A Fuzzy Comprehensive Evaluation Model for Harms of Computer Virus

Cong Zheng Lansheng Han* Jihang Ye Mengsong Zou Qiwen Liu
Laboratory for Information Security, School of Computer Science and Technology
Huazhong University of Science and Technology, Wuhan 430074, China
hanlansheng@hotmail.com, hanlansheng@mail.hust.edu.cn

Abstract

The variety and complexity of virus harm cause there is few practical evaluation methods currently. The paper first presents harms' definition and classification of the virus. Then representative first-level and second-level evaluation indexes are proposed. However, as these evaluation factors could hardly be assigned with definite values and somewhat have fuzziness, the paper constructs a fuzzy evaluation model for the harm of computer virus. Once an obscure value is assigned to each evaluation indexes in the second level, the model can calculate the membership degree to each of the five harm grades, and then obtains a reasonable harm evaluation result. Finally, the paper evaluates the harms of "Worm.WhBoy.h", "AutoRun", "JS.Yamanner.a" and "An-ti-virus" by the fuzzy model; the evaluation result is similar to the statistics of the virus.

1. Introduction

Currently, the harm of computer virus attracts attentions from both software companies and academic institutes [1]. Anti-virus companies and general users have to evaluate the harm of computer virus so that they could take corresponding measures for protection in advance [2]. Precise evaluation for the harm of computer virus is important not only in defending but also in utilizing them [3]. However, due to the variety and the complexity of harms, practical evaluation method for virus' harm has not been established yet [4].

Some famous anti-virus companies respectively applied their own evaluation systems for virus harm; these systems always focus on the direct evidence for harms from specific vicious code of the virus. But harms of the virus are closely related with specific environment of the jeopardized objects and even virus' developers can't make the right predictions. Thus, when a virus floods on the internet, unity action is seldom taken in fighting against the virus.

978-1-4244-5113-5/09/$25.00 ©2009 IEEE

As most computer viruses are long-lasting, it's difficult to make precise prediction about follow-up destructions. In order to get free from the time restrictions, harm evaluation in more thorough aspects is necessary. For instance, virus transmission ways and virus complexity would altogether affect the lifespan of the virus; those influences about destructions can't be ignored [5]. However the harm of computer virus involves many other related factors that reflect different characteristics of the virus. And those factors have fuzziness to some extent since different users' responses to those factors would vary. Thus, in this article, a comprehensive fuzzy model is proposed to evaluate the harms of computer virus.

2. Definition and classification of harms of computer virus

At present, people are familiar with the harms of computer virus but there is no complete definition for it. Harms of computer virus would be presented in different levels and evaluation from different aspect would cause different results. To make accurate and objective evaluation of harms more thoroughly, it's necessary to define the space of the virus' harm and construct its' basic vectors precisely.

2.1. Harms space of computer virus

According to recent studies on computer virus and the related references, the paper describes the harm of computer virus with a four-dimension space: $U=\{U_1,U_2,U_3,U_4\}$, where U_1=Current Infection Scale, U_2=Way of Transmission, U_3=Destructive Behavior, U_4=Self-Complexity.

U_1 (*Current Infection Scale*) — is a direct description about the harm scale caused by the virus. It can be specifically expressed by following indexes: $U_1=\{U_{11},U_{12},U_{13},U_{14}\}$,where U_{11}=Infected Independent Sites, U_{12}=General Infected Computers, U_{13}=Infected Areas, U_{14}=Infected Industries.

U_2 (*Way of Transmission*) — is the index representing how the virus spread its harm. The more transmission ways of the virus would bring more harm to the network users. We collects the most transmitting ways of the virus: U_2={U_{21},U_{22},U_{23},U_{24},U_{25},U_{26},U_{27}, U_{28}}, where U_{21}=File, U_{22}=E-mail, U_{23}=LAN, U_{24}=Internet, U_{25}=System Vulnerability, U_{26}=System Configuration Defect, U_{27}=Social Engineering, U_{28}=Operation Free from Person.

U_3 (*Destructive Behavior*) — is direct evident index to evaluate the virus' harms and also the primary index of harm statistics; the paper collect almost all the behaviors of the current viruses: U_3 = {U_{31}, U_{32}, U_{33}, U_{34}, U_{35}, U_{36}, U_{37}, U_{38}, U_{39}}, where U_{31}=Delete/Modify Files, U_{32}=Trigger Event, U_{33}=Network Blockage, U_{34}=Sustainable Active Dissemination via Network, U_{35}=System Paralysis, U_{36}=Access to Sensitive Information, U_{37}=Reduction in Network Performance, U_{38}=Modify System Configuration, U_{39}=Further Active Attacks.

U_4 (*Self-Complexity*) — is the basic mechanism to display the harms. The more complexity of the virus would also cause more variety harms. Specific indexes about the complexity collected by the paper are U_4 = {U_{41}, U_{42}, U_{43}, U_{44}}, where U_{41}=Computer's Inability to get Latest Patch, U_{42}=Utilizing new Attack Method, U_{43}=Compound Mode of Transmission, U_{44}=Anti-clear.

2.2. Definition of computer virus harm

Definition1: computer virus' harm notates the sum of indexes representing the outbreak scale or impact scope, ways of transmission, destructive behaviors and self-complexity of the virus.

2.3. Classification of harm grade

The paper adopts five-grade classification which is identical with existed classification of harms of computer virus announced by National Computer Virus Emergence Response Center.

Grade 5(*Devastating*) — harms that can cause large scope of blockade in public or special network, it would disenable computers to make data transmission and exchange or cause computer systems suffer large-scope attacks which might lead to system paralysis or information leakage.

Grade 4(*Great*) — harms of computer virus that can transmit via network, cause parts of the network or system suffer paralysis; it would also have potential access to information leakage.

Grade 3(*Serious*) — harms of computer virus that seriously affects normal function of the computer system or cause data loss.

Grade 2(*Ordinary*) — harms of computer virus that affects normal function of computer system or modify its original functions.

Grade 1(*Slight*) — harms of computer virus don't directly destroy network or system, only had the capacity of transmitting and rebirth itself.

3. Comprehensive fuzzy evaluation model together with AHP theory

This section introduces the main procedures of the fuzzy evaluation. But different from the traditional fuzzy theory, the paper adopts Analytical Hierarchy Process (AHP) to get the weight value for the evaluation factors instead of large scale of statistic samples of the fuzzy theory that can save lots of time and space.

3.1. Single-level comprehensive evaluation model

Comprehensive evaluation models can be divided into single-level evaluation model and the multilevel one [6]. Application of single-level model for comprehensive evaluation can be generalized as following procedures [7].

(1) Build the set of factor for the evaluation

It is presumed that factor set $U = \{u_1, u_2, \cdots, u_m\}$ is made up of various factors for the evaluated; $u_i (i = 1, 2, ..., m)$ notates each specific factor for the evaluated and also notates the evaluated object's various attributes which are primary elements in determining evaluation results.

(2) Build the reviews set

We assume that reviews set $V = \{v_1, v_2, \cdots, v_n\}$ contains evaluation levels or grades; $v_i (i = 1, 2, ..., n)$ notates each specific level that is employed for evaluation, it could be fuzzy or non-fuzzy, but they all have definite relationship with reviews set V.

(3) Build the fuzzy evaluation matrix for single factor

To give evaluation from one specific factor set element u_i to ascertain membership degree r_{ij} of the evaluated object to its reviews set element v_j can be appointed as fuzzy evaluation with single factor. Evaluation results R_i for the i[th] factor u_i are named fuzzy evaluation set with single factor.

Similarly to get corresponding fuzzy evaluation set with single factor for every element in factors set and establish fuzzy evaluation matrix from ***U*** to ***V***.

$$R=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix}$$

Thus, (U,V,R) constitutes a comprehensive evaluation model.

(4) Set the weight for each evaluation factor by AHP

Each evaluation factor should be distributed a specific weight value. Traditional fuzzy way to get those weight values is by the large scale of statistic samples, which take lot of spaces and cost lots of time. For the reasons, the paper introduces Analytical Hierarchy Process (AHP) to get the weight value [8]. The AHP's greatest power is that it can compare any two indexed harms from any two different types and at the same time keep the consistency [9].

By the AHP theory, comparison matrix $A=(a_{ij})_{n\times n}$ meets two conditions: （i） $a_{ij}>0$, （ii） $a_{ji}=1/a_{ij}$ （$i,j=1,2,\cdots,n$）, where the term of i[th] row and j[th] column gives the relatives importance of i[th] element compared with j[th] element. In the judgment matrix ***A***, the value of every term is determined by 1-9 scales listing in Table 1.

Table 1. 1-9 Ratio scale values of the comparison

a_{ij}=1	The two elements are equal in importance
a_{ij}=3	*i* is weakly more important than *j*
a_{ij}=5	*i* is strongly more important than *j*
a_{ij}=7	*i* is very strongly more important than *j*
a_{ij}=9	*i* is absolutely more important than *j*
a_{ij}=2,4,6, 8	Lie in the medium between the adjacent values

According to importance of each level's indexes to final review, we can get pairwise comparison matrix ***A*** through mutual comparisons. Figure out matrix ***A***'s maximum eigenvalue $\lambda_{\max}$ and its corresponding eigenvector W.

Then make consistency assay about previous pairwise comparison matrix ***A*** in following steps:

① Assume $CI=\dfrac{\lambda_{\max}-n}{n-1}$ with ***n*** notates the order of matrix ***A*** and give reference parameters ***RI*** for coincidence assay.

Table 2. Random consistency index RI values

N	1	2	3	4	5	6	7
RI	0	0	0.52	0.89	1.12	1.26	1.36
N	8	9	10	11	12	13	14
RI	1.41	1.46	1.49	1.52	1.54	1.56	1.59

② Then assume $CR=\dfrac{CI}{RI}$, and consistency of matrix ***A*** can be accepted when $CR<0.1$.

If the matrix ***A*** can be accepted, the W can be the effective weight value coefficients matrix expressed as:

$$W=[\mu_1,\mu_2,\cdots,\mu_m],\sum_{i=1}^{m}\mu_i=1 \qquad (1)$$

(5) Evaluate algorithm

Through fuzzy transformation of weight value coefficients matrix and fuzzy evaluation matrix, we can get fuzzy evaluation set ***S*** by the following formula.

$$S=W\circ R=[\mu_1,\mu_2,\cdots,\mu_m]\circ\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix}$$

$$=[s_1,s_2,\cdots,s_n] \qquad (2)$$

Thus we get the final review about evaluated object from results matrix $[s_1,s_2,\cdots,s_n]$ according to principle of maximum membership degree.

8.2. Multilevel comprehensive evaluation model

In evaluating the harms of computer virus, it's difficult to define proper weight value distribution. So we can divide set of factors into several groups according to mutual relations among various indexes in the set. It's advisable to first evaluate each group's factor and then make high-level combination of diverse evaluation results from its specific group. Generally, we would employ multilevel comprehensive evaluation system and make evaluations in 5 steps:

(1) Divide original factors set into ***k*** subsets: $U=\bigcup_{i=1}^{k}U_i$ whereas $U_i\bigcap U_j=\phi(i\neq j)$ and assume elements in each subset as $U_i=\{u_{i1},u_{i2},\cdots,u_{im}\}\,(i=1,2,\cdots,k)$.

(2) Distribute weight value to factors in each subset U_i by AHP for U_i: $W_i=[\mu_{i1},\mu_{i2},\cdots,\mu_{im}]$

with $\sum_{j=1}^{m}\mu_{ij}=1$. After ascertaining membership degree, corresponding fuzzy evaluation matrix R_i could be got.

(3) Make evaluation for each subset U_i and get results for each first-level fuzzy comprehensive evaluation: $S_i = W_i \circ R_i$.

(4) Construct evaluation matrix as $R=[S_1,S_2,\cdots,S_k]^{-1}$. Then we obtain weight value coefficients matrix for U_i according to its significance. Thus, we get final second-level fuzzy evaluation by $S = W \circ R$.

(5) Draw the conclusion according to principle of maximum membership degree.

4. Fuzzy Evaluation Models for Computer Virus Harm

By the above analysis, the paper defines reviews set of virus harms as *V= {V1,V2,V3,V4,V5}*,where *V1=Devastating, V2=Great, V3=Serious, V4=Ordinary, V5=Slight*. Harmful factors as first-level factor and second-level factor are sorted in Table3.

Table 3. Classification of factors for virus harm

V	U_i	U_{ij}
V_1	U_1	$U_{11}, U_{12}, U_{13}, U_{14}$
V_2	U_2	$U_{21}, U_{22}, U_{23}, U_{24}, U_{25}, U_{26}, U_{27}, U_{28}$
V_3		
V_4	U_3	$U_{31}, U_{32}, U_{33}, U_{34}, U_{35}, U_{36}, U_{37}, U_{38}, U_{39}$
V_5	U_4	$U_{41}, U_{42}, U_{43}, U_{44}$

By the AHP theory, we get pairwise comparison matrix of the through mutual comparisons.

$$A=\begin{bmatrix} 1 & 2 & 2 & 3 \\ 1/2 & 1 & 1/2 & 4 \\ 1/2 & 2 & 1 & 3 \\ 1/3 & 1/4 & 1/3 & 1 \end{bmatrix}$$

Figure out matrix ***A***'s maximum eigenvalue and its corresponding eigenvector:

$$\lambda_{\max}=4.1833,\ W=[0.40,0.22,0.28,0.10]$$

Then make consistency assay about previous pairwise comparison matrix ***A***. By calculating, we get CR= 0.00687, that means matrix ***A*** can be accepted as weight vector.

Similar methods are used to get weight vectors for second-level indexes:

$$W_1=[0.2,0.4,0.2,0.2]$$

$$W_2=[0.2,0.1,0.1,0.1,0.2,0.1,0.1,0.1]$$

$$W_3=[0.1,0.05,0.1,0.1,0.2,0.2,0.05,0.1,0.1]$$

$$W_4=[0.2,0.3,0.2,0.3]$$

Figure out fuzzy evaluation matrix for each first-level index. Membership degree of each element in U_1 can be determined by following membership function (3).

$$r_{ij}=\frac{\max\limits_{1\le j\le 5}\{|x_i-u_{ij}|\}-|x_i-u_{ij}|}{\max\limits_{1\le j\le 5}\{|x_i-u_{ij}|\}} \quad (3)$$

$$(i=1,2,3,4;\ j=1,2,\cdots,5)$$

Where x_i notates the value of i[th] element in U_1 and u_{ij} notates corresponding value of level ***j*** for the i[th] element. Defined indexes of all first-level factors to different harm grades are shown from Table.4 to Table.7.

Table 4. Standard values of current infection (U_1) to harm grades (V)

U1\V	V5	V4	V3	V2	V1
U11	5	4	2	0	0
U12	500	300	200	100	10
U13	5	3	1	1	1
U14	5	3	1	0	0

Table 5. Standard values of way of transmission (U2) to harm grades (V)

U2\V	V5	V4	V3	V2	V1
U21	IR	IR	Y	IR	IR
U22	IR	Y	IR	N	N
U23	Y	Y	IR	N	N
U24	Y	IR	IR	N	N
U25	Y	IR	IR	N	N
U26	IR	Y	N	N	N
U27	IR	Y	Y	IR	N
U28	Y	N	N	N	N

Table 6. Standard values of destructive behavior (U3) to harm grades (V)

U3\V	V5	V4	V3	V2	V1
U31	IR	IR	Y	N	N
U32	IR	IR	Y	IR	N
U33	Y	IR	N	N	N
U34	Y	IR	N	N	N

U35	IR	N	IR	N	N
U36	IR	Y	N	N	N
U37	IR	Y	Y	N	IR
U38	IR	Y	IR	IR	IR
U39	IR	Y	N	N	N

Table 7. Standard values of self-complexity (U4) to harm grades (V)

U4\V	V5	V4	V3	V2	V1
U41	Y	N	N	N	N
U42	Y	N	N	N	N
U43	Y	Y	N	N	N
U44	Y	Y	Y	N	N

"IR" "Y" and "N" in above tables respectively denotes "Irrelative", "Yes" and "No". Membership for those specific second-level indexes of corresponding harm degrees are shown in Table8. If second-level index value is identical with standard value in some certain corresponding grade, its membership degree to this grade would be larger than any other grade. On the other hand, we can assign the membership degree of "IR" to its corresponding grade as 0.5 since virus would have no direct impact on second-level indexes. More specifically, if the virus corresponding second-level index is "Y", membership degrees of "Y", "N" or "IR" are 0.8, 0.2, 0.5 in accordance with respective grade; if the virus corresponding second-level index is "N", membership degrees of "Y", "N" or "IR" are 0.2, 0.8, 0.5 in accordance with respective grade; if the virus corresponding second-level index is "IR", membership degrees of "Y", "N" or "IR" are 0.2, 0.2, 0.8 in accordance with respective grade shown in Table 8.

Table 8. Values about membership degree of second-level index to its harm grade

Standard Value for Each Grade / Second-level Index Value Degree / Membership	Y	N	IR
Y	0.8	0.2	0.5
N	0.2	0.8	0.5
IR	0.2	0.2	0.8

Finally, according to the above analysis, we choose operator $M(\bullet,\oplus)$ for fuzzy computation with definition of operator $M(\bullet,\oplus)$ as:

$$s_k = \min\left(1, \sum_{j=1}^{m} \mu_j r_{jk}\right), \quad k = 1, 2, \cdots, n$$

We can obtain this level's evaluation results and then make second-level fuzzy comprehensive evaluation about those results, thus we get normalized results. Finally the evaluation results can be figured out according to the principle of maximum membership degree.

5. Evaluation example for the fuzzy model

To test the fuzzy evaluation model, the paper selects "Worm.WhBoy.h" [10] and the other three well-known viruses, which are "AutoRun", "JS.Yamanner.a" and "An-ti-virus" as examples. The evaluation results are listed in Table 9 and their curve are shown in Figure 1.

Table 9. Four viruses' evaluation result

Virus / V	Worm.WhBoy.h	AutoRun	JS.Yamanner.a	An-ti-virus
V_1	0.1110	0.1485	0.2316	0.1358
V_2	0.1150	0.1537	0.2352	0.1411
V_3	0.1601	0.1588	0.1871	0.1614
V_4	0.2860	0.2400	0.2083	0.2740
V_5	0.3278	0.2990	0.1377	0.2877
result	V_5	V_5	V_2	V_5

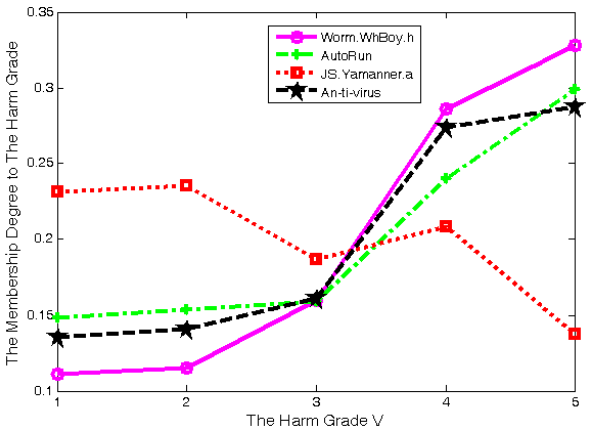

Figure 1. Four viruses' membership degree to the harm grade

From the Table 9, we can conclude that the indexes in U_1 are more important than the other indexes for the final evaluation result. The AutoRun doesn't cause serious damage to the P.C like Worm.WhBoy.h and An-ti-virus but a little slight damage. It is also can be cleaned easily. However, the AutoRun with strong infectious has infected enormous numbers of computers and is the most common virus around us. In addition, many viruses are derived from the AutoRun and make use of its automatic execution technique to spread. That is the reason why it is evaluated as devastating virus.

From the Figure 1, we find that the Worm.WhBoy.h, AutoRun and An-ti-virus's curves increasing monotonically have only one maximum value point .And JS.Yamanner.a's curve has two maximum value point. According to the theory of fuzzy evaluation,

the evaluation result curve that has only one maximum value point is the best. So the evaluation result about JS.Yamanner.a may be inaccuracy. The reason is that perhaps some of the second indexes' data gathered by us are not precise enough.

6. Conclusion

Many harmful factors should be taken into consideration in evaluating harms of computer virus; they can be represented from many aspects and even virus-designers can't predict the harmful results. As general evaluations factors always have somewhat fuzziness [11], the paper constructed a fuzzy evaluation model for virus' harm. Different from the traditional fuzzy theory, the paper applied the AHP to get the weight value instead of large scale of statistics.

The fuzzy evaluation model introduced several representative first-level indexes for virus harm evaluation and got the concrete derivative second-level indexes. From the second-level indexes and by the two steps of fuzzy evaluations, the paper got the final evaluation result for the specific virus. In calculating membership degrees, the paper designed proper membership function and set membership degrees value by the characteristics of virus in each grade.

Finally, the paper selected some specific viruses to test our fuzzy evaluation model and the evaluation results are largely identical with statistics data from authoritative organizations. The fuzzy evaluation model proposed a practical innovative way for the harm evaluation.

7. Acknowledgments

The paper is directly supported by both National Science Foundation of China No: 60703048 and Hubei Province Natural Science Foundation NO: 2007ABA313, NO: 2008CDB352. We would like to thank the supports of National and Hubei Province Science Foundation.

8. References

[1] Stefan Helmreich , Flexible Infections: Computer Viruses, Human Bodies, Nation- States, Evolutionary Capitalism, Science, Technology, & Human Values, Vol. 25 No. 4, Autumn 2000 472-491

[2] V. Skormin, A. Volynkin, D. Summerville, J. Moronski, "Prevention of Information Attacks by Run-Time Detection of Self-replication in Computer Codes" Journal of Computer Security, Volume 15,Number 2/2007

[3] V. Skormin, A. Volynkin, D. Summerville, J. Moronski, "Run-time Detection of Malicious Self-replication in Binary Executables", Information Assurance Workshop, 2006IEEE 21-23 June 2006

[4] Guillaume Bonfante, Matthieu Kaczmarek, and Jean - Yves Marion. A Classification of Viruses through Recursion Theorems. In CIE 2007, volume 4497 of Lecture Notes in Computer Science, Springer -Verlag Berlin Heidelberg, 2007.

[5] Jian Zhang, Hong Liang, Jianming Chen, etc. The evaluation of harms of computer virus. Netinfomation Security, 2005:39-41. New York University

[6] Li Xiao-quan,Ren Jian-bo,Zeng Jiang-dong. Fault Evaluation of Certain Power Vehicle Based on Fuzzy Comprehensive Assessement : The Missile Inst. of Air Force Engineering University ,2008.

[7] Hong Yao, Li-Yin Shen,Jianli Hao and Chi-ho Michael Yam. A fuzzy-analysis -based method for measuring contractors' environmental performance : Management of Environmental Quality:An Inernational Journal,Vol. 18 No.4,2007,pp:442-458.

[8] Saaty, T.L., The Analytic Hierarchy Process, (New York: McGraw Hill, 1980), pp. 78-121.

[9] Saaty, T. L. 1980. The Analytic Hierarchy Process, McGraw Hill, New York., pg xi.

[10] "Worm.WhBoy.h", http://baike.baidu.com/view/697258.htm

[11] Skormin, A. Volynkin, D. Summerville, J. Moronski, "Prevention of Information Attacks by Run-Time Detection of Self-replication in Computer Codes" Journal of Computer Security, Volume 15,Number 2/200.