A Survey of Network Security Situational Awareness Technology

Chen Chen[1(✉)], Lin Ye[1], Xiangzhan Yu[1,2], and Bailang Ding[2]

[1] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
chenchenhit@foxmail.com, {hityelin,yxz}@hit.edu.cn
[2] Institute of Electronic and Information Engineering of UESTC in Guangdong, Dongguan, China
l54024012@qq.com

Abstract. With the increasing importance of cyberspace security, the research and application of network situational awareness is getting more attention. The research on network security situational awareness is of great significance for improving the network monitoring ability, emergency response capability and predicting the development trend of network security. This paper describes the development and evolution of network situational awareness and analyzes the basic architecture of the current situational awareness system. Based on the situational awareness conceptual model, four main research contents of situational awareness are elaborated: network data collection, situational understanding, situational prediction and situational visualization. This paper focuses on the core issues, main algorithms, and the advantages and disadvantages of each method that need to be addressed at each research point. Finally, under the current development trend of big data processing technology and artificial intelligence technology, the application realization and development trend of network situational awareness are analyzed and forecasted.

Keywords: Situational awareness · Network security · Situational visualization

1 Introduction

In recent years, with the development of Internet technology, the attack methods have become more diversified, and the number of security vulnerabilities and security incidents has increased significantly. The research direction of network security has changed, from the research of single security issues to the overall situation of global networks. Network situational awareness is considered to be a new way to solve some current network security problems. It combines the detection of security events by all network sensors to provide real-time visibility into network security conditions and risks. It has become a hot research field at the forefront of the world.

There are two popular definitions of situational awareness. One was the concept of Situation awareness (SA) first proposed by Endsley in 1988 [1]. He defined the cognitive definition from artificial intelligence: situational awareness is the recognition of a

X. Sun et al. (Eds.): ICAIS 2019, LNCS 11635, pp. 101–109, 2019.
https://doi.org/10.1007/978-3-030-24268-8_10

large number of environmental elements in time and space, understanding their meanings, and predicting their status in the near future. Endsley's point of view is mainly on cognitive principles, mainly a top-down driven mental model, divided into three main parts: perception, understanding and projection. Another concept stems from the definition of the Joint Directors of Laboratories (JDL) data fusion model, which provides a more bottom-up, data-centric approach that defines situational awareness as an estimate and prediction of relationships between entities.

Although the network situation is used in different fields, the research on situational awareness in this paper is all about network security. After Endsley's groundbreaking work, research in this area has continued to deepen worldwide. In 1999, Bass believed [2] that in order to create network situational awareness, next-generation cyberspace intrusion detection systems would incorporate data from heterogeneous distributed network sensors. Since then, intrusion detection systems have been combined with situational awareness. And based on the JDL model of data fusion, he proposed a network situational awareness function model based on multi-sensor data fusion. The data fusion model has become a representative study at this stage. In 2000, McGuinness and Foy et al. [3] extended the fourth layer of the situational awareness model called Resolution. Resolution represents the countermeasures needed to deal with the interdependent risks in the network. In 2006, Tadda et al. [4] re-integrated a three-level model consisting of factor extraction, state perception and situational prediction, and proposed evaluation indicators and methods for situational awareness systems.

Therefore, this paper synthesizes the current research on the network security situation in the industry, and gives the following definition: Network security situational awareness refers to the analysis and visualization of various security elements in large-scale networks, and predicts the development trend, and finally assists the follow-up Decision making. The network security situational awareness system analyzes and predicts the current state and development trend of the network by collecting security data such as network traffic, security logs, security alarms, and threat intelligence, and using data analysis and machine learning techniques.

2 Network Security Situational Awareness

According to the different functions of the network situational awareness system, this paper summarizes the research content into four aspects:

- Network element collection
- Situational understanding
- Situation prediction
- Situation visualization.

The related technologies and research contents will be elaborated in four aspects.

2.1 Network Element Collection

Accurate and comprehensive extraction of security situation elements in the network is the basis of network security situational awareness research. Since the network has

developed into a large nonlinear complex system with strong flexibility, it is very difficult to extract the network security situation elements. At present, the security posture elements of the network mainly include static configuration information, dynamic operation information, and network traffic information.

Franke et al. [5] believe that data acquired in network sensors (such as intrusion detection systems) can go directly into the data fusion process or be interpreted by decision makers. But it needs to be combined with other information, such as adding human understanding of security incidents. This combination helps improve overall network situational awareness. In addition, Jajodia et al. [6] assessed the vulnerability of the network by collecting vulnerability information from the network. Wang et al. [7] proposed an anti-attack concept by constructing an attack graph as a measure of security for different network configurations to give an indication of the current operation of the network.

On the other hand, data processing frameworks for large amounts of data are constantly evolving. Hadoop, based on MapReduce [8] technology, makes it possible to process terabytes of data. Spark [9], which focuses on memory computing, combines components such as streaming, machine learning, and graph computing to build a data calculation framework that provides a highly efficient and highly available data processing platform. In addition, the development of components such as Flink and Strom solves many of the difficulties in real-time processing, and people can better extract value from large amounts of data. At the same time, various data fusion algorithms can also be implemented in the big data framework. Security for large amounts of traffic is also evolving [27, 28].

Based on the current research situation, most methods only obtain data from a single aspect, cannot comprehensively consider information, and cannot dig into the internal relationship between data, which poses difficulties for later analysis. However, with the development of data mining technology, rapid processing of massive data becomes possible. Therefore, big data acquisition with intrinsic relevance is a trend of development in the future.

2.2 Situational Understanding

The understanding of network security situation refers to the integration of relevant data to obtain a macro network security situation. Data fusion is at the heart of the understanding of network security posture. According to Haines et al. [11]: Previous results have shown that no single control (such as IDS) can detect all network attacks. The network security situation assessment no longer studies a single event, but studies the overall security status of the network from a macro perspective. At present, data fusion algorithms are divided into the following categories: analytic hierarchy process, logical reasoning, probability analysis, and rule pattern matching.

Analytic Hierarchy Process. Bass first proposed a data fusion method for situational awareness [3]: Using the art and science of multi-sensor data fusion as a design framework, it can identify, track, classify and evaluate network-centric activities in complex infrastructure. The specific implementation method is to layer the data alarms according to the threat level from low to high, and the data at the same level is merged.

The analytic hierarchy process is to comprehensively consider various situational factors affecting the situation and establish several evaluation functions. The most representative rating function is the weighted average. The weighted average method is the most common and simple method of fusion based on mathematical models. Chen et al. [10] proposed a hierarchical cybersecurity threat situation quantitative assessment method. The implementation method is to weight the importance factor of the service and the host itself, calculate the threat index of the computing service, the host, and the entire network system, and then analyze the security posture of the network.

The analytic hierarchy process and the weighted averaging method can intuitively integrate various situational factors, and the implementation process is relatively simple. However, the main problem of this method is: There is no uniform standard for the choice of weights and the basis for stratification, most of which are based on domain knowledge or experience, and lack of objective basis.

Logical Reasoning. Logical reasoning mines the inherent logic between information and integrates information. The logical relationship between alarms is divided into: the similarity of alarm attributes, the relevance in the attack model, the correlation between the premise of the attack and the subsequent conditions. Ning et al. [11] analyzed the threatening situation of the network from the mass alarm information through alarm correlation. Morin et al. [12] used a representation language to normalize each network node based on a topology map of the network nodes, representing each event in a structured manner. This model provides the correlation logic between alerts to describe security events.

A typical algorithm for logical reasoning is fuzzy logic. Fuzzy logic is a mathematical method for people to reason about uncertain things, using fuzzy sets and fuzzy rules. In the network situation assessment, firstly, the single source data is locally evaluated, then the corresponding model parameters are selected, and the membership function is established for the local evaluation result, which is divided into corresponding fuzzy sets to realize the fuzzification of specific values, and the results are carried out. Quantify. After quantification, if a state attribute value exceeds a predetermined threshold, the local evaluation result is used as input for causal reasoning. Finally, the situational classification is identified by fuzzy rule reasoning, thus completing the assessment of the current situation.

The biggest advantage of logical reasoning is that it is easy to understand, and it can reflect the network security situation very intuitively. However, the limitation of this method is that it requires a detailed analysis of the type of attack, and this analysis is very difficult. And some unknown alarms can't make some judgments.

Probabilistic Analysis. The probability and statistics method makes full use of the statistical characteristics of prior knowledge and combines the uncertainty of information to establish a model of situation assessment. Bayesian networks and hidden Markov models are the most common methods of probability and statistics.

Bayesian network was defined by Pearl [1] in 1988 and became a research hotspot in the field of knowledge representation and reasoning for more than 20 years. In the network situation assessment, the Bayesian network uses a directed acyclic graph representation, nodes represent different situations and events, each node contains a conditional probability allocation table, and nodes use edges to connect, indicating the

situation and events. Interdependence, after some nodes obtain evidence information, the Bayesian network spreads and fuses the information between nodes to obtain new situation information. Chen et al. [15] proposed a new method using Bayesian inference tools. The specific implementation is to use the layered model to re-define the decision fusion problem, and propose a Gibbs sampler to perform the fusion based on the posterior probability. Park et al. [16] added time factors to the original network. They propose a multi-instance Bayesian network that can be analyzed over time and incorporate high-level languages to handle complex situations and uncertainties.

The hidden Markov model is equivalent to a dynamic Bayesian network. Damarla et al. [17] proposed a situational awareness framework based on hidden Markov models. HMM is equivalent to a dynamic Bayesian network. In the network situation assessment, the transfer process of the network security state is defined as the implicit state, and the security events at different time points are defined as the sequence of observation values, and the HMM model is trained using the sequence of observation values and the implicit state. Finally, the model is used to evaluate the situation.

Probabilistic analysis based methods can fuse all data and prior knowledge, and the reasoning process is clear and easy to understand. However, this model requires a large amount of data and takes a long time in the training process. This method has some difficulties in feature extraction and model construction.

Rule Pattern Matching. The rule pattern method is to build an evaluation model based on expert knowledge and experience, and analyze the security posture of the entire network through pattern matching. At present, the D-S evidence combination method and the gray correlation algorithm are the research hotspots.

The D-S evidence fusion method was first proposed by DEMPSTER [18] and then refined by SHAFER [19]. The D-S evidence fusion method is a measure of the support for each possible decision of the single source data, that is, the degree of support for the decision using the data information as evidence. Then look for a synthetic rule of evidence. By repeatedly applying the synthesis rules, the algorithm finally achieves the degree of support for a certain decision and completes the process of data fusion. Sabata et al. [20] proposed a multi-source evidence fusion method to complete the fusion of distributed real-time attack events and realize the perception of network situation. Zhang et al. [21] improved the D-S theory by introducing Bhattacharyya distance, evidence confidence and modified combination rules, and effectively solved the conflict of evidence in DS evidence theory. The grey system theory was first proposed by Deng [22] and is a theoretical method for dealing with uncertain information. The basic idea of gray correlation analysis is to judge whether the connection is tight according to the similarity of the geometry of the sequence curve. The closer the curve is, the greater the degree of association between the corresponding sequences, and vice versa. Hu et al. [23] proposed an improved adaptive grayscale model to analyze the situation that the network security situation is “S” curve.

Algorithms based on rule patterns generally need to mine the intrinsic patterns between data, which can be adapted to uncertain situations without prior information. However, this method may have problems with the explosion of the number of associated patterns and the conflict of evidence, so it will have a great impact on the results.

2.3 Situation Prediction

The prediction of the network security situation refers to predicting the development trend of the network in the future according to the historical information and current state of the network security situation. Due to the randomness and uncertainty of cyber-attacks, predicting the change of security situation is a complex nonlinear process, thus limiting the use of traditional prediction models. At present, network security situation prediction generally adopts methods such as neural network and time series prediction.

Neural network is a commonly used network situation prediction method. The algorithm uses multiple associated neurons as the structure of the model, linking the inputs to the output. The algorithm adjusts the parameters by means of gradient descent and other methods to construct a prediction model. Ying et al. [24] improved the BP neural network by using wavelet neural network (WNN) to predict the network situation. The neural network model has many parameters, strong adaptability, and good nonlinearity fitting, so it has strong robustness. However, since the effect of the model depends on the quality of the feature engineering, and the model requires a large amount of computational power for training, it cannot be used in some environments.

Time series prediction method reveals the law of network situation change with time through time series, and predicts the future situation according to this law. The prediction process uses the top N values of the sequence to predict the next M values. Commonly used for time series are HMM algorithm, autoregressive moving average method and so on.

The current situation prediction method mainly uses machine learning, which has good convergence and fault tolerance, and can handle large-scale data. However, in the real network environment, there is often a mutual game between the attacker and the defender. Therefore, the hybrid model based on game theory is the future development direction.

2.4 Situation Visualization

Visualization is an important part of situational awareness. The data collected from various network sensors is abstract and fragmentary, and people cannot understand the analysis very quickly. Therefore, we need to perform aggregate analysis on multi-source data to extract high-level situational results and display them using some visualization techniques. Many visualization systems are data driven. Host and server monitoring is one of the manifestations for a single data source. Users can get information about system load, network link status, abnormal traffic, and so on. These are all visual methods for basic information. When multiple servers are interconnected to form a large-scale cluster, it is necessary to display the host connection diagram, monitor the traffic status between the hosts, and monitor the traffic between the internal cluster and the external network. However, these methods simply count the basic data, and the invisible threat situation and attack mode cannot form an intuitive display.

Current research focuses on hierarchical visualization, attack visualization, and interactive visualization. Chen et al. [10] conducted hierarchical analysis of multiple data sources and presented them according to different categories. Beaver et al. [25] visualize the attack by screening the information generated by the IDS and collecting the results of

multi-source cleaning. Phan et al. [26] proposed a time-centric visualization system that classifies events by interacting with humans and iteratively produces visual charts.

With the advancement of computer graphics and visual technology, many modeling methods continue to evolve. Visualization methods are fully applied at every stage of network situational awareness. Future interactive situational visualization technology will become a trend.

3 Outlook

Through the narrative of the four aspects of the network situational awareness field, we discuss the advantages and disadvantages of various algorithms. At present, the research on network situational awareness is still in the development stage, so many problems need to be further solved. This paper believes that there are several aspects to the future research direction.

Fusion of Massive Data. At present, the data sources on which situational awareness depends are increasing, making the integration method more and more difficult. At the same time, the network situation understanding requires high-quality data alarm as a data source. Due to the detection level of IDS and various firewall systems, the alarm quality is not very high. How to extract high-quality data sources and carry out efficient and rapid integration is the future development trend. At present, deep learning has received extensive attention, and its performance and accuracy have reached a very good level. Therefore, data fusion technology based on deep learning will develop rapidly.

Situational Understanding of Incomplete Warning. The robustness of the network situational awareness system will be tested when there is an error, omission, or new type of alarm in the device that generated the data. Therefore, how to face the unknown missing data is also a problem that needs to be solved. At present, some methods based on logical reasoning, probability and so on require prior knowledge, which requires experts to analyze the entire network and summarize the relevant laws. Methods that do not require prior knowledge are often severely affected by data quality issues, and therefore require more efficient algorithms.

Visualization of Situation. Situational visualization is the most intuitive way to understand the situation, and the results of data fusion and situational understanding need to be represented using appropriate representations. At present, many papers tend to study the content of fusion and prediction, ignoring the importance of visualization in situational awareness. Therefore, with the development of visualization technology, situational awareness will become more direct and concise.

Situation Prediction in a Complex Environment. The randomness and uncertainty of network attacks determine that the change of security situation is a complex non-linear process. At present, the methods of situation prediction are mainly probabilistic models and machine learning models, which have good effects on regular time series. But it does not reflect the complex trend changes very well. Therefore, the mathematical model based on causality needs further research.

4 Conclusion

As the scale of the Internet continues to expand, the number of cyber threats continues to increase. How to comprehensively and accurately detect network situation is a problem that needs to be solved. This paper introduces the definition and development status of current network situational awareness, and describes four aspects of situational awareness: network security situation factor collection, situation understanding, situation prediction and situation visualization. The current algorithms and related advantages and disadvantages of these four aspects are analyzed. With the development of deep learning and big data technology, people have made significant progress in dealing with multi-source data, and there are still some problems to be solved. Finally, this paper gives the problems that the situational awareness needs to solve in the future and the direction of development. I hope that through the development of related algorithms, it will bring more benefits to human beings.

Acknowledgement. This work was supported by National Key Research & Development Plan of China under Grant 2016QY05X1000, National Natural Science Foundation of China under Grant No. 61872111, and Dongguan Innovative Research Team Program under Grant No. 201636000100038.

References

1. Endsley, M.R.: Design and evaluation for situation awareness enhancement. In: Proceedings of the Human Factors Society Annual Meeting, vol. 32, no. 2, pp. 97–101. SAGE Publications, Los Angeles (1988)
2. Bass, T.: Multisensor data fusion for next generation distributed intrusion detection systems (1999)
3. McGuinness, B., Foy, L.: A subjective measure of SA: the crew awareness rating scale (CARS). In: Proceedings of the First Human Performance, Situation Awareness, and Automation Conference, Savannah, Georgia, vol. 16 (2000)
4. Tadda, G., Salerno, J.J., Boulware, D., et al.: Realizing situation awareness within a cyber environment. In: Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2006, vol. 6242, p. 624204. International Society for Optics and Photonics (2006)
5. Franke, U., Brynielsson, J.: Cyber situational awareness – a systematic review of the literature. Comput. Secur. **46**, 18–31 (2014)
6. Jajodia, S., Noel, S., O'Berry, B.: Topological analysis of network attack vulnerability. In: Kumar, V., Srivastava, J., Lazarevic, A. (eds.) Managing Cyber Threats, pp. 247–266. Springer, Boston (2005). https://doi.org/10.1007/0-387-24230-9_9
7. Wang, L., Singhal, A., Jajodia, S.: Measuring the overall security of network configurations using attack graphs. In: Barker, S., Ahn, G.-J. (eds.) DBSec 2007. LNCS, vol. 4602, pp. 98–112. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-73538-0_9
8. Bu, Y., Howe, B., Balazinska, M., et al.: HaLoop: efficient iterative data processing on large clusters. Proc. VLDB Endowment **3**(1–2), 285–296 (2010)
9. Zaharia, M., Xin, R.S., Wendell, P., et al.: Apache spark: a unified engine for big data processing. Commun. ACM **59**(11), 56–65 (2016)

10. Chen, X.Z., Zheng, Q.H., Guan, X.H., et al.: Quantitative hierarchical threat evaluation model for network security. J. Softw. **17**(4), 885–897 (2006)
11. Ning, P., Cui, Y., Reeves, D.S., et al.: Techniques and tools for analyzing intrusion alerts. ACM Trans. Inf. Syst. Secur. (TISSEC) **7**(2), 274–318 (2004)
12. Morin, B., Mé, L., Debar, H., et al.: A logic-based model to support alert correlation in intrusion detection. Inf. Fusion **10**(4), 285–299 (2009)
13. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Elsevier, Amsterdam (2014)
14. Mahoney, S.M., Laskey, K.B.: Constructing situation specific belief networks. In: Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, pp. 370–378. Morgan Kaufmann Publishers Inc., (1998)
15. Chen, B., Varshney, P.K.: A Bayesian sampling approach to decision fusion using hierarchical models. IEEE Trans. Sig. Process. **50**(8), 1809–1818 (2002)
16. Park, C.Y., Laskey, K.B., Costa, P.C.G., et al.: Predictive situation awareness reference model using multi-entity bayesian networks. In: 2014 17th International Conference on Information Fusion (FUSION), pp. 1–8. IEEE (2014)
17. Damarla, T.: Hidden markov model as a framework for situational awareness. In: 2008 11th International Conference on Information Fusion, pp. 1–7. IEEE (2008)
18. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. In: Yager, R.R., Liu, L. (eds.) Classic Works of the Dempster-Shafer Theory of Belief Functions, vol. 219, pp. 57–72. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-44792-4_3
19. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
20. Sabata, B., Ornes, C.: Multisource evidence fusion for cyber-situation assessment. In: Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2006, vol. 6242, p. 624201. International Society for Optics and Photonics (2006)
21. Zhang, W., Ji, X., Yang, Y., et al.: Data fusion method based on improved DS evidence theory. In: 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 760–766. IEEE (2018)
22. Deng, J.L.: Properties of relational space for grey system. Grey Syst. (1988)
23. Hu, W., Li, J., Chen, X., et al.: Network security situation prediction based on improved adaptive grey Verhulst model. J. Shanghai Jiaotong Univ. (Sci.) **15**(4), 408–413 (2010)
24. Jibao, L., Huiqiang, W., Xiaowu, L., et al.: A quantitative prediction method of network security situation based on wavelet neural network. In: ISDPE, pp. 197–202. IEEE (2007)
25. Beaver, J.M., Steed, C.A., Patton, R.M., et al.: Visualization techniques for computer network defense. In: Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense X, vol. 8019, p. 801906. International Society for Optics and Photonics (2011)
26. Phan, D., Gerth, J., Lee, M., Paepcke, A., Winograd, T.: Visual analysis of network flow data with timelines and event plots. In: Goodall, J.R., Conti, G., Ma, K.L. (eds.) VizSEC 2007, pp. 85–99. Springer, Heidelberg (2008)
27. Cheng, J., Ruomeng, X., Tang, X., Sheng, V.S., Cai, C.: An Abnormal Network Flow Feature Sequence Prediction Approach for DDoS Attacks Detection in Big Data Environment, CMC: Computers. Materials & Continua **55**(1), 095–119 (2018)
28. Xiaonian, W., Zhang, C., Zhang, R., Wang, Y., Cui, J.: A distributed intrusion detection model via nondestructive partitioning and balanced allocation for big data, CMC: computers. Mater. Continua **56**(1), 61–72 (2018)