International Journal of Current Advanced Research
ISSN: O: 2319-6475, ISSN: P: 2319-6505, Impact Factor: SJIF: 5.995
Available Online at www.journalijcar.org
Volume 6; Issue 12; December 2017; Page No. 7944-7949
DOI: http://dx.doi.org/10.24327/ijcar.2017.7949.1258

Research Article

BEHAVIORAL ANALYSIS OF USER IN CYBER CRIMES BY PREDICTION TECHNIQUES

Swapna S*

Aurora's Technological Research Institute

ARTICLE INFO

Article History:
Received 15th September, 2017
Received in revised form 25th October, 2017
Accepted 23rd November, 2017
Published online 28th December, 2017

Key words:

Behavior Analysis, Cyber Security, Cyber Attacks, Prediction Analysis

ABSTRACT

With the increasing usage of Internet and computing devices with network competence, the Internet crimes and cyber attacks are increasing exponentially. Most of the existing detection and protection systems rely on signature based methods and are unable to detect sophisticated and targeted attacks like advanced persistent threats (APTs). In order to protect Internet users and cyber infrastructure from various threats, proactive defense systems are required, which have the capability to make intelligent decisions in real time. This paper reviews various predictive techniques that can be used for predicting the cyber user whether he is an attacker or not depending on his behavioral parameters, It also highlights the challenges, which can be explored by researchers for future studies.

INTRODUCTION

The information and communications technology (ICT) industry has evolved greatly over the last half century. The technology is ubiquitous and increasingly integral to almost every facet of modern society. ICT devices and components are generally interdependent, and disruption of one may affect many others. Over the past several years, experts and policymakers have expressed increasing concerns about protecting ICT systems from cyber-attacks, which many experts expect to increase in frequency and severity over the next several years. The act of protecting ICT systems and their contents has come to be known as cyber security. A broad and arguably somewhat fuzzy concept, cyber security can be a useful term but tends to defy precise definition. It is also sometimes inappropriately conflated with other concepts such as privacy, information sharing, intelligence gathering, and surveillance. However, cybersecurity can be an important tool in protecting privacy and preventing unauthorized surveillance, and information sharing and intelligence gathering can be useful tools for effecting cyber security. The management of risk to information systems is considered fundamental to effective cyber security.

The risks associated with any attack depend on three factors: threats (who is attacking), vulnerabilities (the weaknesses they are attacking), and impacts (what the attack does). Most cyberattacks have limited impacts, but a successful attack on some components of critical infrastructure (CI)-most of which is held by the private sector-could have significant effects on national security, the economy, and the livelihood and safety of individual citizens. Reducing such risks usually involves removing threat sources, addressing vulnerabilities, and lessening impact.

Cybersecurity Issues and Challenges

In Brief Congressional Research Service to that person. Thus, good cybersecurity can help protect privacy in an electronic environment, but information that is shared to assist in cybersecurity efforts might sometimes contain personal information that at least some observers would regard as private.

Cybersecurity can be a means of protecting against undesired surveillance of and gathering of intelligence from an information system. However, when aimed at potential sources of cyberattacks, such activities can also be useful to help effect cybersecurity. In addition, surveillance in the form of monitoring of information flow within a system can be an important component of cybersecurity.

Management of Cybersecurity Risks

The risks associated with any attack depend on three factors: threats (who is attacking), vulnerabilities (the weaknesses they are attacking), and impacts (what the attack does). The management of risk to information systems is considered fundamental to effective cybersecurity.

The Cyber Security Threats

People who actually or potentially perform cyberattacks are widely cited as falling into one or more of five categories: criminals intent on monetary gain from crimes such as theft or extortion; spies intent on stealing classified or proprietary information used by government or private entities; nation-state warriors who develop capabilities and undertake

Corresponding author:* **Swapna S
Aurora's Technological Research Institute

cyberattacks in support of a country's strategic objectives; "hacktivists" who perform cyberattacks for nonmonetary reasons; and terrorists who engage in cyberattacks as a form of non-state or state-sponsored warfare.

The Cyber Security Vulnerabilities

Cybersecurity is in many ways an arms race between attackers and defenders. ICT systems are very complex, and attackers are constantly probing for weaknesses, which can occur at many points. Defenders can often protect against weaknesses, but three are particularly challenging: inadvertent or intentional acts by insiders with access to a system; supply chain vulnerabilities, which can permit the insertion of malicious software or hardware during the acquisition process; and previously unknown, or zero-day, vulnerabilities with no established fix. Even for vulnerabilities where remedies are known, they may not be implemented in many cases because of budgetary or operational constraints.

The Impacts of Cyber Security

A successful attack can compromise the confidentiality, integrity, and availability of an ICT system and the information it handles. Cybertheft or cyberespionage can result in exfiltration of financial, proprietary, or personal information from which the attacker can benefit, often without the knowledge of the victim. Denial-of-service attacks can slow or prevent legitimate users from accessing a system. Botnet malware can give an attacker command of a system for use in cyberattacks on other systems.

Managing the risks from cyberattacks usually involves

1. Removing the threat source.
2. Addressing vulnerabilities by hardening ICT assets.
3. Lessening impacts by mitigating damage and restoring functions

LITERATURE REVIEW

Behavior analysis

Behavior analysis is rooted in the behaviorist tradition and utilizes learning principles to bring about behavior change. Some branches of psychology strive to understand underlying cognitions, but behavioral psychology is not concerned with mentalistic causes of behavior and instead focuses on the behavior itself.

Behavioural analysis can use number of factors depending on the type of user, these include:

People's work rhythm-when they are working

Every workday is the same! Lots of bored employees complain about that. They wake up at the same time, do the same routines before going to work and arrive to their workplace at approximately at the same time. Furthermore, they try to have lunch and snacks every day at the same time, and leave the office around the end of their working hours. So they behave very similarly in at least 90 percent of their workdays. For example, I start to work around 8:30 AM every morning and leave the office at 5:30 PM. Based on that, logging in in the middle of the night would be highly unusual!

Used applications-what do they run

Most of us are using the same applications day by day. For example, I'm usually running MS Word & Excel, Google Chrome, File Explorer, Evernote and sometimes Paint. But I never use SAP, Jupyter Notebook or Emacs, which are applications frequently used by our finance department, data scientists or developers. It means that the usage of these-or other unusual-apps would be highly suspicious.

Accessed files and servers-what are we working on

Although I'm a curious person, I usually access only the marketing server in my work, where I can find all of the files I need. I'm sure that the HR director would ask me why I downloaded and opened the Excel-sheet, which contains the salary of all of my colleagues.

Work environment-from where and what device do we use

I work in the same office day by day and spent only a several workdays far away from it, when I visit one of the major IT security conferences of the world. Furthermore, I always bring my corporate notebook when I travel to somewhere, as most of the office workers do that. In other words, I log in to the corporate network from Vietnam, from a host, that I never used before, would be an obvious anomaly.

Keystroke dynamics-how do we type?

Fingerprint reading and retina scanning are the most well-known forms of biometrics authentication-but these are not the only ones. The way we type is also very idiosyncratic. Not only the speed, but the mistakes that we make as we type, and the duration between the pushing of two specific characters varies from person to person.

Security intelligence and threat knowledge

There is much information out there about the type of attacks being perpetrated and how they are being initiated. Security companies build up profiles of attack vectors and malware instances and use these to predict next moves and identify incoming threats

Profile analysis

To understand and determine any changes in behaviour, you have to understand the behaviour first. Behavioural analysis works by analysing normal behavioural patterns. A simple example, known as credential behavioural monitoring, would be to apply user specific questions to a login attempt that looks like it may be a brute force, or is coming in from an unusual location, etc. An example of its implementation could be if the monitored user is genuine, instead of locking their account, which is both annoying and can lead to DOS attacks, you can ask them some personal questions; if they answer correctly they are logged in. Another example is analysing a particular action, say a database query; Malware presents a very different profile when extracting data, than a human being performing the same operation.

Monitoring, analysis and detection

This involves understanding your baseline of expected behaviour on a network, for example, knowing which are trusted sites, the types of files accessed by individuals, the types of access to servers, external sites and as and so on that are normal for that network. You can use the profile analysis information as a basis for your monitoring and detection of potential cyber attacks. Traffic behaviour is one area that can give a lot of information and allow early detection of anomalies. It can also help in the fight against Botnets, which are typically difficult to detect.

Prediction Issues

The major issue is preparing the data for Prediction. Preparing the data involves the following activities-

Data Cleaning- Data cleaning involves removing the noise and treatment of missing values. The noise is removed by applying smoothing techniques and the problem of missing values is solved by replacing a missing value with most commonly occurring value for that attribute.

Relevance Analysis-Database may also have the irrelevant attributes. Correlation analysis is used to know whether any two given attributes are related.

Data Transformation and reduction-The data can be transformed by any of the following methods.

- **Normalization**-The data is transformed using normalization. Normalization involves scaling all values for given attribute in order to make them fall within a small specified range. Normalization is used when in the learning step, the neural networks or the methods involving measurements are used.
- **Generalization**-The data can also be transformed by generalizing it to the higher concept. For this purpose we can use the concept hierarchies.

Note-Data can also be reduced by some other methods such as wavelet transformation, binning, histogram analysis, and clustering.

The criteria for comparing the methods of Prediction

- **Accuracy**-Accuracy of classifier refers to the ability of classifier. It predict the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data.
- **Speed**-This refers to the computational cost in generating and using the classifier or predictor.
- **Robustness**-It refers to the ability of classifier or predictor to make correct predictions from given noisy data.
- **Scalability**-Scalability refers to the ability to construct the classifier or predictor efficiently; given large amount of data.
- **Interpretability**-It refers to what extent the classifier or predictor understands.

Predictive Models

Function	Algorithm
Classification	Naive Bayes Adaptive Bayes Network Support Vector Machine
Regression	Support Vector Machine
Attribute Importance	Minimal Descriptor Length

Classification

In a classification problem,typically we have historical data (labeled examples) and unlabeled examples. Each labeled example consists of multiple predictor attributes and one target attribute (dependent variable). The value of the target attribute is a class label. The unlabeled examples consist of the predictor attributes only. The goal of classification is to construct a model using the historical data that accurately predicts the label (class) of the unlabeled examples.

A classification task begins with build data for which the target values are known. Different classification algorithms use different techniques for finding relations between the predictor attributes' values and the target attribute's values in the build data. These relations are summarized in a model, which can then be applied to new cases with unknown target values to predict target values. A classification model can also be used on build data with known target values, to compare the predictions to the known answers; such data is also known as *test data* or *evaluation data*. This technique is called testing a model, which measures the model's predictive accuracy. The application of a classification model to new data is called *applying the model*, and the data is called *apply data* or *scoring data*. Applying data is often called *scoring the data*.

Classification is used in customer segmentation, business modeling, credit analysis, and many other applications. For example, a credit card company may wish to predict which customers will default on their payments. Each customer corresponds to a case; data for each case might consist of a number of attributes that describe the customer's spending habits, income, demographic attributes, etc. These are the predictor attributes. The target attribute indicates whether or not the customer has defaulted; that is, there are two possible classes, corresponding to having defaulted or not. The build data is used to build a model that you then use to predict, for new cases, whether these new customers are likely to default.

Costs

In a classification problem, it may be important to specify the costs involved in making an incorrect decision. Doing so can be useful when the costs of different misclassifications vary significantly.

For example, suppose the problem is to predict whether a user will respond to a promotional mailing. The target has two categories: YES (the customer responds) and NO (the customer does not respond). Suppose a positive response to the promotion generates $500 and that it costs $5 to do the mailing. If the model predicts YES and the actual value is YES, the cost of misclassification is $0. If the model predicts YES and the actual value is NO, the cost of misclassification is $5. If the model predicts NO and the actual value is YES, the cost of misclassification is $500. If the model predicts NO and the actual value is NO, the cost is $0.

The row indexes of a cost matrix correspond to *actual values*; the column indexes correspond to *predicted values*. For any pair of actual/predicted indexes, the value indicates the cost of misclassification.

Classification algorithms apply the cost matrix to the predicted probabilities during scoring to estimate the least expensive prediction. If a cost matrix is specified for apply, the output of the scoring run is prediction and cost., rather than predication and probability

Priors

In building a classification model, priors can be useful when the training data does not accurately reflect the real underlying population. A priors vector is used to inform the model of the true underlying distribution of target classes in the underlying population. The model build adjusts its predicted probabilities for Adaptive Bayes Network and Naive Bayes or relative Complexity factor for Support Vector Machine.

Naive Bayes Algorithm

The Naive Bayes algorithm (NB)N can be used for both binary and multiclass classification problems to answer questions such as "Which customers will switch to a competitor? Which transaction patterns suggest fraud? Which prospects will respond to an advertising campaign?" For example, suppose a bank wants to promote its mortgage offering to its current customers and that, to reduce promotion costs, it wants to target the most likely prospects. The bank has historical data for its customers, including income, number of household members, money-market holdings, and information on whether a customer has recently obtained a mortgage through the bank. Using NB, the bank can predict how likely a customer is to respond positively to a mortgage offering. With this information, the bank can reduce its promotion costs by restricting the promotion to the most likely candidates.

NB affords fast model building and scoring for relatively low volumes of data.

NB makes predictions using Bayes' Theorem, which derives the probability of a prediction from the underlying evidence.Bayes' Theorem states:
$P(A \mid B) = (P(B \mid A) P(A))/P(B)$

That is, the probability of event A occurring given that event B has occurred is equal to the probability of event B occurring given that event A has occurred, multiplied by the probability of event A occurring and divided by the probability of event B occurring.

NB assumes that each attribute is conditionally independent of the others: given a particular value of the target, the distribution of each predictor is independent of the other predictors.

In practice, this assumption, even when violated, does not degrade the model's predictive accuracy significantly, and makes the difference between a fast, computationally feasible algorithm and an intractable one.

Naive Bayes lets you, using cross-validation, test model accuracy on the same data that was used to build the model, rather than building the model on one portion of the data and testing it on a different portion. Not having to hold aside a portion of the data for testing is especially useful if the amount of build data is relatively small.

"Leave-one-out cross-validation" is a special case of cross-validation in which one record is left out of the build data when building a model. The number of models built equals the number of records (omitting a different build record for each model), which makes this procedure computationally expensive. With Naive Bayes models, however, the approach can be modified such that all build records are used for building a single model. Then, the model is repeatedly modified to quickly remove the effects of one build record, incrementally "unbuilding" the model for that record, as though that record had been omitted when building the model in the first place. The accuracy of the prediction for each build record can then be assessed against the model that would have been built from all the build records except that one, without having had to actually build a separate model for each build record.

Adaptive Bayes Network Algorithm

Adaptive Bayes Network (ABN) is an Oracle proprietary algorithm that provides a fast, scalable, non-parametric means of extracting predictive information from data with respect to a target attribute. (Non-parametric statistical techniques avoid assuming that the population is characterized by a family of simple distributional models, such as standard linear regression, where different members of the family are differentiated by a small set of parameters.)

ABN, in single feature build mode, can describe the model in the form of human-understandable rules. The rules produced by ABN are one of its main advantages over Naive Bayes. The business user, marketing professional, or business analyst can understand the basis of the model's predictions and can therefore be comfortable acting on them and explaining them to others. In addition to explanatory rules, ABN provides performance and scalability, which are derived via a collection of user parameters controlling the trade-off of accuracy and build time.

ABN predicts binary as well as multiclass targets. Binary targets are those that take on only two values, for example, *buy* and *not buy*. Multiclass targets have more than two values, for example, products purchased (product A or product B or product C). Multiclass target values are not assumed to exist in an ordered relation to each other, for example, hair brush is not assumed to be greater or less than comb.

ABN Model Types

An ABN model is an (adaptive conditional independence model that uses the minimum description length principle to construct and prune an array of conditionally independent Network Features. Each Network Feature consists of one or more Conditional Probability Expressions. The collection of Network Features forms a product model that provides estimates of the target class probabilities. There can be one or more Network Features. The number and depth of the Network Features in the model determine the model mode. There are three model modes for ABN:

- Pruned Naive Bayes (Naive Bayes Build)
- Simplified decision tree (Single Feature Build)
- Boosted (Multi Feature Build)

ABN Rules

Rules can be extracted from the Adaptive Bayes Network Model as Compound Predicates. Rules form a human-interpretable depiction of the model and include statistics indicating the number of the relevant training data instances in support of the rule. A record apply instance specifies a pathway in a network feature taking the form of a compound predicate.

For example, suppose the feature consists of two training attributes: Age {20-40, 40-60, 60-80} and Income {<=50K, >50K}. A record instance consisting of a person age 25 and income $42K is expressed as
IF AGE IN (20-40) and INCOME IN (<=50K)

Suppose that the associated target (for example, response to a promotion) probabilities are {0.8 (no), 0.2 (yes)}. Then we have a detailed rule of the form
IF AGE IN (20-40) and INCOME IN (<=50K) =>Prob = {0.8, 0.2}

In addition to the probability distribution, there are the associated training data counts, e.g. {400, 100}. Suppose there is a cost matrix specifying that it is 6 times more costly to incorrectly predict a no than to incorrectly predict a *yes*. Then the cost of predicting *yes* for this instance is 0.8 * 1 = 0.8 (because the model is wrong in this prediction 80% of the time) and the cost of predicting no is 0.2 * 6 = 1.2. Thus, the minimum cost (best) prediction is *yes*.

Without the cost matrix and the decision is reversed. Implicitly, all errors are equal and we have: 0.8 * 1 = 0.8 for *yes* and 0.2 * 1 = 0.2 for *no*.
The order of the predicates in the generated rules implies relative importance.

When you apply an ABN model for which rules were generated, with a single feature, you get the same result that you would get if you wrote an external program that applied the rules.

ABN Build Parameters

To control the execution time of a build, ABN provides the following user-settable parameters:

***Maximum Network Feature Depth**:* NetworkFeatures are like individual decision trees. This parameter restricts the depth of any individual NetworkFeature in the model. At each depth for an individual NetworkFeature there is only one predictor chosen. Each level built requires an additional scan of the data, so the computational cost of deep feature builds is high. The range for this parameter consists of the positive integers. The NULL or 0 value setting has special meaning: unrestricted depth. Builds beyond depth 7 are rare. Default is 10.

Maximum Consecutive Pruned Network Features: The maximum number of consecutive pruned features before halting the stepwise selection process. Default is 1.

Maximum Build Time: The maximum build time (in minutes) parameter allows the user build quick, possibly less accurate models for immediate use or simply to get a sense of how long it will take to build a model with a given set of data. To accomplish this, the algorithm divides the build into milestones (model states) representing complete functional models (see ABNModelBuildState for details). The algorithm completes at least a single milestone and then projects whether it can reach the next one within the user-specified maximum build time. This decision is revisited at each milestone achieved until either the model build is complete or the algorithm determines it cannot reach the next milestone within the user-specified time limit. The user has access to the statistics produced by the time estimation procedure (see ABNModelBuildState for details). Default is NULL (no time limit).

MaximumPredictors: The maximum number of predictors is a feature selection mechanism that can provide a substantial performance improvement, especially in the instance of training tables where the number of attributes is high (but less than 1000) and is represented in single-record format. Note that the predictors are rank ordered with respect to an MDL measure of their correlation to the target, which is a greedy measure of their likelihood of being incorporated into the model. Default is 25.

Number Predictors In NB Model: The number of predictors in the NB model. The actual number of predictors will be the minimum of the parameter value and the number of active predictors in the model. If the value is less than the number of active predictors in the model, the predictors are chosen in accordance with their MDL rank. Default is 10.

Model Types: You can specify one of the following types when building an ABN model:

- ***MultiFeature Build:*** The model search space includes an NB model and single and multi-feature product probability models. Rules are produced only if the single feature model is best. No rules are produced for multi-feature or NB models.
- ***Single Feature Build:*** The model search space includes only a single feature model with one or more predictors. Rules are produced.
- ***NaiveBayesBuild:*** Only a single model is built, an NB model. It is compared with the global sample prior (the distribution of target values in the sample). If the NB model is a better predictor of the target values than the global prior, then the NB model is output. Otherwise no model is output. No rules are produced.

ABN Model States

When we specify MaxBuildTime for a boosted mode ABN model, the ABN build terminates in one of the following states:

- **CompleteMultiFeature:** Multiple features have been tested for inclusion in the model. MDL pruning has determined whether the model actually has one or more features. The model may have completed either because there is insufficient time to test an additional feature or because the number of consecutive features failing the stepwise selection criteria exceeded the maximum allowed or seed features have been extended and tested.
- **Complete SingleFeature:** A single feature has been built to completion.
- **Incomplete SingleFeature:** The model consists of a single feature of at least depth two (two predictors) but the attempts to extend this feature have not completed.
- **NaiveBayes:** The model consists of a subset of (single-predictor) features that individually pass MDL correlation criteria. No MDL pruning has occurred with respect to the joint model.

Linear models

This Model works most naturally with *numeric attributes*. The standard technique for numericprediction is *linear regression*.Predicted class value is *linear combination* of attribute values (a_i): $C = w_0*a_0 + w_1*a_1 + w_2*a_2 + ... + w_k*a_k$. For k attributes we have *k+1* coefficients. To simplify notation we add a_0 that is always ,*Squared error*: Sum through all instances (actual class value - predicted one). Deriving the coefficients (w_i): *minimizing squared error* on training data. Using standard numerical analysis techniques (matrix operations). Can be done if there are more instances than attributes (roughly speaking).

Classification by linear regression

Multi-response linear regression (learning a membership function for each class)

Training*:* perform a regression (create a model) for each class, setting the output to 1 for training instances that belong to the class, and 0 for those that do not.

Prediction*:* predict the class corresponding to the model with largest output value

Pairwise regression (designed especially for classification)

Training: perform regression for every pair of classes assigning output 1 for one class and -1 for the other.

Prediction: predict the class that receives most "votes" (outputs > 0) from the regression lines.

More accurate than multi-response linear regression, however more computationally expensive.

Fuzzy Set Approaches

Fuzzy Set Theory is also called Possibility Theory. This theory was proposed by LotfiZadeh in 1965 as an alternative the two-value logic and probability theory. This theory allows us to work at a high level of abstraction. It also provides us the means for dealing with imprecise measurement of data.

Research Objectives and Approach

The main objective of the research is to find the better solution to the problem by avoiding the cyber attacker not to confidential information from the network.

The main approach of this research is to study about the attacker carefully how he is going to attack the serving the behavior of the user.

This involves the study of many case studies regarding each and every user in spiteof their profession or qualification only he is accessing internet or not.

Usefullness of Research

“Prevention is better than cure” ,This solution may avoid the hackers to attack the users information because the previous solutions are only after the attacker attack the users information but not before that.To protect the user from any type of hacking is the main goal of this research.

Work Plan

1. The first step is to study about the users information accurately.
2. Then identifying each and every parameter from the information after behavioral analysis of the user’s data.
3. Then we can apply the preprocessing of data by using different techniques.
4. Analyzing to select the prediction algorithm which is applicable.
5. Then the resulthave to be evaluated whether it is accurate or not.
6. If the prediction of attacker is true then we can save some users information.

CONCLUSION

This research is very much use full in providing security to the users information from the cyber attackers in various areas. Its scope is not to an enterprise or an organization, it only handles the information of the user.

References

1. Kusiak, Z. Song, "Combustion efficiency optimization and virtual testing: A data-mining approach", *IEEE Trans. Ind. Inf.*, vol. 2, no. 3, pp. 176-184, Aug. 2006.
2. D. A. Spera, Wind Turbine Technology: Fundamental Concepts of Wind Turbine Engineering, ASME.

How to cite this article:

Swapna S (2017) 'Behavioral analysis of User in Cyber Crimes by Prediction Techniques ', *International Journal of Current Advanced Research,* 06(12), pp. 7944-7949. DOI: http://dx.doi.org/10.24327/ijcar.2017.7949.1258

****