue that secure, authenticated access should be a hardware and software requirement for all systems in this critical environment, reducing the likelihood of such an attack.

Under-defined process. The remaining 20 issues involve processes that are not sufficiently detailed for a secure implementation. We assessed that Section 2.1 has an extremely high risk, as written, due to the critical severity and frequent probability of a security concern occurring within critical environments. The issue involves the implementation of a patch management program for improving the security of systems. Throughout all of the NERC CIP documents, we were unable to find any mandate that organizations maintain a representative test environment for patch evaluation. Applying patches directly to live systems that provide power — including to critical infrastructure such as hospitals — could result in outages and corresponding loss of life; one such incident occurred in March 2008 and caused a nuclear power plant to shutdown [32]. Testing patches prior to live deployment allows administrators to observe potential effects within their environment and

reduce the likelihood that unforeseen outages will occur as the result of the patch [57].

We discovered a potential loophole in Sections 2.1 and 2.2, which rely upon validated sources for patches against known vulnerabilities. If the entity responsible for patching systems does not provide sources, then there is no requirement for patching. Additionally, CIP 007-6 does not account for patches from external sources beyond the list of valid providers. Do administrators have a requirement to apply a patch for a known vulnerability if it is from an outside source? According to Cardenas, there are instances where applying a patch may violate the certification of certain control systems [10]. We deem this loophole to present a high risk due to the critical severity associated with unpatched systems in these environments and the occasional probability of their presence.

Section 5.3 requires administrators to "identify individuals who have authorized access to shared accounts." We assessed shared accounts as a moderate-risk threat, as administrators are unable to deploy granular controls on a by-need basis. Shared accounts also inhibit auditing, as the compromise of a privileged shared account could lead to the spread of malware or outages that administrators cannot positively attribute to one individual. Researchers from Sandia National Lab identified this security concern in 2003 [51].

Section 5.4 outlines provisions that allow systems to retain their default usernames and passwords if documentation supports that the "vendor passwords were generated pseudo-randomly and are thereby unique to the device." Our auditors believe that vendor-generated pseudorandom credentials can present a threat to BPSs if the pseudorandom algorithm is predictable (e.g., basing its seed on a unique identifier such as a serial number). This type of exploit requires in-depth knowledge about the vendor's algorithm and might seldom occur despite posing a moderate risk to the environment. We recommend eliminating this provision entirely and mandating that administrators change all system credentials before allowing a system to communicate with a BPS.

We identified a high-risk issue in Section 4.3 concerning event log retention. CIP 007-6 requires facilities to retain 90 days of consecutive logs and demonstrate proof of such practice over a three year period. This relatively short-term rolling requirement can interfere with incident investigations, given that advanced persistent threats can operate within networks for years before being detected [1], [62]. We recommend mandating that organizations ship logs to a data warehouse for long-term storage and investigation support if needed.

C. Expert validation

We partnered with a government organization that focuses on national security issues to validate our CIP 007-6 findings. Expert E4, as the organization's representative, has 20 years of experience conducting digital security assessments against BPSs. E4 confirmed first-hand utilization of NERC CIP standards as a checklist for past audits. E4 has served on numerous executive councils and federal-level panels addressing cybersecurity concerns within industrial control systems. Most notably, E4 was a contributing author to many of the NERC CIP standards.

Due to the complexity of NERC CIP, our 60- to 90-minute survey could include only nine audit findings (43%). We included the extremely-high risk issue and all four high-risk issues, and we randomly sampled two moderate-risk and two low-risk issues. Of these, E4 confirmed one issue and one broader trend, rejected one issue, and categorized the remaining seven issues as plausible.

When comparing our auditors' risk estimates to those of E4, there was no statistical difference between severity estimates ($p = 0.18$),but our auditors assessed the issues to be statistically more likely with a large effect ($p = 0.01, = 0.603$). E4, addressing these comparison differences, indicated that CIP 007-6 relies heavily on the broader framework of CIP standards and that security controls in other CIP documents help harden the overall environment. Like E1, E4 commented that he was unable to assess CIP 007-6 only "if standard is followed as written and nothing else," as directed (Appendix B). As such, E4 indicated that he rated each issue as less likely given his broader understanding of the compliance framework.

The issue E4 rejected involves the loophole we identified in Sections 2.1 and 2.2 for patch management. E4 stated that "each item in the [system] baseline needs a source identified or evidence that a source no longer exists." In his experience, he never encountered an external source that could provide a trusted, proprietary patch. However, E4 acknowledged that if a component is no longer supported or a source no longer exists, it is highly likely that the component will remain unpatched against all future publicly-disclosed vulnerabilities.

E4 confirmed the log-retention issue we identified in Section 4.3, attributing the known gap between log retention and investigation windows to two factors. Primarily, the specification is written to account for the limited log retention capacity on most devices within a BPS environment. Second, most administrators and BPS owners are unwilling to connect to and aggregate event logs on an external platform. Placing an additional device within the environment (for logging) increases the number of devices an attacker can exploit and is one more device potentially subject to financial sanctions.

E4 also confirmed the risks of not specifying a responsible party for tasks, a trend our researchers identified, and referenced the aforementioned Duke Energy fine as an example.

Additional defenses. E4 noted that the best additional defense for mitigating the issues we identified was to upgrade system components to more modern devices that can implement up-to-date best practices (e.g., multi-factor authentication, strong passwords, limiting login attempts. As with P1075 and PCI DSS, organizations that only meet the minimum required by the standard will not be able to take advantage of these defenses. E4 confirmed that while some facilities exceed this "minimum baseline" and systematically replace obsolete devices, he has also audited facilities that only follow the standard exactly as written.

Other recommendations. E4 described additional security concerns that our auditors did not identify. Subsets of NERC CIP security controls apply to BPSs based on how much power the system produces, creating three tiers of compliance: the highest tier of power producers are subject to all security controls, while the lower tiers of power producers must comply

with decreasing subsets. E4 believes this perversely allows attackers to use publicly-available information to locate facilities that must adhere to fewer security controls and then systematically target the controls that may not be present. E4 therefore argues that NERC must standardize controls across all facilities to mitigate the targeting of smaller stations.

Additionally, E4 stated that the zero-defect culture and high fines associated with NERC's sanctions program can incentivize minimum-effort security. Organizations that undertake additional security precautions beyond NERC CIP mandates may discover vulnerabilities that would not otherwise be identified. NERC levies fines for non-compliance even when organizations self-report such vulnerabilities, potentially punishing organizations for transparency. E4 believes this behavior inhibits sharing of information across the power sector and collectively lowers security for all facilities. He argues that NERC could reverse this trend by eliminating fines associated with self-reporting and providing "credits" to organizations that contribute to the overall health of the power sector.

When discussing concerns with log retention, E4 recommended that all facilities should contribute toward a common log aggregation center, where security professionals could conduct in-depth security-breach investigations spanning all NERC-compliant facilities.

VII. Reporting

We made an effort to disclose our findings responsibly. Compliance standards typically have a request-for-comment (RFC) period that allows for the submission of comments, concerns, and recommendations during a fixed window. During this study, none of the standards we assessed had an open RFC, and we found that no clearly defined channel existed for reporting security concerns, either directly to affected organizations or at the federal level. Using our partners as mediators, we turned over all of our findings to the IRS; the PCI Security Standards Council; a contributing author of the NERC CIP standards (E4); the United States Computer Emergency Readiness Team (US-CERT); the MITRE Corporation's Common Vulnerabilities and Exposures (CVE) team; and the Department of Homeland Security. We had varying levels of success with these disclosures, as described below.

IRS P1075. We contacted the IRS, NIST National Vulnerability Database (NVD), US-CERT, and the MITRE Corporation to disclose our P1075 findings. US-CERT was the first organization to respond to our disclosure attempt; their technicians concluded that "CVEs are assigned for specific vulnerability in implementations. Each issue that requires a 'separate patch' can get a CVE [58]." We argued that each of the recommendations we provided are "patches" for the vulnerable portions of the compliance standards, but US-CERT stated that the "patches" we identified must be tied to a specific piece of software. Both NIST NVD and the MITRE Corporation indicated that compliance documents are outside their scope of responsibility, with MITRE stating "that a reasonable person can conclude that IRS Publication 1075 was never intended to have a level of abstraction that was sufficient to direct secure coding [36]." Contradicting this argument, our partners at NYC3 confirmed that auditors are indeed using P1075 as a line-by-line checklist to confirm controls at levels as granular as access control lists on firewalls.

Document	Controls	Total Issues	Extr. High	High	Moderate	Low
IRS	309	81	2	13	32	34
PCI	851	46	0	8	22	16
NERC	79	21	1	4	6	10

TABLE II: Security concerns, by document and assessed risk

We attempted to disclose our findings directly to the IRS nine times via personal contacts, emails, and phone calls over the span of three months. To date, we have not received any form of acknowledgment other than the automated responses.

PCI DSS. Unlike P1075, we had success in responsibly disclosing our findings to members of the PCI SSC. We established a memorandum of understanding with a PCI SSC member organization; in turn, this organization provided our findings to the PCI DSS v4 Working Group.

We received notification that our recommendation for improving the "Network Segmentation" section of PCI DSS has already been implemented within Version 4, prior to the opening of their RFC submission window. This change will apply PCI DSS guidelines to the entire networked environment and not only an isolated subnet with cardholder data. Additionally, the v4 Working Group is considering incorporating all feedback associated with our ambiguous specification findings.

NERC CIP 007-6. Expert E4, after providing feedback, noted that our recommendations would be included at future working groups for CIP revisions. However, it will be years before the next CIP update (potentially taking our recommendations into account) is released. Additionally, our partnered organization for CIP disclosure is incorporating our feedback into a comprehensive evaluation of electric grid security.

Federal-level recognition. To approach problems with federal-level compliance standards in a top-down manner, we met with representatives from the NIST National Cybersecurity Center of Excellence (NCCoE) to discuss our findings [38]. We highlighted that IRS P1075 Section 9 (which contains 49% of the P1075 security concerns we discovered) is copied from older versions of NIST SP 800-53 (NIST has since updated SP 800-53 twice). NCCoE offered to incorporate our findings into future document revisions. In ongoing revisions that began before our meeting, NIST acknowledged in draft SP 800-53v5 that organizations may inherit risk when implementing mandated security controls; that is, standards may create security problems [43]. Specifically, NIST describes deliberate efforts to remove ambiguity, improve understanding of responsibility, and keep controls up to date, corroborating many findings from our study.

Next, we contacted the Department of Homeland Security (DHS) National Protection and Programs Directorate. Several personnel within the Federal Network Resilience Division expressed interest in assisting with our findings; however, the DHS Office of External Affairs for Cybersecurity and Communications directed our contacts to cease communication and did not provide any alternative mechanisms for disclosure.

VIII. Discussion and Conclusion

We provide the first structured evaluation of security issues within digital-security compliance standards. In our study, we find that using compliance standards as checklists, with "by-the-letter" implementation of security controls, can create security concerns. Our systematic approach identified security issues spanning multiple root causes and varying levels of risk (shown in Table II). In this section, we discuss common issues across the audited compliance standards, potential mitigations, recommendations for reconsidering compliance programs, and opportunities for future work.

Common issues. When considering our findings, some common issues become apparent. All standards we assessed exhibit under-defined processes and vague writing. While issues of vague writing may not seem as immediately dangerous as, for example, failing to identify passwords as sensitive data requiring protection, they have important implications when standards are treated like point-by-point checklists.

Many issues stem from passive voice, creating ambiguity concerning who is responsible for exactly what actions. Using the active voice to construct compliance controls is a best practice that helps eliminate uncertainty and ensure there is a responsible party for requisite actions [30]. If it is not feasible to eliminate passive voice (perhaps because it would prescribe organizational structure too strongly), standards authors could perhaps include supplemental best-practice recommendations for identifying responsible personnel. In addition, the standard might require each implementing organization to create a written plan identifying who is responsible for each requirement.

Further, we observed that numerous compliance controls did not have clear deadlines for action. Compliance standards should define expected periodicity (e.g., every 30 days) or thresholds for action (e.g., within 12 hours of an event). These issues with deadlines seem especially concerning in light of observations by several auditors we spoke with that many problems are only mitigated during an immediate run-up to a compliance audit, as part of preparations to pass.

Terms such as "when feasible" and optional guidelines create confusion about what is actually required and may provide an illusion of more security than what is actually provided. We recognize that in some cases, this wording reflects practical limitations: for example, updating legacy power systems to include modern security controls (NERC CIP) could require multi-million-dollar equipment investments and degrade near-term power availability. Nonetheless, we argue that categorizing clearly insecure systems as "compliant" simply because there is no feasible alternative is counterproductive. Instead, compliance standards could adopt a third category that does not punish the affected organization but still indicates clearly to administrators and auditors that the situation is suboptimal and further precautions are needed. We also recommend, for clarity, moving optional guidelines into supplemental documents separate from mandatory compliance.

We also noted that each compliance standard has weak controls for user-access review and revocation procedures. To mitigate insider threats, compliance standards could mandate frequent review of active user accounts, as well as access termination before formally notifying an employee who is terminated.

Lastly, and perhaps most concerning, none of the compliance standards we assessed have mechanisms for reporting security concerns. Without a direct line of communication with a governing body, it is likely many discovered security concerns will remain unaddressed. The lack of a centralized CVE database-like construct for reporting problems with compliance standards affects both governing bodies and compliant organizations. Governing bodies do not have a reference for common mistakes when developing compliance standards, meaning issues are likely to repeat across multiple standards. Additionally, this lack of transparency prevents industry-wide alert notifications for issues within a compliance standard; if a researcher discovers a valid security concern, all affected parties should be notified. Further, no standard could be expected to perfectly capture all needed security controls; as several of our experts noted, strong security practices often require going beyond the minimum established by a standard. A centralized repository would also present an opportunity to recommend additional best practices to build upon compliance and mitigate any reported gaps.

Recommendations. Our work highlights difficulties than can arise when compliance standards are used as checklists, regardless of their original intent. This approach seems inevitable when a standard is associated with potentially large penalties for non-compliance, but little or no incentive for going beyond the minimum requirements. This state of affairs suggests a need for rethinking the compliance paradigm more broadly.

First, authors of compliance standards should take into consideration that their standards might be used as an audit checklist. Whenever possible, guidelines should be broadly applicable across a particular domain but concrete enough that line-by-line compliance will provide meaningful security. Of course, writing guidelines that achieve this ideal is difficult and may sometimes be impossible; standards authors should explicitly consider tradeoffs between generalizability and secure implementation when making choices. Providing supplemental documents describing potential such issues could help standards implementers manage resulting risks.

Secondly, authors should identify opportunities to craft compliance standards that improve audits beyond checklist assessments and consider an organization's overall security culture. Provisions for a rewards program could incentivize organizations to bolster security. As examples, organizations that take proactive measures beyond minimum requirements or organizations that publish digital security lessons learned could receive some limited safe harbor against future sanctions. As discussed during our audit of NERC CIP standards, an organization that responsibly discloses and remedies a vulnerable condition is still liable for financial sanctions. Allowing organizations to self-report issues with less fear of sanctions could incentivize better behavior and increase transparency, with potential benefits for the entire associated sector [5].

Another consideration for standards authors is that rapidly changing technology necessitates rapidly updated security mechanisms. An effective standards update mechanism should allow easy reporting of issues and enable fast revision of the standard itself, while avoiding imposing costs on organizations that cannot immediately meet the new requirement. Newly updated standards could provide suggestions for transitioning

and require organizations to provide a plan for becoming compliant with the updated requirement within some specified time period.

Future work. To validate the issues we identified, we developed close collaborations with organizations that implement compliance standards and conduct associated audits. Future work should investigate how standards organizations generate compliance standards, as well as how a broader range of compliance standards are applied in real-world environments. Drawing on common themes identified across standards, researchers could design guidelines or even templates for avoiding common issues when writing standards.

ACKNOWLEDGMENTS

The authors would like to thank Colin Ahern, Tim Conway, and the leadership team at NIST for their advice and expertise in shaping this study, as well as Lujo Bauer for feedback on a draft version.

This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA). The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

REFERENCES

[1] L. Ablon and T. Bogart, "Zero Days, Thousands of Nights," 2017. [Online]. Available: https://www.rand.org/pubs/research_reports/RR1751.html

[2] S. Abraham and I. Chengalur-Smith, "An overview of social engineering malware: Trends, tactics, and implications," *Technology in Society*, vol. 32, no. 3, pp. 183–196, 2010.

[3] Amazon, "IRS Publication 1075," 2018. [Online]. Available: https://aws.amazon.com/compliance/irs-1075/

[4] Amazon Web Services, "Compliance and top security threats in the cloud – are you protected?" 2018. [Online]. Available: https://www.youtube.com/watch?v=Rc55aYODnMI&feature=youtu.be&t=18m10s

[5] A. Arora, R. Telang, and H. Xu, "Optimal policy for software vulnerability disclosure," *Management Science*, vol. 54, no. 4, pp. 642–656, 2008.

[6] H. Assal and S. Chiasson, "Motivations and amotivations for software security," 2018.

[7] ——, "Security in the software development lifecycle," pp. 281–296, 2018.

[8] A. Beautement, M. A. Sasse, and M. Wonham, "The compliance budget: managing security behaviour in organisations," in *Proceedings of the 2008 New Security Paradigms Workshop*. ACM, 2009, pp. 47–58.

[9] B. Beyer, C. Jones, J. Petoff, and N. R. Murphy, *Site Reliability Engineering: How Google Runs Production Systems*. " O'Reilly Media, Inc.", 2016.

[10] A. A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems." in *HotSec*, 2008.

[11] A. Carlson, "3 Reasons Anti-Virus Software Alone Is No Longer Enough," 2013. [Online]. Available: https://www.lawtechnologytoday.org/2013/03/3-reasons-anti-virus-software-alone-is-no-longer-enough/

[12] R. Clark, "Compliance != security (except when it might be)," in *Enigma 2018 (Enigma 2018)*. Santa Clara, CA: USENIX Association, 2018. [Online]. Available: https://www.usenix.org/node/208142

[13] G. M. Coates, K. M. Hopkinson, S. R. Graham, and S. H. Kurkowski, "A trust system architecture for scada network security," *IEEE Transactions on Power Delivery*, vol. 25, no. 1, pp. 158–169, 2010.

[14] G. W. Corder and D. I. Foreman, *Nonparametric statistics for non-statisticians: a step-by-step approach*. John Wiley & Sons, 2009.

[15] S. Elky, "An introduction to information system risk management," *SANS Institute InfoSec Reading Room*, 2006. [Online]. Available: https://www.sans.org/reading-room/whitepapers/auditing/introduction-information-system-risk-management-1204

[16] D. Geer and J. Harthorne, "Penetration testing: A duet," in *Computer Security Applications Conference, 2002. Proceedings. 18th Annual*. IEEE, 2002, pp. 185–195.

[17] A. Ghosh, P. K. Gajar, and S. Rai, "Bring your own device (byod): Security risks and mitigating strategies," *Journal of Global Research in Computer Science*, vol. 4, no. 4, pp. 62–70, 2013.

[18] Glassdoor, "Compliance Auditor Salaries," 2019. [Online]. Available: https://www.glassdoor.com/Salaries/compliance-auditor-salary-SRCH_KO0,18.htm

[19] Government Services Administration, "Federal Risk and Authorization Management Program," 2018. [Online]. Available: https://www.fedramp.gov/about/

[20] P. Grassi, J. Fenton, E. Newton, R. Perlner, A. Regenscheid, W. Burr, and J. Richer, "NIST Special Publication 800-63B Digital Identity Guidelines Authentication and Lifecycle Management," 2017. [Online]. Available: https://pages.nist.gov/800-63-3/sp800-63b.html

[21] A. F. Hayes and K. Krippendorff, "Answering the call for a standard reliability measure for coding data," *Communication methods and measures*, vol. 1, no. 1, pp. 77–89, 2007.

[22] R. Heidorn, "Nerc seeks $10m fine for duke energy security lapses," Feb 2019. [Online]. Available: https://www.rtoinsider.com/nerc-fine-duke-energy-cip-110308/

[23] N. Hiroshima, "How I lost my $50,000 Twitter Username," 2014. [Online]. Available: https://arstechnica.com/information-technology/2014/01/how-i-lost-my-50000-twitter-username/

[24] Q. Hu, T. Dinev, P. Hart, and D. Cooke, "Managing employee compliance with information security policies: The critical role of top management and organizational culture," *Decision Sciences*, vol. 43, no. 4, pp. 615–660, 2012.

[25] L. Hugick and J. Best, "Questionnaire length," *Encyclopedia of Survey Research Methods*, 2008.

[26] J. Humble, "Continuous delivery sounds great, but will it work here?" *Queue*, vol. 15, no. 6, p. 70, 2017.

[27] Internal Revenue Service, "Tax Information Security Guidelines For Federal, State and Local Agencies," 1998. [Online]. Available: http://www.unclefed.com/ForTaxProfs/irs-drop/1998/pub1075.pdf

[28] ——, "Publication 1075: Tax Information Security Guidelines For Federal, State and Local Agencies," 2016. [Online]. Available: https://www.irs.gov/pub/irs-pdf/p1075.pdf

[29] International Organization for Standardization, "Are you safe online? New ISO standard for cybersecurity," 2012. [Online]. Available: https://www.iso.org/news/2012/10/Ref1667.html

[30] ——, "How to Write Standards: Tips for standards writers," 2016. [Online]. Available: https://www.iso.org/publication/PUB100335.html

[31] K. Julisch, "Security compliance: the next frontier in security research," in *Proceedings of the 2008 New Security Paradigms Workshop*. ACM, 2009, pp. 71–74.

[32] B. Krebs, "Cyber incident blamed for nuclear power plant shutdown," *Washington Post, June*, vol. 5, p. 2008, 2008.

[33] K. Krippendorff, "Reliability in content analysis," *Human Communication Research*, vol. 30, no. 3, pp. 411–433, 2004.

[34] M. Lombard, J. Snyder-Duch, and C. C. Bracken, "Content analysis in mass communication: Assessment and reporting of intercoder reliability," *Human communication research*, vol. 28, no. 4, pp. 587–604, 2002.

[35] J. Long, B. Gardner, and J. Brown, *Google hacking for penetration testers*. Elsevier, 2011, vol. 2.

[36] MITRE Corporation, Personal communication, 2018.

[37] S. Nangle, Private Communication, Feb 2019.

[38] National Institute of Standards and Technology, Personal communication, 2018.

[39] M. Nieles, K. Dempsey, and V. Yan Pillitteri, "NIST Special Publication 800-12: An Introduction to Information Security," 2017.

[40] North American Electric Reliability Corporation, "NERC Sanction Guidelines," 2012. [Online]. Available: https://www.nerc.com/FilingsOrders/us/RuleOfProcedureDL/Appendix_4B_SanctionGuidelines_20121220.pdf

[41] ——, "CIP-007-6 Cyber Security Systems Security Management," 2014.

[42] ——, "Critical Infrastructure Protection Committee," 2018. [Online]. Available: http://www.nerc.com/comm/CIPC/Related%20Files%20DL/CIPC%20Roster%20as%20of%20February%202018.pdf

[43] N. I. of Standards and Technology, "Sp 800-53 rev. 5 (draft) security and privacy controls for information systems and organizations," *Special Publications*, 2019. [Online]. Available: https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/draft

[44] M. T. Orne, "On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications." *American psychologist*, vol. 17, no. 11, p. 776, 1962.

[45] PCI Security Standards Council, "Payment Card Industry Data Security Standard: Requirements and Security Assessment Procedures v3.2," 2016. [Online]. Available: https://www.pcisecuritystandards.org/pci_security/

[46] A. Peterson, *Cracking Security Misconceptions*. O'Reilly Media, Inc., 2013.

[47] C. Potts, "Software-engineering research revisited," *IEEE software*, vol. 10, no. 5, pp. 19–28, 1993.

[48] P. Puhakainen and M. Siponen, "Improving employees' compliance through information systems security training: an action research study," *Mis Quarterly*, pp. 757–778, 2010.

[49] R. D. O. A. E. A. V. Z. Y. K. B. S. R. Kasturi, Y. Sun, "TARDIS: Rolling Back The Clock On CMS-Targeting Cyber Attacks," in *Proceedings of the IEEE Symposium on Security and Privacy 2020*. IEEE, 2020.

[50] M. J. Schwartz, "Ransomware victims who pay cough up $6,733," Feb 2019. [Online]. Available: https://www.bankinfosecurity.com/ransomware-victims-who-pay-cough-up-6733-on-average-a-11994

[51] J. Stamp, J. Dillinger, W. Young, and J. DePoy, "Common vulnerabilities in critical infrastructure control systems," *SAND2003-1772C. Sandia National Laboratories*, 2003.

[52] R. Stevens, C. Ahern, D. Votipka, E. Redmiles, P. Sweeney, and M. Mazurek, "The battle for new york: A case study of applied digital threat modeling at the enterprise level," USENIX Association, 2018.

[53] A. Strauss, J. Corbin *et al.*, *Basics of qualitative research*. Newbury Park, CA: Sage, 1990, vol. 15.

[54] S. Subashini and V. Kavitha, "A survey on security issues in service delivery models of cloud computing," *Journal of network and computer applications*, vol. 34, no. 1, pp. 1–11, 2011.

[55] L. Tam, M. Glassman, and M. Vandenwauver, "The psychology of password management: a tradeoff between security and convenience," *Behaviour & Information Technology*, vol. 29, no. 3, pp. 233–244, 2010.

[56] T. W. Thomas, M. Tabassum, B. Chu, and H. Lipford, "Security during application development: an application security expert perspective," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018, p. 262.

[57] S. Tom, D. Christiansen, and D. Berrett, "Recommended practice for patch management of control systems," *DHS control system security program (CSSP) Recommended Practice*, 2008.

[58] United States Computer Emergency Readiness Team, Personal communication, 2018.

[59] United States Federal Reserve, "The Federal Reserve Payments Study: 2017 Annual Supplement," 2017. [Online]. Available: https://www.federalreserve.gov/paymentsystems/2017-December-The-Federal-Reserve-Payments-Study.htm

[60] U.S. Department of Health and Human Services, "IRS Safeguards and Publication 1075 Update," 2018. [Online]. Available: https://www.acf.hhs.gov/css/resource/irs-safeguards-and-publication-1075-update

[61] U.S. Department of the Army, "Field Manual 100-14 Risk Management," 1998.

[62] N. Virvilis, D. Gritzalis, and T. Apostolopoulos, "Trusted computing vs. advanced persistent threats: Can a defender win this game?" in *Ubiquitous intelligence and computing, 2013 IEEE 10th international conference on and 10th international conference on autonomic and trusted computing (uic/atc)*. IEEE, 2013, pp. 396–403.

[63] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

APPENDIX A
DATA SET

As a reference, all of our data can be viewed at https://ter.ps/hackcompli. There are five tabs within the spreadsheet: (1) all of the data that external experts validated, (2) all IRS P1075 findings, (3) all PCI DSS findings, (4) all NERC CIP 007-6 findings, and (5) a subset of the findings we analyzed for specified time durations and/or responsible roles.

APPENDIX B
EXPERT SURVEY

Participant is presented with consent form; Please check all that apply (you may choose any number of these statements): I am age 18 or older; I have read this consent form or had it read to me; I voluntarily agree to participate in this research and I want to continue to the survey.

Introduction: This survey will ask for you to assess the validity of an independent evaluation of [standard name]. Please be as candid and detailed as possible.

For each issue, please confirm:

1) If your organization followed the standard as written and nothing else, would your organization be vulnerable to this issue? (Yes/No/Possibly)
2) If yes or possibly ⇒ In your opinion, what is the likelihood of this vulnerability being exploited if standard is followed as written and nothing else? (Frequent - Occurs often, continuously experienced; Likely - Occurs several times; Occasional - Occurs sporadically; Seldom - Unlikely, but could occur at some time; Unlikely - Can assume it will not occur)
3) If yes or possibly ⇒ In your opinion, what is the severity associated with exploitation if standard is followed as written and nothing else? (Catastrophic - Complete system loss, major property damage, full data breach, corruption of all data; Critical - Major system damage, significant property damage, significant data breach, corruption of sensitive data; Moderate - Minor system damage, minor property damage, partial data breach; Negligible - Minor system impairment)
4) If yes or possibly ⇒ Is there past evidence of this vulnerability within your organization? (Yes/No/Maybe)
5) If yes or possibly ⇒ What would you recommend, based on your experience, to remedy this issue? (Open response)
6) If no ⇒ What additional policies, procedures, or defensive techniques does your organization use to mitigate this issue? (Open response)

End of survey: Does your organization allow waivers to the compliance standard? If yes, how frequently are they used? If no, does frequently does this create issues for your organization?

Demographics: What is the highest level of school you have completed or the highest degree you have received? Please estimate the number of years experience you have in the compliance and information technology fields. Please describe your work role and your interaction with compliance standards. Please estimate the organization size that you work in.