Computational and Mathematical Organization Theory (2022) 28:335–351
https://doi.org/10.1007/s10588-021-09341-0

S.I.: SOCIAL CYBERSECURITY IN TIMES OF CRISIS

Survival analysis for insider threat

Detecting insider threat incidents using survival analysis techniques

Elie Alhajjar[1] · **Taylor Bradley**[1]

Accepted: 17 July 2021 / Published online: 24 July 2021

Abstract
In the current information era, we rely on cyber techniques and principles to protect the confidentiality, integrity, and availability of everything from personally identifiable information and intellectual property, to government and industry information systems. Despite persistent efforts to protect this sensitive information, security breaches continue to occur at alarming rates, the most common of them being *insider threats*. Over the past decade, insider threat detection has attracted a considerable amount of attention from researchers in both academia and industry. In this paper, we develop a novel insider threat detection method based on survival analysis techniques. Specifically, we use the Cox proportional hazards model to provide more accurate prediction of insider threat events. Our model utilizes different groups of variables such as activity, logon data, and psychometric tests. The proposed framework has the ability to address the challenge of predicting insider threat instances as well as the approximate time of occurrence. This study enables us to perform proactive interventions in a prioritized manner where limited resources are available. The criticality of this issue in the insider threat problem is twofold: not only correctly classifying whether a person is going to become a threat is important, but also the time when this is going to happen. We evaluate our method on the CERT Insider Threat Test Dataset and show that the proposed Cox-based framework can predict insider threat events and timing with high accuracy and precision.

Keywords Insider threat · Survival analysis · Kaplan–Meier curve · Cox proportional hazards model

✉ Elie Alhajjar
elie.alhajjar@westpoint.edu

Taylor Bradley
taylor.bradley@westpoint.edu

[1] United States Military Academy, West Point, USA

1 Introduction

In today's digital age, it is essential for individuals, industries, and government agencies to protect themselves against cyber threats. As such, cybersecurity has become an essential practice in almost every organization around the world. In order to protect organizations' information, intellectual property and security, measures must be taken to ensure their data is secure. However, due to the vast growth and perpetually changing nature of technology, this task gets more difficult with each passing day.

Insider threats are malicious events from people within the organization, which usually involve intentional fraud, the theft of confidential or commercially valuable information, or the sabotage of computer systems. The subtle and dynamic nature of insider threats makes detection extremely difficult. The 2018 U.S. State of Cybercrime Survey indicates that 25% of the cyberattacks are committed by insiders, and 30% of respondents indicate incidents caused by insider attacks are more costly or damaging than outsider attacks (U.S. State of Cybercrime 2018).

Insider threats are one of the most challenging threats in cyberspace, usually causing significant loss to organizations. While the problem of insider threat detection has been studied for a long time in both the security and data mining communities, it remains difficult to accurately capture the behavior difference between insiders and normal users due to various challenges related to the characteristics of underlying data, such as high-dimensionality, complexity, heterogeneity, sparsity, lack of labeled insider threats, and the subtle and adaptive nature of insider threats.

According to the latest technical report (Costa et al. 2016) from the CERT Coordination Center, a malicious insider is defined as "a current or former employee, contractor, or business partner who has or had authorized access to an organization's network, system, or data, and has intentionally exceeded or intentionally used that access in a manner that negatively affected the confidentiality, integrity, or availability of the organization's information or information systems".

While cyber attacks can be attributed to anything from phishing attacks to ransomware, malware, and denial of service, insiders by far pose the greatest threat to organizations' assets and information. Compared to the external attacks whose footprints are difficult to hide, the attacks from insiders are hard to detect because malicious insiders already have the authorized power to access the internal information systems. In general, there are three types of insiders: (i) traitors who misuse their privileges to commit malicious activities, (ii) masqueraders who conduct illegal actions on behalf of legitimate employees of an institute, and (iii) unintentional perpetrators who innocently make mistakes (Liu 2018b). Based on the malicious activities conducted by the insiders, the insider threats can also be categorized into three types: (i) IT sabotage which directly uses IT to make harm to an institute, (ii) theft of intellectual property which steals information from the institute, and (iii) fraud which indicates unauthorized modification, addition, or deletion of data (Homoliak 2019).

Given the high number of insider incidents, insider threat detection has become a central, yet very challenging, task. The complexity of such task stems from

many perspectives: first, insiders perform unauthorized actions by the use of their trusted access, which renders external network security devices, like firewalls and anti-viruses, useless. Second, the diversity of insider attack scenarios hinders the possibility of one-solution-fits-all approaches: insider attacks can take the form of a disgruntled employee planting a logic bomb to disrupt systems, stealing intellectual property, acquiring financial information, etc. Third, most insider threat activities are performed during working hours, which makes them hard to detect since they are spread out among normal routine behavior (Yuan et al. 2018).

In response to this growing problem, the US National Insider Threat Policy, written in response to Executive Order 13587, "Structural Reforms to Improve the Security of Classified Networks and the Responsible Sharing and Safeguarding of Classified Information" (Obama 2011), sets expectations and identifies best practices for deterring, detecting, and mitigating insider threats. It calls for program establishment, training of program personnel, monitoring of user and network activity, and employee training and awareness. The more recent Insider Threat Guide and maturity framework (Belk and Hix 2018) from the National Insider Threat Task Force (NITTF) continues this trend.

Insider threat research constitutes one of the facets of the new and emerging field of *social cybersecurity* (Carley 2020). Social cybersecurity is a computational social science with a large foot in the area of applied research. It uses computational social science techniques to identify, counter, and measure (or assess) the impact of communication objectives. The methods and findings in this area are critical, and advance industry-accepted practices for communication, journalism and marketing research.

Although existing approaches demonstrate good performance on insider threat detection, the traditional shallow machine learning models are unable to make full use of the user behavior data due to their high-dimensionality, complexity, heterogeneity, and sparsity. On the other hand, deep learning as a representation learning algorithm is able to learn multiple levels of hidden representations from the complicated data based on its deep structure. Hence, it can be used as a powerful tool to analyze the user behavior in an organization to identify the potential malicious activities from insiders. In this paper, we take a different approach to tackle the insider threat detection problem, namely we introduce a novel method to identify potential insider threat events using survival analysis techniques.

The analysis of insider incidents include all related aspects and behaviors of a malicious insider before, during, and after conducting an incident. Efforts in this realm focused on behavioral frameworks, formalization frameworks, psychological and social theory, criminology theories, simulation research, system dynamics, game theory, and many other fields. The aim of this paper is not to survey the vast amount of existing literature, the interested reader is referred to the recent survey (Homoliak 2019) and the pointers therein. To the best of our knowledge, survival analysis techniques were not used in the context of insider threat and the current paper aims at initiating that direction of research.

Survival analysis is a subfield of statistics where the goal is to analyze and model data for which the outcome is the time until an event of interest occurs. One of the main challenges in this context is the presence of instances whose

event outcomes become unobservable after a certain time point or are not experienced during the monitoring period. This so-called censoring can be handled most effectively using survival analysis techniques.

The rest of the paper is organized as follows. After this brief introduction, we give an overview of the field of survival analysis. We set the notations and definitions that will be used throughout the paper. In Sect. 3, we review the most relevant and up to date literature that deals with insider threat detection, as well as survival analysis applications. Section 4 describes the methodology used to analyze the CERT dataset. We explain the intuition as well as the technical details of our chosen survival methods. In Sect. 5, we show the performance of our method and discuss our findings. Finally, Sect. 6 concludes the work and offers directions for future research.

2 Survival analysis

In this section, we introduce the definitions and terminologies used in the remainder of the paper. We adopt a simplistic approach for the sake of exposition, the reader interested in the technical details is referred to the books (Klein and Zhang 2005; Miller 2011).

Survival analysis is defined as a collection of statistical methods which contains the time of a particular event of interest as the outcome variable to be estimated. It is useful whenever we are interested not only in the frequency of occurrence of a particular type of event, but also in estimating the time for such an event occurrence. During the study of a survival analysis problem, it is possible that the events of interest are not observed for some instances. This may be because of either the limited observation time window, or missing traces caused by other events – a concept known as *censoring*. In general, censoring is categorized into three groups: (i) right-censoring, where the observed survival time is less than or equal to the true survival time, (ii) left-censoring, where the observed survival time is greater than or equal to the true survival time, and (iii) interval censoring, where we only know that the event occurs during a given time interval. Note that the true event occurrence time is unknown in all three cases. We will restrict our study to right-censored data for the remainder of the paper and we will refer to it as censored for brevity.

Broadly speaking, survival analysis methods can be classified into two main categories: statistical methods and machine learning based methods. Statistical methods share a common goal with machine learning methods in that both are expected to make predictions of the survival time and estimate the survival probability at the estimated survival time. However, the former focus more on characterizing both the distributions of the event times and the statistical properties of the parameter estimation by estimating the survival curves, while the latter focus primarily on the prediction of event occurrence at a given time by combining the power of traditional survival analysis methods with various machine learning techniques (Wang et al. 2019).

2.1 Survival functions

In survival analysis, a data point consists of a triple (X_i, y_i, δ_i) where X_i is the feature vector, y_i is the observed time—it is equal to the survival time T_i for an uncensored instance, the censored time C_i for a censored instance, and δ_i is an indicator function, i.e., $\delta_i = 0$ for a censored instance and $\delta_i = 1$ otherwise. Figure 1 illustrates the concept of censored data. It contains six observed instances over a 12-month period. Note that the fourth and sixth subjects are the only ones that experience the "event", while the others are considered censored either due to withdrawal from the experiment or because no event occured during the study time.

Next we define the three main functions used in survival analysis: the survival function, the cumulative death distribution function, and the death density function. Figure 2 depicts the relationships between these three functions. The *survival function* is the probability that the time to the event of interest is not earlier than a specified time t: $S(t) = Pr(T \geq t)$. As a function, $S(t)$ monotonically decreases with t with $S(0) = 1$. This means that at the beginning of the observation, none of the events of interest has already occurred. The *cumulative death distribution function* is the probability that the event of interest occurs earlier than t, i.e., $F(t) = 1 - S(t)$. The *death density function* $f(t)$ is the derivative of $F(t)$ in the continuous case and the rate of change of $F(t)$ in the discrete case.

Another commonly used function in survival analysis is the *hazard function*, also known as the instantaneous death rate. It measures the likelihood of the event occurring at time t given that no event has occurred before time t, and can be written as $h(t) = \frac{f(t)}{S(t)}$. As a function, $h(t)$ is non-negative and can have a variety of shapes (not necessarily monotone).

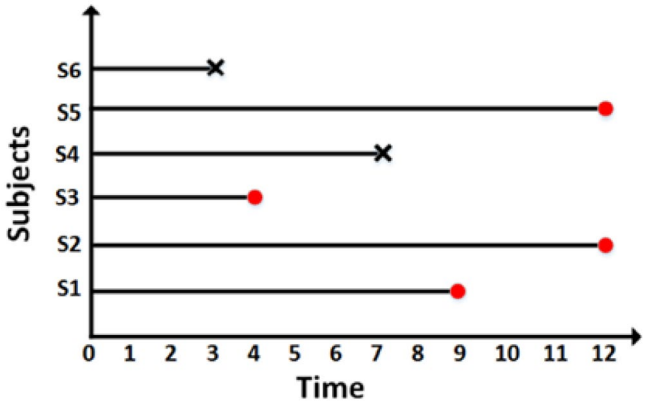

Fig. 1 Censored vs. uncensored data (Wang et al. 2019)

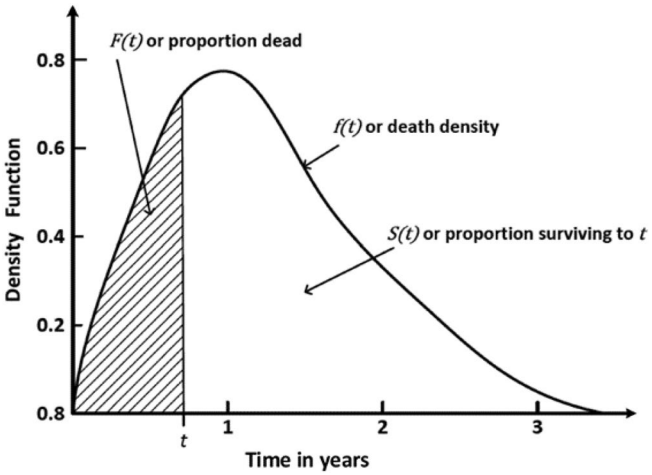

Fig. 2 Commonly used survival functions (Wang et al. 2019)

2.2 The Kaplan–Meier curve

Among the functions discussed in the previous section, the survival function is the most widely used one in survival analysis. To better represent this function, the *Kaplan–Meier curve* is used to estimate the survival function based on the actual length of observed time.

In mathematical terms, let $T_1 < T_2 < \cdots < T_k$ be a set of distinct ordered event times observed for N instances where $k \leq N$ (there are $N - k$ censored times in this case). For each $i = 1, 2, \ldots, k$, we denote by d_i the number of observed events in time T_i and by r_i the number of instances whose event time or censored time is greater than or equal to T_i. The two terms are related via the recursion: $r_i = r_{i-1} - d_{i-1} - c_{i-1}$, where c_{i-1} denotes the number of censored instances during the time period between T_{i-1} and T_i. With this setting, the conditional probability of surviving beyond time T_i is defined as:

$$p(T_i) = \frac{r_i - d_i}{r_i} = 1 - \frac{d_i}{r_i}. \tag{1}$$

Based on the conditional property in Eq. (1), the survival function can be estimated by

$$\widetilde{S}(t) = \prod_{T_i < t} p(T_i) = \prod_{T_i < t} \left(1 - \frac{d_i}{r_i}\right). \tag{2}$$

2.3 The Cox proportional hazards model

The Cox proportional hazards model is the most commonly used model in survival analysis due to the fact that it does not require knowledge of the underlying

distribution. The baseline hazard function in this model can be an arbitrary non-negative function, but the baseline hazard functions of different individuals are assumed to be the same. The Cox model provides a useful and easy way to interpret information regarding the relationship of the hazard function to predictors.

For each data point (X_i, y_i, δ_i), the hazard function $h(t)$ in the Cox model follows the proportional hazards assumption given by

$$h(t) = h_0(t)e^{\beta.X_i}, \tag{3}$$

where the baseline hazard function $h_0(t)$ can be an arbitrary non-negative function, X_i is the feature vector, β is the corresponding coefficient vector, and $\beta.X_i$ is the vector scalar product. From Eq. (3), we can deduce that the survival function can be computed as

$$S(t) = S_0(t)e^{\beta.X_i}, \tag{4}$$

where $S_0(t)$ is the baseline survival function given by $S_0(t) = e^{-\int_0^t h_0(x)dx}$.

Parameters of the Cox regression model are estimated by maximizing the partial likelihood. Based on the Cox regression formula, a partial likelihood can be constructed as

$$L(\beta) = \prod_{\delta_i=1} \frac{e^{\beta.X_i}}{\sum_{t_j \geq t_i} e^{\beta.X_j}}. \tag{5}$$

By setting the derivative of Eq. (5) with respect to β equal to zero, we can estimate the coefficients and hence the baseline hazard function. Simply stated, the parameter estimates represent the increase in the expected logarithm of the relative hazard for each one unit increase in the feature, holding other features constant:

$$ln\left(\frac{h(t)}{h_0(t)}\right) = \beta.X_i = \beta_1.X_{i1} + \beta_2.X_{i2} + \cdots \tag{6}$$

3 Related work

In this section, we survey briefly some of the relevant literature on two fronts, namely on the insider threat detection problem and on a handful of survival analysis applications. For more details, the reader is encouraged to consult the survey papers (Homoliak 2019; Mohammed Nasser 2020; Wang et al. 2019; Yuan and Wu 2021) and the references therein.

3.1 Insider threat detection

The problem of insider threat detection is usually framed as an anomaly detection task. In general, the purpose of anomaly detection is to find patterns in data that do not conform to the expected behavior. The key problem in this field is the difficulty to

model a user's normal behavior, as explained in the comprehensive survey (Chandola et al. 2009). In the same realm, the authors in Rashid et al. (2016) make use of Hidden Markov Models to learn what constitutes normal behavior, and then use them to detect significant deviations from that behavior.

Several studies have proposed the use of deep feedforward neural networks for insider threat detection. The paper Liu et al. (2018a) uses deep autoencoder to detect potential insider threats. A deep autoencoder consists of an encoder and a decoder, where the encoder encodes the input data to hidden representations while the decoder aims to reconstruct the input data based on the hidden representations. The objective of the deep autoencoder is to make the reconstructed input close to the original input. Because the majority of activities in an organization are benign, the input with insider threats should have relatively high reconstruction errors.

Recurrent neural networks (RNN) are mainly used for modeling sequential data, which maintain a hidden state with a self-loop connection to encode the information from a sequence. The user activities on a computer can be naturally modeled as sequential data. As a result, many RNN-based approaches have been proposed to model user activities for insider threat detection. The basic idea is to train an RNN model to predict a user's next activity or period of activities. The paper Tuor et al. (2017) proposes a stacked LSTM structure to capture the user's activities in a day and adopts negative log-likelihoods of such activities as the anomalous scores to identify malicious sessions.

Convolutional Neural Networks (CNN) have achieved great success in computer vision. A typical CNN structure consists of a convolutional layer, followed by a pooling layer, and a fully connected layer for prediction. A recent study on insider threat detection proposes a CNN-based user authentication method by analyzing mouse bio-behavioral characteristics (Hu et al. 2019). The proposed approach represents the user mouse behaviors on a computer as an image. If an ID theft attack occurs, the user's mouse behaviors will be inconsistent with the legal user. Hence, a CNN model is applied on images generated based on the mouse behavior to identify potential insider threats.

Graph convolutional networks (GCN), able to model the relationships between nodes in a graph, have gained increasing popularity for graph analysis. They use graph convolutional layer to extract node information. The paper Jiang et al. (2019) adopts a GCN model to detect insiders. Since users in an organization often make connections to each other via email or operation on the same devices, it is natural to use a graph structure to capture the inter-dependencies among users. Besides taking the adjacency matrix of structural information as input, GCN also incorporates the rich profile information about the users as the feature vectors of nodes. After applying the convolutional layers for information propagation based on the graph structure, GCN adopts the cross-entropy as the objective function to predict malicious nodes (users) in a graph.

3.2 Survival analysis applications

Survival analysis aims to model data where the outcome is the time until the occurrence of an event of interest. It was originally used in health data analysis and has since been employed in many applications, such as predicting student dropout time (Ameri 2016).

Over the past few years, a number of advanced machine learning methods have been developed to deal with and make predictions based on censored data. Ensemble learning methods (bagging, boosting, etc.) generate a committee of classifiers and then predict the class labels for new data points as they arrive by taking a weighted vote among the prediction results from all these classifiers (Dietterich 2000). It is often possible to construct good ensembles and obtain a better approximation of the unknown function by varying the initial points, especially in the presence of insufficient data. Such ensemble models have been successfully adapted to survival analysis whose time complexity mainly follows that of the base-learners (survival trees, random survival forests, etc.).

Active learning based on data containing censored observations allows the opinions of an expert in the domain to be incorporated into the models. Active learning mechanisms allow the survival model to select a subset of subjects by learning from a limited set of labeled subjects first, and then querying the expert to confirm a label for the survival status before considering including new data in the training set. The feedback from the expert is particularly useful for improving the model in real-world application domains. The goal of active learning for survival analysis problems is to build a survival regression model by utilizing the censored instances completely, without deleting or modifying the instances (Vinzamuri et al. 2014).

Collecting labeled information for survival problems is very time consuming, as it is necessary to wait for the event to occur in a sufficient number of training instances to build robust models. A naive solution for this insufficient data problem is to merely integrate the data from related tasks into a consolidated form and build prediction models on this integrated dataset. However, such approaches often do not perform well because the target task becomes overwhelmed by auxiliary data with different distributions. In such scenarios, knowledge transfer between related tasks usually produces much better results. The authors in Li et al. (2016b) propose the use of a regularized Cox model to improve the prediction performance of the Cox model in the target domain through knowledge transfer from the source domain in the context of survival models built on multiple high-dimensional datasets.

In Li et al. (2016a), the authors reformulated the survival time prediction problem as a multitask learning problem. In survival data, the outcome labeling matrix is necessarily incomplete since the event label of each censored instance is unavailable after its corresponding censoring time. This means that it is not possible to handle censored information using the standard multitask learning methods. To address this problem, the Multitask Learning Model for Survival Analysis (MTLSA) translates the original event labels into an indicator matrix to capture the dependency between the outcomes at various time points by using a shared representation across the related tasks in the transformation, which will reduce the prediction error for each task.

4 Methodology

In this section,we give a technical description of the features in the datasets used in our experiments. We then explain the details of the techniques employed for survival analysis. This leads to the layout of our computational setting.

4.1 Dataset description

There is no comprehensive real world dataset publicly available for insider threat detection, unfortunately. In this study, we adopt the synthetic CMU CERT Insider Threat Test Dataset (Glasser and Lindauer 2013) to create our own variation, modified to suit the need of our experiments. The CERT division of Software Engineering Institute at Carnegie Mellon University maintains a database of more than 1000 real case studies of insider threat and has generated a collection of synthetic insider threat datasets using scenarios containing traitor instances and masquerader activities.

The dataset captures 17 months of activity logs of 1000 users with 70 insiders in a simulated organization, resulting in roughly 33 million log lines. It is divided into five activity categories: logon, device, http, email, and file. Each of these subsets of data contain recorded information relating to employees in the organization. More precisely, the *logon* file records the logon and logoff operations of all employees, the *device* file records the usage of a thumb drive (connect or disconnect), the *http* file records all the web browsing operations (visit, download, or upload), the *email* file records all the email operations (send or receive), and the *file* file records activities involving a removable media device (open, write, copy or delete).

In addition to the employees' activity data on computers, the CERT dataset also includes the psychometric score for each employee, known as "Big Five personality traits". These traits are Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism (OCEAN) and are defined as follows. Openness measures the level of creativity and desire for knowledge and new experiences. Conscientiousness measures the level of care in someone's life and work. Extraversion measures the level of sociability. Agreeableness measures the level of helpfulness and tendency to compromise toward other people. Neuroticism measures emotional reactions to good/bad news.

In this paper, we specifically focus on the device and logon data subsets. These specific areas are chosen due to their relevance in historic insider threat instances. For example, a recent survey (Maddie 2020) conducted by *Tessian* revealed that 45% of employees download, save, and exfiltrate work-related documents before leaving or after being dismissed from a job (see Fig. 3). The preprocessing steps of the data consisted of separating the two columns that represent the device and logon attributes, crossing any incomplete rows for accuracy, and running the survival analysis rubrics in Python. All computations were performed on a personal laptop (64 GB RAM Core i7) over several periods of time, a couple of days each.

4.2 Survival analysis techniques

In order to generate the Kaplan–Meier curves, we first divide the study time into intervals $t_0 < t_1 < \ldots < t_m$, where t_0 and t_m are the starting and the ending times, respectively. We then define d_i as the number of malicious system events at time

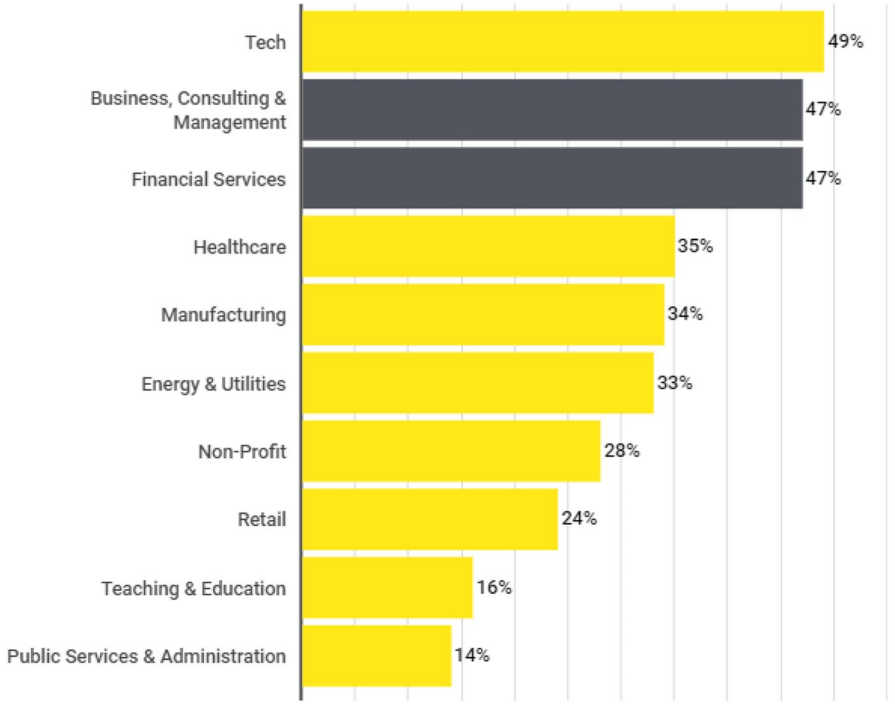

Fig. 3 Employees' actions before leaving a job, sorted by industry (Maddie 2020)

t_i and n_i as the number of non-malicious users left at time t_i, for $i = 0, 1, \ldots, m$. Based on Eqs. (1–2), we can iteratively estimate the survival function as follows

$$S(t_0) = 1 \quad S(t_{i+1}) = S(t_i) * \left(1 - \frac{d_i}{n_i}\right) \tag{7}$$

On the other hand, the Cox proportional hazards model requires a priori the identification of the risk factors associated with the prediction of the outcome. To this end, we extract seven variables in the datasets in question and we define them as follows. The first variable is the time of action, represented as hours since midnight (i.e., 2 : 30 am = 2.5). The next five variables are the respective scores in the big five traits explained in the previous section (OCEAN). The last variable is the activity binary status of the user; in the device file 0 = disconnect, 1 = connect, and in the logon file 0 = logoff, 1 = logon. Finally, each data point is labeled malicious (1) if it is an insider and benign (0) otherwise. Table 1 records a snippet of the logon file.

Table 1 Example logon data points

ID	Time	O	C	E	A	N	Activity	
JKS2444	7.378	48	16	19	34	30	1	0
CBA1023	7.528	20	42	23	21	28	1	0
GNT0221	7.558	43	19	21	24	27	1	0

5 Results and discussion

The resulting Kaplan–Meier curves are recorded in Fig. 4 for the device dataset and in Fig. 5 for the logon dataset. From these curves, we can estimate the probability that an employee "survives" past 300 days, for example, by locating 300 days on the x-axis and reading up and over to the y-axis. Moreover, at the end of the study period, the proportion of employees surviving is 99.35% based on the logon data and 99.5% based on the device data.

The results of the Cox proportional hazards model shed light on the effects of the characteristics of each system event on the overall likelihood of insider threats. Tables 2 and 3 show the parameter estimates corresponding to the variables of the model in the logon and device datasets, respectively (CI stands for confidence interval). For interpretability, we compute the hazard ratios by exponentiating the parameter estimates. If the hazard ratio is less than 1, then the predictor is protective (i.e., associated with improved survival) and if the hazard ratio is greater than 1, then the predictor is associated with increased risk (or decreased survival). For example, in the logon dataset, there is a 7% increase in the expected hazard relative to a one unit increase in neuroticism, holding all other factors constant, while there is a 4% decrease in the expected hazard relative to a one unit increase in agreeableness.

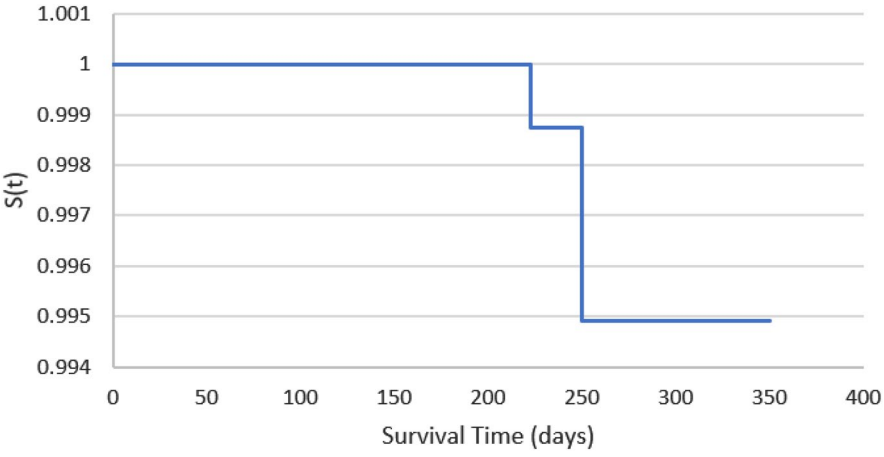

Fig. 4 The Kaplan–Meier curve for the device dataset

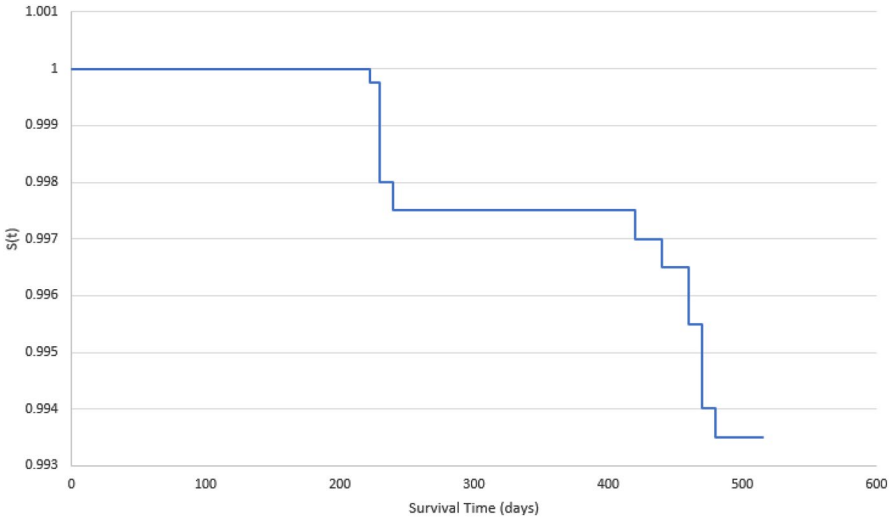

Fig. 5 The Kaplan–Meier curve for the logon dataset

Table 2 Parameter estimates for the logon dataset

Risk factor	Parameter estimate	95% CI	Hazard rate	95% CI
Time	0.28	[0.19,0.36]	1.32	[1.21,1.44]
Openness	− 0.03	[− 0.07,0.00]	0.97	[0.93,1.00]
Conscientiousness	0.03	[0.00,0.07]	1.03	[1.00,1.07]
Extraversion	− 0.01	[− 0.04,0.03]	0.99	[0.96,1.03]
Agreeableness	0.06	[0.02,0.10]	0.96	[1.02,1.11]
Neuroticism	− 0.04	[− 0.12,0.03]	1.07	[0.89,1.03]
Activity	0.34	[− 0.83,1.50]	1.40	[0.44,4.48]

Table 3 Parameter estimates for the device dataset

Risk factor	Parameter estimate	95% CI	Hazard rate	95% CI
Time	0.08	[− 0.13, 0.29]	1.08	[0.88, 1.33]
Openness	− 0.16	[− 0.26, − 0.07]	0.85	[0.77, 0.94]
Conscientiousness	0.02	[− 0.04, 0.08]	1.02	[0.96, 1.09]
Extraversion	− 0.04	[− 0.11, 0.04]	0.96	[0.90, 1.04]
Agreeableness	− 0.04	[− 0.11, 0.04]	0.96	[0.90, 1.04]
Neuroticism	0.07	[− 0.10, 0.23]	1.07	[0.91, 1.26]
Activity	0.02	[− 1.37, 1.40]	1.02	[0.25, 5.07]

In order to analyze the relationship between the predictors and the risk of insider threat, we take a closer look at the hazard ratios in Tables 3 and 4. We summarize the findings below:

1. Openness, extraversion, and agreeableness are negatively correlated with the relative hazard. This can be explained by the fact that employees who are creative, social, and helpful to others are less likely to pose an insider threat.
2. Conscientiousness and neuroticism are positively correlated with the relative hazard. One way to justify this is that nosy employees and those with a high emotional spectrum have more tendency to become insiders within their organizations.
3. In an obvious sense, it is no surprise that the level of activity performed by an employee (logon/logoff, inserting a flash drive, etc.) has the potential of increasing the risk of insider threat.
4. In both datasets, there is a positive association between the time of activity and the relative hazard of insider events. This is indicative that it is highly probable that malicious events will take place later in the day, even after working hours.

Despite the small amount of insider events and the consequent skewed nature of the CERT dataset, the results above are able to capture the intricate relations between insider threat risk and different attributes of an employee. Similar findings were previously achieved using machine learning techniques in general, and deep learning in particular Lu and Wong (2019). The advantages of our methods here lie in the natural extension of survival analysis to the insider threat domain, which to date remained surprisingly unexplored, and in the accurate pinpointing of the specific attributes that correlate (positively or negatively) with the risk of insider threat incidents.

6 Conclusion

The insider threat problem is one the most challenging security threats and the main concern of organizations of all sizes and in all industries. Hence, understanding and gaining insights into insider threat detection is an important research direction that remains underexplored. In this paper, we introduce a survival analysis based framework for the problem of estimating high risk employees in an organization based on their personality traits, as well as the time and the type of activities they perform on their computers. We use the CERT Insider Threat Dataset to validate the correlations between the predicting factors and the relative risk of insider threats. Our results show a negative association between openness, extraversion, and agreeableness and the relative hazard, while the remaining factors (conscientiousness, neuroticism, time, and activity) exhibit a positive correlation with the potential risk induced. To the best of our knowledge, our work is the first to deploy survival analysis techniques in the study of the insider threat problem.

There remain a lot of challenges in this field of study, we mention a couple of them herein. First, the major issue in insider threat detection research is the lack of real data for assessing the dynamics of the threat as well as the possible defense solutions. Most work to date relies on a handful of synthetic datasets that are often criticized for not correctly modeling real environments. Second, more factors need to be taken into consideration when evaluating an employee's activity within an organization. For example, current datasets should be enriched by adding additional predictors like previous history, promotion tracker, performance reviews, etc. while being cautious in quantifying them. Third, even with full blown datasets, the community still faces the difficulty of drawing a clear line between what is legitimate behavior and what is malicious behavior. This raises unnecessary false alarms in handling anomaly detection cases and makes detection of insider activity even harder. Fourth, the ability of capturing logs for the activities is an advantage that may provide insight into employee actions. Despite this advantage, the analysis of activity logs continues to be difficult for analysts because of the sheer volume of activities that employees produce every day. The high dimensionality of the monitored activities results in the massive needs for data to be processed and creates an extra hurdle to be overcome. Last but not least, the complexity of detecting insider threats is increasing due to the failure of current defense systems, the diversity of possible insider attacks, and the ability of employees to work from anywhere and be connected to any network outside their organization's servers.

As technological advances provide better tools to detect and prevent insider threat attacks, they also introduce new threats. They not only make it easier for adversaries to engage trusted human actors in a network but also introduce new, nonhuman trusted agents such as mobile devices, internet-connected devices, and artificial intelligence (AI). Indeed, many researchers have recently called for the definition and treatment of insider threat to be expanded to include technology that acts as trusted agents within networks (Cybersecurity 2019). More precisely, technology has resulted in an increase in external adversary use of unwitting insiders to gain a digital foothold in an organization. The crux of the argument is that a continued human-centric approach that focuses solely on malicious actors is myopic and dangerous; the insider is the trusted actor on a network, whether that actor is human, an embedded device, the software, the network, or the AI, and its risk should be considered regardless of whether the action is volitional or non-volitional and whether the motive is malicious or non-malicious.

Finally, we hope the survival analysis techniques used in this study could be transfered to other pressing issues where early prediction is essential such as network traffic anomalies, sex offender registries, organized crime, etc. Survival analysis techniques provide a relatively simple and effective way to predict the occurrence of specific events of interest at future time points. Due to the widespread availability of survival data from various domains, combined with the recent developments in various machine learning methods, there is an increasing demand for methods that can help understand and improve the way survival data might be handled and interpreted.

Acknowledgements The authors acknowledge the support of COL Paul Goethals and the Insider Threat Research Center at the United States Military Academy in West Point, NY.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Al-Mhiqani MN, et al (2020) A review of insider threat detection: classification, machine earning techniques, datasets, open challenges, and recommendations. Appl Sci 10(15):5208

Ameri S, et al (2016) Survival analysis based framework for early prediction of student dropouts. In: Proceedings of the 25th ACM international on conference on information and knowledge management. pp 903–912

Belk RW, Hix TD (2018) Insider threat program: maturity framework

Carley KM (2020) Social cybersecurity: an emerging science. Comput Math Org Theory 26(4):365–381

Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: A survey. ACM Comput Surv (CSUR) 41(3):1–58

Costa Daniel L, Albrethsen Michael J, Collins Matthew L (2016) Insider threat indicator ontology. Tech. rep. Carnegie Mellon University, Pitts burg, PA

David RA, Sproull RF (2019)Cybersecurity: a growing challenge for engineers and operators. In: The bridge: linking engineering and society vol 49(3)

Dietterich TG (2000) Ensemble methods in machine learning. In: International workshop on multiple classifier systems. Springer. pp 1–15

Glasser J, Lindauer B (2013) Bridging the gap: a pragmatic approach to generating insider threat data. In: 2013 IEEE security and privacy workshops. IEEE. pp 98–104

Homoliak I et al (2019) Insight into insiders and IT: a survey of insider threat taxonomies, analysis, modeling, and countermeasures. ACM Comput Surv (CSUR) 52(2):1–40

Hu T, et al (2019) An insider threat detection approach based on mouse dynamics and deep learning. In: Security and communication networks 2019

Jiang J, et al (2019) Anomaly detection with graph convolutional networks for insider threat and fraud detection. In: MILCOM 2019-2019 IEEE military communications conference (MILCOM). IEEE. pp 109–114

Klein JP, Zhang M-J (2005) Survival analysis, softwar. In: En-cyclopedia of biostatistics 8

Li Y, et al (2016a) A multi-task learning formulation for survival analysis. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp 1715–1724

Li Y, et al (2016b) Transfer learning for survival analysis via efficient l2, 1-norm regularized cox regression. In: 2016 IEEE 16th international conference on data mining (ICDM). IEEE, pp 231–240

Liu L, et al (2018a) Anomaly-based insider threat detection using deep au-toencoders. In: 2018 IEEE international conference on data mining workshops (ICDMW). IEEE. pp 39–48

Liu L et al (2018b) Detecting and preventing cyber insider threats: a survey. IEEE Commun Surv Tutor 20(2):1397–1417

Lu J, Wong RK (2019) Insider threat detection with long short-term memory. In: Proceedings of the Australasian Computer Science Week Multiconference. pp 1–10

Maddie R (2020) Insider threat statistics you should know. https://www.tessian.com/blog/insider-threat-statistics/. accessed 10 June 2020

Miller RG Jr (2011) Survival analysis, vol 66. Wiley, Hoboken

Obama B (2011) Structural reforms to improve the security of classified networks and the responsible sharing and safeguarding of classified information - executive order 13587

Rashid T, Agrafiotis I, Nurse JRC (2016) A new take on detecting insider threats: exploring the use of hidden markov models. In: Proceedings of the 8th ACM CCS international workshop on managing insider security threats. pp 47–56

Tuor A, et al (2017) Deep learning for unsupervised insider threat detection in structured cybersecurity data streams. arXiv:1710.00811

U.S. State of Cybercrime (2018) Tech. rep. CERT Division of SRI-CMU, and ForcePoint

Vinzamuri B, Li Y, Reddy CK (2014) Active learning based survival regression for censored data. In: Proceedings of the 23rd ACM international conference on conference on information and knowledge management. pp 241–250

Wang P, Li Y, Reddy CK (2019) Machine learning for survival analysis: a survey. ACM Comput Surv (CSUR) 51(6):1–36

Yuan S, Wu X (2021) Deep learning for insider threat detection: review, challenges and opportunities. In: Computers & Security, pp 102221

Yuan F, et al (2018) Insider threat detection with deep neural network. In: International conference on computational science. Springer. pp 43–54

Elie Alhajjar is a research scientist at the Army Cyber Institute (ACI) and jointly an Assistant Professor in the Department of Mathematical Sciences at the United States Military Academy (USMA) in West Point, NY, where he teaches and mentors cadets from all academic disciplines. His research interests include mathematical modeling machine learning and network analysis, from a cybersecurity viewpoint. He has presented his research work in international meetings in North America, Europe, and Asia. Before coming to West Point, Dr. Elie Alhajjar had a research appointment at the National Institute of Standards and Technology (NIST) in Gaithersburg, MD. He holds a Master of Science and a PhD in mathematics from George Mason University, as well as master's and bachelor's degrees from Notre Dame University.

Taylor Bradley is a cyber lieutenant in the US Army. She is currently completing her masters degree in Cybersecurity at Johns Hopkins University in Baltimore, MD. Her research interests include network architecture amd cyber operations.