Information & Security

N. Stoianov & M. Bozhilova
vol. 46, no. 2 (2020): 182-197
https://doi.org/10.11610/isij.4613

Published since 1998

ISSN 0861-5160 (print), ISSN 1314-2119 (online)

Research Article

A Model of a Cyber Defence Awareness System of Campaigns with Malicious Information

Nikolai Stoianov (✉), ***Maya Bozhilova***

Bulgarian Defence Institute "Prof. Tsvetan Lazarov",
2 "Prof. Tsvetan Lazarov" Blvd., Sofia 1594, Bulgaria
https://www.di.mod.bg/

ABSTRACT:

Many organizations experience cyberattacks with the aim of the dissemination of malicious information. Situational awareness is a tool to counteract the campaigns of malicious information and reduce its dissemination. This article proposes a conceptual model for a cyber defence awareness system, which aims to support human operators to avoid this type of threat. The system will identify (classify) three campaign types of malicious information operations – malicious information injections in web content, malicious information injections in fake social network accounts, and malicious information dissemination via email messages. A model for identification of the type of campaign of malicious information operations based on Dempster-Shafer evidence theory is proposed. The work presented here is a part of the Cyber Rapid Analysis for Defence Awareness of Real-time Situation - CyRADARS project.

ARTICLE INFO:

RECEIVED: 08 MAY 2020
REVISED: 12 AUG 2020
ONLINE: 21 AUG 2020

KEYWORDS:

cyber defence, situational awareness, malicious information, CyRADARS

Creative Commons BY-NC 4.0

✉ E-mail: n.stoianov@di.mod.bg

Introduction

The number of organizations that have experienced cyberattacks with the aim of dissemination of malicious information and malware is growing up every day. These attacks do not reside in borders of the country, they cross borders and are distributed in global cyberspace. In this context, situational awareness is an important tool for countermeasures and threat avoidance.

Traditionally, Situational Awareness (SA) is an activity for human decision making. Nowadays, because of the spread of information technologies in our life, big data and activities in the cyberspace, humans would not be able to analyse a huge amount of data and timely react to malicious attacks. SA activities should be assisted by a system, analysing constantly stored statistical data and/or other information for determining malicious information attacks.

One of the key priorities of NATO Science for Peace and Security Programme [20] is cyber defence situation awareness and Cyber Rapid Analysis for Defence Awareness of Real-time Situation project (CyRADARS) is an awarder to support research activities in this area.

The Cyber Rapid Analysis for Defence Awareness of Real-time Situation project – CyRADARS

The goal of the CyRADARS project,[19] as it is described in the Project plan, is to develop theoretical foundations, methods, and recommendations, as well as software tools for Situational Awareness (SA) that will enable, in an almost online mode, friendly security forces to:

- monitor Cyberspace to detect malicious information injections and give timely notification of an information attack;
- create conditions necessary for decision making about prevention or timely response to enemy's information injections.

The research tasks are related to discovering with machine learning techniques direct network attacks associated with disinformation campaigns or unusual network behaviour.

During the project, models, metrics, algorithms, and information technologies, including associated software are developed, which implement all three Situation Awareness (SA) levels: perception, comprehension, projection.[3, 4] The developed software is created to a large extent on the basis of open source technologies, so it will be possible to integrate the system with other information systems already in operation. This will provide an important, currently missing capability to protect against hostile information operations

Cyber Defence Situational Awareness

The focus of this work is one of the main modes of adversarial unarmed influence, namely, the spreading of malicious information in cyberspace. It has a complicated character due to a complex problem domain and complex behaviour.

There are huge numbers of publications on the current experience in SA based on the ideas and definition from.[3, 4] Endsley's definition of situational

awareness is "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future." According to the Project plan, assuring SA in area of detecting and tracking of malicious information operations implies performing activities at the following three levels:

Level 1: Perception of elements in the environment

In our case, systems of online information acquisition from multiple channels (such as Ukrainian, Russian, and English news and information websites, social networks, emails).

Level 2: Comprehension of the current situation

A human operator determines the presence of particular information threats. During this activity, the system assists him in analysing constantly stored statistical data or other information by determining information attacks and applying contrast analysis or other techniques.

Level 3: Projection of future status

In online mode, conclusions are being made about the evolution of the situation and possible future actions of the enemy, decisions are being made about resisting information injections:

- an order is issued to prepare a counterpropaganda message;
- distracting messages (prepared beforehand) are being broadcast;
- channels transmitting fake messages are being tracked, in order to terminate them afterward, and so on.

Cyber security situational awareness implies information to be collected from many sources. To analyse and evaluate the impact of security incidents on a networked system in [17] authors propose cyber security situation assessment model, based on multi-heterogeneous sensors. Security data obtained from multi-sensors are fused according to the network topology and the importance of services and hosts, using Dempster-Shafer evidence theory. Based on this theory in [18] is suggested a model of network security situation awareness, which obtains the values of security situation awareness of network by data source level fusion, host level fusion and system level fusion.

A multi-attribute decision-making trust evaluation model based on Dempster-Shafer evidence theory in a multi-source and heterogeneous environment is developed and described by Zheng.[16] By collecting, processing, and evaluating attack event information from many data sources of security devices, it evaluates the possibility of network intrusion for aims of a network security situation assessment.

In the next sections we propose a model for a cyber defence situational awareness system of campaigns with malicious information operations. The approach for recognition of a type of campaign of malicious information operations is based on Dempster-Shafer evidence theory, and the next section presents briefly the basics of this theory.

Dempster-Shafer Evidence Theory

Dempster-Shafer theory was first introduced by Arthur Dempster [13] in 1967 as a mechanism for reasoning under knowledge uncertainty, and subsequently extended by Glenn Shafer. In 1976 Shafer published 'A Mathematical theory of evidence.'[15]

Dempster-Shafer theory [10, 12, 14, 16] is a mathematical theory of evidence based on belief functions and plausible reasoning.

Definition 1: Assume a set of *n* mutually exclusive and exhaustive propositions (hypothesizes) $\Theta=\{X_0, X_1, \ldots, X_n\}$, Θ is called a frame of discernment, 2^{Θ} is the set of all the subset of Θ, called power set of Θ.

A mass value *m* between 0 and 1 is assigned to each subset of the power set. i.e.

$$m: 2^{\Theta} \rightarrow [0, 1] \quad (1).$$

The function (1) is called the mass function (or the basic probability assignment – BPA) whenever it satisfies the equation (2) and (3):

$$m(\emptyset) = 0 \quad (2)$$

$$\sum_{A\subseteq\Theta} m(A) = 1 \quad (3).$$

In other words, $m(A)$ is a measure of belief assigned by a given evidence to *A*, where *A* is any element of 2^{Θ}.

Definition 2: Belief function. The Belief function based on BPA *m* on the frame of discernment is defined as:

$$Bel: 2^{\Theta} \rightarrow [0, 1] \quad (4)$$

$$Bel\ (X)=\sum_{Y\subseteq X} m(Y)\ , for\ each\ X \subseteq \Theta \quad (5),$$

i.e *Bel(X)* is the degree of support for the proposition *X.*

Definition 3: Plausibility function. The Plausibility function based on BPA *m* on the frame of discernment is defined as:

$$Pls(X)=\sum_{Y\cap X\neq\emptyset} m(Y)\ , for\ each\ X \subseteq \Theta \quad (6)$$

Given mass assignments for the power set, the upper and lower bounds of a probability interval can be determined since these are bounded by two measures that can be calculated from the mass, the degree of belief (Bel) and the degree of plausibility (*Pls*).

It is always true that:

$$m(X) \leq Bel(X) \leq Pls(X). \quad (7).$$

Definition 4: Dempster rule of combination:

Dempster rule of combination is concerned with uniting two mass functions on a frame of discernment, for example m_1 and m_2. In this case, the combined mass function of m_1 and m_2 is denoted by $m_{1,2}$ where:

$$m_{1,2}(\emptyset) = 0 \quad (8).$$

When $X\neq\emptyset$, then $m_{1,2}(X) = (m_1 \oplus m_2)(X)$ (9),

and

$$(m_1 \oplus m_2)(X)=\frac{1}{1-K}\sum_{Y\cap Z=X} m_1(Y)m_2(Z) \quad (10),$$

where

$$K=\sum_{Y\cap Z=\emptyset} m_1(Y)m_2(Z), K \neq 1 \quad (11).$$

Equation (10) emphasizes the agreement between multiple sources of information and ignores conflicting evidence by using a normalization factor, which is equal to 1-*K*.

A Model for a Cyber Defence Situational Awareness System

A proposed model of a cyber defence situational awareness system is intended to provide situational awareness of campaigns with malicious information based on three existing systems for network behaviour awareness, attacks against the simulated system (services) with honeypot awareness, and malicious web content awareness.

Campaigns with Malicious Information in the Context of the CyRADARS Project

The definition of malicious information is dependent on the context. In the Council of Europe's Information Disorder Report,[22] authors propose a conceptual framework for examining information disorder, identifying three different types: mis-, dis- and mal-information. The difference between the three types of information is based on the intention of malicious information – just falseness or harm.

During the course of the CyRADARS project (i.e. as an activity of Task 2.1), an expert study was conducted. Results from this study about the meaning of malicious information,[11] show that for more than 52 % of the responders, malicious information means malicious software. For over 38 % of the experts, malicious information is equal to disinformation (i.e. false information is knowingly shared to cause harm [25] or fishing scam, or spam.

We distinguish the following campaign types of malicious information operations, which will be addressed by the developed cyber defence situational awareness system:

- Malicious information injections in web content;
- Malicious information injections in fake social network accounts;
- Malicious information dissemination via email messages – fishing scam or spam.

The Proposed Conceptual Model

The aim of situation awareness in our case is to display information and alarm the responsible staff about certain types of campaigns of malicious information operations, in order to support decision maker to define the countermeasures. To detect potential malicious information operations in near to real time, the system should implement data pre-processing, feature extraction, data fusion,

situation assessment, and situation visualization. Based on this and keep in mind the definition of SA in section 1.2 which imply activities at three levels, we proposed a layered conceptual model for an architecture of the CyRADARS cyber defence situational awareness (CCDSA) system.

Figure 1 presents the proposed conceptual model that the system for situational awareness of campaigns with malicious information operations, identified in section 2.1, will be based on. The aim is to integrate and fuse the information captured by three systems – System for discovering unusual network behaviour, System for primary analysis of information flows and anomalies, content monitoring and hidden patterns detection, and Honeypot systems, in order to provide a broader and more complete and reliable situational awareness. The CyRADARS partners will share the report on a discovered and confirmed type of campaign of malicious information activities, as well.

The suggested conceptual model consists of the following layers:

- Data layer;
- Integration layer;
- Interpretation layer;
- Presentation layer.

Data layer. The data layer is responsible for data collection from input interfaces, which are fed by the three source systems. Information for detected unusual network behavioural profiles, records from honeypot log file analysers, and analyser of malicious information injections in web content and/or fake social network accounts form a raw dataset. This layer also provides the metadata descriptions of the input data. For example, the metadata in case, when Honeypots are a source of data, could be HTTP headers, the content of various protocol-specific fields, md5 sums of downloaded files, etc.

Integration layer. Because of the huge amount of different information from the three different sources, the information needs to be normalized. This level is responsible for semantic integration of data from the sources for additional operations on them, which include features extraction and attributing. The integration layer is the layer that transforms and completes refinement of target data.

Interpretation layer. This layer is responsible for the analysis and assessment of data, provided by the Integration layer. The analysis could include correlation methods, data fusion, machine learning, etc. The correlation analysis should be conducted of characteristic parameters of the information collected by the three sources. The layer should define the primary suggestion of current state of situational awareness.

Presentation layer. Functions of the layer are related with representation of the information derived at the previous layer and provided a current state of situational awareness in order to support decision makers. This includes visualization graphics, alarming messages, etc. The analysis of information at this level will be supported by a cyber security expert. If the expert confirms the

recognized malicious activities by the previous level, the report will be generate and shared with other systems/organizations.

The CCDSA system will be developed based on the conceptual model, presented in figure 1. The CCDSA system will collect data, generated by the three system. Each of these systems is designed to monitor and detect traces and evidences of campaigns with malicious information operations. The three systems are:

- Honeypot systems;
- System for discovering of unusual network behaviour;
- System for discovering malicious information injections in web content and/or fake social network accounts.

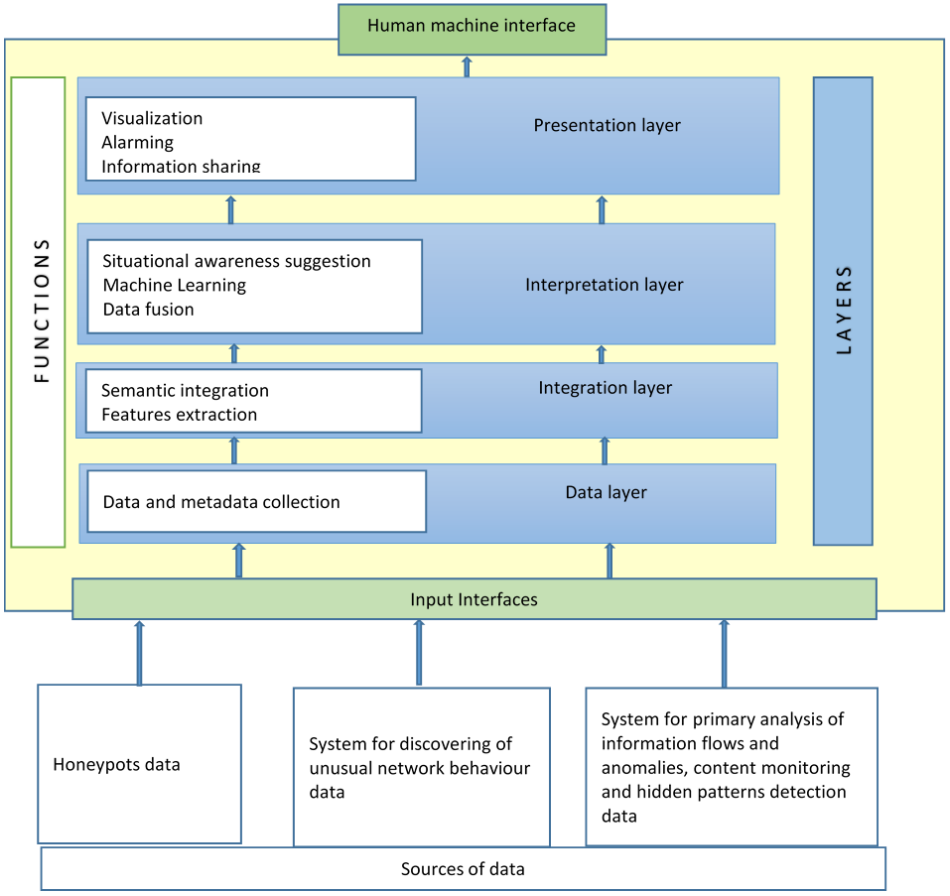

Figure 1: A conceptual model of a cyber defence situational awareness system of campaigns with malicious information.

Honeypot is a computer system or application developed to be intentionally vulnerable, so an attacker to detect and try to exploit the vulnerabilities. Honeypots could be used for the scanning activity of worms or bots, learning about compromised nodes, capturing new malware, studying hacker behaviour, looking for internal infections or attacks from insiders, etc.[21] For the goals of the CyRADARS project, we plan to use a low interaction open source honeypot systems such as Glastopf,[23] Dionaea,[24] or similar ones.

A software system for recognition fake//phony social network accounts and malicious content insertion in a web [5] is output from the CyRADARS project work.

A software tool that is able to identify unusual network behaviour [6, 7] is developed also as the project result.

A workflow process of the CCDSA system is presented with a flowchart diagram in Fig. 2.

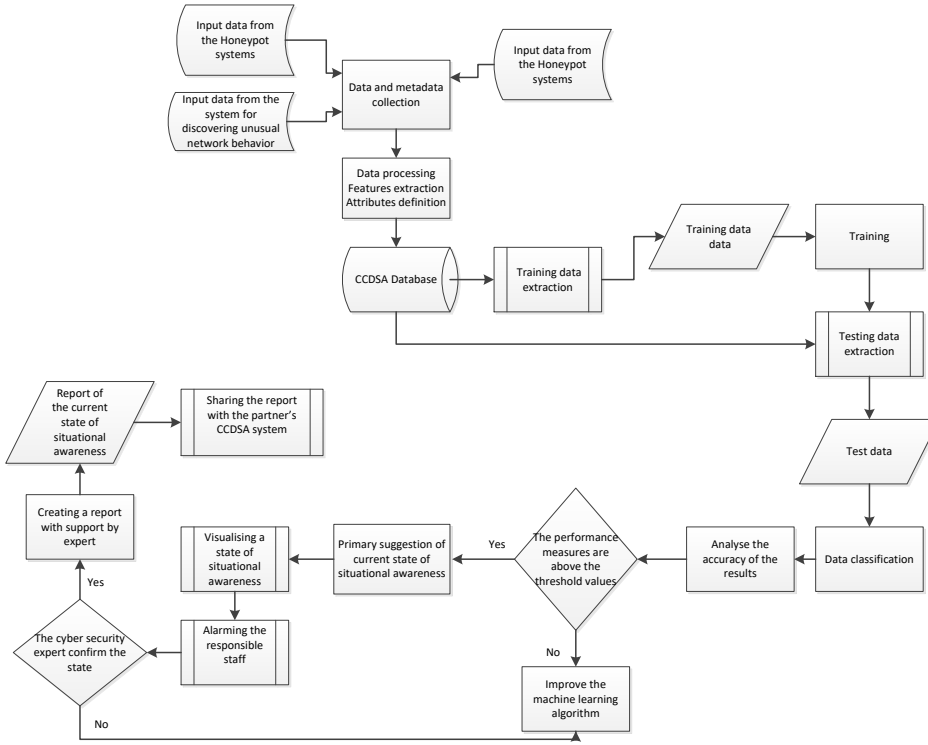

Figure 2: A flowchart of the CCDSA system.

The input data for the CCDSA system are data from the three source systems. These different types of data will be collected, analysed and attributed and the results should be integrated, and fused to present the current state of cyber defence situational awareness about the campaign types of malicious

information operations and support the countermeasure process. The decision about campaign types of malicious information operations will be supported by the cyber security expert. The output will be reports on discovered types of campaign of malicious information operations in the monitored network infrastructure. Each of the reports will include (at a minimum) the following information about suspicious event (some parts could be generated by the CCDSA system, others could be described with the support by a human operator):

- Type of malicious information operations;
- Source of the malicious operation (ex. a source IP address and/or an identifier of the account);
- Event time;
- Short description of the event (malicious information operations).

The report will be shared with partner's CCDSA systems through a specific interface.

The most important component of the system is the model for identification (recognition) of the type of campaign of malicious information operations. Section 2.3 presents an approach for recognition of a type of campaign of malicious information operations.

A Model for Identification of the Type of Campaign of Malicious Information Operations Based on Dempster-Shafer's Evidence Theory

The problem for recognitions of type of campaign of malicious information operations, could be considered as a classification problem.[8] Each campaign of malicious information operations needs to be recognized, so the implemented countermeasures and defence actions should be initiated. The following algorithm is proposed to solve the classification problem for the type of campaign of malicious information operations.

Step 1: Data collection

Collect the information from the three source systems for a predefined time period (for example one hour/day/week/month). The collected data will be a raw dataset.

Step 2: Feature extraction, attributes definition

The first step is data preparation, i.e. processing data to produce meaningful information. This means that influential features should be extracted and a dataset with meaningful attributes and their values should be built. The built dataset will include all the possible hypotheses of the Dempster-Shafer system, as a given campaign will be assigned only to one class (type).
The collected data will be processed automatically to extract meaningful features and produce the needed dataset.

Step 3 Split the dataset into two subsets – training and test datasets

The dataset should be divided into two datasets – training dataset and test dataset. This division is dependent on the dataset. For example, if the dataset is

imbalanced, it would be suitable to use a *k*-fold cross validation approach,[9] or in another case splitting, based on good practices (ex. 60%-80%/40%-20%).

Step 4. Derive mass values using the training dataset

A mass value should be assigned for each attribute in the following way:

A probability is used to assign mass values.[12] The probability is a value in the interval [0, 1] which defines the probability of a given type of campaign of malicious information operations conducted with an attribute *k*.

$$\text{Let } P_{ik} = \frac{N_{ik}}{N_k},$$

where

- P_{ik} is the probability of the campaign of type *i* of malicious information operations with an attribute *k.*
- N_{ik} is the number of the records for the campaigns of type *i* of malicious information operations with the attribute *k*.
- N_k is the total number of the records for the campaigns of malicious information operations with the attribute *k.*
- $i \in \Theta$={*type 1, type 2, type 3, other*}
- where,
  - *type 1* – Malicious information injections in web content;
  - *type 2* – Malicious information injections in a fake social network accounts;
  - *type 3* – Malicious information dissemination through fishing scam or spam.
- *k* is a set of values, which ∀ attribute can have.

The values of N_{ik} and N_k are calculated from the training dataset, so that the general probabilities (P_i, *i*=1, 2, 3) of the three types of campaigns of malicious information operations can be obtained for each attribute and used to define the mass functions.

Consequently,

$$m_{ik}(\Theta)=1, \text{ if } P_{ik}=P_i,$$

$$m_{ik}(i) = \frac{P_{ik}-P_i}{1-P_i} \text{ and } m_{ik}(\Theta)=1\text{-}m_{ik}(i), \text{ if } P_{ik}>P_i.$$

$$m_{ik}(\Theta\backslash\{i\}) = \frac{P_i-P_{ik}}{P_i} \text{ and } m(\Theta\backslash\{i\})=1\text{-}m(\Theta\backslash\{i\}), \text{ if } P_{ik}<P_i,$$

$$(m_{ik}(X) = 0, \text{ for all undefined subsets of } \Theta).$$

For each attribute of each type of the campaign, if the probability for type *i* with value *k* for this attribute *is* larger than the general probability for type *i*, then the type is more likely to be type *i* with value *k* for that attribute, *k* is a specific value of that attribute. Contrary, if the probability for type *i* with value *k* for this attribute is smaller than the general probability, then the campaign is not likely to be type *i* with the value *k* for that attribute, it is more likely to be

'not type *i*'. If the probability for type *i* with value *k* for this attribute is equal to the general probability, then no evaluation (deduction) is possible based on that information, i.e., the uncertainty is equal to 1.

The defined in that way mass function meets the conditions (1) – (3).

Step 5. Assign mass values for each attribute

Assigning the mass values for each attribute based on the defined in the previous step mass functions. The dataset is processed three times:

- For each record in the dataset, combine (see the formulas (8) – (11)) the mass values of all the attributes of type 1 of the campaign (Malicious information injections in web content), i.e $m_1 = \oplus_k m_{1k}$
- For each record in the dataset, combine (see the formulas (8) – (11)) the mass values of all the attributes of type 2 of the campaign (Malicious information injections in a fake social network accounts), i.e. $m_2 = \oplus_k m_{2k}$
- For each record in the dataset, combine (see the formulas (8) – (11)) the mass values of all the attributes of type 3 of the campaign (Malicious information dissemination through fishing scam or spam), i.e $m_3 = \oplus_k m_{3k}$.

Step 6. Combine the mass values for the three types of the campaign of malicious information operations

For each record in the dataset, combine (according to the formulas (8) – (11)) the mass values for the three types of the campaign of malicious information operations, i.e $m = m_1 \oplus m_2 \oplus m_3$.

Step 7. Detect the result

In this step, a decision rule should be applied in order to identify the type of campaign of malicious information operations.

Step 8. Analyse the recognition accuracy using performance measures

Depending on the dataset (its balance) different performance metrics should be used in parallel to analyse the accuracy of the recognition process. Below are listed some of the metrics that could be used to analyse the usability of the proposed model.

- Accuracy *A* is a relevant measure when the data set is balanced, when a dataset is imbalanced, accuracy could be a misleading indicator.

$$A = \frac{TP+TN}{TP+FP+FN+TN} \quad (12).$$

- Precision *P* represents the confidence of recognition of the type of the campaign.

$$P = \frac{TP}{TP+FP} \quad (13).$$

- Recall *R* is a measure that represents the true positive rate, i.e the ability of the model to recognize correctly the type of campaign.

$$R=\frac{TP}{TP+FN}. \quad (14).$$

- *F-measure* is defined as the harmonic mean of the Precision and Recall.

In case of an imbalanced dataset, it is proposed to use a *g*-mean metric:[12]

$$g\text{-}mean=\sqrt{\frac{TP}{TP+FN}\frac{TN}{TN+FP}} \quad (15).$$

The metric *g-mean* is the geometric mean of the true positive and true negative rates.

In all expressions (12) – (15), *TP*, *TN*, *FP*, and *FN* are the numbers of true positives, true negatives, false positives and false negatives recognitions (classifications) respectively.

A model for Identification of the Type of Campaign of Malicious Information Operations Based on Dempster-Shafer's Evidence Theory

The problem for recognitions of type of campaign of malicious information operations, could be considered as a classification problem.[8] Each campaign of malicious information operations needs to be recognized, so the implemented countermeasures and defence actions should be initiated. The following algorithm is proposed to solve the classification problem for the type of campaign of malicious information operations.

Step 1: Data collection

Collect the information from the three source systems for a predefined time period (for example one hour/day/week/month). The collected data will be a raw dataset.

Step 2: Feature extraction, attributes definition

The first step is data preparation, i.e processing data to produce meaningful information. This means that influential features should be extracted and a dataset with meaningful attributes and their values should be built. The built dataset will include all the possible hypotheses of the Dempster-Shafer system, as a given campaign will be assigned only to one class (type).

The collected data will be processed automatically to extract meaningful features and produce the needed dataset.

Step 3 Split the dataset into two subsets – training and test datasets

The dataset should be divided into two datasets – training dataset and test dataset. This division is dependent on the dataset. For example, if the dataset is imbalanced, it would be suitable to use a *k*-fold cross validation approach,[9] or in another case splitting, based on good practices (ex. 60%-80%/40%-20%).

Step 4. Derive mass values using the training dataset

A mass value should be assigned for each attribute in the following way:

A probability is used to assign mass values.[12] The probability is a value in the interval [0, 1] which defines the probability of a given type of campaign of malicious information operations conducted with an attribute *k*.

$$\text{Let } P_{ik} = \frac{N_{ik}}{N_k},$$

where

- P_{ik} is the probability of the campaign of type *i* of malicious information operations with an attribute *k.*
- N_{ik} is the number of the records for the campaigns of type *i* of malicious information operations with the attribute *k*.
- N_k is the total number of the records for the campaigns of malicious information operations with the attribute *k.*
- $i \in \Theta$={*type 1, type 2, type 3, other*}
- where,
  - *type 1* – Malicious information injections in web content;
  - *type 2* – Malicious information injections in a fake social network accounts;
  - *type 3* – Malicious information dissemination through fishing scam or spam.
- *k* is a set of values, which ∀ attribute can have.

The values of N_{ik} and N_k are calculated from the training dataset, so that the general probabilities (P_i, *i*=1, 2, 3) of the three types of campaigns of malicious information operations can be obtained for each attribute and used to define the mass functions.

Consequently,

$$m_{ik}(\Theta)=1, \text{ if } P_{ik}=P_i,$$

$$m_{ik}(i) = \frac{P_{ik}-P_i}{1-P_i} \text{ and } m_{ik}(\Theta)=1-m_{ik}(i), \text{ if } P_{ik}>P_i.$$

$$m_{ik}(\Theta\backslash\{i\}) = \frac{P_i-P_{ik}}{P_i} \text{ and } m(\Theta\backslash\{i\})=1-m(\Theta\backslash\{i\}), \text{ if } P_{ik}<P_i,$$

$$(m_{ik}(X) = 0, \text{ for all undefined subsets of } \Theta).$$

For each attribute of each type of the campaign, if the probability for type *i* with value *k* for this attribute *is* larger than the general probability for type *i*, then the type is more likely to be type *i* with value *k* for that attribute, *k* is a specific value of that attribute. Contrary, if the probability for type *i* with value *k* for this attribute is smaller than the general probability, then the campaign is not likely to be type *i* with the value *k* for that attribute, it is more likely to be 'not type *i*'. If the probability for type *i* with value *k* for this attribute is equal to the general probability, then no evaluation (deduction) is possible based on that information, i.e., the uncertainty is equal to 1.

The defined in that way mass function meets the conditions (1) – (3).

Step 5. Assign mass values for each attribute

Assigning the mass values for each attribute based on the defined in the previous step mass functions. The dataset is processed three times:

- For each record in the dataset, combine (see the formulas (8) – (11)) the mass values of all the attributes of type 1 of the campaign (Malicious information injections in web content), i.e. $m_1 = \oplus_k m_{1k}$
- For each record in the dataset, combine (see the formulas (8) – (11)) the mass values of all the attributes of type 2 of the campaign (Malicious information injections in a fake social network accounts), i.e. $m_2 = \oplus_k m_{2k}$
- For each record in the dataset, combine (see the formulas (8) – (11)) the mass values of all the attributes of type 3 of the campaign (Malicious information dissemination through fishing scam or spam), i.e. $m_3 = \oplus_k m_{3k}$.

Step 6. Combine the mass values for the three types of the campaign of malicious information operations

For each record in the dataset, combine (according to the formulas (8) – (11)) the mass values for the three types of the campaign of malicious information operations, i.e. $m = m_1 \oplus m_2 \oplus m_3$.

Step 7. Detect the result

In this step, a decision rule should be applied in order to identify the type of campaign of malicious information operations.

Step 8. Analyse the recognition accuracy using performance measures

Depending on the dataset (its balance) different performance metrics should be used in parallel to analyse the accuracy of the recognition process. Below are listed some of the metrics that could be used to analyze the usability of the proposed model.

- Accuracy *A* is a relevant measure when the data set is balanced, when a dataset is imbalanced, accuracy could be a misleading indicator.

$$A = \frac{TP+TN}{TP+FP+FN+TN} \quad (12).$$

- Precision *P* represents the confidence of recognition of the type of the campaign.

$$P = \frac{TP}{TP+FP} \quad (13).$$

- Recall *R* is a measure that represents the true positive rate, i.e the ability of the model to recognize correctly the type of campaign.

$$R = \frac{TP}{TP+FN}. \quad (14).$$

- *F-measure* is defined as the harmonic mean of the Precision and Recall.

In case of an imbalanced dataset, it is proposed to use a *g*-mean metric [12]:

$$g\text{-}mean = \sqrt{\frac{TP}{TP+FN}\frac{TN}{TN+FP}} \quad (15).$$

The metric *g-mean* is the geometric mean of the true positive and true negative rates.

In all expressions (12) – (15), *TP*, *TN*, *FP*, and *FN* are the numbers of true positives, true negatives, false positives and false negatives recognitions (classifications) respectively.

Conclusion and Next Steps

The proposed Cyber Defence Situational Awareness system model for cyber defence awareness system of campaigns with malicious information is currently implemented for validation in a software system. The three existing systems, now are working independently for providing the raw dataset and associated metadata. The next step is practical verification and validation of the proposed model for recognition of the campaign type of malicious information operations on the collected raw data. After successful validation the three source systems will be integrated into the CCDSA system.

The developed system will be deployed at three sites – Chernihiv National University of Technology (Ukraine), "Igor Sikorsky" Kyiv Polytechnic Institute (National Technical University of Ukraine), and Bulgarian Defence Institute. The system at one site will work independently from systems at other sites but will share the information via the special interface.

Acknowledgements

This work has been funded by the NATO Science for Peace and Security Programme under award no. G5286, the Cyber Rapid Analysis for Defense Awareness of Real-time Situation (CyRADARS) project.

References

1 Alexander Kott, Cliff Wang, Robert F. Erbacher (eds.), *Cyber Defense and Situational Awareness*, 1st edition (Switzerland: Springer International Publishing, 2014).

2 Ulrik Franke and Joel Brynielsson, "Cyber situational awareness - A systematic review of the literature," *Computers & Security* 46 (2014): 18-31.

3 Mica Endsley, "Design and Evaluation for Situation Awareness Enhancement," *Proceedings of the Human Factors Society Annual Meeting* 32, no. 2 (Oct. 1988): 97–101.

4 Mica Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37, no. 1 (1995): 32-64.

5 Oleg Chertov, Taras Rudnyk, and Oleksandr Palchenko, "Search of phony accounts on Facebook. Ukrainian case," *International Conference on Military Communications and Information Systems (ICMCIS2018), Warsaw, Poland, 22-23 May 2018, IEEE Xplore,* https://doi.org/10.1109/ICMCIS.2018.8398725.

6 V. V Let and Ihor S. Skitter, "Determination of non-standard behavior of the network by methods of statistical analysis," *Thirteenth International Scientific and Practical Conference "Mathematical and Simulation Modeling of Systems MODS 2018", Chernihiv-Kiev-Zhukin, June 25-29,* 2018, pp. 321-326.

7 Ihor S. Skiter, "Methods of forming an image of a normal computer behavior Networks," *III International Conference "Problems of Decommissioning of Objects*

Nuclear Energy and the Restoration of the Environment," INUDECO, Slavutych, 2018, pp. 96-99.

[8] Alex Smola and S.V.N. Vishwanathan, *Introduction to Machine Learning* (Cambridge University Press, 2008).

[9] Peter Wlodarczak, *Machine Learning and its Applications* (CRC Press, 2019).

[10] Xavier Gros, *NDT Data Fusion* (Great Britain: Arnold, 1997).

[11] Nikolai Stoianov and Maya Bozhilova, "Expert's Study on Situational Awareness of Operations Directed at the Wide Dissemination of Malicious Information," *Conference Proceedings, MODS2020, Chernihiv, Ukraine, 2020* (in print).

[12] Qi Chen, Amanda Whitbrook, Uwe Aickelin, and Chris Roadknight, "Data classification using the Dempster–Shafer method," *Journal of Experimental & Theoretical Artificial Intelligence* 26, no. 4, (2014): 493-517.

[13] Arthur P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Ann. Math. Statist.* 38 (1967): 325-339.

[14] Jean Gordon and Edward H. Shortliffe, *The Dempster-Shafer theory of evidence. Readings in uncertain reasoning* (San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1990), 529–539.

[15] Glenn Shafer, *A Mathematical Theory of Evidence* (Princeton University Press, 1976).

[16] Weifa Zheng, "Research on situation awareness of network security assessment based on Dempster-Shafer," *2019 International Conference on Computer Science Communication and Network Security (CSCNS2019), MATEC Web of Conferences,* Vol. 309 (2020).

[17] Yan Zhang, Shuguang Huang, Shize Guo, and Junmao Zhu, "Multi-sensor Data Fusion for Cyber Security Situation Awareness," *2011 3rd International Conference on Environmental, Science and Information Application Technology (ESIAT*) *2011,* https://www.sciencedirect.com/science/article/pii/S1878029611003604, last accessed May 30, 2020.

[18] Gang Chen, Jun Ping Cai, and Jun Yang, "Network Security Situation Awareness Based on Multi-source Data Fusion," *Advanced Materials Research* 989-994 (2014): 4885-4888.

[19] The CyRADARS web site, https://www.cyradars.net/ (last accessed 2020/05/20).

[20] NATO Science for Peace and Security Programme Homepage, https://www.nato.int/cps/en/natolive/78209.htm, last accessed May 20, 2020.

[21] "Proactive detection of security incidents II – Honeypots," ENISA, 2012, https://www.enisa.europa.eu/publications/proactive-detection-of-security-incidents-II-honeypots, last accessed May 25 2020.

[22] Claire Wardle and Hossein Derakhshan, "Council of Europe's Information Disorder Report: Toward an interdisciplinary framework for research and policymaking," 2nd revised edition, August 2018, https://rm.coe.int/information-disorder-report-version-august-2018/16808c9c77, last accessed May 20, 2020.

[23] Glastopf, https://github.com/mushorg/glastopf, last accessed May 30, 2020.

[24] Dionaea, https://dionaea.readthedocs.io/en/latest/, last accessed May 31, 2020.

[25] EUvsDisinfo, https://euvsdisinfo.eu, last accessed June 6, 2020.