ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 323-330 (2020)

www.astesj.com

Special Issue on Multidisciplinary Innovation in Engineering Science & Technology

ASTESJ

ISSN: 2415-6698

Multi-Model Security and Social Media Analytics of the Digital Twin

Jim Scheibmeir[1], Yashwant Malaiya[2]

[1]*Colorado State University, Systems Engineering, 80523, United States*

[2]*Colorado State University, Computer Science, 80523, United States*

ARTICLE INFO

Article history:
Received: 28 August, 2020
Accepted: 28 October, 2020
Online: 10 November, 2020

Keywords:
Internet of Things
Digital Twin
Cybersecurity
Contact Tracing
Social Media Analytics

ABSTRACT

Digital twins act through application programming interfaces to their physical counterparts to monitor, model, and control them. Beyond these traditional functions of digital twins, they must also act to secure their physical counterparts. A multi-model scheme is presented to help digital twins towards the task of securing the physical system. Additionally, this work includes an analysis of more than four hundred thousand tweets each relating to digital twin technology and cybersecurity which were collected during June and July in 2020. Of the first corpus of tweets collected by searching for #digitaltwin during the research period, only a small population of 10% reference security concepts. In the second and larger corpus of collected tweets, the top mentioned industries were health, education, and public. A naïve Bayes model reached a 70.3% accuracy at differentiating tweets that were either related to cybersecurity or the internet of things. The study also indicates that cybersecurity tweets are consistently more negative in many areas of sentiment when compared to tweets about the internet of things. The sentiment findings of cybersecurity tweets will reinforce the need to address culture in cybersecurity posture while the security multi-model schema contributes to the state of the art.

1. Introduction

An API development model for digital twins has been proposed recently [1]. This paper augments that work to address the cybersecurity concerns of the digital twin within the context of the various implementations and industries where the twin may be utilized. Furthermore, the adoption of current security models, such as state machine, lattice, non-interference, and information flow models are proposed for digital twins.

In the original paper, a development model for the APIs of the digital twin was created. The model was proposed to establish the context of the environment, the system, relationships between the two, and the functional and non-functional requirements.

This research includes the mining of Twitter tweets using appropriate search criteria for digital twins, the internet of things (IoT), and cybersecurity. The analysis includes mention analysis of industries, exploration of the sentiment, natural language processing (NLP) implementation towards the classification of tweets, and modeling to predict the twitter user behavior.

*Corresponding Author: Jim Scheibmeir, Email: jimscheibmeir@gmail.com

Three sets of tweets were collected for this study with the latter two sets being joined into a combined corpus. Each set was collected through the execution of scripts written in R. The first collection is smaller and has 3,102 tweets about digital twins. The second and third collections are larger. In total, 422,963 industrial internet of things (IIoT) tweets and 497,174 cybersecurity tweets were collected. These larger sets were combined into one corpus for analysis. The search criteria for the tweet collections are listed in Table 1.

Table 1: Tweet collection search criteria

Script Topic	Criteria used in actual R code
Digital Twin	"#digitaltwin"
IIoT	"#industry40 OR #IIOT OR #IOT"
Cybersecurity	"#cybersecurity OR #infosec OR #hacking"

This study is organized as follows. Section 2 reviews the literature and the background information on the subject. Section 3 describes our research method, including the collection of tweets as well as the processing. Sections 4 and 5 present the results from

social media analysis. Section 6 proposes a multi-model security scheme for digital twins. Finally, section 7 presents the conclusion of this study.

2. Background and related work

Digital twins are a logical conduit between physical entities and humans, geared towards monitoring and controlling a system [2]. The design of the interfaces themselves may utilize structured standards such as OpenAPI to list the various objects, paths, and operations [3].

Digital twins can be used during system design for vulnerability detection or during physical system operations to assist with intrusion detection [4]. Digital twins can utilize and manage the Internet of Things (IoT) devices and IoT systems of systems [5]. Wireless Sensor Networks (WSN) and attached IoT sensor components can be energy and cost-effective monitoring approaches [6]. WSNs are also frequently constrained to limited computing and power resources and thus security in WSNs can be challenged [7]. To maintain the integrity of the sensor data and overall system state awareness, secure links with proper authentication must be established [8].

Vulnerable digital twins offer hackers a blueprint of the physical counterpart as well as other backend systems that may be a part of system integration [9]. The digital twin can help secure the physical systems or be yet another vulnerability as IoT growth has created a large cyber-attack surface [10]. IoT has revealed security vulnerability ranging from trust management, authentication, privacy, to access control of embedded systems [11]. While digital twins may present such risk, they are still getting popular. The smart factory market, an economy of integrated automation solutions adopted to streamline manufacturing, is predicted to be worth approximately $205 billion-dollar by 2022 [9]. Hearn and Rix also cite another 2019 study whereof 220 security leaders in industrial and manufacturing who were a part of the research survey, 79% indicated they had experienced an IoT cyberattack within that past year. IoT and digital twins-based systems can be attractive targets for malicious actors.

Hearn and Rix identify advantages, such as the prevention of downtime and monitoring attacks against cyber-physical systems, as potential benefits of digital twins [9]. Risks of digital twins in cyber-physical systems (CPS) include the intellectual property incorporated in the digital twins as well as critical information about the CPS itself. For such reasons, the digital twin itself must go through rigorous software hardening routines, such as fuzz and penetration testing [9]. Security of the digital twin must start from the ground-up, inception to retirement. Security concerns must be incorporated with the organization's culture [9].

IoT systems and CPS are being deployed in the energy industry, and there have been instances of cyber-attacks on them. Some known attacks include the US electrical grid in 2009, Ukraine power outage in 2015, and Stuxnet attacks in 2005 and 2010 [10]. Common problems can facilitate such attacks, such as low power IoT devices have minimal resources and thus they are not resilient to denial of service attacks. A digital twin may create yet another vulnerability of an industrial or energy system.

Atalay & Angin in their 2020 study proposes that the digital twin can help to secure a cyber-physical system by establishing a security framework. The framework should contain an extensible digital twin that represents the smart grid, a cyber-threat database for applicable attack vectors, attack simulation tools, and an analysis and reporting module [10].

The framework suggested by Atalay & Angin starts with the creation of the digital twin through specification and reuse of other models. The framework separates the modeling of the CPSs into two phases, identification of the elements and their relationships, and building the network model using simulation. By creating the digital twin and using models, it is possible to achieve standardization through model reuse, utilization of a central threat intelligence database of smart grid attacks, and to achieve continuity of security evaluation [10].

Digital twins may secure physical systems through rule-based controls as well as intrusion detection as proposed by Eckhart and Ekelhart [4]. They identify a wide range of definitions for digital twins. Some definitions omit simulation and focus on visualization, such as implementing augmented reality (AR). AR applications have their security and quality concerns. Eckhart and Ekelhart also explain how a digital twin be used to implement intrusion detection for an industrial control system (ICS) or CPS. They identify concerns beyond intrusion detection, such as:

- the origin of digital twins was in the simulation of systems, although some authors omit simulation as a benefit of digital twins
- digital twins may have fidelity issues which may limit their use
- digital twins have a lifecycle that follows the lifecycle of their physical counterpart
- proper retirement of a digital twin system is imperative as it may contain private and confidential data
- digital twins may be used for security personnel training or in a cyber range as the attack target
- cost of digital twin creation can still be a prohibitive factor

Eckhart and Ekelhart refer to their 2018 publication, where, it was suggested that previously established specifications of the physical system could be modeled with AutoML, virtualized within Mininet and other tools, which can reduce the cost of creating digital twins [12]. They further proposed that digital twins can assist in intrusion detection by offering a baseline of normal behavior that can be analyzed and compared to the behavior of the physical system to detect outlying and potentially malicious activity [12].

3. Methodology

3.1. Collection

This paper focuses on the analysis of collections of Twitter tweets, as well as academic papers. Scripts have been written in R and executed throughout the research period. More R code has been utilized to clean the collected tweets, transform new classification data fields, and analyze them for n-grams, mention,

and sentiment analysis. The scripts use various libraries, such as the rtweet library as an API to the Twitter social media platform. Figure 1 below omits various keys and passwords yet illustrates the basics of library import and the implementation of a search and data frame store of tweets.

```
library(rtweet)
library(ggplot2)
library(dplyr)
library(tidyr)

c_key <- ''
c_secret<- ''
access_token <- ''
access_secret <- ''
appdt <- ''

twitter_token <- create_token(
  app = appdt,
  consumer_key = c_key,
  consumer_secret = c_secret,
  access_token = access_token,
  access_secret = access_secret)

setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)

digitaltwin_tweets <- search_tweets(q = "#digitaltwin",since = '2020-06-08', n = 50000)
```

Figure 1: Screenshot of example usage of rtweet library within R Studio

Three scripts were written to collect tweets for this research. The first used a single search criterion, "#digitaltwin", which has collected 3,102 tweets within 21 days of June 2020. The second and third scripts ran for the entire month of July 2020. 422,963 industry internet of things (IIoT) tweets and 497,174 cybersecurity tweets were collected. The tweets are stored with metadata made available via the rtweet API, but not all data has been retained. Additional data has been generated and stored such as the industry a tweet may mention and the sentiment score (positive or negative) of the text within the tweet. Table 2 lists the acquired and generated data.

Table 2: Tweet data fields

Field Name	Short Description
created_at	The date the tweet was originally posted to the Twitter platform
text	Character contents of the tweet
favorite_count	An integer representing the number of times a tweet has been marked as a favorite by Twitter users
retweet_count	An integer representing the number of times a tweet has been retweeted by Twitter users
country	Frequently NA within the dataset and sometimes providing the country of origin of the Twitter user posting the tweet
retweet_location	Frequently NA within the dataset and sometimes providing the country of origin of the Twitter user retweeting a tweet
Industry and count	A list of industries and the count of mentions from the gathered tweets
Sentiment	The individual sentiment score of a tweet's text
Retweet	Binomial value indicating if retweet_count > 0
Favorite	Binomial value indicating if favorite_count is > 0

3.2. Data Preparation and Processing

Data preparation activities may account for 80 percent of the time invested in data science effort [13]. In this study, the collection took many days, but the Twitter data was usable practically from the beginning. This is due to the limited search criteria used and the nature of the required data (such as we did not need the country or retweet_location fields for this research, both having many empty or not applicable entries). The longest time in preparation was determining industry mentions and labeling the tweets. Industry and identification keywords for cross-reference and classification efforts were collected from the International Labor Organization's web site [14]. Preprocessing was done cross-referencing the ILO keywords to the tweet's text to determine the tweet's topics. This enabled mention analysis by the labor industry.

In the second corpus of tweets, the labeling of tweets allowed the naïve Bayes models to be trained and then tested. Tweets were labeled based upon the search criteria that was used to collect the tweets. This label would consist of either IoT or cybersecurity. Eighty percent of the data was used to train the models and twenty percent was utilized to test the model accuracy. Various models and parameters were used to predict the correct tweet label. Other preprocessing for the model input parameters included sentiment analysis and word lemmatization. The package textstem was utilized for word lemmatization and the packages tidytext and sentimentr were utilized for the sentiment analysis.

4. Results of First Corpus Analysis

The word cloud illustrated in Figure 2 highlights common bigrams found within the first corpus of collected tweets, those 3,102 found with search criterion #digitaltwin, after removing common stop words. While concerning the word 'security' is not paired with another common word frequently enough to make this bi-gram chart, we do see that conversations on Twitter around defense, risk, and the industrial use of IoT are occurring, each having relevance towards this research.

Figure 2: Word Cloud of Bigrams within the Tweet Collection Text

The word cloud used in figure 2 omits counts of the actual instances of the bigrams. Table 3 outlines a few common security terms and the number of instances they occur within our first

corpus, those 3,102 tweets collected by searching #digitaltwin. Such terms may occur in one tweet multiple times. Overall, terms such as "encrypt", or "vulnerability" have almost no mention in such tweets and do not appear as a part of the social media conversation which is a concerning observation.

Table 3: Common security terms and their occurrence

Terms	Count of occurrences
"security"|"secure"	88
"risk"	212
"defense"|"defensive"	179
"iiot"|"industrial iot"	750
"encrypt"|"decrypt"|"cipher"	1
"vulnerability"|"threat"	1

Looking at this first corpus of tweets, the percentage of tweets containing common security terms make up 10% of the total volume of tweets collected using the #digitaltwin search criteria within the specified time range, as illustrated in Figure 3. We observe a dearth of conversation within the social media analysis towards securing digital twins, which could be used to implement the security models needed.

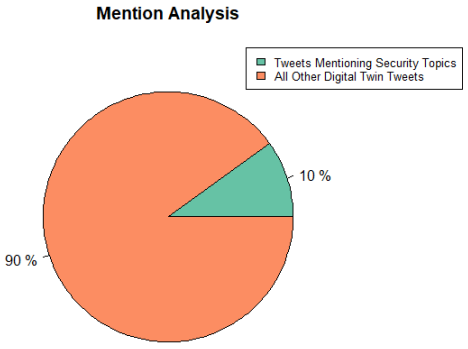

Figure 3: Most IoT tweets do not mention security topics

4.1. Most Favored and Most Retweeted Tweets about Digital Twins

Tweets can be favored as well as re-tweeted. The number of favorites indicates the count of unique user accounts that like or agree with the content of the tweet [15]. In the first corpus, the most favored tweet is also the most retweeted. It will be discussed later in this section. To provide more breadth of the findings, the second most favored tweet is illustrated in Figure 4 [16]. This tweet refers to Intel technology and links to a 2015 white paper where Intel published best practices to help with IoT sensor technology implementation against challenges such as cost of ownership, maintaining security, and designing for scalability [17].

Within the scope of this first corpus, the most retweeted tweet is about the use of digital twin used in contact tracing to help prevent the spread of Covid-19. The digital twin was specifically built and used to help contain the spread of the virus in Dharavi, a location within Mumbai that houses nearly one million residents. Many of these residents rely on daily wages. Living and working conditions may reduce or prevent social distancing. During the time of the tweet creation, epidemiologists had credited the Dharavi community for their efforts towards containment [18]. Figure 5 is a screen capture of this most favorited tweet within the first corpus [19]. While we cannot confirm the accuracy of the tweet, we can recognize from the tweet contents the broad application of digital twin usage to monitor, model, and control a pandemic. The nature of such a system and the associated data also provide support for why such the system requires a mature security posture, to protect individuals and their privacy rights.

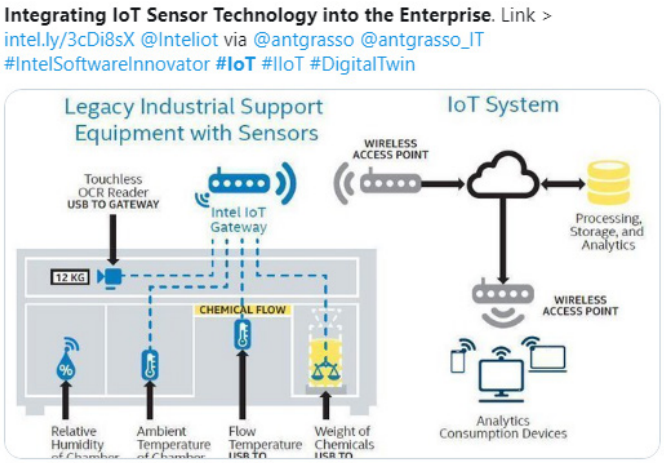

Figure 4: Second Most Favored Tweet within the Scope of this Research

Figure 5: Most retweeted tweet within the research scope

4.2. Mention and Sentiment Analysis

Studies show digital twins being utilized across industries; mention analysis from social media provides an additional perspective. To analyze industry mention among the first tweet collection, we crossed referenced the tweets' text with a list of known industry categories [14]. The simple algorithm reads the tweet texts and counts industry mentions, including when multiple

industries are mentioned within a single tweet's text. Figure 6 illustrates mention analysis results with health, education, and public as the top three topics with most mentions within the tweets' text.

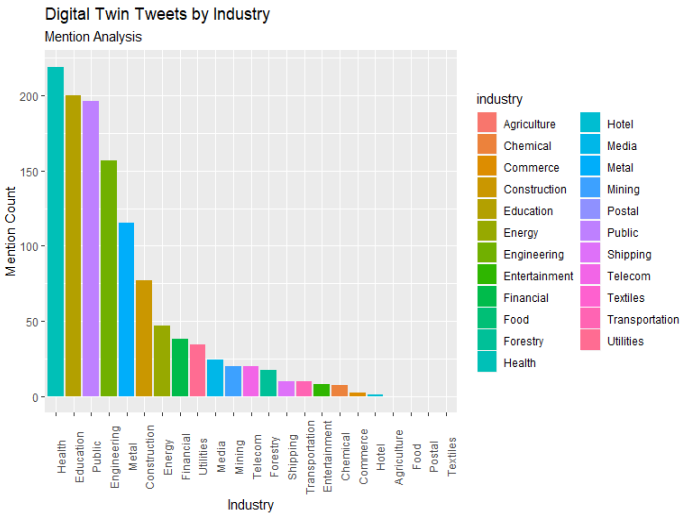

Figure 6: Industry mention analysis: health, education, and public are the top three mentioned industries

Utilizing the sentimentr package, each industry group has the individual tweet text sentiment scores averaged and placed on the chart found in Figure 7. The sentiment scores from the sentimentr package may range from a negative 2.0 to a positive 2.0. While the health and public industries were high in the mention-analysis, the sentiment scores for those industry groups are negative. The three industry groups having the highest average sentiment scores are engineering, commerce, and energy.

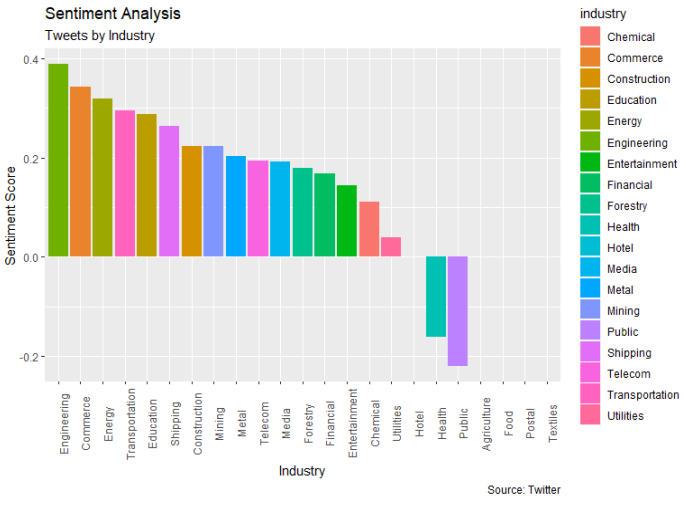

Figure 7: Aggregated sentiment scores grouped by industry

The tweet in the engineering group having the highest sentiment score is given in Figure 8 [20]. This tweet indicates the usage of digital twins to assist in the simulation of electric vehicles for optimal energy management.

Today's electric vehicles market is high stakes. The current pandemic threatens the automotive supply chain [21], a Tesla truck recently marketed towing a competitor's vehicle (F-150) up a hill [22], and Tesla competitor Nikola Motors saw a one-day stock jump of 104% [23]. Security posture is required around the intellectual property in this dynamic and competitive market, and therefore must be applied to such a digital twin.

Figure 8: Tweet having highest sentiment score within the engineering group

5. Results of Social Media Analytics of IIoT and Cybersecurity Tweets

Multiple prediction models were created to further inspect the tweets. This included attempts to predict whether a given tweet was towards the topic of cybersecurity or IIoT. Methods such as word stems, lemmatization, term frequency-inverse document frequency (TFIDF) were utilized in naïve Bayes and general linear models.

A naïve Bayes model that attempted to predict the label, cybersecurity or IIoT, of a given tweet using word lemmatization and sentiment analysis could predict the label correctly with a 70.3% accuracy. While this prediction accuracy is not high, perhaps the sentiment coefficient correlations are more telling. As previously stated, security must start with the organization's culture [9]. An organization's culture consists of the beliefs and values of the employees. Communication incorporating the sentiment of the topic may be more influential towards influencing behavior or belief. The sentiment level of the cybersecurity tweets, when compared to the IIoT tweets, was higher in the areas of anger, disgust, fear, negative, and sadness, whereas the IIoT tweets sentiment level, when compared to cybersecurity tweets, was higher in anticipation, joy, positive, surprise, and trust. Table 4 lists the average sentiment scores for each classification of sentiment by the cybersecurity and IIoT labels.

Given a tweet is about Cybersecurity, there is a 14.5% probability that the tweet is negative, an 11.6% probability that the tweet provokes anticipation, and an 11.3% probability that the tweet has the sentiment of fear. This information may be key for communicating cybersecurity strategies and plans within an organization. We should consider the sentiment of our security posture messaging, and how it fits into the culture, behaviors, and values of the communication's target audience.

The Term Frequency-Inverse Document Frequency (TFIDF) score utilization within a naïve Bayes model had less accuracy,

56.3%, compared with the model using both word lemmatization and sentiment analysis. TFIDF does allow us to see important words in a corpus, we observe that "bugbountytip" had the highest TFIDF score within the cybersecurity corpus at 0.00127.

Table 4: Sentiment coefficient probabilities on tweet labels

Conditional Probabilities of IIoT and Cybersecurity Tweet Sentiment		
Sentiment	Cybersecurity	IIoT
Anger	0.07145	0.03765
Anticipation	0.11663	0.14300
Disgust	0.01868	0.01400
Fear	0.11321	0.06652
Joy	0.05369	0.07048
Negative	0.14541	0.07757
Positive	0.25503	0.35884
Sadness	0.04404	0.02833
Surprise	0.03586	0.03683
Trust	0.14597	0.16677

6. Application of a Digital Twin as a Multi-Model Security Architecture

Digital twins are commonly used to help monitor, model, and control cyber-physical systems. Engineering a digital twin should be done using a systematic approach by performing a needs analysis from the physical device and context of the operating environment [1]. Needs analysis and environmental context must include non-functional requirements, including cybersecurity. Securing information confidentiality of these cyber-physical systems is a growing concern [24]. Beyond the common monitor, model, and control mechanisms, digital twins may also be utilized for incorporating multiple security models.

Several security control models exist. The information flow model is a common mechanism to maintain information confidentiality [25]. We outline in Table 5 how the information flow model may also work to maintain the integrity of a digital twin and its physical counterpart. Multiple models will need to be implemented in the system engineering effort to establish a secure posture. Each model has a strength, such as Bell-LaPadula working towards the confidentiality of information [26]. Furthermore, in Table 5, we indicate how the lattice model, incorporating Biba, can work to ensure information integrity.

The digital twin may also incorporate concepts and techniques such as automated policy exchange (APEX) to prevent data leakage. APEX technology works to keep honest people working within the guidance of security best practices and policies, as large systems and enterprises may have complex policies that not everyone is aware of [27]. Using APEX, the digital twin would monitor labeled files throughout their lifecycle and use control mechanisms, such as deployed agents within the cyber-physical system, to prevent or warn on file activities by users.

Another possible avenue would be for the monitoring portion of the digital twin to utilize a System Call Intercept (SCI) framework. SCI works to reduce data leakage by reviewing policies of subjects and objects, trapping some specific activities before an operating system completing the execution of the activity and perhaps suspending the activity completely [28]. Using the SCI framework, the monitor and modeling functions of the digital twin would use policies to review specific subjects or objects (perhaps all users and system components within a zero-trust approach) and trap some actions for review before the OS or another system component completing the action. An inventory of objects, subjects, and system calls that could be trapped would be required to implement such a framework, which will improve data protection. This inventory could be complemented with behavior analysis by machine learning models to identify malicious activity. Machine learning can be used for countering cybersecurity threats and vulnerability mitigation using regression, prediction, and classification techniques.

Table 5 provides high-level examples of how a digital twin implementation may satisfy or utilize common security models such as State Machine, Lattice, Interference, and Information Flow models.

Table 5: Application examples of a digital twin as a multi-model security schema

Security Model	**Example Digital Twin Application**
State Machine	Digital twin control input is first executed within the model. Validation occurs within the model for an appropriate output state. The input may be rejected before entering the digital twin control mechanism for actual processing on the cyber-physical system.
Non-Interference Model	Monitoring and control mechanisms must obfuscate users reading (monitoring) and writing (controlling) across domains. This reduces the probability of users being influenced by the actions of subjects with greater/lesser clearance. This approach may utilize common discretionary access control (DAC) across file systems, mandatory access control (MAC) through rule-based systems, and role-based control. DAC and MAC limit the availability of information objects to certain subjects to retain the confidentiality of information [29].
Lattice Model	Biba protects integrity by preventing a subject from reading lower integrity or lower-level access data [29]. Using the Biba Model approach, the digital twin modeling process accesses only known secured sensors (having a level of integrity through controls). While we may monitor many sensors, decision dashboards and predictive analytics should utilize only trusted and secured sensors. A matrix of such processes and secure sensors would instruct inputs for such high integrity algorithms.
Information Flow Model	Sensor information flow is one way. The Digital twin may read sensor data for monitoring or modeling purposes. No data is transmitted to a sensor to be later sent back, as sensor-originated data. This prevents levels of spoofing sensor data, by not allowing writes to the sensor memory. Any PUT type of action to sensors APIs would be rejected and logged as malicious intent.

Figure 8 illustrates a situation where a population, Bluetooth sensors from mobile devices, network paths, and system logic are used for the creation of a population's digital twin. In this case, the digital twin of the population would be used by public health officials and the application subscribers for contact tracing. The diagram's purpose is to provide examples illustrating the concepts given in Table 5 applied to a hypothetical contact tracing system.

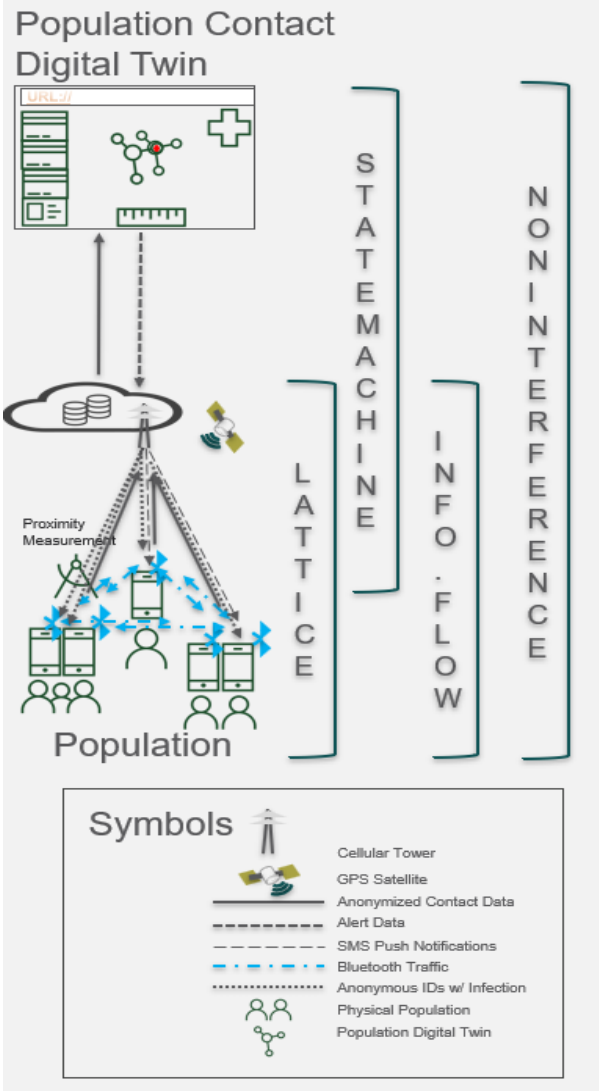

Figure 8: Applying the multi-model in digital twin security

If contact tracing systems were injected with malicious data, it could alter the behaviors of a population. Applying the lattice model helps assure that a system having integrity does not read from a system or sensor with low integrity. The lattice model can prevent a malicious actor's device farm from spoofing the contact system. To validate the integrity of a sensor or mobile device in this case, the system may validate a history of GPS or use human interaction to assure the device is not a simulator or otherwise driven by a malicious bot.

Public health officials may instruct the system to deploy warning notifications to participants who have opted in to use the feature. However, before deploying the notification, the digital twin should utilize the state machine model. Again, a notification could be a malicious message being sent to alarm a population. The logic should first model that notification within the digital twin. The modeling should evaluate the message accuracy and may even predict how the population may behave to help public health officials craft response plans. If the notification messages were not related by a contact tracing scheme, the message may be malicious and should be stopped.

The information flow model is utilized to protect the population and their privacy. The mobile device's application logic allows Bluetooth connections to write to ledgers indicating contacts. This data is anonymized and does not indicate a person's identity. Writing to this ledger is only allowed by the local mobile device which is in the possession of the user. The centralized system does not allow writing to this contact ledger. The mobile device application will pull anonymized alert data from the central system. This data is written to a separate ledger. The only mechanism for the centralized system to send data to the host devices is through the SMS push notification mechanism, which has levels of security such as users being required to opt-in and the state machine model.

Users of the Population Contact Digital Twin dashboard will utilize roles, authentication, and authorization to access the digital twin of the population. Low entitlement users would not have the ability to interfere with the population by spying on them using inference techniques. Entitled users, such as medical doctors, may be granted access to information having more fidelity to help protect the population's health. This protects a population using the non-interference model with access controls.

7. Conclusions and Future Work

7.1. Conclusions

Digital twins operate from and create new cyber-physical system information. A proper design approach for a digital twin should model its cyber-physical system, including its future states. In addition to monitoring, modeling, and controlling the cyber-physical system, a digital twin must also provide security. This fourth element as applied to the prior three digital twin requirements allows for a layered approach to securing cyber-physical systems by implementing multiple security models. While past researchers have focused on intrusion detection, this article proposes that common security control models such as state machine, non-interference, lattice, and information flow models can be utilized in digital twin implementation to secure itself the associated cyber-physical system.

In the first corpus of tweets collected in this study by searching for #digitaltwin, those referencing security concepts represented a small population of 10%. In the second and larger corpus of the collected tweets, the top mentioned industries or business sectors were health, education, and public in that order. The top three industries having the highest tweet sentiment were engineering, commerce, and energy. While naïve Bayes models reached only a 70.3% accuracy at differentiating a tweet that was about cybersecurity versus a tweet about digital twins, the sentiment findings of the messages leaned towards negative messaging from the cybersecurity tweets.

7.2. Future Work

Most of our research utilized social media, which can be time-consuming to prepare such data and has limits in what it can

contribute. Future research should include detailed case studies of direct implementation which would strengthen the approaches proposed in this work. The sentiment findings of cybersecurity tweets deserve further research as other researchers have stated that security must start with an organization's culture. The implication of negative messaging on an organization's cybersecurity posture and culture needs to be studied.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] J. Scheibmeir, Y. Malaiya, "An API Development Model for Digital Twins," in 2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C), Sofia, Bulgaria, 518-519, 2019, doi: 10.1109/QRS-C.2019.00103.

[2] E. A. Saddik, "Digital Twins: The Convergence of Multimedia Technologies," IEEE Multimedia Mag., 25, 87-92, Apr. 2018.

[3] B. Martino. M. Di, M. Rak, A. Ficco, S. Esposito, S. Nacchia, "Internet of Things Reference Architectures, Security and Interoperability: A Survey." Internet of Things, 1-2, 99-112, 2018, doi:10.1016/j.iot.2018.08.008.

[4] M. Eckhart, A. Ekelhart, "Digital Twins for Cyber-Physical Systems Security: State of the Art and Outlook." Security and Quality in Cyber-Physical Systems Engineering. Springer, 2019, doi:10.1007/978-3-030-25312-7_14.

[5] A. Canedo. "Industrial IoT Lifecycle via Digital Twins." Proceedings of the Eleventh IEEE/ACM/IFIP International Conference on Hardware/software Codesign and System Synthesis. ACM, 1-1, 2016, doi: 10.1145/2968456.2974007

[6] D. Yadav, S. Jayanthu, S. Das, S. Chinara, P. Mishra, "Critical Review on Slope Monitoring Systems with a Vision of Unifying WSN and IoT," IET Wireless Sensor Systems, 2019, doi:10.1049/iet-wss.2018.5197.

[7] Y. Wang, A. Garhan, R. Byrav. "A Survey of Security Issues in Wireless Sensor Networks." IEEE Communications Surveys & Tutorials, **8**(2), 2-23, 2006, doi:10.1109/comst.2006.315852.

[8] P. Porambage, C. Schmitt, P. Kumar, A. Gurtov, M. Ylianttila, "Two-phase Authentication Protocol for Wireless Sensor Networks in Distributed IoT Applications," IEEE Wireless Communications and Networking Conference, 2014, doi:10.1109/WCNC.2014.6952860.

[9] M. Hearn, S. Rix, "Cybersecurity Considerations for Digital Twin Implementations report". IIC Journal of Innovation, 2019

[10] M. Atalay, P. Angin, "A Digital Twins Approach to Smart Grid Security Testing and Standardization," 2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT" 435-440, 2020, doi: 10.1109/MetroInd4.0IoT48571.2020.9138264.

[11] A. Girma, "Analysis of Security Vulnerability and Analytics of Internet of Things (IoT) Platform," in Information Technology - New Generations. Advances in Intelligent Systems and Computing, 738, 2018, doi:10.1007/978-3-319-77028-4_16

[12] M. Eckhart, A. Ekelhart, "Towards Security-Aware Virtual Environments for Digital Twins," in Proceedings of the 4th ACM Workshop on Cyber-Physical System Security, ACM, 61–72, 2018, doi:https://doi.org/10.1145/3198458.3198464

[13] A. Cirillo, RStudio for R Statistical Computing Cookbook. Birmingham Packt Publishing, 2016.

[14] International Labor Organization. "Industries and Sectors." Industries and sectors, https://www.ilo.org/global/industries-and-sectors/lang--en/index.htm.

[15] R.S. Perdana, A. Pinandito, " Combining Likes-Retweet Analysis and Naive Bayes Classifier within Twitter for Sentiment Analysis", Journal of Telecommunication, Electronic, and Computer Engineering, **10**(1-8), 41-46, 2018.

[16] A. Grasso [@antgrasso]. (2020, June 8). Integrating IoT Sensor Technology into the Enterprise. [Twitter moment]. Retrieved from: https://twitter.com/antgrasso/status/1269962528530071552

[17] K. H. Gaiser, V. P. Harrington, G. T. Loughrin, S. J. Meyer, J. M. Sartini. "Integrating IoT Sensor Technology into the Enterprise." December 2015. https://www.intel.com/content/www/us/en/it-management/intel-it-best-practices/integrating-iot-sensor-technology-into-the-enterprise-paper.html.

[18] P. M.N., S. Bengali. "How an Indian Slum Became an Unlikely COVID-19 Success Story." Los Angeles Times. June 24, 2020. https://www.latimes.com/world-nation/story/2020-06-24/dharavi-slum-in-mumbai-india-contained-covid-19.

[19] Zuppa [@ZuppaGeoNavTech]. (2020, June 17). "Offence Is The Best Defense" [Twitter moment]. Retrieved from: https://twitter.com/ZuppaGeoNavTech/status/1273449554797408257

[20] Simcenter [@Simcenter_] (2020, June 12). Learn more about #simulation for optimal vehicle energy management in our new on-demand webinar series. [Twitter moment] Retrieved from: https://twitter.com/Simcenter_/status/1271442013586624512

[21] L. Harbour, "The Coronavirus' Impact On The Global Automotive Supply Chain." Forbes. March 14, 2020. https://www.forbes.com/sites/laurieharbour1/2020/03/13/the-coronavirus-impact-on-the-global-automotive-supply-chain/#17fd6b9f444e.

[22] M. Wayland, L. Kolodny, "Tesla Cybertruck Tug-of-war with a Ford F-150 Is High Stakes and Highly Unlikely." CNBC. November 27, 2019. https://www.cnbc.com/2019/11/26/tesla-cybertruck-tug-of-war-with-a-ford-f-150-is-high-stakes-unlikely.html.

[23] M. Fox, "Tesla Competitor Nikola Motors Skyrockets 104% after Setting Reservation Date for Its Electric Truck (NKLA, TSLA) | Markets Insider." Business Insider. June 8, 2020. https://markets.businessinsider.com/news/stocks/nikola-motors-stock-price-soars-setting-truck-reservation-date-competitor-2020-6-1029290291.

[24] A. Humayed, J. Lin, F. Li, B. Luo, "Cyber-physical systems security—A survey", IEEE Internet Things J., **4**, 1802-1831, 2017, doi: 10.1109/JIOT.2017.2703172.

[25] A. Sabelfeld, A. C. Myers, "Language-based information-flow security", IEEE J. Sel. Areas Commun., **21**(1), 5-19, 2003, doi: 10.1109/JSAC.2002.806121.

[26] X. Ma, Y. Huang, D. Li, "A Security Model Based on Lattice," International Conference on Electrical and Control Engineering, 4958-4961, 2010, doi: 10.1109/iCECE.2010.1199.

[27] S.J., Simske, and H. Balinsky. "Apex." Proceedings of the 10th ACM Symposium on Document Engineering, 10, 2010, doi:10.1145/1860559.1860587.

[28] H. Balinsky, D. S. Perez, and S. J. Simske. "System Call Interception Framework for Data Leak Prevention." 2011 IEEE 15th International Enterprise Distributed Object Computing Conference, 2011. doi:10.1109/edoc.2011.19.

[29] C. Schou, & S. Hernandez, Information assurance handbook:effective computer security and risk management strategies. New York: McGraw-Hill Education, 2015.