Detection of behavior violating Ethics Rules in computer networks

Yang Fan

Ministry of Social Science Teaching, Hebei Finance University, Baoding, 071051, China

Keywords: ethics rules; similarity estimation; computer network

Abstract: Most of the human destruction can be attributed to the ethical rules sabotage. Existing methods are mainly conduct to property calculation or structure measure of the malicious program code, lack of weighing from ethics rules angle. This paper proposes a network threaten detection algorithm to measure how much a behavior violating the ethics rules. It weighs the malicious behaviors through improved ethics rules and comprehensively estimates the structure similarity of intentional programs. The simulation results show that the proposed algorithm can make a more accurate estimation for the intentional similarity thus measuring the behavior violates ethic rules.

1 Introduction

The similarity measurement for violating ethic rules is to detect and access the malicious code intends to damage network ethic rules. It is well applied in preventing software piracy and protecting intellectual property [1~3]. The difficulty in this similarity measurement is how to deal with the flexibility of malicious code. However, no matter how the code changes, the destructive nature to ethic rules never change[4-5]. The similarity measure mainly includes property calculation and structure assessment where the property calculation refers to extract the corresponding characters, calculate the matrix used to form a spatial vector for the similarity measurement. This paper proposes a network threaten detection algorithm to measure how much a behavior violating the ethics rules. It weighs the malicious behaviors through improved ethics rules and comprehensively estimates the structure similarity of intentional programs. The simulation results show that the proposed algorithm can make a more accurate estimation for the intentional similarity thus measuring the behavior violates ethic rules.

2 Similarity estimation with considerations of ethical damage

In the measurement for violating ethic rules, we first normalize the malicious code in a functional domain, and then generate a Token sequence whose sub-sequence is treated as the characteristic value of the whole string which also known as the ethic rules.

The (A,B) similarity measurement can be described as following formula when one network damaged behavior is converted into one violating ethic rules: $sim(A,B)=2*\frac{|A\cap B|}{|A|+|B|}$ Wherein, $(|A|,|B|)$ is the number of ethic rules of malicious code (A,B), and $|A\cap B|$ is the common ethic rules of (A,B).

Suppose code A can be divided into L functional domains, and B divided into F functional domain. Then we get $A=\{A_1,A_2,...,A_L\}$, $B=\{B_1,B_2,...,B_F\}$. A is the source code of malicious program, and B is the comparison code. The similarity of ethic rules and damaging rules is accessed as:

$$sim(A,B)=\frac{\sum_{i=1}^{L}\sum_{j=1}^{F}|A_i\cap B_j|}{\sqrt{\sum_{i=1}^{L}|A_i|^2}+\sqrt{\sum_{j=1}^{F}|B_j|^2}}$$

A_i is the ethic rule of the ith functional domain of A_i, B_j is the ethic rule of the jth

functional domain of B, $\sum_{i=1}^{L}\sum_{j=1}^{F}\left|A_i \cap B_j\right|$ is the common ethic rules of (A,B) in range (L,F). So, $\sum_{i=1}^{L}\left|A_i\right|$ and $\sum_{j=1}^{F}\left|B_j\right|$ are respectively the total ethic rules in program A and B.

Conducting code B change to code A is a random sequence change based on structure feature of C, which can not guarantee the corresponding changes to the functional domain. The string is regarded as some sub-strings, then there are k longest common strings in (A,B), where the value of k is related to the similarity particle size threshold T.

When measuring the weight of longest common string, we use the C language to conduct pre-process, the similarity $sim(A,B)'$ is

$$sim(A,B)' = \frac{2 * MatchLen}{|A| + |B|}$$

Wherein $MatchLen = \sum_{match(i,j,length) \in Tiles} length$; $match(i,j,length)$ is the common string with length of $length$ and start position of i in A and j in B. $Tiles$ is the set of common string.

Therefore, if $\exists A, \forall B$, (A,B) is the malicious code, B stands for the malicious code which converted from source code A through C language, and there exit longest common string in (A,B), then the similarity of (A,B) is $Sim(A,B) = \alpha * sim(A,B) + (1-\alpha) * sim(A,B)' = \frac{\alpha * \sum_{i=1}^{L}\sum_{j=1}^{F}\left|A_i \cap B_j\right|}{\sqrt{\sum_{i=1}^{L}|A_i|^2} + \sqrt{\sum_{j=1}^{F}|B_j|^2}} + \frac{2(1-\alpha) * MatchLen}{|A|+|B|}$ wherein α is the adjustment factor comes from the weights of two similarity methods.

The value of $Sim(A,B)$ reflects the similarity level of (A,B), whose threshold is corresponding Quantitative Analysis. Large threshold results in less similar code, otherwise, small threshold results in more similar code. Therefore, the threshold is a significant in ethic rules similarity measurement and need to be chosen properly.

3 Simulation

In the testing, we choose dichotomy algorithm prog.c as the source, and then modify the source code many times according to the malicious modification method presented in literature [2], thus obtaining the test set with sample modified by different extent. Suppose K is the smallest particle size generating Token string, w is the window scale of fingerprint extraction in literature [3]. Let $K = 5$ $w = 4$, we get the best result.

We compare the proposed algorithm with the ones in literature [4] and [5] to get measure effectiveness and efficiency. The code is modified 5 times forming 5 groups. We calculate the similarity between each sample and the source code prog.c, and then find the average, which shown in figure 1

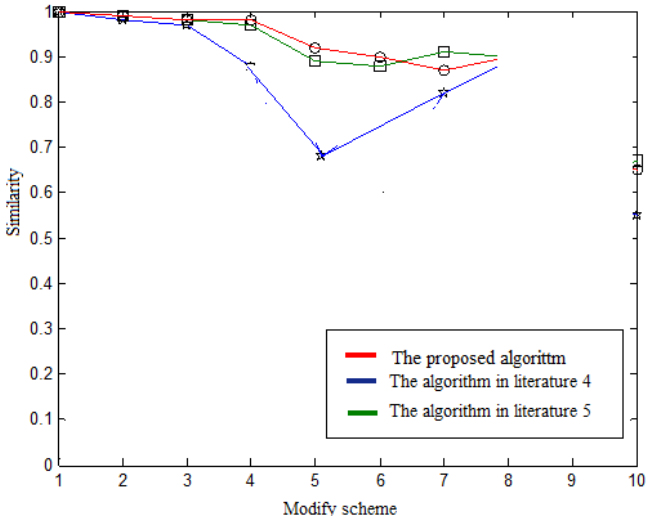

Figure 1 The similarity calculated by three algorithms

It can be inferred from the result that the proposed algorithm has a better effect to calculate the similarity of $\{4,5,6,7,9,10\}$ samples than the traditional method that is mainly because the proposed algorithm realize the calculation based on the similarity weight of ethic rules which makes the detection easier.

4 Conclusions

Based on studying the similarity of the malicious programs, this paper proposes a network threaten detection algorithm to measure how much a behavior violating the ethics rules. It weighs the malicious behaviors through improved ethics rules and comprehensively estimates the structure similarity of intentional programs. The simulation results show that the proposed algorithm can make a more accurate estimation for the intentional similarity thus measuring the behavior violates ethic rules.

Reference

[1] Baker B S, Giancarlo R. Sparse Dynamic Programming for Longest Common Subsequence from Fragments[J]Journal of Algorithms, 2002, 42(2): 231-254.

[2] Schleimer S, Wilkerson D S, Aiken A. Winnowing: Local Algorithms for Document Fingerprinting[C]//Proc. of ACM SIGMOD International Conference on Management of Data. San Diego, California, USA: [s. n.], 2003.16-18

[3] Karp R M, Rabin M O. Efficient Randomized Pattern-matching Algorithms[J]IBM Journal of Research and Development, 1987,31(2): 249-260.

[4] Andrew Granville. Detecting plagiarism in Java code[D]Supervisor : Yorick Wilks,2002.(2)56-58

[5] Alex Aiken, Saul Schleimer, Daniel S Wilkerson. Winnowing: Local algorithms for document fingerprinting[C]Proceedings of the ACM SIGMOD International Conference on Management, 2003.45-46