ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)

https://doi.org/10.17559/TV-20210122095638
Original scientific paper

A Security, Privacy and Trust Methodology for IIoT

Lifeng CHEN, Zhixiao YE, Shanyue JIN*

Abstract: The implements of IoT and industrial IoT (IIoT) are increasingly becoming the consensus with Industry 4.0. Relevant data-driven methodologies are typically concentrated on the scoring systems of CVE prioritization schemes, the scoring formulas of CVSS metrics, and other vulnerability impact factors. However, these prioritized lists such as the CWE/SANS Top 25 suffer from a critical weakness: they fail to consider empirical evidence of exploits. Considering the distinct properties and specific risks of SCADA systems in IIoT, this paper overcomes the inherent limitation of IIoT empirical research which is the sample size of exploits by collecting data manually. This study then developed an exploits factors-embedded regression model to statistically access the significant relationships between security, privacy, and trust-based vulnerability attributes. Through this data-driven empirical methodology, the study elucidated the interactions of security, privacy, and trust in IIoT with professional quantitative indicators, which would provide grounds for substantial further related work. In addition to the security privacy and trust regression analysis, this study further explores the impact of IoT and IIoT by difference-in-difference (DID) approach, applying bootstrap standard error with Kernel option and quantile DID test to evaluate the robustness of DID model. In general, the empirical results indicated that: 1) the CVSS score of vulnerability is irrelevant to the disclosure of exploits, but is positively correlated with CWEs by Density and CVE year, 2) among the exploits of SCADA-related authors, the more identical CWEs that exist in these exploits, the higher the CVSS score of the exploit CVE will be, and CVE year has a negative moderating effect within this relationship; 3) the CVSS scores of SCADA exploits have significantly decreased in comparison with non-SCADA after the promulgation of Industry 4.0.

Keywords: CVSS scores; SCADA; industrial IoT (IIoT); security privacy and trust

1 INTRODUCTION

1.1 Research Background

The proposal of Industry 4.0 was first publicly introduced in 2011 under an initiative to enhance the industrial competitiveness of German, and the official promulgation of this concept at the Hannover Messe in 2013 has attracted worldwide attention from different fields. The academic research on this topic focused more on Cyber-Physical Systems (CPS) than on the Internet of Things (IoT) in early stages [1]. Since factory hardware upgrades and return on investment have reached a certain equilibrium, industrial manufacturing capacity has approached saturation especially in last decade, thus relevant researchers begin to lean towards the field of IoT, including RFID, big data, fog/cloud computing, machine learning, prevention and predictive analysis. The Industrial Internet of Things (IIoT) refers to the application of IoT in manufacturing industries, the difference between IIoT and other IoT applications is that it focuses more on connecting machines in industries such as oil, gas, healthcare and electric utilities. IIoT system may be as complex as a fully automated large-scale production line that can track maintenance, production, even order and distribution through a huge multi-layer network. With the maturation of SCADA and related technologies, IIoT would logically become a new development trend in the era of Industry 4.0. Therefore, this paper is to analyze the security privacy and trust of IIoT, providing an empirical basis for various subsequent algorithms.

1.2 Security, Privacy, and Trust in Industrial IoT Overview

Industrial Internet of Things (IIoT) security represents a fusion of the physical and digital worlds. Major challenges will involve the complexity of integrating different IIoT systems, which in turn results in the multiplication of potential vulnerabilities and exploits [2, 3]. Impact factors represent another important challenge, compared to attacks on conventional computer systems; there are more physical assets while human lives would be directly affected by IIoT attacks [4]. Privacy has an impact on security and trust; Henderson and Charles [5] define privacy as an individual's right to control the allocation and utilization of personal information. Chen and Alan [6] define privacy concerns regarding unauthorized invasion and sharing of personal information. Despite the diversity of definitions of security and privacy, trust management in IoT and IIoT can be examined in two main structural styles, including social IoT [7-9] and service-oriented architecture (SOA)-based IoT [10, 11]. First, SOA-based IoT is a typical interaction between different IOT entities, and both service consumers and service providers could provide services and share resources within this system, representing the physical trust relationships between device entities [12, 13]. Second, trust-based social IoT includes traditional peer-to-peer social networks, structures which reflect the virtual trust relationships between people and devices [14, 15].

Some recent studies of security, privacy, and trust in cloud computing used network vulnerabilities and hacking attacks to analyze the subject [16, 17]. Many mathematical studies in IoT involved network vulnerabilities and hacker attacks, especially for the topic of IIoT, such as [18-20]. In addition, there have been numerous proposals for unifying vulnerabilities, exploits, and related security information with different abstract frameworks [21-23]. G. Stonebumer et al. [24] described the vulnerability as defects or weaknesses in security program, internal control and designing implementation that could be used to violate the security system. Extensive research on the current trend of security vulnerabilities and exploits would help to prevent and mitigate the impact of attacks [25]; previous security tools such as Fortify, Sonar Qube and Penetration testing could professionally evaluate related risks, but they fail to make an overall assessment from the perspective of management. Therefore, we have decided to use vulnerabilities and exploits as professional quantitative indicators for the empirical research on security, privacy, and trust in this study.

1.3 SCADA and Industrial IoT

The supervisory control and data acquisition (SCADA) system collects data from sensors and industrial equipment at remote area, thereby these data can be transferred to the central station for the system monitoring [26]. This system is the core remote monitoring technology of IIoT; it is closely related to the lives of citizens, and is widely used for networks in various fields of industry, such as power, oil, metallurgy, natural gas, railways, water supply, chemical industry, and other basic industries. These critical infrastructures are interrelated with each other, forming a massive, complex system that provides irreplaceable materials and services for national security and economic operation. Now, with the current wave of industrial revolution including Industry 4.0, Intelligent Manufacturing 2025, and IIoT, hacker attacks have shifted from ordinary networks to critical infrastructure. The damage caused by an attack on SCADA systems would be difficult to estimate, which may involve the following three types of serious impact:

1) If there is a safety problem in the industrial control system or the equipment is damaged, the important information from the operation technology (OT) system will be intercepted or interfered;

2) In the industrial control environment, the security and reliability of remote control system may be attacked by criminals or terrorists;

3) While the substantial control of target system is required, that may result in a real-time impact on the availability of operational system.

SCADA system security is a new challenge in the field of IIoT risk assessment, with the increasing level of automation, and network attacks will become easier than before when these systems are exposed to the internet or office networks. Security, privacy, and trust in IIoT will become a major issue affecting industrial production, people's lives, and even national defense. SCADA systems run roughly 80% of US utilities [27]. In 2015, a number of transformer substations in Ukraine were damaged, resulting in rolling blackouts affecting 225,000 people [28]. In 2017, many of Canada's critical infrastructures, including a series of government systems such as power plants and power transmission grids, encountered 25 hacking attacks within 2 months. Other governments may be covertly responsible for some of these attacks [29]. In Ukraine and Canada, SCADA comprises a huge and complex IIoT system that is responsible for controlling the smart grid, making it impossible to patch all vulnerabilities in the entire network. According to the 2020 IBM X-Force Threat Intelligence Index Report [30], ICS and OT infrastructure attacks have made a year-on-year increase of 2000% in the previous three years. Monitoring results show that most of these attacks utilized both SCADA and ICS hardware vulnerabilities to apply password spraying technology. North America and Asia have been the most severely attacked areas in the past year, with the greatest data losses consisting of 5 billion and 2 billion records, respectively. There are many taxonomic methods and network security frameworks that can reduce network risks, but most data-driven research in this field is focused on various scoring systems. SCADA is a unique software subclass with unique potential attack targets [31]. Recently, Gregory et al. [31] from the Massachusetts Institute of Technology (MIT) created a data-driven priority scoring schema based on empirical analysis on SCADA exploits to guide IIoT research in 2018. This paper aims to research on technical exploits publicly available in the databases searched, and these data could also account for the prevalence of exploits found in the wild, which would help to mitigate risk of IIoT vulnerabilities endangering social public safety.

1.4 Variable Identification and Classification

The Common Vulnerability Scoring System (CVSS) is published and updated by the National Vulnerability Database (NVD). It is designed to assess the severity of vulnerabilities, which can assist to determine the risk of required response. CVSS scoring system rates the urgency and security of vulnerabilities with a comprehensive value calculated by three matrix data modules: Base score metrics, Temporal score metrics, and Environmental score metrics. CVSS versions 1.0 and 2.0 were proposed in 2005 and 2007, respectively, then CVSS 3.0 was released in 2015. The basic score of CVSS is not only an authoritative score used to determine the severity, but also a comprehensive score that includes time and environmental indicators, so it is widely used both in academic research and practical operations. Some studies indicated that CVSS is not effective enough to fully reflect risks and vulnerabilities under certain circumstances, therefore scholars have created other scoring systems to supplement and improve CVSS scores. Allodi and Massacci [32] and Nayak et al. [33] published analyses on software vulnerabilities, with results suggesting that CVSS is not an accurate indication of software exploitability, but their research did not include the subcategories of IIoT. Although CVSS is a free and open industry standard, few works have taken the application of CVSS for IIoTi nto account; Gregory et al. [31] researched SCADA systems and noted that CVSS risk metrics could be adopted to reflect the severity of exploits involving the software subcategory of IIoT. Recently in 2019，Rehman et al. [34] analyzed the use of CVSS for IoT embedded systems, named $\mathrm{CVSS}_{\mathrm{IoT}}$, applying this approach to the actual supply chain network. They found that $\mathrm{CVSS}_{\mathrm{IoT}}$ is effective and capable of vulnerability assessment for traditional network systems; furthermore, it could also exploit the unique attributes of IoT and IIoT devices. In short, CVSS scores were employed to quantitatively analyze the modeling attack scenarios and suggest mitigation techniques, which are more effective than the alternatives. These scores provide a foundation for developing and measuring IT security metrics; thus, CVSS is the main professional quantitative parameter and official authoritative index for measuring IIoT security and risk.

CVSS scores are also useful for measuring the potential impact of the exploitation of these vulnerabilities [35, 36]. Gallon and Bascou [37] suggested using attack graph theory in conjunction with the CVSS framework to analyze vulnerabilities. An exploit occurs when a hacker takes advantage of a flaw or error; a vulnerability does not necessarily indicate the possibility of an exploit, but there must be a vulnerability when exploitation occurs. Vulnerabilities are flaws or weaknesses in programs, algorithms, and protocols, but not all vulnerabilities are exploitable. The theoretical existence of a vulnerability is not necessarily sufficient for an attacker to threaten your system. An exploit is a piece of code that controls a target system by triggering one or several vulnerabilities. The attack code usually releases an attack payload that contains the code that the attacker aims to execute. Exploit code could be carried out locally or remotely, allowing an attacker to manipulate a computer at long distance and execute arbitrary code under ideal conditions. Therefore, an attacker can remotely control a host device even without the owner visiting a website, clicking on files, or opening email attachments. To deal with the problem of privacy leakage, a number of exploits scenes and corresponding solutions for preserving privacy have recently been introduced [38]. In general, exploits should be the core professional indicators of IIoT security and privacy. In this paper, CVSS and exploits will be the key quantitative parameters of our data-driven research on security and privacy in IIoT.

1.5 CVE and CWE

Common Vulnerabilities & Exposures (CVE) is the most famous dictionary of vulnerability security; it is also a list of standardized coding for known vulnerabilities. The database of CVE is an international organizational project with the participation of enterprises, government and academia. It could perform a preliminary classification to define the common software defects. The Common Weakness Enumeration (CWE) database is a strategic software security project established by MITRE and funded by the National Computer Security Department (NCSD) of the US government. CWE is the most authoritative research project of source code defect, and it further identifies and classifies the security code based on the CVE database. MITRE is a non-profit organization that mainly provides systems engineering, information and research methodology to support the US government [39]. The NVD recognizes CWE as a classification mechanism to differentiate CVEs by the type of vulnerability [40]. CWE includes all types of flaws and weaknesses in software security. We should regard CWE as an analytical summary of all vulnerabilities, in which the source of a given CVE can be found. From the perspective of CWE, a CVE entry results from defects described by CWE in the code layer and the application layer. MITRE regularly releases CVE entries to the software vulnerability database and updates the CWE database in real time. Both the CVE/CWE databases and the related common standards for identification, mitigation, and prevention of software vulnerabilities are widely cited in the industry. In addition, MITRE and the SANS institute created a ranking list of the most dangerous CWEs, named "CWE/SANS Top 25". These prioritized lists are intended to identify the most serious types of software vulnerabilities and guide professional research in relevant fields.

2 DATA AND METHODOLOGY

2.1 Data Collection

The National Vulnerability Database (NVD) [41] is a national database in America that supports the standards of CVE. The Exploit Database (EDB) [42] is considered as a general database for existing exploits; which is adopted by extensive vulnerability professionals and security researchers [43]. NVD and EDB represent the top 1 of vulnerabilities and exploits database, respectively, and they also demonstrate how CVSS metrics can provide better insight into exploits [44]. This study chose these two of the most authoritative databases, NVD and EDB, to obtain vulnerabilities and exploits. If EDB has a vulnerability number that is missing in NVD, we then go to MITRE's CVE [45] for confirmation.

According to the keyword search method of Gregory et al. [31], there was only one SCADA CVE before the year of 2008 in NVD. Thus, our research interval was set to be 2008-2019, and we extracted a total of 229 SCADA-related CVEs from NVD. The empirical IIoT research by Gregory et al. [31] had the limitation that they were only able to find 52 SCADA-related exploits, which led to few samples in their research. When we used keywords to search from the home page of EDB, the results of 48 exploits were consistent with the search results of 52 exploits from Gregory et al. [31] in 2018. They used a web scraper to acquire data from three databases; the other two databases, CVE Details and the Metasploit code database, have few additional exploits. However, we found 130 exploits while further searching the EDB website using the "Content" filter. Comparing these 130 exploits with the first search results by CVE number, it was found that all CVE numbers from the 48 exploits obtained in the first search were included. The 82 exploits newly discovered in the second search were reconfirmed by manually searching the internal contents again using the keyword SCADA. As a result, we found that more than 50% of these exploits directly affect SCADA systems and 99% are SCADA-related, except for the EDB-ID of 8954, whose details included the keyword incidentally as part of an unrelated phrase, but the CVE number of this entry is one of 229 SCADA-related CVEs extracted from NVD. After another comparison with the CVE content, it was confirmed that this exploit is also SCADA-related, and we also reconfirmed the data and comparison results during this process, thus yielding the conclusion that these 130 exploits are all SCADA-related. We used the method of manual data searching to bypass the limitation of the first data-driven research on IIoT from MIT in 2018, and also cleared the most significant obstacle regarding samples for our further empirical research. After removing the CVE-default exploits, the total number of 90 SCADA CVEs was discovered to have 99 associated exploits. These 99 exploits included 35 authors, 5 types, and 9 platforms. Combining these 90 CVEs extracted from EDB and 229 CVEs extracted from NVD, we finally acquired 291 SCADA CVEs as the SCADA group after removing duplicate values and merging.

The non-SCADA data acquisition method of Gregory et al. [31] was random. In order to ensure that our research is traceable and reproducible, and that exploits from specific authors should be more closely compared, we used the names of the authors from the previous 99 exploits as keywords to obtain related non-SCADA exploits. Finally, after removing the CWE-default data (as in the work of Gregory et al. [31]), and removing the 5 abnormal authors with the largest number of exploits, thus counting the exploits of 30 authors, a total number of 226 CVEs with exploits were obtained as the non-SCADA control group. In comparison with the previous random methodology, our method could be considered more scientific and rigorous.

2.2 Comparison and Analysis

Firstly, according to the variable setting of Gregory et al. [31] and Allodi and Massacci [32], we used 291 SCADA CVEs collected from NVD, EBD, and MITRE's CVE to regain the variables of SCADA CWEs by density and CWEs by exploit density. The Top 5 comparison of these two variables is shown below:

Table 1 Top 5 SCADA CWEs by density comparison

Gregory et al.			Our research		
RANK	CWE-ID	DENSITY	RANK	CWE-ID	DENSITY
1	CWE-119	0.244	1	CWE-119	0.223
2	CWE-200	0.105	2	CWE-22	0.11
3	CWE-20	0.1	3	CWE-20	0.072
4	CWE-79	0.063	4	CWE-79	0.062
5	CWE-22	0.062	5	CWE-200	0.055

The metric of CWEs by density, i.e. the vulnerability density of each CWE, is the key to establish a prioritization list. For instance, there were 65 CVEs in CWE-119; this number was divided by the total number of SCADA vulnerabilities, 291, to determine the CWE density of 22.3%. The difference of this variable is due to the discrepancy in SCADA CVE samples. There is an updating lag phenomenon between these two CVE databases, since vulnerabilities appear faster than the databases could be updated [46]. In addition to the discrepancy between different CVE databases, our dependent variable of CVSS is published and updated by NVD, so we chose the official data from the government website for our quantitative analysis. From the distribution of CWEs by density, our SCADA vulnerabilities are largely consistent with the data of Gregory et al. [31]. Then, we made another comparison of exploits:

Table 2 Top 5 SCADA CWEs by exploit density comparison

Gregory et al.			Our research		
RANK	CWE-ID	DENSITY	RANK	CWE-ID	DENSITY
1	CWE-119	0.615	1	CWE-119	0.333
2	CWE-200	0.115	2	CWE-200	0.081
3	CWE-20	0.058	3	CWE-22	0.121
4	CWE-79	0.039	4	CWE-89	0.061
5	CWE-22	0.039	5	CWE-20	0.051
			5	CWE-79	0.051

In consideration of these exploits being readily available for attackers, the density of CWE exploits would provide a better evaluation of operational risk [31]. A similar calculation for CWEs by exploit density is demonstrated as follows: there were 33 exploits associated with CVEs in CWE-119; this number was divided by the total number of SCADA exploits, 99, yielding an exploit density for CWE-119 of 33.3%. The difference in this variable lies in the fact that our 99 SCADA exploits represent a much larger sample size than that of the 52 exploits considered by Gregory et al.; thus, our samples are newer and more comprehensive than theirs. After further comparison with specific CWE numbers, the research of Gregory et al. [31] may have missed some data for CWE-22 and CWE-89, and the biggest difference appears with CWE-119, ranked at 1st place. Then, observing the data on specific CWEs, we discover that more than 90% of CWE-119's CVE-Year statistic was before 2015; thus, we infer that Gregory et al. missed significant data from the past 5 years, and we should introduce CVE-Year as the control variable while establishing the model. Overall, their Top 5 list of CWEs by exploit density is essentially in line with ours. Furthermore, we downloaded the latest CWE Top 25 2019 [47] from the official website and compared the top 6 of CWE-119, CWE-79, CWE-20, CWE-200, CWE-125, and CWE-89 with the Top 5 in Tab. 1 and Tab. 2. Surprisingly, the top 6 list of the CWE prioritization from 2019, calculated from a series complex scoring formulas made by MITRE, is very similar to the top 5 results given by our SCADA variables. Therefore, by following our Top 5 of CWEs by density and CWEs by exploit density, equivalent to the features of the CWE prioritization list, SCADA developers could focus on these weaknesses occurring in IIoT environments. Moreover, educators and scholars should use this approach in multiple methods, such as in the situation of teaching IIoT weaknesses, an educator can concentrate on the Top 5 SCADA CWEs by density and CWEs by exploit density.

In the second step, we refer to the research on all software vulnerabilities from Allodi and Massacci's work [32] and the research on IIoT using the linear regression of CVSS scores on SCADA exploits from Gregory et al. [31] to analyze the relationships between SCADA exploits, CWEs by density, and CVSS. We introduce CVE-Year as the control variable for Model 1 based on our previous data analysis; the research objects of Model 1 are 291 SCADA CVEs.

Table 3 Parameter list of Model 1

Variable type	Variable name	Variable code	Variable definitions
Dependent variable	CVSS	*CV*	If 2008-2015, only V2; If 2015-2019, take the average of V2 and V3
Independent variable	Exploit_ Dummy	*ED*	If exploit = 1; If no exploit = 0
	CWEs by Density	*CD*	The number of certain CWEs divided by the total number of CVEs
Control variable	CVE-Year	*YE*	2020-X; X is the CVE year of certain vulnerability

Model 1: $CV = \beta_0 + \beta_1 ED + \beta_2 CD + \beta_3 YE + \varepsilon$;
The regression results are as follows:

Table 4 Regression results of Model 1

	Estimate	Std. Error	*t*-value	$Pr > \|t\|$
Intercept	6.2510***	0.219	28.51	< 0.0001
CD	3.9630***	1.406	2.82	0.005
YE	0.1060***	0.037	2.9	0.004
ED	0.3330	0.237	1.41	0.16
Adj. R^2	0.0899		No. of Obs	291

Note: ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

The regression results illustrate that the linear relationship between exploit disclosure and CVSS is not significant, but the linear relationship between CWES by density and CVSS is very significant and positive, indicating that the more identical CWE vulnerabilities exist in SCADA vulnerabilities, the higher the CVSS score of the CVE will be.

The Adj. R^2 of this model is 0.0899, in comparison with the Adj. R^2 found by Gregory et al. [31] of 0.074 and the Adj. R^2 found by Allodi and Massacci [35] of 0.098; the R^2 values of these three models are consistent with each other. Since the disclosure of exploits is a sensitive topic involving national security and privacy, some governments have already invested extensive financial resources to protect the network security from the expose of individuals and other nations [48]. And the R-squared results above have demonstrated a low correlation between the disclosure of exploits and CVSS scores, which is also a major obstacle for current empirical research on IIoT. Previously, we have explained that exploits are core factors for IT professional research, and also the key parameters that affect security and privacy in IIoT. Tran and Wei [49] used the model of "Trustiness = Privacy Concern + Security Concern"to quantitatively analyze the correlation between those factors, and found that privacy concern usually plays an important role in trust, using a sociological questionnaire. Our Model 1 has made an empirical attempt, using the CVSS and exploits as quantitative indicators of security and privacy in IIoT. We convert exploits based on quantitative data into variables of trust in IIoT for our data-driven empirical research. Our research objects are 99 SCADA exploits and 226 non-SCADA exploits found by searching by the authors of the SCADA exploits in EDB, yielding a total of 325 exploits.

2.3 IIoT Security, Privacy, and Trust Modeling for SCADA Systems

Due to the limitation that not all exploits are disclosed, we attempt to find more quantitative variables and establish a model to explore this relationship from the limited available data. If the attackers could understand the system better than the humans in the system do, they could easily find a weakness to exploit. Schneier [50] indicated that the human in the network often subordinated to the machine part on account of "The computer is always right". The literature review in this paper has summarized that trust in IoT includes social IoT and SoC IoT; to provide a tentative definition, the quantitative indicators of trust in IIoT could be defined based on the physical trust relationship between machine entities and the virtual trust relationship between human and machine entities. Wang et al. [51] introduced an improved CVSS framework by adding the system information to gain more comprehensive and reliable score, this kind of information being not limited to server type, operating system and so on. Since our research objects are 325 SCADA-related exploits, referring to the recommendations of the above [50, 51], we use the same variable setting method that we used for CWEs by exploit density to create a new variable from human and machine entities, then bring these variables of trust in IIoT into the model. In this study, we created "Platforms (operating systems) by exploit density" and "Types by exploit density" as the professional quantitative indicators to measure social IIoT. Then, we regard "Authors by exploit density" as the quantitative indicator of SoC in IIoT. Finally, we added a dummy variable of SCADA_Dummy to further explore the impact of IIoT on security, privacy, and trust.

Model 1 has proven a significant relationship between CD and CVSS. Since our non-SCADA data were found by searching by the authors of SCADA exploits in EDB, the metric of CWES by exploit density (CED) is equal to CD in the control group of non-SCADA exploits. As a result, the overall correlation coefficient between the CED and CD is 0.961, which would indicate multiple linear relationships. Therefore, in Model 2, we replaced CD with CED, assuming that CED should have a similar relationship with CVSS. Moreover, we use the SCADA_Dummy to compare the CVSS scores between SCADA and non-SCADA exploits. Finally, we set Year_Dummy as the adjustment variable for the auxiliary analysis of CED and CVSS.

Table 5 Parameter list of Model 2

Variable type	Variable name	Variable code	Variable definitions
Dependent variable	CVSS	*CV*	If 2008-2015, only V2; If 2015-2019, take the average of V2 and V3
Independent variable	CWES by Exploit Density	*CED*	The number of certain CWE divided by the total number of exploits
	Types by Exploit Density	*TED*	The number of certain types divided by the total number of exploits
	Platforms by Exploit Density	*PED*	The number of certain platforms divided by the total number of exploits
	Authors by Exploit Density	*AED*	The number of certain authors divided by the total number of exploits
	SCADA_ Dummy	*SD*	If SCADA = 1; If non-SCADA = 0
Control variable	Year_ Dummy	*YD*	If year within 2014-2019, 1; If year within 2008-2014, 0

Model 2: $CV = \beta_0 + \beta_1 TED + \beta_2 PED + \beta_3 AED + \beta_4 CED + + \beta_5 YE + \beta_6 CED \cdot YD + \beta_7 SD + \varepsilon$;

The regression results are as follows:

Table 6 Regression results of Model 2

	Estimate	Std. Error	*t*-value	*Pr* > |*t*|
Intercept	8.988***	0.63	14.27	< 0.0001
TED	−5.849***	1.544	−3.79	0.0002
PED	1.563***	0.576	2.71	0.007
AED	−4.112**	2.027	−2.03	0.0434
CED	4.483***	1.073	4.18	< 0.0001
YD	−0.13***	0.041	−3.14	0.0018
CED_YD	−1.134***	0.42	−2.7	0.007
SD	0.755***	0.252	3	0.003
Adj. R^2	0.22		No. of Obs	325

Note: ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

We took the values of CVSS V2 and V3 as the dependent variables to make the other two confirmatory regressions; both results were the same as the regression results of Model 2. Moreover, we tested the variance inflation factor (VIF) of variables, all values are less than

2, and the results show that there was no multi-collinearity between variables. Finally, we reached the following conclusions:

1) Among the exploits of SCADA-related authors, the more exploits from the identical platform and the fewer exploits from the same author and type, the higher CVSS score of the exploit's CVE will be.

2) Among the exploits of SCADA-related authors, the more identical CWEs that exist in these exploits, the higher the CVSS score of the exploit CVE will be, and this relationship is more significant in older exploits.

3) Among the exploits of SCADA-related authors, the vulnerability of SCADA exploits is higher than that of non-SCADA exploits.

2.4 Further Verification

The concept of Industry 4.0 and IoT was officially introduced in 2013 in order to verify whether the application of IoT has an impact on CVSS scores of SCADA. This paper used the DID model and the 325 exploits from Model 2 to test the robustness of our IIoT Security, Privacy, and Trust Model. The dependent variable being CVSS scores, we took the values of CVSS V2 and V3 as the dependent variables to make the other two confirmations. The dummy variable of the experimental group is treated (treated = 1 indicates SCADA exploits; treated = 0 indicates non-SCADA exploits). Since the conception of IIoT proposed in recent 3 years is too new to be recognized, the dummy variable in the experiment of period which should is consistent with the moderator of Year_Dummy from Model 2 (period = 1 is when CVE-year within 2014-2019; period = 0 is when CVE-year within 2008-2013). Finally, we utilize the TED, PED, CED and AED as control variables in the DID test. This estimation assumes that time-invariant unobserved heterogeneity and exclusively contaminates the identification strategy.

Table 7 Regression results of DID Model

	CVSS V2	CV(V2 & V3)	CVSS V3
Baseline	1.455*** (4.71)	1.453*** (4.84)	1.452*** (4.8)
Follow-up	−0.033 (0.07)	−0.016 (0.04)	0.002 (0)
Diff-in-Diff	−1.488*** (2.72)	−1.469*** (2.76)	−1.45*** (2.7)
TED	−6.97*** (−4.381)	−6.404*** (−4.135)	−5.839*** (−3.743)
PED	0.999 (1.618)	0.935 (1.555)	0.871 (1.437)
CED	2.85*** (2.921)	2.963*** (2.969)	3.004*** (2.989)
AED	−5.063** (−2.432)	−4.972** (−2.454)	−4.882** (−2.392)
No. of Obs.	325	325	325
R^2	0.25	0.23	0.21

Note:***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. *T* statistic in parentheses

The results in Tab. 7 verified the proposal of Industry 4.0 and IoT is negatively related to the CVSS scores. The baseline rows contain information on the mean outcome for each group as well as each group's single difference is around 4.8, and the same information is displayed for the follow-up period with no significance of CVSS scores after the proposal of IoT. The coefficient of Diff-in-Diff is the DID treatment-effects estimate, implying a decrease in the CVSS scores by 1.45 in comparison of SCADA with non-SCADA after the implementation of Industry 4.0. The *R*-square of this model is quite high, and all of significance levels from DID test support the establishment of our Security, Privacy, and Trust Model, except the PED. According to the time the concept of IIoT was formed, we changed the baseline of period to 2018, PED has significance at 10% level accompanied by a star in this case, and other significances have not changed at the same time, which further proves our IIoT model is valid. Moreover, we make other optional DID tests to evaluate the robustness of this model, including bootstrapping the standard error using Kernel test and quantile DID (QDID); although the results are unstable, the significance is consistent with the above table. These results indicate the implementation of Industry 4.0 has a significant impact on CVSS, TED, PED and CED, thus further confirming the robustness of our security privacy and trust model

3 CONCLUSION

This methodology represents more of a managerial exploit than technical scoring analysis; our unique contributions are significant for these critical infrastructure operators working in context of IIoT, and security, privacy, trust researchers investigating SCADA systems. There are three major groups that can benefit from this paper: IT experts, scholars, and critical urban infrastructure system architects. Furthermore, data-driven research methodology in IIoT should not be limited to the current scoring systems: the vulnerability impact factors formula, CWE prioritization, and ranking in IT professional fields. The current advanced statistical science should also be introduced to quantitatively evaluate SCADA software risk for IIoT devices. We also provide an effective example and lay a foundation for subsequent research on security, privacy, and trust in IIoT. Future works have the following research opportunities related to this topic:

1) There are 35 SCADA authors in EDB. Our non-SCADA data removed the authors within the top 5 quantity of exploits as outliers to obtain 226 non-SCADA exploits as the control group. The authors of LiquidWorm, Luigi Auriemma and Metasploit all have more than 400 exploits each. Even if the CVE-default and CWE-default exploits are eliminated, the number of exploits per author should be sufficient as a separate control group to perform empirical analysis on 99 SCADA exploits. Follow-up research could also choose these 5 authors or their recent exploits to perform empirical analysis to confirm our findings. If the conclusions are different, our research results will be outdated [52].

2) Our vulnerability attributes include platform, type, and author; in addition, language is another basic requirement for device-to-device communication. Low-level programming languages such as C are prevalent in SCADAoperating environments, and Buffer overflows (CWE-119) are common in operating environments of C language system. However, Rust is a safe language that can improve memory efficiency; if future SCADA designers will be programming with Rust language, CWE-119 can be

solved effectively, thus reducing the potential attack possibility in IIoT SCADA systems [53]. Therefore, when designing future SCADA systems, it is imperative to use a memory-safe programming language. Due to limitations of data acquisition, we did not introduce programming languages into the model. Subsequent research could take this variable as another SOA-based parameter to further analyze trust in IIoT.

3) The CWE/SANS Top 25 list is intended to identify the greatest software vulnerability types. Our research has proved that the factors "Types by exploit density" and "CWEs by exploit density" have significant relationships with CVSS. Thus, a new model should be established to investigate the relationship between "Types by exploit density", "CWEs by exploit density" and the CWE/SANS Top 25 in the IIoT subclass.

4) Future research could apply our results and empirical evidence from exploits to the scoring mechanisms applied for the CVE/CWE ranking system in IIoT, or to a data-driven research for evaluating security, privacy, and trust in SCADA systems.

4 OUR CONTRIBUTION

Previous data-driven methodologies of IoT research have mainly focused on the scoring systems of CWE/CVE prioritization schemes, the scoring formula of CVSS metrics, and other vulnerability impact factors. However, the principal weakness of CWE/CVE prioritized lists is that they fail to consider empirical evidence of exploits. Prioritization using statistics is more effective than ranking by professional algorithms, because data-driven research can explain the universality of vulnerabilities existing in real world [31]. Through manual data acquisition, we surpassed the key limitation of the first data-driven empirical study on IIoT from MIT in 2018, that is, the sample size of SCADA exploits. Our study confirms their research and evaluates SCADA software risks for IIoT devices more comprehensively and systematically. Through the analysis of empirical results, we statistically validate the significant relationship between CWEs by density and CVSS score in SCADA systems. Therefore, future CVSS scoring of SCADA vulnerabilities could use the CWE database as a reference to update the data and scoring systems according to our empirical research. Furthermore, based on the parameter settings of previous research, we establish the model of security, privacy, and trust to fill the vacancy of empirical research on the IIoT, further evaluating the correlation of exploit and CVSS scores for the SCADA software subclass. Because the quantity of SCADA exploits available is very limited, we use the various operating parameters from exploits to establish the model, with the intention of exploring interactions between vulnerability attributes, including CWE, type, platform, author, CVE year, and CVSS. With this dynamic model, we statistically evaluate the correlation of exploit-based vulnerability attributes with CVSS scores, which lays the empirical basis for professional research on scoring systems for IIoT. Due to the existence of significant relationships between these variables, this quantitative method also shows the interactions and influences between security, privacy, and trust in IIoT. In conclusion, we make the following contributions in this paper:

1) By following our Top 5 of CWEs by density and CWEs by exploit density, SCADA developers will be able to significantly reduce the number of flaws and weaknesses that occur in the context of IIoT.

2) Our research provides an empirical foundation for future scoring systems of vulnerability impact factors and CWE/CVE prioritized lists in the IIoT subclass. Moreover, according to our empirical research results, future CVSS and CWE of SCADA vulnerabilities could also use Exploit Database as a reference to update the data and scoring systems of SCADA-related vulnerabilities.

3) This article creates an empirical basis for the research of scoring systems and prioritized lists for the IIoT software subclass, and also serves as an example for research on other subclasses. This methodology provides grounds for substantial further work to evaluate the correlation of security, privacy, and trust.

4) SCADA IIoT system developers and designers can use our model to easily identify the principal attributes of vulnerabilities based on exploitation risk, such as vulnerability types by exploit density, enabling them to design systems without these vulnerabilities in the context of security, privacy, and trust.

5 RESEARCH IMPLICATIONS

This study explained that the disclosure of exploits is irrelevant to CVSS score because information disclosure is a sensitive issue. On the other hand, most hacking attacks are motivated by the potential for commercial gains. CVSS is an official index obtained by integrating massively impact factors, and both the V2 and V3 versions are calculated based on three matrix modules with continuous adjustments of the algorithm, but the relationship between CVSS and exploits is not clear yet. As previously mentioned, more and more network attacks have shifted from general targets to public infrastructures, enterprises, and governments, especially now that electronic wallets are becoming more and more popular. CVSS is widely embraced and applied, and it is a mandated metric adopted worldwide to evaluate the security of payment card systems [54]. In the era of IIoT, if CVSS could add a new commercial matrix within the context of security, privacy, and trust based on empirical evidence, the comprehensive reference value of the CVSS score would definitely increase. If future versions of CVSS could take business factors and financial data into calculation, it would also provide a critical basis for IIoT developers and designers to ensure appropriate security, privacy, and trust in their environments.

Our research objects are taken from authoritative databases of NVD and CVSS, while the updating of these two databases is performed by the US government in cooperation with world-renowned IT companies. The United States has already achieved the leading position in the era of Industry 4.0 as a result of its strong IT foundation. The vulnerability databases of other countries, such as CNNVD in China, are based on those of the United States. The database of CNNVD has not been developed for a considerable period, and its data can only be applied by related enterprises. There is limited information available for both companies and individuals, and enterprises can only obtain limited professional information, such as vulnerability reports, patches, and so on. Moreover, the low enthusiasm of giant IT companies and academia for this project will relegate the database to becoming a government image project in the long term. SCADA systems are becoming the lifeblood of countries, as well as the safety guarantee of most industrial facilities in the era of Industry 4.0 with the development of IIoT; therefore, other countries should ideally establish SCADA subclass databases based on the US database in advance, in order to ensure the security, privacy, and trust of their national networks. The security vulnerabilities of IoT devices have attracted hackers to develop sophisticated tools to hack connected systems for financial and political gain [55]. Disregarding political factors, the CVSS of SCADA vulnerability should add a business data matrix as a supplement and improvement to the current scoring system. A valuable vulnerability database and evaluation system would be established based on the national background of each respective country, rather than simply adding a country code prefix to CVE numbers from the US. To ensure free and open industry standards, CVSS would not necessarily involve national security, but the lack of commercial data or financial evidence in the risk evaluation system may compromise the accurate assessment of potential impacts. Because commercial value is the motivation for most hacker attacks, in the future CVSS could play an important role in security, privacy, and trust environments with the development of online banking. This would be a government project involving considerable professional knowledge and software skills, therefore also requiring the interdisciplinary cooperation of experts and scholars in various professional fields.

Acknowledgments

We would like to thank the editor and the anonymous reviewers for their valuable comments and suggestions.

6 REFERENCES

[1] Chen, L. F. & Jin,S. Y.(2020). A study on the direction of China's industry 4.0. *Asia Life Sciences, 29*(1), 379-387.

[2] Hubbard, D. W. & Seiersen, R. (2016). *How to measure anything in cybersecurity risk.* Chichester, UK: Wiley. https://doi.org/10.1002/9781119162315

[3] See https://www.gsma.com/iot/future-iot-networks/iot-security-guidelines/

[4] See https://www.schneier.com

[5] Henderson, S. & Snyder, C. (1999). Personal information privacy: implications for MIS managers. *Information & Management*, 213-220. https://doi.org/10.1016/S0378-7206(99)00019-1

[6] Chen, K. & Rea, A. (2004). Protecting personal information online: a survey of user privacy concerns and control techniques. *The Journal of Computer Information, 85.*

[7] Chen, I., Bao, F., & Guo, J. (2016a). Trust-based service management for social internet of things systems. *IEEE Transactions on Dependable and Secure Computing*, 684-696. https://doi.org/10.1109/TDSC.2015.2420552

[8] Kogias, E. K., Voutyras, O., & Varvarigou, T. (2016). TRM-SIoT: a scalable hybrid trust & reputation model for the social Internet of Things. *Emerging technologies and factory automation (ETFA), 2016 IEEE 21st international conference*. Berlin. https://doi.org/10.1109/ETFA.2016.7733612

[9] Nitti, M., Girau, R., & Atzori, L. (2014). Trustworthiness management in the social internet of things. *IEEE Transaction on Knowledge and Data Engineering*, *26*(5), 1253-1266. https://doi.org/10.1109/TKDE.2013.105

[10] Chen, I., Guo, J., & Bao, F. (2016b). Trust management for SOA-based IoT and its application to service composition. *IEEE Transactions on Services* Computing, *9*(3), 482-495. https://doi.org/10.1109/TSC.2014.2365797

[11] Guinard, D., Trifa, V., Karnousko, S., Spiess, P., & Savio, D. (2010). Interacting with the SoAbased internet of things: discovery, query, selection, and on-demand provisioning of web services. *IEEE transactions on Services Computing*, *3*(3), 223-235. https://doi.org/10.1109/TSC.2010.3

[12] Xuan, S. & Kim, D. (2018). A hierarchical IoT federation in architecture based on local cloud, platform and middleware for massive context data acquisition in multiple sensor networks. *International Journal of Grid and Distributed Computing, 11*(3), 1-10. https://doi.org/10.14257/ijgdc.2018.11.3.01

[13] Bhattacharyya, D. (2018). Space and security issues in cloud computing: A review. *International Journal of Security and Its Applications, 12*(6), 37-46.

[14] Verma, G. & Sushil, R. (2018). Secured identity management system for preserving data privacy and transmission in cloud computing. *International Journal of Future Generation Communication and Networking, 11*(1), 23-36. https://doi.org/10.14257/ijfgcn.2018.11.1.03

[15] Ullah, I. & Kim, D. (2018). IoT resource management using direct discovery mechanism in OCF framework. *International Journal of Grid and Distributed Computing, 11*(5), 1-10. https://doi.org/10.14257/ijgdc.2018.11.5.01

[16] Xiao, Z. & Xiao, Y. (2013). Security and privacy in cloud computing. *IEEE Communications Surveys & Tutorials, 15*(2), 843-859. https://doi.org/10.1109/SURV.2012.060912.00182

[17] Robinson, N., Valeri, L., Cave, J., Starkey, T., Graux, H., Creese, S., & Hopkins, P. (2010). *The cloud: Understanding the security, privacy and trust challenges.* Santa Monica, CA: Rand. https://doi.org/10.2139/ssrn.2141970

[18] Thirukkumaran, R. & Muthu, P. (2018). Survey: security and trust management in internet of things. *2018 IEEE global conference on wireless computing and networking (GCWCN*), 131-134. Lonavala, India. https://doi.org/10.1109/GCWCN.2018.8668640

[19] Zhan, J., Fan, X., Cai, L., Gao, Y., & Zhuang, J. (2018). TPTVer: A trusted third party based trusted verifier for multi-layered outsourced big data system in cloud environment. *China Commumity, 15*(2), 122-137. https://doi.org/10.1109/CC.2018.8300277

[20] Dua, A., Tyagi, V., Patel, N., & Mehtre, B. (2019, November 21-22). IISR: A secure router for IoT Networks. *2019 4th international conference on information systems and computer networks (ISCON)*, 636-643. Mathura, India. https://doi.org/10.1109/ISCON47742.2019.9036313

[21] Cheng, F., Roschke, S., Schuppenies, R., & Meinel, C. (2009). Remodeling vulnerability information. *Proceedings of the international conference on information security and cryptology*, 324-336. Beijing. https://doi.org/10.1007/978-3-642-16342-5_24

[22] Moreira, E. S., Martimiano, L. A. F., Brandão, A. J., & Bernardes, M. C. (2008). Ontologies for information security management and governance. *Information Management & Computer Security, 16*(2), 105-165. https://doi.org/10.1108/09685220810879627

[23] Tian, H., Huang, L., Zhou, Z., & Zhang, H. (2003).Common vulnerability markup language. *Proceedings of the First International Conference on Applied Cryptography and*

Network Security, Lecture Notes in Computer Science, 2846, 228-240. Kunming, China. https://doi.org/10.1007/978-3-540-45203-4_18

[24] Stoneburner, G., Goguen, A., & Feringa, A. (2002). *Risk management guide for information technology systems*, Recommendations of the national institute of standards and technology (NIST), special publication 800-30. Washington, DC: US Government Printing Office, 54. Retrieved from http://csrc.nist.gov/publications/nistpubs/800-30/sp800-30.pdf

[25] Chang, Y., Zavarsky, P., Ruhl, R., & Lindskog, D. (2011). Trend analysis of the CVE for software vulnerability management. *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, 1290-1293. Boston, MA. https://doi.org/10.1109/PASSAT/SocialCom.2011.184

[26] National Communications System (2004). *Supervisory control and data acquisition (SCADA) systems* (Technical information bulletin 04-1). Arlington, VA: Author.

[27] Slay, J. & Miller, M. (2008).Lessons learned from the maroochy water breach. *Critical infrastructure protection*, 73-82. Boston, MA: Springer. https://doi.org/10.1007/978-0-387-75462-8_6

[28] Electricity Information Sharing Analysis Center (2016). *Analysis of the cyber attack on the Ukrainian power grid, defense use case*. Washington, DC: Author.

[29] See https://www.freebuf.com/news/94873.html

[30] See https://www.ibm.com/

[31] Gregory, F., Caldera, C., & Shrobe, H. (2018). IoT cyber security risk modeling for SCADA systems. *IEEE Internet of Things Journal, 5*(6), 4486-4495. https://doi.org/10.1109/JIOT.2018.2822842

[32] Allodi, L. & Massacci, F. (2012). A preliminary analysis of vulnerability scores for attacks in wild: The ekits and sym datasets. *2012 ACM workshop on building analysis datasets gathering experience returns for security* (BADGERS), 17-24. Raleigh, NC. https://doi.org/10.1145/2382416.2382427

[33] Nayak, K., Marino, D., Efstathopoulos, P., & Dumitra, T. (2014). Some vulnerabilities are different than others. *Research in attacks, intrusions and* defenses, 426-446. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-11379-1_21

[34] Rehman, A., Gondal, I., Kamruzzuman, J., & Jolfaei, A. (2019). Vulnerability modelling for hybrid IT systems. *2019 IEEE international conference on industrial technology (ICIT)*, 1186-1191. https://doi.org/10.1109/ICIT.2019.8755005

[35] Ali, A., Zavarsky, P., Lindskog, D., & Ruhl, R. (2010). *A software application to analyze affects of temporal and environmental metrics on overall CVSS v2 Score.* London: IEEE. https://doi.org/10.1109/WorldCIS17046.2011.5749893

[36] Mell, P., Scarfone, K., & Romanosky, S. (2006). *CVSS: A complete guide to the common Vulnerability scoring system (Ver. 2.0).* Washington, DC: National Institute of Standards and Technology

[37] Gallon, L. & Bascou, J. J. (2011). Using CVSS in attack graphs. *2011 Sixth International Conference on Availability, Reliability and Security*, 59-66. Vienna, Austria. https://doi.org/10.1109/ARES.2011.18

[38] Yin, D., Shen, Y., & Liu, C. (2017). Attribute couplet attacks and privacy preservation in social networks. *IEEE Access, 5*, 25295-25305. https://doi.org/10.1109/ACCESS.2017.2769090

[39] See https://cwe.mitre.org/index.html

[40] See http://nvd.nist.gov/cwe.cfm

[41] See https://nvd.nist.gov/

[42] See https://www.exploit-db.com/

[43] Younis, A. & Malaiya, Y. K. (2015). Comparing and evaluating CVSS base metrics and microsoftrating system. *2015 IEEE International Conference on Software Quality, Reliability and Security,* 252-261. http://doi: 10.1109/QRS.2015.44

[44] Feutrill, A., Ranathunga, D., Yarom, Y., & Roughan, M. (2018). The effect of common vulnerability scoring system metrics on vulnerability exploit delay. *2018 Sixth International Symposium on Computing and Networking* (CANDAR), 1-10. http://doi: 10.1109/CANDAR.2018.00009

[45] See https://www.cvedetails.com/

[46] See https://cve.mitre.org/about/index.html

[47] See https://cwe.mitre.org/data/definitions/1200.html

[48] Marczak, W. R. & Paxson, V. (2017). Social engineering attacks on government opponents: Target perspectives. *Proceedings on Privacy Enhancing Technologies, 2.* https://doi.org/10.1515/popets-2017-0022

[49] Tran, H. & Wei, J. (2011). Impact of privacy and security on users' trust in ubiquitous commerce. *2011 International Conference on Business Management and Electronic Information,* 7-10. https://doi.org/10.1109/ICBMEI.2011.5914418

[50] Schneier, B. (2013). Trust in man/machine security systems. *IEEE Security & Privacy, 11*(5), 96. https://doi.org/10.1109/MSP.2013.128

[51] Wang, R., Gao, L., Sun, Q., & Sun, D. (2011). An improved CVSS-based vulnerability scoring mechanism. *Multimedia Information Networking and Security (MINES), International Conference*, 352-355. Shanghai, China. https://doi.org/10.1109/MINES.2011.27

[52] Chen, L., Jin, S., & Ye, Z. (2020). IIOT security privacy and trust modeling for scada systems. *Journal of Science and Engineering Management, 1*(2), 25-32. https://doi.org/10.1145/2692956.2663188

[53] Matsakis, N. D. & Klock, F. S, II. (2014). The rust language. *ACM SIGAda Ada Letters, 34*(3), 103-104.

[54] Ibidapo, A. O., Zavarsky, P., Lindskog, D., & Ruhl, R. (2011). An analysis of CVSS v2 environmental scoring. *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, 1125-1130. http://doi: 10.1109/PASSAT/SocialCom.2011.121

[55] Cheng, P., Wang, L., Jajodia, S., & Singhal, A. (2012). Aggregating CVSS base scores for semantics-rich network security metrics. *2012 IEEE 31st Symposium on Reliable Distributed Systems*, 31-40. https://doi.org/10.1109/SRDS.2012.4

Contact information:

Lifeng CHEN
Department of Global Business Administration,
Gachon University,
Seongnam 13120, South Korea

Zhixiao YE
College of Economics and Social Welfare,
Zhejiang Shuren University,
Hangzhou 310015, China

Shanyue JIN
(Corresponding Author)
Department of Global Business Administration,
Gachon University,
Seongnam 13120, South Korea
E-mail: jsyrena0923@gachon.ac.kr