International Journal of Innovative
Computing, Information and Control
Volume **17**, Number **5**, October **2021**

ICIC International ©2021 ISSN 1349-4198
pp. 1717–1730

NETWORK ATTACK PATH PREDICTION BASED ON VULNERABILITY DATA AND KNOWLEDGE GRAPH

Yifan Wang*, Zhi Sun and Ye Han

Advanced Technology Research Institute
The 30th Research Institute of China Electronic Technology Group Corporation
No. 8, Chuangye Road, Chengdu 610093, P. R. China
*Corresponding author: kingnop@mail.ustc.edu.cn
sunzhi@alu.uestc.edu.cn; hany12202@cetcsc.com

Received March 2021; revised July 2021

Abstract. *With the increase in network security problems, accurately predicting the possible attack path of an attacker and fixing bugs has become a concern of network security administrators. To overcome the shortcomings of existing methods that mainly focus on the path prediction under the ideal attack scenario and ignore the key position of nodes in the network, a network attack path prediction method based on knowledge graph and attack graph model is proposed, which takes the knowledge graph as the core, uses CVSS's quantitative indicators for a single vulnerability, and combines the network security evaluation method to calculate the possible path. Experimental results show that this method can evaluate the security risk value of networks and nodes, which can point out the possible attack path of the attacker and calculate the risk value of the corresponding path. It can also rank the network node on the path and give repair suggestions. Hence, this method can provide a basis for the implementation of security protection strategies.*
Keywords: Network security, Knowledge graph, Attack graph, Attack path prediction, CVSS

1. **Introduction.** With the current rapid development of the Internet and the information technology industry, computer network systems have penetrated all aspects of social life, and knowledge economy has been developed rapidly [1]. While bringing convenience to the public, the various security risks it faces have gradually increased. In recent years, the frequent occurrence of various security incidents and security issues has also reminded the importance of maintaining network system security. According to the 2019 China Internet Cyber Security Report [2] issued by the CNCERT, DDoS (Distributed Denial of Service) and APT (Advanced Persistent Thread) attacks have become the mainstream of cyber-attacks; the number of vulnerabilities and the scope of influence have increased significantly, and the number of newly included vulnerabilities in CNVD (China National Vulnerability Database) has increased by 14%, ranging from software systems such as operating system, OA (Office Automation) system to hardware, VPN (Virtual Private Network) and routers; the emerge in large numbers of new technologies such as 5G, IPv6 and blockchain also makes network security face new challenges.

When an attacker invades a network system outside the network, he usually needs to use vulnerabilities and other means to attack a node (such as boundary router or DMZ (Demilitarized Zone) area) on the boundary of the network system, and then use the node as a springboard to further invade the node that stores important data in the network system. In this process, attackers need to continuously exploit vulnerabilities or system vulnerabilities, such as weak passwords or guest logins, to obtain node permissions, and

DOI: 10.24507/ijicic.17.05.1717

attack layer by layer to form a clear attack link. Therefore, a method that can accurately assess the possible attack path of an attacker is needed. This method combines the status and connection of each node in the network, as well as the vulnerability information, and gives the attack path with the lowest overall score by scoring the security of each node in the network.

The attack graph model is based on the attacker's point of view to exploit the possible vulnerabilities of each node in the network system to form one or more possible attack paths, and the path with the least attack cost is the most dangerous place of system. Different from traditional intrusion detection technology, intrusion detection is to alert according to specific rules after an attack occurs, while the attack graph model can dynamically and actively evaluate possible risks in the network system.

Since the knowledge graph is a networked knowledge base formed by entities with attributes linked through relationships [3], which is similar in form to the attack graph model, each network node in the attack graph can be used as an entity in the knowledge graph. The vulnerability of network nodes can be used as attributes of entities, and the connection topology between nodes can be used as the relationship between entities. Therefore, the knowledge graph can be used to completely store the information contained in the attack graph, and the natural retrieval advantage of the knowledge graph can be used to evaluate the security risk of any node in the attack graph. On the other hand, information such as vulnerabilities and CVSS (Common Vulnerability Scoring System) scores can be used as attributes of nodes or as entities that exist independently, and there are many relationships between node entities and vulnerability entities, which can form a complex network structure.

Based on the attack graph model, this paper discloses the node information in the network system and the required Common Vulnerabilities and Exposures (CVE) [4], Chinese National Vulnerability Database (CNNVD) [5] and CVSS [6]. Information related to network system security is stored in the knowledge graph. By extracting information into "entity-relation-entity" triples and key-value pairs of attributes in entity, the topology of the target network system is constructed and the connection between network nodes and vulnerabilities is established. Use the powerful graph storage, calculation and reasoning capabilities of the knowledge graph to improve the generation and analysis capabilities of the attack graph model for complex network systems, and referring to the CVSS scoring standard, combined with the attack graph path transition probability and the shortest path algorithm calculate the possible attack path of the attacker and the key vulnerable nodes in the network.

The main contributions of this paper can be summarized as follows. Use structured vulnerability information data to construct a vulnerability knowledge graph, including 610,445 nodes and 2,463,352 relationships between nodes. The relationship between the vulnerability, the type of vulnerability, and the device affected by the vulnerability is established. And a method of using network detection data to construct a knowledge graph model of the target network and combining the vulnerability knowledge graph to calculate the risk value of each node and generate a network attack graph is proposed. The new attack graph generation method is used to assess the network security status and calculate the attacker's possible attack path, use the graph algorithm to calculate the key nodes in the attack path. An algorithm of prediction of attack path and the risks value of nodes is also proposed. Use two knowledge graphs and the generated attack graph to evaluate the nodes in the network and calculate the most likely attack path.

The rest of this paper is organized as follows. Section 2 reviews related work. Section 3 describes our attack path prediction method. Specifically, it describes the method of generating attack graph by constructing knowledge graphs for vulnerabilities and the

target network. It also describes the network security assessment algorithm and the path prediction algorithm. Experimental data structure, experimental process and the analysis of results are described in Section 4. We conclude this paper in Section 5.

2. **Related Work.** In recent years, many researchers have used attack graph models as tools to analyze the security of network systems and have also proposed many methods to assess network security and predict attack intentions from different perspectives. In this section, we will review these works.

The concept of attack graph was first proposed by Phillips and Swiler [7] in the 1990s and applied it to the field of network vulnerability analysis. It solves the defect that the previous technology cannot dynamically evaluate the position of the node in the network. It has received widespread attention and recognition from the academic and industrial circles.

With the deepening of research, to solve the problems in different practical scenarios, the academic community has proposed various types of attack graphs. Sheyner et al. [8, 9] first proposed the concept of state attack graphs constructed with model detection technology. However, problems such as state explosion are prone to occur for large-scale networks. At present, there are few researches on state attack graphs. To solve the problem of state explosion, Homer et al. proposed an attribute attack graph [10], which uses secure nodes in the network as the attribute vertex. Therefore, the attribute attack graph has the advantages of fast construction speed and simple structure. It has been widely used in network risk assessment [11, 12], alarm correlation [13], etc.; the Bayesian attack graph model was developed by Liu and Man [14], this model uses the Bayesian network proposed by Pearl [15] to estimate the probability of the attack path, thereby inferring the attacker's possible attack path based on the prior knowledge. In recent years, some scholars have also proposed models such as 0day attack graph and social engineering attack graph [16] to solve the unpredictable problems of 0day vulnerabilities and social engineering.

For predicting future network attacks with more precision and dynamically adapts to the changes of network. GhasemiGol et al. [17] proposed a method using additional information, and pointed out the limitation of using static vulnerability probability to predict risk. In the process of prediction, it was used to extract and analyze information elements such as real-time intrusion alert and dependency graph and proposed a network attack prediction method based on information integration.

[18] proposed an attack path generation algorithm based on the attack graph model. If it is expected to obtain K paths with the largest reachable probability, the probability distribution is dynamically adjusted according to the K value, and the path prediction is realized by calculating the cumulative reachable probability of the attack path. However, this algorithm only considers the vulnerability utilization factor when quantifying the reachability probability of nodes and lacks modeling and analysis of the attack event itself.

[19] presented an attack path prediction method based on causal knowledge net, which detected the current attack actions by mapping the alarm sets to the causal knowledge net. By analyzing the attack actions, the capability grade of the attacker was inferred, according to which adjust the probability knowledge distribution dynamically. With the improved Dijkstra algorithm, the most possible attack path was computed.

[20] maps the attack graph to a Markov chain, and proposes a state transition probability algorithm for nodes in the Markov chain. However, the calculation of the state transition probability relies only on the static factor of the vulnerability availability index in CVSS, which does not consider the dynamic impact of the attacker's ability on the vulnerability availability, so the prediction accuracy of the attack path is low.

[24] aimied at the problem of the difficulty of predicting the network attack behavior and proposed an NAPG (Network Attack Profit Graph) model, which intuitively reflected the feasibility of attack behavior through attack cost and attack profit. And they use the attack feasibility analysis algorithm to eliminate redundant paths, and introduce the attack profit into the evaluation.

For prediction of intrusion intention of abnormal information, Liu [25] built a model which considered the path of the network nodes involved in the attack behavior. They defined a network attack graph based on the theory of attack graph, detected the right state of attacker and obtained the connection matrix. The attack path graph is used to describe the transfer relationship between nodes, map the process of the attack from one host or vulnerability to the next host or vulnerability and finally give the shortest path of attacker.

To sum up, the methods in these papers have disadvantages in some aspects. These methods do not consider the location of the nodes in the network or rely on a single CVSS availability score and ignore the attacker's ability, resulting in complex calculation methods and low accuracy of attack path prediction.

3. Attack Path Prediction Method.

In this section, we describe an attack path prediction model based on the knowledge graph. First, we introduce the construction method of the knowledge graph, including the vulnerability knowledge graph and the target network knowledge graph; then we generate the attack graph based on the knowledge graph and graph algorithm, and calculate the security score of each node in the network on this basis, and finally calculate the possible attack path of the attacker. The overall system architecture of the proposed models is shown in Figure 1.

3.1. Construction of knowledge graph.

The knowledge graph is a structured semantic knowledge database used to describe concepts and their relationships in the physical world in symbolic form. According to the knowledge acquisition process, the construction method of the knowledge graph is mainly divided into three processes: information extraction, knowledge fusion and knowledge processing. Among them, there are two main ways to extract information: top-down and bottom-up. The top-down refers to the extraction of ontology and relationship information from high-quality data such as structure and semi-structure, and bottom-up refers to the processing of semi-structured or unstructured information through technical means such as entity extraction, relationship extraction and attribute extraction. And among them the more credible information is selected as facts and stored in the knowledge base. Figure 2 shows the construction process of the knowledge graph.

Since the collected data is basically structured and semi-structured data, the top-down approach is chosen when constructing the knowledge graph. The collected data can be divided into two categories:

1) Topology of target network system from network detection;
2) Vulnerability information from CVE, CWE and CVSS.

We build a knowledge graph based on the open source neo4j database, which can accept a CSV file in a specified format as input and must contain two definitions of nodes and their attributes and the relationship between nodes. Therefore, it is necessary to generate a CSV file that meets the requirements for the two types of data collected and input neo4j to construct a knowledge graph. The public vulnerability information and vulnerability information contained in the nodes are the links connecting these two graphs.

According to the different detection methods, the obtained network topology information is also different. We use the NMAP software to scan network topology information.

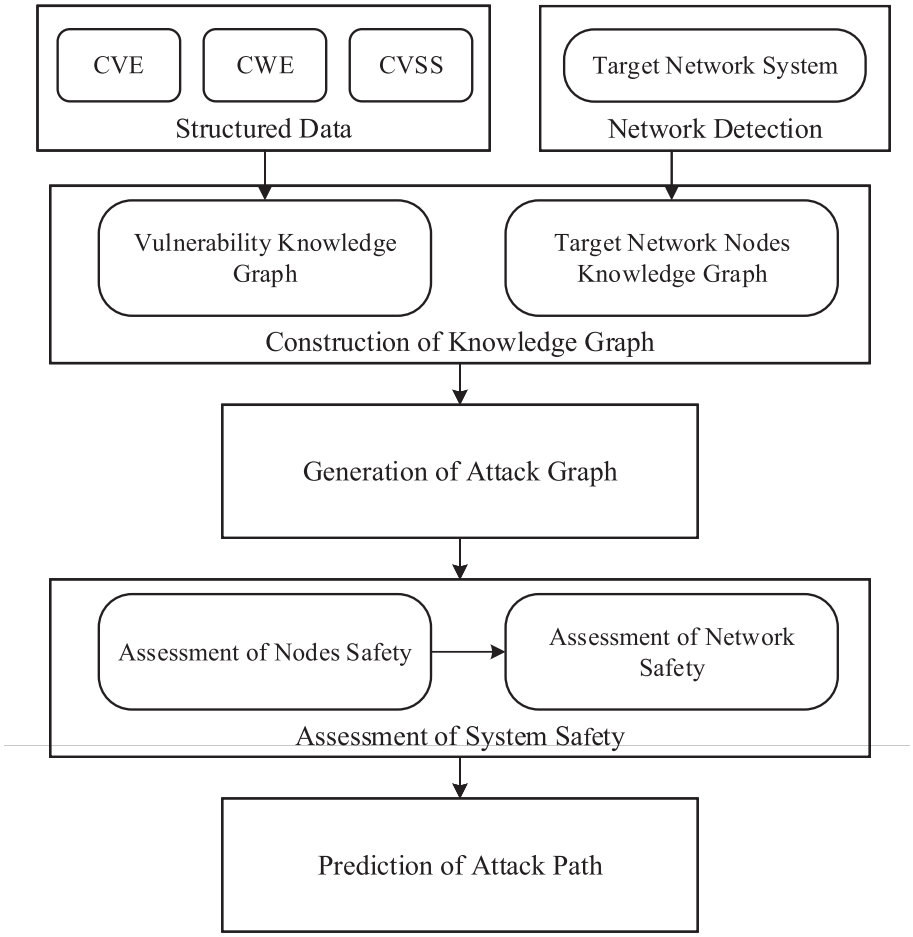

FIGURE 1. Network risk assessment process

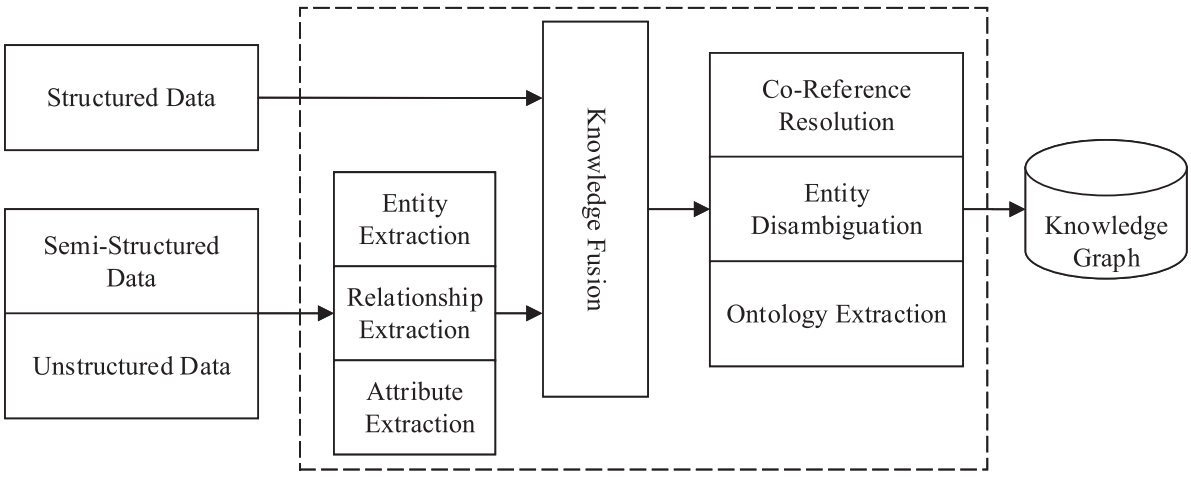

FIGURE 2. Knowledge graph creation process

Nodes and their attributes can be defined by a 7-tuple <type, ipAddr, macAddr, openPorts, OS, risk, networkSegment>, where "type" corresponds to the type of the device, such as a firewall, router, database or server. The different device types show the importance of the device in the network. "ipAddr" is the IP address of the device in the

network, and "macAddr" is the MAC address of the device, which are attributes that must be included in the nodes in the network and are important attribute to determine the nodes. "openPorts" is the open ports and possible services of the device, and the range is 1-65,535, which indicates the services and possible weaknesses that the device is running. "OS" is the type of operating system of the device, which includes Linux, Windows and other embedded systems, "risk" represents the vulnerabilities and risks of the device, which contain all known flaws and vulnerabilities on the device, and the risk score can be further calculated by the relationship between CVE and CVSS. "networkSegment" is the network segment where the device is located. The detailed definition is shown in Figure 3. All devices in the network are represented in this way.

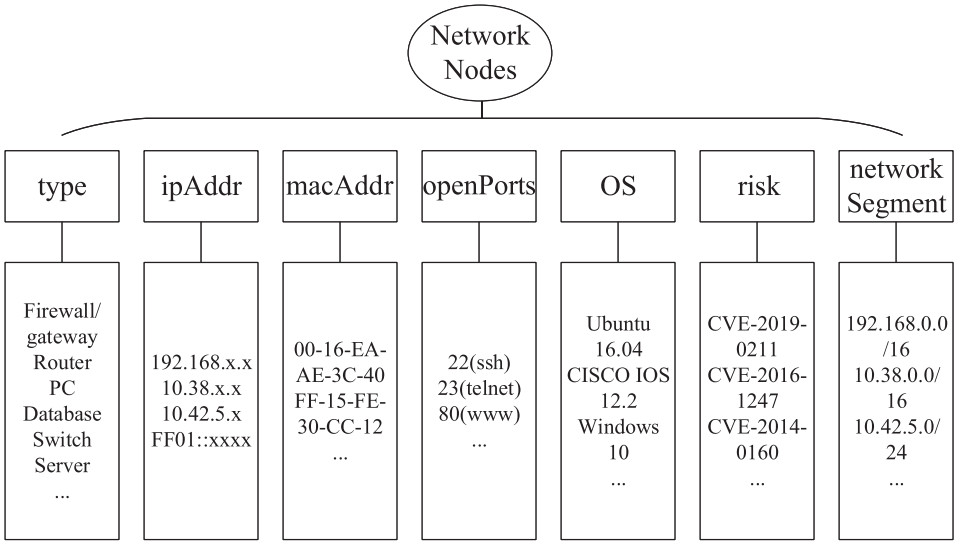

FIGURE 3. Network nodes' attributes and definitions

The relationship between network nodes is represented by a 3-tuple <start, end, type>, "start" represents the starting node, "end" represents the target node, and "type" represents the connection type between the two nodes. Connection types include "Connected" and "Linked". Connected means that two devices are logically connected and can access each other, and Linked means that two devices are physically connected and may not be able to access each other.

Through the above processing of the original data, several CSV files that can be imported into neo4j can be obtained, including nodes and relationships. Finally, a knowledge graph centered on network topology is generated. Figure 4 shows part of the expanded content of a node in the knowledge graph, the central node is the network device node with ip 10.38.7.96, which, like the other nodes in yellow, represents a device in the network with the properties depicted in Figure 3. "Connected" means the two nodes can communicate with each other, for example, nodes with ip 10.38.7.96 can access nodes with ip 10.42.1.5. Nodes in green indicate special devices in the network, such as switches and firewalls, and red nodes represent the risks with different numbers that the central nodes have.

3.2. **Generation of attack graph.** In a real network attack, due to the existence of the border firewall, the attacker usually cannot directly attack the target. The attack is often initiated through the DMZ area exposed to the external network, by using the vulnerability of the DMZ area and the connection of internal network, penetrate, and finally take control of the target machine. The more vulnerabilities a node contains, the

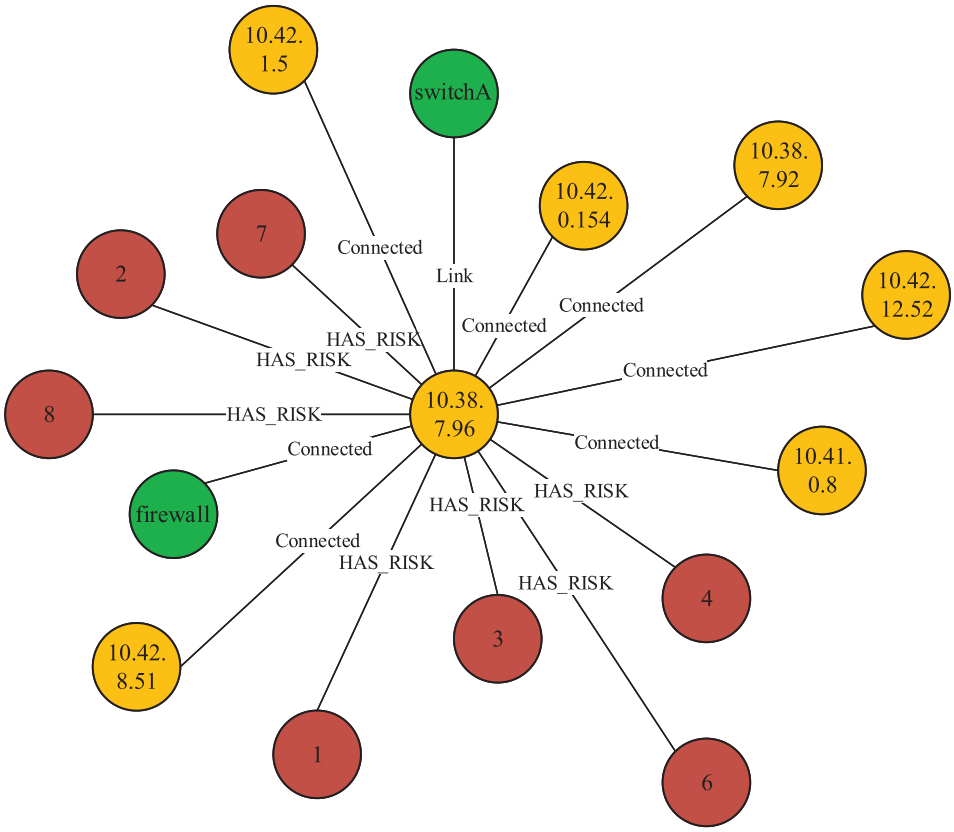

FIGURE 4. (color online) Example of knowledge graph nodes

higher the vulnerability risk is, and the easier the node will be hacked, which may further affect the security of adjacent nodes connected to the node. For a certain attack graph, it can be expressed as follows:

$$A = (N, S, P_\alpha) \tag{1}$$

where

- A is the model of attack graph.
- N represents a set of nodes.
- S represents a set of state of nodes.
- P_α is the probability of state transition, which is the probability of the attacker moving from this node to next node.

Further, the probability of state transition is based on a SoftMax function, which can be represented as

$$P_\alpha = \frac{e^{f(C_i,S_i,D_i)/10}}{\sum_{j=1}^{n} e^{f(C_j,S_j,D_j)/10}} \quad (i, j \in n) \tag{2}$$

where

- Function $f(C_i, S_i, D_i)$ is the risk score of the ith node.
- C_i is the function of number of vulnerabilities in the node.
- S_i is the function of CVSS vector of all vulnerabilities.
- D_i is the centrality of the node in knowledge graph.

In the generated knowledge graph, there is already information about the topological connection of the network, the vulnerability of each node and the CVSS score of vulnerability. In neo4j, Cypher language or Gremlin language can be used to traverse all nodes

and related attributes in the graph. Therefore, the attacker is used as the starting node to traverse the entire graph and calculate the centrality of each node. If the transition probability of a node is greater than the set threshold, the node is added to the attack graph as the next node, and retain the original connection relationship, and then continue to traverse, until there are no child nodes, return to the parent node to continue traversal, and finally generate a complete attack graph.

Since it is necessary to find out all possible nodes to form an attack graph, this article uses a depth-first search algorithm (DFS) to traverse the knowledge graph.

3.3. **Assessment algorithm of node risk.** We comprehensively consider the four dimensions of the vulnerabilities and risks contained in the node, the centrality of the node, the importance of the node, and the ability of the attacker to describe the security risk of the network node.

The higher the risk score of a node, the greater the probability of transferring to a compromised node.

The concrete realization of function $f(C_i, S_i, D_i)$ is

$$f(C_i, S_i, D_i) = V_i \times C_i \times I_i \times Cap_i \tag{3}$$

where

- V_i is the comprehensive vulnerability risk score, which is a sigmoid normalization function with adjusted input parameters. The realization of V_i is

$$V_i = 100 \times \frac{1}{1 + e^{2-0.05\times S_i}} \tag{4}$$

  where
  - Hyper-parameter 2 and 0.05 limit the range of output values to $(0, 100)$, and standardize the distribution of output.
  - S_i divides the vulnerabilities in the node into three grades of high risk, medium risk and low risk according to the basic score of CVSS vector, and weighs the sum. H_i, M_i and L_i correspond to the score of the all high, medium and low risk vulnerabilities in the node.

$$S_i = \sum H_i + 0.8 \times \sum M_i + 0.2 \times \sum L_i \tag{5}$$

- C_i is the centrality of the node, which is the in-out degree of the node calculated in the attack graph using the closeness centrality algorithm, indicating the connectivity relationship between the node and others. The bigger the C_i, the closer the node to the center, and means that there are more nodes connected to the node in the network, and the greater the impact of hacking the node.

$$C_i = \frac{N-1}{\sum_{i \neq v} d_{vi}} \tag{6}$$

  where
  - N is the number of nodes in the attack graph.
  - d_{vi} is the distance from the node to others.
- I_i is the importance of the node, representing the type of the node and software service that the node runs. Table 1 shows the importance of some device and service types. We quantified the importance as a value between 0 to 10. Consider the devices to be protected in the network as the most important and give them an importance score of 10, such as the database. The importance score of other device is determined by the number of CPE (Common Platform Enumeration, which is a structured naming scheme for information technology systems, software, and packages.) owned

TABLE 1. Importance comparison table

Device/Service Type	Importance
Firewall/Gateway	7
Router	5
Switch	5
General PC	2
Server	6
Database	10
Web service	6

by the device, and the more CPE owned by the device, the more importance it is in the network.

- Cap_i is an additional subjective indicator that indicates the ability of the attacker. This parameter can be determined by historical experience and traces left by the attacker. The range of Cap_i is $(0, 1)$.

3.4. **Prediction of attack path.** With the above data, we get the risk value of the node and the transition probability of each path. We follow two principles in forecasting.

- Attackers always give priority to nodes with higher risk.
- The attacker always attacks along a path with a greater probability of transition.

Therefore, we can predict the attacker's possible attack path (the maximum probability), and the shortest path.

4. Experiment Details.

4.1. **Experimental scene construction.** In order to verify the feasibility and effectiveness of the attack path prediction method, we use VMware virtualization software to build a local network experiment environment as shown in Figure 5. The network is mainly divided into two parts: the DMZ and the Intranet. The DMZ is responsible for providing services to the external network. The part consists of a web server D and a mail server E; there are three machines in the Intranet, consisting of a server F responsible for providing services to the Intranet, a user host G, and a database server H storing important data. The DMZ area and the Intranet area are connected to a firewall with routing function through a switch. Under normal circumstances, visitors from the external network can only access the servers in the DMZ zone of the network and cannot access the internal network, while in the internal network, they can freely access the servers in the DMZ zone. The machines in the DMZ area can only access some machines in the 40 network segment of the Intranet, and cannot access server E and database G. Attacker A launched an attack on the network through the Internet.

The open ports and existing vulnerability information of each node in the network are shown in Table 2.

Through the vulnerability information data stored in neo4j, the CVSS information of the vulnerabilities contained in the nodes in the network can be queried, as shown in Table 3.

4.2. **Experiment procedure.** According to the network topology, the attack graph of a network attack targeting at database G is obtained using our attack graph generation algorithm, as shown in Figure 6, where L means the link of the nodes. There are many paths an attacker can take to start from node A and hijack node MYSQL_G, such as A-C-F-E-G or A-D-F-E-G. Because this is only an experimental attack scene, the attack

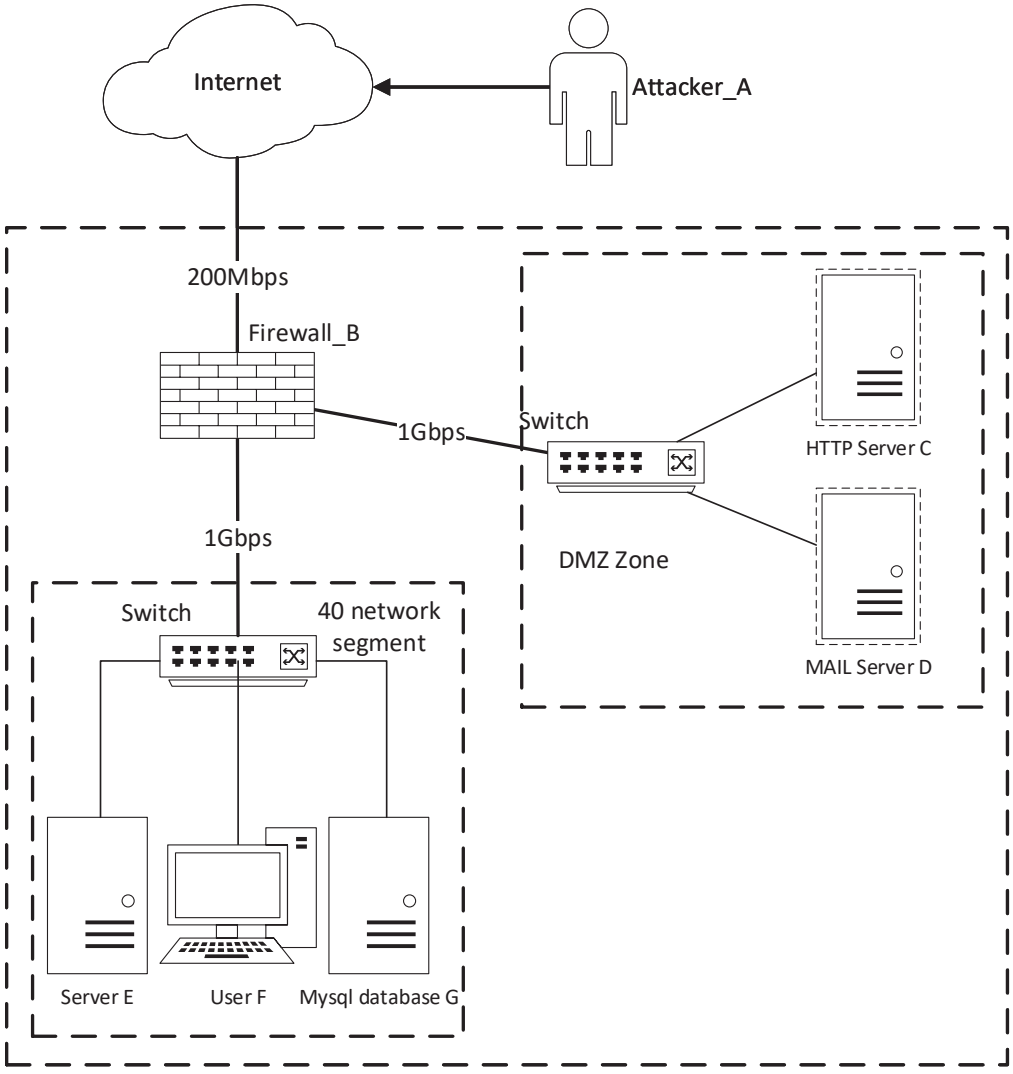

FIGURE 5. Experimental network environment

TABLE 2. Ports and vulnerability of nodes

Host	IP address	Open ports	Vulnerability
B	192.169.10.7	23	Unencrypted Telnet Server
C	10.41.19.1	23,80,443	CVE-2019-0211 CVE-2016-1247 CVE-2014-0160
D	10.41.19.2	25,80,109,110,143,443	CVE-2020-8794 CVE-2019-10149 CVE-2017-16943
E	10.40.0.83	22,23,80,443,139,445,902,912	CVE-2016-0778 CVE-2017-12615 Weak Password
F	10.40.0.84	22,23,80,443	Guest Account
G	10.40.0.251	22,902,3306	CVE-2016-6663 CVE-2016-6664

graph looks simpler. If real network topology data is used to generate the attack graph, then the graph will be very complex and the attack path will be even more exponentially increased.

Using the single node risk calculation formula, the risk value of each node in the network can be calculated, as shown in Table 4. Since the vulnerability contained in node D has the highest degree of harm, where CVE-2020-8794 has a CVSS score of 10, and the remaining two vulnerabilities are also 7.5, the calculated V is 32.08. However, node F has only one vulnerability with a CVSS score of 4.9, so its V is only 14.14. On the other hand, node

TABLE 3. CVSS information of vulnerability

Vulnerability	CVSS3.0 vector	Base score
Unencrypted Telnet Server	AV:N/AC:M/Au:N/C:P/I:P/A:N	5.8
CVE-2019-0211	AV:L/AC:L/Au:N/C:C/I:C/A:C	7.2
CVE-2016-1247	AV:L/AC:L/Au:N/C:C/I:C/A:C	7.2
CVE-2014-0160	AV:N/AC:L/Au:N/C:P/I:N/A:N	5.0
CVE-2020-8794	AV:N/AC:L/Au:N/C:C/I:C/A:C	10.0
CVE-2019-10149	AV:N/AC:L/Au:N/C:P/I:P/A:P	7.5
CVE-2017-16943	AV:N/AC:L/Au:N/C:P/I:P/A:P	7.5
CVE-2016-0778	AV:N/AC:H/Au:S/C:P/I:P/A:P	4.6
CVE-2017-12615	AV:N/AC:M/Au:N/C:P/I:P/A:P	6.8
Weak Password	AV:N/AC:L/Au:N/C:C/I:C/A:C	10.0
Guest Account	AV:A/AC:M/Au:S/C:P/I:P/A:P	4.9
CVE-2016-6663	AV:L/AC:M/Au:N/C:P/I:P/A:P	4.4
CVE-2016-6664	AV:L/AC:M/Au:N/C:C/I:C/A:C	6.9

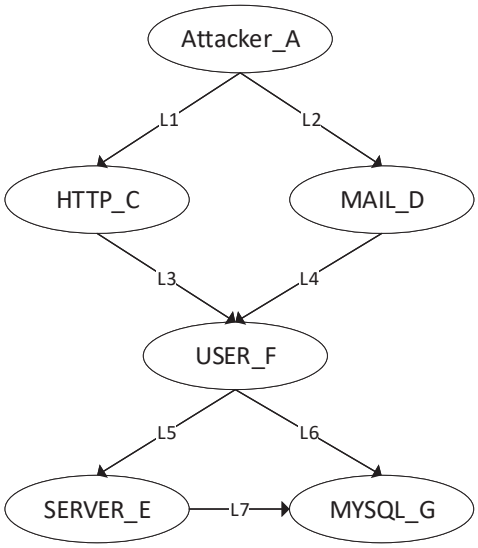

FIGURE 6. Attack graph

TABLE 4. Calculation for nodes risk

	C	D	E	F	G
V	25.35	32.08	26.03	14.14	17.54
C	0.71	0.71	0.56	0.83	0.56
I	6.00	6.00	6.00	2.00	10.00
Cap	0.5	0.5	0.5	0.5	0.5
R	**54.00**	**68.33**	**43.73**	**11.74**	**49.11**

F is a mandatory route to node G and has the most connections to other nodes, so the centrality C is 0.83, which is the highest among all nodes. The importance of node F for the attack can also be seen in Figure 6. For node G, its importance I is 10 because it runs an important database service, while the other nodes in Figure 6 do not have such a high score. To facilitate the experimental calculation, we set the hacking capability Cap of all nodes to 0.5, and in the actual calculation, the score is set based on historical experience.

According to the calculation results, node D has the highest risk value R of 68.33, while node F has the lowest risk value R of 11.74. This shows that in this attack graph, node D is the highest risk node and node F is the safest node. The attacker may start the attack from node D.

With the risk value R of each node, the transition probability can be calculated as

$$P_{L1} = \frac{e^{5.4}}{e^{5.4} + e^{6.833}} = 0.193 \tag{7}$$

$$P_{L2} = \frac{e^{6.833}}{e^{5.4} + e^{6.833}} = 0.807 \tag{8}$$

$$P_{L3} = 1 \tag{9}$$

$$P_{L4} = 1 \tag{10}$$

$$P_{L5} = \frac{e^{4.373}}{e^{4.373} + e^{4.911}} = 0.369 \tag{11}$$

$$P_{L6} = \frac{e^{4.911}}{e^{4.373} + e^{4.911}} = 0.631 \tag{12}$$

$$P_{L7} = 1 \tag{13}$$

The probability of an attacker from node A to node D is 0.807, while to node C is only 0.193, so there is a greater possibility of reaching node D via L2. Also, since there is only one path from node D to node F, the transfer probability is 1. So, the most likely attack path of the attacker is A→L2→D→L4→F→L6→G, shown in red in Figure 7, which contains the riskiest nodes in the network; in fact, these nodes are also the ones with the highest risk of flaws and vulnerabilities.

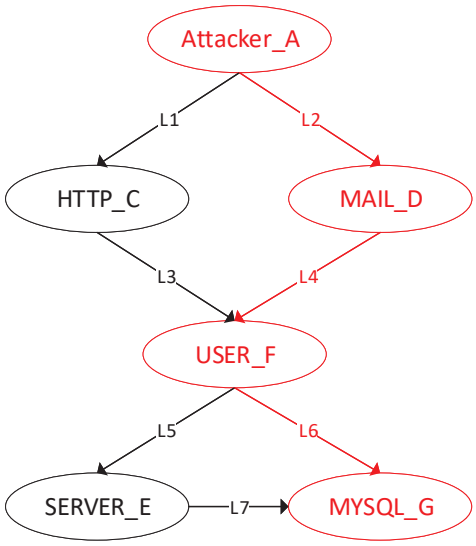

FIGURE 7. (color online) Possible attack path for attacker

4.3. **Experimental results.** This experiment proves that the attack graph constructed by the knowledge graph and CVSS vector can describe the possible attack path of the attacker from the entry node to the target node, and can reasonably reflect the risk value of each node, which is convenient to compare the security between different nodes and networks.

Through the above analysis, the method in this paper can quantitatively calculate the threat probability and impact of the nodes and possible attack paths in the current

network. For example, in the above experiment, node F has fewer vulnerabilities and is also an ordinary computer that does not carry important services, so its risk value is low, while node D has the most vulnerabilities and is a server running mail services. Its risk of being attacked is even higher. Although the risk value of other nodes in the network is high, the risk value of node F is low, so that the overall risk value of the network is not high, and attackers do not have much attack paths. On the contrary, if node F has a new vulnerability or adds a new node H to connect the DMZ and the Intranet area, the risk value of the network will increase, and the possible attack path will also increase.

The comparison between the method in this paper and other related research is shown in Table 5. Most researches give the maximum probability attack path and probability value. The method in this paper also gives the shortest attack path (not necessarily the maximum probability). Some methods do not use vulnerability information (such as CVSS) to assess the attack path.

TABLE 5. Comparison of method

Method	[21]	[22]	[23]	This method
Shortest path	✓	×	✓	✓
Most possible path	✓	✓	✓	✓
Using CVSS	×	✓	×	✓
Probability of successful attack	✓	×	✓	✓
Construct with knowledge graph	×	×	×	✓

5. **Conclusion.** The network topology has many node entities and connection relationships due to its complex mesh structure, which is suitable for storage and analysis using graph databases and knowledge graphs. This paper uses the knowledge graph method to design an attack path prediction method. The state transition probability of the node is calculated by using the risk score of the node in the network, and the possible attack path is calculated according to the transition probability. This article mainly discusses from the following aspects.

- Construction of knowledge graph. Using a top-down method to generate a knowledge map after processing the collected structured and semi-structured data, which generates vulnerability knowledge graph and target knowledge graph respectively.
- Generation of attack graph. From the perspective of the attacker, the DFS is used to traverse the network scene nodes and relationships stored in the knowledge graph, combined with CVSS metrics, to generate an attack graph.
- Assessment of node risk. Combining the CVSS metrics, the close centrality of the node in the network and the services that the node runs and other information to calculate the quantitative risk value of the node, and predict attack path.

In the future, we may consider adding more security-related data on the basis of the existing knowledge graph, such as logs or IDS alarms, so that attack path prediction can be performed in multiple dimensions to improve the accuracy of prediction.

Acknowledgment. This work is partially supported by Excellent Youth Foundation of Sichuan Scientific Committee under Grant 2019JDJQ0058 and by Sichuan Science and Technology Program under Grant 2019YJ0643. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] D. Bai, J. Yin and W. Yang, Research on the mechanism of university-enterprise knowledge transfer on enterprise innovation performance, *ICIC Express Letters, Part B: Applications*, vol.11, no.10, pp.979-986, 2020.

[2] National Computer Network Emergency Technology Coordination Center, *China Internet Network Security Report 2019*, People's Posts and Telecommunications Press, 2020.

[3] G. L. Qi, H. Gao and T. X. Wu, Research progress of knowledge graph, *Information Engineering*, vol.3, no.1, pp.4-25, 2017.

[4] *CVE*, https://cve.mitre.org/, Accessed in March 2021.

[5] *CNNVD*, http://www.cnnvd.org.cn/, Accessed in March 2021.

[6] *CVSS*, https://nvd.nist.gov/vuln-metrics/cvss, Accessed in March 2021.

[7] C. Phillips and L. P. Swiler, A graph-based system for network-vulnerability analysis, *Proc. of the 1998 Workshop on New Security Paradigms*, 1998.

[8] O. Sheyner, J. Haines, S. Jha et al., Automated generation and analysis of attack graphs, *Proc. of 2002 IEEE Symposium on Security and Privacy*, 2002.

[9] S. Jha, O. Sheyner and J. Wing, Two formal analyses of attack graphs, *Proc. of the 15th IEEE Computer Security Foundations Workshop (CSFW-15)*, 2002.

[10] J. Homer, S. Zhang, X. Ou et al., Aggregating vulnerability metrics in enterprise networks using attack graphs, *Journal of Computer Security*, vol.21, no.4, pp.561-597, 2013.

[11] C. Xiao and J. Z. Li, Research on quantitative evaluation of network security based on Bayesian attack graphs, *Application Research of Computers*, vol.30, no.9, pp.2763-2766, 2013.

[12] D. Wu, Y.-F. Lian, K. Chen and Y.-L. Liu, A security threat identification and analysis method based on attack graph, *Chinese Journal of Computers*, vol.35, no.9, 2012.

[13] S. Noel and S. Jajodia, Optimal IDS sensor placement and alert prioritization using attack graphs, *Journal of Network and Systems Management*, vol.16, no.3, pp.259-275, 2008.

[14] Y. Liu and H. Man, Network vulnerability assessment using Bayesian networks, *Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security*, 2005.

[15] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Elsevier, 2014.

[16] K. Beckers, L. Krautsevich and A. Yautsiukhin, Analysis of social engineering threats with attack graphs, in *Data Privacy Management, Autonomous Spontaneous Security, and Security Assurance. DPM 2014, QASA 2014, SETOP 2014. Lecture Notes in Computer Science*, J. Garcia-Alfaro et al. (eds.), Cham, Springer, 2015.

[17] M. GhasemiGol, A. Ghaemi-Bafghi and H. Takabi, A comprehensive approach for network attack forecasting, *Computers & Security*, vol.58, pp.83-105, 2016.

[18] B. Kun, D. Z. Han and J. Wang, K maximum probability attack paths dynamic generation algorithm, *Computer Science and Information Systems*, vol.13, no.2, pp.677-689, 2016.

[19] W. Shuo, G. M. Tang and G. Yan, Attack path prediction method based on causal knowledge network, *Journal of Communications*, vol.37, no.10, pp.188-198, 2016.

[20] H. Hu, Y. Liu, H. Zhang, Y. Yang and R. Ye, Route prediction method for network intrusion using absorbing Markov chain, *Journal of Computer Research and Development*, vol.55, no.4, pp.831-845, 2018.

[21] S. Carlos, G. Richarte and J. L. Obes, An algorithm to find optimal attack paths in nondeterministic scenarios, *Proc. of the 4th ACM Workshop on Security and Artificial Intelligence*, 2011.

[22] I. Nwokedi and B. Bhargava, Extending attack graph-based security metrics and aggregating their application, *IEEE Trans. Dependable and Secure Computing*, vol.9, no.1, pp.75-86, 2010.

[23] S. Sendi, Alireza, M. Dagenais and L. Wang, Realtime intrusion risk assessment model based on attack and service dependency graphs, *Computer Communications*, vol.116, pp.253-272, 2018.

[24] S. Moskal, S. J. Yang and M. E. Kuhl, Cyber threat assessment via attack scenario simulation using an integrated adversary and network modeling approach, *J. Def. Model Simul.*, vol.15, no.9, pp.13-29, 2018.

[25] X. Liu, A network attack path prediction method using attack graph, *J. Ambient. Intell. Human Comput.*, https://doi.org/10.1007/s12652-020-02206-5, 2020.