A Trusted Virtual Network Construction Method Based on Data Sources Dependence

Xiaorong Cheng[1] and Tianqi LI[2]
[1]Department Computer Science, North China Electric Power University, China
[2]C-Epri Electric Power Engineering CO, LTD, China

Abstract: *At present, the isolated and single data source cannot meet the needs of system security. Based on the research of the trusted computing theory, this paper creatively put forward a method to construct a trusted virtual network based on data source dependency. Firstly, the credibility of data source is calculated by the NEWACCU algorithm, and then, the trusted virtual network which is composed of the entity of data source is built dynamically by calculating the credibility between data sources, which will provide technical support for future credibility assessment and further research on information security. Taking the data of e-commerce platform as an example, the experimental results verify the effectiveness of the method.*

Keywords: *Data source, credibility, trusted virtual network, dynamics, modeling and simulation.*

Received June 22, 2016; accepted April 11, 2017

1. Introduction

With the growing scale of the Internet, there are more and more data sources in the network [7, 9]. Whether it is a web site, personal space, blog, or networking hardware and equipment, etc., at present, they are only used as an isolated, single entity to generate data from time to time, but in fact, there is a very close relationship between them. If we can establish a virtual network of all the data sources through a certain relationship, then we can provide a new idea for further research on credibility evaluation and information security [2, 4, 16, 17].

At present, there are many research methods and some results for the trusted measurement and evaluation of data. The methods of credibility analysis are mainly divided into two categories, one is subjective trust analysis based on belief, which is a cognitive phenomenon which is the subjective judgment of the specific characteristics or behavior of the object of trust, and this kind of judgment, which has ambiguity, uncertainty and can't be accurately described, verified and speculated, is relatively independent of the subject's characteristics and behavior [3]. Documents [3, 6, 10, 11, 12] have proposed different subjective methods based on Probability, Fuzzy Set Theory, Cloud Theory and so on. The other is the objective trust analysis based on the evidence theory, which can be accurately described, verified and speculated. The trust relationship between the two is strictly defined by appropriate evidence. D-S evidence theory is used to calculate the credibility by documents [8, 15].

At present, the research of trusted network is focused on the traditional network. The research of the theory and technology of trusted network is mainly on the user level, emphasizing the protection of the security and survivability of the service; On the basis of the original network security theory and technology, the security thought of increasing the behavior is credible; At the design level, the emphasis is on ensuring the controllability of the network [1]. Compared with the research of the trusted network, the research on the dependence relation between the data sources is lack, and the research of the trusted network which is constructed by the data sources is seldom. Therefore, it has stronger practical significance to establish between the data source and the trusted network research and application. In this paper, we propose a new method of Constructing Hierarchical trusted network dynamically based on the relationship between data sources. This method considers the influence of time attenuation coefficient and the penalty factor on the credibility calculation. A hierarchical virtual network is established by computing the credibility of the data source and the credibility between data sources, which is different from the traditional network, each of the hardware and software of the data can be used as a node in the network [5, 13, 14].

2. A Description of the Trusted Network

In order to facilitate the understanding, this paper gives the definition of the method presented, to explain the basic problems of the trusted network analysis.

In this paper, the network is different from the traditional network and a hierarchical and directional virtual network by computing the credibility of the data sources. The traditional network node is composed of network equipment, and the virtual network node is a software and hardware entity that can provide data.

The value of network link weight is the value of the credibility of the data source. The data provided by the data source is composed of multiple attributes, expressed as v={d_1, d_2, d_3, ..., d_n}. There into, d_i refers to the data of the “i” attributes.

- *Definition* 1. $\mu(t)$ is the time decay factor at the “*t*” moment. If the credibility value calculated at the *t* and *t-1* moment is the same, then it is punished by the time decay factor. As is shown below in Equation (1).

$$\mu(t) = 1 - \frac{\Delta t \cdot \xi}{t - t_0} \quad (0 \le \mu(t) \le 1) \tag{1}$$

There into, Δt is the time difference of calculation between two times. t_0 is the starting moment of the current calculation, *t* is the current moment. ξ is used to adjust the attenuation rate, and the scale is [0,1]. The time decay factor takes into account the influence of the time decay of the credibility, and the accuracy of the calculation is improved.

- *Definition* 2. $\lambda(t)$ is the penalty coefficient of credibility of the model at the “t” moment. As is shown below in Equation (2).

$$\lambda(t) = \begin{cases} 1 & ,\Delta LocalT \ge 0 \\ 0 \le x < 1 & ,\Delta LocalT < 0 \end{cases} \tag{2}$$

There into, *Local T* is the local credibility. As is shown below in Equation (3).

$$\Delta LocalT = LocalT(t) - LocalT(t-1) \tag{3}$$

When its value is $\Delta LocalT < 0$, it is to impose a penalty on the local credibility coefficient, the smaller the value, the greater the intensity of punishment. Penalty factor for the provision of false information of the entity, can be quickly reduced to a very low credibility, so as to effectively detect malicious entities.

- *Definition* 3. $\Delta Context(A,B,t)$ is whether between data source A and data source B has a new context of direct interaction at the “t” moment. As is shown below in Equation (4).

$$\Delta\text{Context(A, B, t)} = \text{Context(A, B, t)} - \text{Context(A, B, t} - 1) \tag{4}$$

- *Definition* 4. *Accept(A,B,t)* is the degree of recognition of the similarity degree of data source B to the data source A at the “t” moment, which is evaluating the expected value of the similarity of the data provided by the two data sources. As is shown below in Equation (5).

$$Accept(A,B,t) = \frac{\sum\limits_{\substack{data_a \in Data(A) \\ data_b \in Data(B)}} Sim(data_a, data_b)}{Data(A) \cap Data(B)} \tag{5}$$

There into, *Data* (*A*) is a data set provided by data source A. $Data(A) \cap Data(B)$ is the number of data with the same theme in the data set provided by the data source A and B. Data value is composed of a set of semantic original words. *data_a*, is decomposed into {d_{a1}, d_{a2}, d_{a3}, ..., d_{am}}, $data_b$ is decomposed into {d_{b1}, d_{b2}, d_{b3}, ..., d_{bn}}. $Sim(data_a, data_b)$ is the similarity degree between $data_a$ and $data_b$. As is shown below in Equation (6).

$$Sim(data_a, data_b) = \frac{\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} Sim(d_{ai}, d_{bj})}{(m + n - \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} Sim(d_{ai}, d_{bj}))} \tag{6}$$

3. Trusted Network Model

According to the above description, this paper presents a method to dynamically construct the credibility analysis network of hierarchical structure. In the initial time, the node in the trusted virtual network is composed of a distributed data source with initial value or a data source with expert experience, and the network topology is dynamic; Because the data source has the characteristic of dynamic and dependence. In the course of operation, the correlation credibility between data sources is calculated, and the link between the two nodes of the network is established, and the trusted virtual network of the model is dynamically reconstructed. As the specific process is described below.

3.1. Trusted Network Analysis Model

When the credibility value between data sources is more than a certain threshold, the link between the two data sources can be established, which indicates the credibility of a data source to another data source. With the expansion of the scale of the network, the data source trusted network is more and more stable. If the data provided by the data source is not reliable, the method can quickly assign the provider (data source) to the penalty coefficient, so that it can reduce the credibility of the provider for a period of time; If the data source can continue to provide reliable data in the later period, the credibility penalty of which will be weakened, and the original credibility will be restored in the trusted network. In the credibility of the virtual network, if there is no new context direct interaction between the data sources in a computing interval, the credibility value is not changed, and the method will also impose a time penalty to the data source.

- *Definition* 5. Credibility of Data Source**:** It is the credibility value of the data source calculated by using the number of votes in the “NEWACCU”[2] algorithm in the whole trusted network. Its notation is denoted as *T*(*S*,*t*), the meaning behind which is the credibility of data source S at the “t” moment. As is shown below in Equation (7).

$$T(\mathrm{S},t)=\frac{\sum\limits_{v\in V(S)} Vote(v\bullet,\mathrm{t})}{m} \tag{7}$$

There into, m is the number of values provided by the data source *S*; *v*(*s*) is a collection of data values provided by the data source S. *T(S,t)* is the credibility value of the data source at the "*t*" moment; *Vote(v,t)* is the number of v's votes at the "*t*" moment. The merit of the formula is that the value of the votes is represented by the accuracy of each data value.

- *Definition* 6. Local Credibility: It is composed of the credibility of direct context interaction and the credibility of the similarity between data sources. When there is a direct context interaction between data sources or the similarity of data or behaviors provided by between data sources exceeds a certain threshold, we believe that between data sources have a local credibility. Its notation is denoted as $Local\mathrm{T}_A(B,t)$, the meaning behind which is the local credibility of data source A relative to data source *B* at the "t" moment. As is shown below in Equation (8).

$$LocalT(B,t)=\begin{cases} Random()\ or\ 0 & ,t=0 \\ LocalT(B,t-1)\cdot\mu(t) & ,\Delta Context(A,B,t)=0 \\ \left[\alpha\cdot DirT(A,B,Context(A,B,t),t)+\beta\cdot Accept(A,B,t)\right]\cdot\lambda(t)\,\text{�},other\,\text{�} \end{cases} \tag{8}$$

There into, the initial value is a random number or 0, which indicates that data source A has some trust or no trust for data source B. α_1 and β_1 respectively represent the degree of recognition of the weight coefficients and the degree of similarity in the local credibility, and $\alpha_1+\beta_1=1$. $DirT(A,B,Context(A,B,t),t)$ is the trusted value of data source A relative to data source B in the circumstances of context interaction at the "t" moment.

- *Definition* 7. Credibility Between Data Sources: It is composed of the local credibility between data sources and the credibility of data sources. Its notation is denoted as $T_A(B,t)$, the meaning behind which is the comprehensive credibility of data source A relative to data source B at the "t" moment. As is shown below in Equation (9).

$$\mathrm{T}_{\mathrm{A}}(B,t)=\alpha_2\cdot T(B,t)+\beta_2\cdot LocalT_A(B,t) \tag{9}$$

There into, α_2 and β_2 respectively represent the dsata source credibility weight coefficient and the local confidence weight coefficient, and $a_2+\beta_2=1$.

From the above definitions, the relationship between the credibility definitions between the data sources can be obtained, as shown in Figure 1.

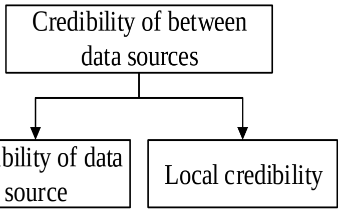
Credibility of between data sources
Credibility of data source
Local credibility

Figure 1. The correlation of the credibility definition between data sources in the trusted network.

3.2. Trusted Network Algorithm Analysis Process

According to the foregoing, the original data are pre processed, and each data collected will be processed into a corresponding set of the original words; Then the Equation (7) is used to calculate the credibility of the data source; the Equation (8) is used to calculate the local credibility between the data sources, which needs to calculate the local credibility of the direct credibility and similar degree of recognition, and the weights are combined to synthesize the overall reliability; If there is a context interaction between the two data sources or the similarity of the data provided by the same subject changes, then the local credibility need to be updated; If there is no new behavior, then it is punished by the time decay factor; Finally, the Equation (9) is used to calculate the credibility between the data sources, this requires the synthesis of the credibility of the data source and the local credibility. At this time, it is only the evaluation of the credibility of a data source to another data source. The algorithm uses an iterative method to obtain the credibility between each data source and all other data sources. If the credibility value exceeds the threshold "η" set by the system, a directed link between the two data sources is increased, the value of which is the credibility of the data source, and the data value is changed with the change of network.

In the experimental verification, if a data source provides some malicious and non real data, the credibility can be punished by the method according to the influence degree, so that the credibility can be reduced quickly. If the data source can continue to provide reliable data in the later period, the credibility penalty will be weakened. The program flow chart of the whole algorithm is shown in Figure 2.

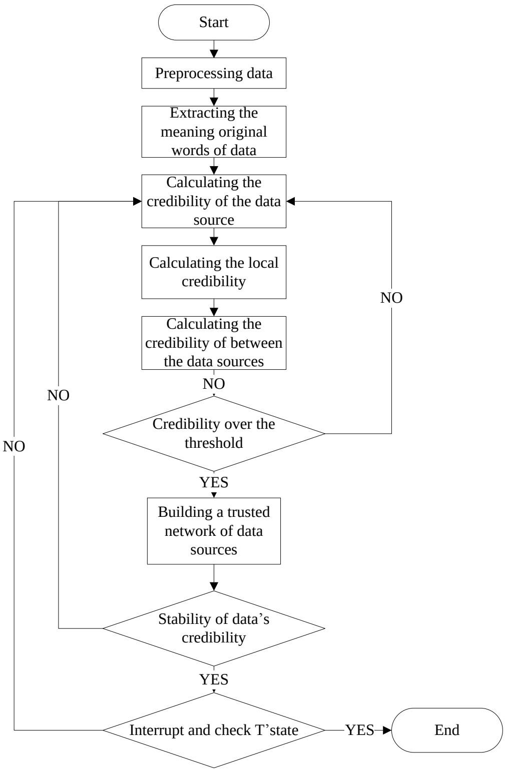
Start
Preprocessing data
Extracting the meaning original words of data
Calculating the credibility of the data source
Calculating the local credibility
Calculating the credibility of between the data sources
NO
Credibility over the threshold
NO
NO
NO
YES
Building a trusted network of data sources
Stability of data's credibility
YES
Interrupt and check T'state
YES
End

Figure 2. The program flow of the construction method of the trusted virtual network.

4. Case Analysis and Verification

In this paper, the selection of the simulation tool is MATLAB, the object of simulation experiment is selected as the data in electronic commerce.

4.1. Simulation Experiment Design

The experimental data was collected through the web crawler technology to the commercial website of the current e-commerce platform, especially the evaluation information of commodity information acquisition, and the data of some samples were labeled artificially. Data set includes 4 categories of goods, different brands of goods data, from a random sample of 7972 pieces of goods, the number of users can reach about 100000 people, the evaluation of information up to ten million. The algorithm is applied to the credibility analysis of user evaluation parameters. In this experiment, a subject refers to a commodity, the entity refers to the customer, the data refers to a product assessment information made by the user.

The collected data is divided into two parts, which part is as the establishment of the trusted network, the network sample repeated training, while adjusting the updated parameter values to adapt to the environment changes, another part of the data is the verification of the model stability, accurate accuracy. The parameters involved in the model are quite a few, and the parameters are used in an artificial setting. In the later period, the parameters are modified to adapt to different scenarios. The setting of parameters is shown in Table 1.

Table 1. The default parameters' list for simulation experiments.

Parameter	Default value	Description
N	302412	the number of data sources
$\mu(t)$	$1-\frac{\Delta t\cdot\xi}{t-t_0}$	the time decay coefficient
$\lambda(t)$	$0\le\lambda(t)\le 1$	the penalty coefficient
α_1	0.735	the local direct credibility weight
β_1	0.265	the similarity weight of local credibility
η	0.314	the data or behavior similarity threshold
Δt	1s	the time difference of calculation
α_2	0.735	The credibility weight of the data source
β_2	0.265	the local credibility weight

4.2. Experimental Results and Analysis

At a certain time, the hierarchical data source is composed of a trusted virtual network topology diagram shown below.

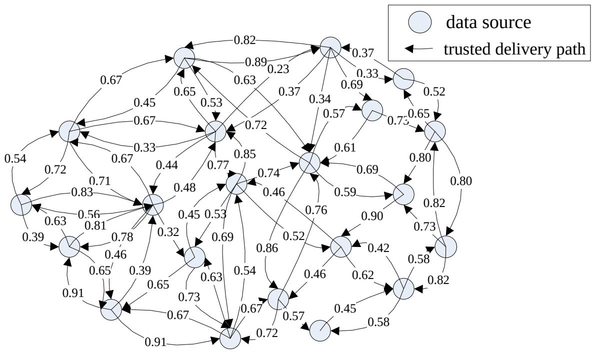

Figure 3. Partial topology of a trusted virtual network.

By definition 7, using the Equation (9) calculates the local credibility between the data sources, the trusted network of data sources can be constructed. A network node device is only a part of the entire trusted network, which cannot be considered separately. The role of the credibility of a single network node is to provide data to calculate the credibility between the data sources. The credibility of the network node equipment is restricted by the credibility of data that provided. Figure 1 shows that the credibility between the data sources is restricted by the credibility of the data source and the local credibility. By iteratively calculating the credibility of the three, the credibility of the virtual network is dynamically constructed, and the stability of the credibility of the network further improves network security. In Figure 3, a hierarchical data source network topology is given, and the direction of the arrow is the evaluation of the credibility of a data source to point to another data source, and the weight is the credibility value of the data source.

5. Conclusions

In this paper, we propose a new method to construct the trusted network based on the data source dependency by studying the demand of the trusted virtual network and the related theory of the trusted computing. Data quantity provided by data source is more, the more stable trusted network construction, more to provide the credibility of data measurement of accurate analysis, so it can well meet the demand of information security and credibility. There are still some places which need to be optimized in this paper. These are the focus of the study in the future, and need to be further improved. In this paper, a simple data instance is selected to verify the feasibility of the model, which provides a way to solve the problem of further research on the credibility evaluation method in the future.

Reference

[1] Chuang L. and Xue-hai P., "Research on Trustworthy Networks," *Chinese Journal of Computers*, vol. 28, no. 5, pp. 751-758, 2005.

[2] Chandran K., Shanmugasudaram V., and Subramani K., "Designing a Fuzzy-Logic Based Trust and Reputation Model for Secure Resource Allocation in Cloud Computing," *The International Arab Journal of Information Technology*, vol. 13, no. 1, pp. 30-37, 2016.

[3] Deng-guo F., Min Z., and Wei L., "Big Data Security and Privacy Protection," *Chinese Journal of Computers*, vol. 37, no. 01, pp. 246-258, 2014.

[4] Feng Z., Jian W., Yan-fei Z., and He D., "Trust Model Based on Group Recommendation in Social Network," *Computer Science*, vol. 41, no. 5, pp. 168-172, 2014.

[5] Hongquan L. and Wei W., "Research of Trust Evaluation Model Based on Dynamic Bayesian Network," *Journal on Communications*, vol. 34, no. 9, pp. 68-76, 2013.

[6] Huang H. and Wang R., "Subjective Trust Evaluation Model Based on Membership Cloud Theory," *Journal on Communications*, vol. 29, no. 4, pp. 13-19, 2008.

[7] Jun-feng T., Rui-zhong D., and Yu-ling L., "Trust Evaluation Model Based on Node Behavior Character," *Journal of Computer Research and Development*, vol. 48, no. 6, pp. 934-944, 2011.

[8] Lin Z., Yuwen L., Weichuan W., and Haiyan W., "Trust Evaluation Model Based on Improved D-S Evidence Theory," *Journal on Communications*, vol. 34, no. 7, pp. 167-173, 2013.

[9] Meng X., Zhang G., Liu C., Kang J., and Li H., "Research on Subjective Trust Management Model Based on Cloud Model," *Journal of System Simulation*, vol. 19, no. 14, pp. 3310-3317, 2007.

[10] Mehenni M. and Moussaoui A., "Data Mining from Multiple Heterogeneous Relational Databases Using Decision Tree Classification," *Pattern Recognition Letters*, vol. 33, no. 13. pp. 1768-1775, 2012.

[11] Qiu-yue Z., Wan-li Z., Zhong-sheng T., and Ying W., "A Method for Assessment of Trust Relationship Strength Based on the Improved D-S Evidence Theory," *Chinese Journal of Computers*, vol. 37, no. 4, pp. 873-883, 2014.

[12] Tang W. and Chen Z., "Research of Subjective Trust Management Model Based on the Fuzzy Set Theory," *Journal of Software*, vol. 14, no. 8, pp. 1401-1408, 2003.

[13] Xiao-yong L. and Xiao-lin G., "Trust Quantitative Model with Multiple Decision Factors in Trusted Network," *Chinese Journal of Computers*, vol. 32, no. 3, pp. 405-16, 2009.

[14] Xiuquan Q., Chun Y., Xiaofeng L., and Junliang C., "A Trust Calculating Algorithm Based on Social Networking Service Users, Context," *Chinese Journal of Computers*, vol. 34, no. 12, pp. 2403-2413, 2011.

[15] Zhao Q., Zuo W., Tian Z., and Wang Y., "A Method for Assessment of Trust Relationship Strength Based on the Improved D-S Evidence Theory," *Chinese Journal of Computers*, vol. 37, no. 4, pp. 873-883, 2014.

[16] Zhiqiang Z., Lixia L., Xiaoqin X., and Yixiang F., "Information Evaluation Based on Sources Dependence," *Chinese Journal of Computers*, vol. 35, no. 11, pp. 2392-2402, 2012.

[17] Zhou J., Wang Q., Hung C., and Yi X., "Credibilistic Clustering: The Model and Algorithms," *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems*, vol. 23, no. 4, pp. 545-564, 2015.

Xiaorong Cheng Ph.D. Professor at School of control and computer engineering, North China Electric Power University. Hers research interest covers Computer network application and Network information security. North China Electric Power University, Hebei Province.

Tianqi LI engineer, master, mostly focusing on the design and research of WAMS advanced application system, computer network.