文章编号:1671-4598(2022)04-0165-07 DOI:10.16526/j.cnki.11—4762/tp.2022.04.028 中图分类号:TP274 文献标识码:A

基于校验和法比较法与夹角余弦公式的变形病毒检测算法

朱俚治
(南京航空航天大学 信息化处,南京 210016)

摘要:长度比较法、校验和法以及基于行为的检测算法是3种经典的病毒检测算法,因此将这3种算法相互结合而提出一种新的病毒检测算法,该算法的思路是:首先通过相应的算法检测某个程序的校验和与程序的长度是否发生了变化;如果发生了变化,则采用计算机病毒代码权值计算公式,判断该程序是否被未知病毒感染了;如果成了未知病毒的宿主,则在虚拟机中将该代码进行运行,判断未知病毒的功能属性,同时采用夹角余弦公式对未知病毒进行了相似性计算,根据检测算法来判断该未知病毒属于那种类的病毒,从而达到对计算机未知病毒检测的目的。

关键词:校验和;权值;病毒;夹角余弦;比较法

Deformation Virus Detection Algorithm Based on Checksum Comparison Method and Included Angle Cosine Formula

ZHU Lizhi
(Information Office, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: Length comparison method, checksum method and behavior-based detection algorithm are three classic virus detection algorithms. Therefore, a new virus detection algorithm is proposed by combining three algorithms. The idea of algorithm first detect whether the program checksumand the programlength have changed. If there is any change, the weightcalculation formula for computer virus codeis used to determine whether the program is infected by an unknown virus. If it becomes the host of an unknown virus, the virtual machinecodeis runto determine the functional attributes forthe unknown virus. At the same time, the angle cosine formula is used to calculate the similarity of unknown virus, and the detection algorithm is used to determine whether the unknown virus belongs to whichkind of virus, thus the purpose of detecting unknown viruses on the computer is achieved.

Keywords: checksum; weight; virus; similarity; comparison method

0 引言

计算机病毒在互联网中肆意地传播,给互联网造成了重大破坏,严重威胁着网络正常运行。由于计算机新技术不断地出现,从而有了多态病毒,加壳病毒和变形病毒。文献[1]指出多态病毒,加壳病毒,变形病毒是不同种类的病毒,其中变形病毒对网络的危害最大,也是最难查杀的病毒。

根据文献[2]可知病毒主要的感染对象是.com,.exe,.ovl等可执行文件,文献[3-4]指出计算机病毒可以通过电子邮件,网络文件和网页传播,还可以通过跨站脚本漏洞,以及跨站请求伪造攻击等这些渠道传播。文献[3]还指出新型病毒还可以通过无线局域传播等。

目前计算机病毒的检测算法主要有特征码,特征字,校验和法,启发式扫描等[5-7],长度检测法,病毒标记检测法,校验和法,行为检测法和感染实验法是常用的病毒检测算法,长度检测法和校验和法只能判断某个程序可能感染了病毒,不能确定该程序是否确实感染了病毒[4,8]。

文献[5]指出特征码,校验和法,特征字检测不能有效应对多态病毒和变形病毒。启发式病毒扫描技术的不足之处在于准确度较差,错误率较高,检测效果较差等[5]。目前基于行为的病毒检测算法可以有效地应对变形病毒,多态病毒[4]。然而目前病毒变种速度快,传播速度快等特点,因此有必要对基于行为病毒的检测算法进行改进[4]。文献[2-5]还指出当今的病毒检测算法需要智能化,根据文献[14]可以知道智能光谱扫描检测技术是一种新的病毒检测技术。

1 当今计算机病毒检测技术与变形病毒

1.1 变形病毒

根据文献[8-10,12-13]可以知道,一维变形病毒,二维变形病毒,三维变形病毒和四维变形病毒具有如下特征:

1)一维变形病毒的特征是:一维变形病毒进行传染时,通过病毒样本分析,可以知道前一个病毒的代码与后一个病毒的代码几乎没有连续的3个字节是相同的,然而字节的相对位置也没有发生变化。

收稿日期:2021-08-22; **修回日期**:2021-10-06。
作者简介:朱俚治(1980-),男,江苏宜兴人,大学本科,工程师,主要从事计算机技术和网络安全方向的研究。
引用格式:朱俚治.基于校验和法比较法与夹角余弦公式的变形病毒检测算法[J].计算机测量与控制,2022,30(4):165-171.

2）二维变形病毒的特征是：二维变形病毒具有一维变形病毒的特征，二维变形病毒进行复制时能够改变字节与字节的相对位置，变化代码之间的相对位置的变化也是相对变化。

3）三维变形病毒的特征是：三维病毒具有二维变形病毒的特征，三维变形病毒除了能对字节的本身，字节的相对位置进行变化外，它具有最大的特征就是能将整个病毒代码进行分散隐藏，将分解的代码隐藏到不同的文件中，也可以在引导扇区的不同位置，当病毒被激后把分散的病毒代码重新成一个完整的病毒。

4）四维变形病毒的特征是：四维变形病毒具有一维，二维和三维变形病毒的特征，四维变形病毒最大的特点就是病毒的代码结构能随着时间的变化而变化，四维变形病毒不进行复制时也能使代码处于随机变化中。

1.2 基于比较法与校验和病毒检测算法

基于长度比较法，校验和法是病毒经典的检测算法，但这两种方法只能判断某个程序十分有可能被病毒感染了，但不能确定该程序是否真的被病毒感染了[9-10]，这是这两种检测算法的不足之处。

比较法和校验和法是两种不同的算法，比较法有基于长度的比较法，基于内容的比较法和基于中断的比较法[4,8]。基于程序外观的比较法就是一种基于长度的比较法，基于长度的比较法，也就是对程序体大小进行检测，校验和法中的校验和值是程序体中数据项的和，是一种基于程序内容的检测算法[4,8]。

一维变形病毒，二维变形病毒，三维变形病毒和四维变形病毒采用了不同的变形机制，不同的变形原理，因此构成 4 种不同的变形病毒，这也就代表着 4 种变形病毒的校验和值不同，病毒校验和值的不同将导致病毒体长度大小的不同，但如果是变形病毒感染文件后，存在原位替换的变形方式，被感染的宿主程序长度有可能不发生变化，然而比较法无法检测该病毒，因此这时需要校验和法来检测出病毒。

基于变形病毒的特点，单独采用的比较法，校验和法是不理想的。为了有效地应对变形病毒，本文将比较法，校验和法与基于行为特征的病毒检测算法结合在一起对变形病毒进行检测，从而提出一种新的检测算法。

2 校验和法，比较法与变形病毒的检测

2.1 病毒寄生与校验和值

病毒写入其它程序后就能改变宿主体中的数据，宿主的数据发生了变化，那么程序的校验和值也发生了变化。变形病毒是一种特殊的程序，能不断地复制，每传染一次都改变一次代码中的数据，同时也改自身的校验和值，校验和值发生了改变，这就意味对该程序进行了写操作，因此变形病毒的复制，传染时是一个写的过程，是个随机动态变化的过程。

校验和值是程序体中数据项的和[4,8]，因此当某个程序只读不写，程序的数据不会发生变化的，程序的校验和值也不会发生变化。然而病毒对宿主程序进行感染时，就是写入的过程，因此当病毒寄生程序后，就能改变宿主的数据，也能使原程序的校验和值发生变化。

2.2 宿主副本与其校验和值

对当某个程序仅仅只是复制，不进行写操作，程序的校验和值不会发生变化。具有变形能力的病毒，每复制一次就是一次写的过程[4,8]，复制时会改变一次代码的结构，产生一个新病毒，此时变形病毒的校验和值也发生了改变。然而不具备变形能力的病毒在复制时，校验和值则不发生变化。

宿主程序进行子自动，自我复制时，宿主体内的病毒也进行了相应的复制，如果宿主程序复制了自身的副本，并且副本的校验和值发生了变化，其原因是该程序被变形病毒感染了，因此在本文提出的算法中，比较运行前后宿主的校验和值，来发现某个程序是否有可能被变形病毒感染了。

如果宿主感染的病毒是非变形病毒，那么当宿主进行自身复制时，被病毒感染的宿主的校验和值不会发生变化。

2.3 由校验和，比较法而提出的检测算法

当病毒寄生于宿主后，就改变了宿主的校验和值，宿主的校验和值发生了改变是由体内病毒寄生而引起的。然而病毒寄生于宿主后，与宿主就融为了一体，因此此时宿主校验和值并不是病毒体本身校验和值，而是病毒与宿主校验和的总值，同时这就给计算宿主体内病毒的校验和值造成了困难。

当病毒寄生于宿主后也改变了原程序的长度[4,8]，原程序的大小发生了变化。因此当宿主程序进行复制时，宿主体中的变形病毒随着宿主的复制而复制，并且随着宿主的复制而变形，宿主程序每复制一次都将改一次宿主程序的大小，但如果是变形病毒感染文件后，存在原位替换的变形方式，被感染的宿主程序长度有可能不发生变化，然而比较法无法检测该病毒，因此这时需要校验和法来检测出病毒。而变形病毒每复制一次，也都将改变一次病毒体的代码结构[8-12]，此时变形病毒的校验和值发生了变化的，同时变形病毒体的大小也发生了变化，变形病毒进行复制时引起病毒体大小的变化，必然引起宿主体大小的变化。

文件的最小单位是字节，在本文提出的算法中采用字节作为计算宿主以及病毒体大小的基本单位。因此在本文中将原始文件的大小保存在数据库中，由于文件只是进行了复制，没有写操作，那么在检测过程中将原始文件的大小与宿主的大小进行比较，就能得出文件大小的变化值，这个变化值就是寄生于宿主体内病毒的大小。本文将未知病毒大小与样本病毒体大小的值进行对比，从而检测寄生于宿主中的病毒。

在本文提出的算法中，如果通过比较法检测出宿主，副本的大小发生了变化，则采用下一步，通过计算机病毒权值计算公式来计算该文件是否被病毒感染了。

3 计算机病毒代码权值计算公式与病毒的检测

3.1 计算机病毒与权值计算公式

传染性是计算机病毒的重要特征[2-7,14]，如果某个程序

具备了复制，传播和寄生这 3 个属性，那么就可以判断程序是病毒[2-5,14]，根据文献[2－5，14]可以知道：传染性是计算机病毒的重要属性。

计算机病毒的传染性可以分为两个过程：复制过程、寄生过程。

本文中采用计算机病毒的复制，寄生这两种功能作为检测病毒的重要标准，其中寄生为病毒最显著的特征，因此权重值为最大的值，传播性是所有恶意代码都具有的属性，作为次重要特征，权重值的次弱。其次复制特征，作为计算机病毒最次要的特征属性，将其权重值设为最小值。

在文献［15］中提到了恶意代码权重的计算公式，而由本文也提出了计算机病毒代码权重计算公式，该公式为：

$$g(x) = \alpha g_i(x) + \beta g_j(x) + \gamma g_m(x)$$

g_i（x），g_j（x），g_m（x），分别表示病毒：复制，传播和寄生 3 种行为属性，$\alpha\beta\gamma$ 则分别表示病毒：复制，传播，寄生 3 种属性的权重值。

3.2　计算机病毒判断公式与宿主程序

综上所述，采用计算机病毒代码权值计算公式进行病毒行为检测时，首先判断一个程序是否具有复制特性，如果具有了复制性，则继续判断该程序是否具有传播性和寄生性。

当某个程序被病毒感染了，病毒就寄生于宿主体内，与宿主融为一体。当宿主体内的病毒没有发作时，则该宿主与善意的程序没有差别，然而当宿主体内的病毒发作时，该宿主就具有了恶意性，因此当某个程序具有了复制，传播和寄生的 3 种特性后，就可以判断该程序为病毒或该程序被病毒感染了。

计算机代码权值计算公式对某个程序行为属性上进行计算和判断，该过程是一种基于行为准则的病毒检测过程。经过计算机代码权值计算公式对程序进行属性计算后，就可以判断该善意程序是否成为宿主程序或该程序就是一个病毒。

如果某个程序成为宿主，则采用夹角余弦公式对宿主体内的病毒对进行相似性计算，从而完成对未知病毒的检测。

4　变形计算机病毒相似计算与检测

4.1　夹角余弦公式相关原理的描述

为了描述两个实例的相似度，通常采用相似性系数表示它们的相似性，以下采用夹角余弦公式来度量两个实例的相似度。

夹角余弦公式[16]：

$$r_{ij} = \frac{\left|\sum_{k=1}^{m} x_{ik}x_{jk}\right|}{\sqrt{(\sum_{k=1}^{m} x_{ik}^2)(\sum_{k=1}^{m} x_{jk}^2)}}$$

对于 m 维空间中的两个对象 x_i 和 x_j，r_{ij} 表示对象 j 和 i 的相似性系数，r_{ij} 则满足以下条件[22]：

1）相似性系数的绝对值不大于 1：即对任意对象 i 和 j，恒有 $|r_{ij}|\leqslant 1$；当且仅当 $x_i = x_j$ 时，$r_{ij} = 1$。

2）用两个向量的余弦为作为相似性系数，范围［－1，1］，当两个向量正交时取值为 0 表示完全不相似。

3）相似性系数的值越大，则表示这两个实例的相似程度就越强。

4.2　变形病毒分类与检测

在前面两个检测过程完成后，就能判断该程序被病毒感染了或就是病毒，如果判断该程序被病毒感染了或就是病毒，则以下采用相似算法来判断该未知病毒属于那种病毒。

4.2.1　相似性系数程度度量与变形病毒

本文选择已知变形病毒的样本作为比较的实例，未知病毒则为被比较的实例。

在夹角余弦公式中有：r_{ij} 为相似性系数，经过对夹角余弦公式研究后，有以下结论：相似性系 r_{ij} 与 1 差值为：δ。

1）如果未知病毒与样本病毒的相似性系数值有：r_{ij} 与 1 的差值 δ 为 0 时，则表明 r_{ij} 与 1 的偏差概率为百分之零，这时两个病毒相似程度最高，进行比较的这两种病毒实体最为相似，为同一种病毒。

2）如果未知病毒与样本病毒的相似性系数值有：r_{ij} 与 1 的差值 δ 小于 0.5，则表明 r_{ij} 与 1 的偏差概率小于 50％时，则表明相似性系数 r_{ij} 与 1 的值越接近，进行比较的这两种病毒实体的相似程度就越强，两者就越相似。

3）如果未知病毒与样本病毒的相似性系数值有；与 1 的差值大于 0.5，则表明与 1 的偏差概率大于 50％时，则表明相似性系数与 1 的值偏离程度越大，进行比较的这两种病毒实体时的相似程度就越弱，不相似程度就越明显。

4.2.2　相似区间与变形病毒的分类检测

1）变形病毒可以产生一维，二维，三维和四维的变形，并且二维变形病毒具有一维变形病毒的特征，三维变形病毒具有二维变形病毒的特征，四维变形病毒具有三维变形病毒的特征[11-13]，可见这四种类型的变形病毒，其代码结构是不一样的，它们的校验和值也不同。不同变形病毒体的校验和值存在差异，同时这也导致病毒体大小的差异，这里病毒体大小存在差异，随着校验和值的变化而变化，但如果是变形病毒感染文件后，存在原位替换的变形方式，被感染的宿主程序长度有可能不发生变化，及文件的长度不变，然而比较法无法检测该病毒，因此这时需要校验和法来检测出病毒。如果是一维变形的病毒，采用一维病毒变形的方式，那么未知病毒体大小值的变化应该是相似的。对于二维变形病毒，三维变形病毒和四维变形病毒依次类推可以得出类似的结论。

2）变形病毒每复制一次，程序的大小必定发生一次变化，在本文中设定的变形病毒的变化区间，需要一定数量的一维变形病毒，二维变形病，三维变形病毒和四维变形病毒的样本大小作为划分的依据，最后经过相似性计算从 4 个不同区间中，选择一个最优的区作为未知病毒的检测区间。

作为 4 个变形病毒的不同区间：这 4 个区间在于 a 的值不同而不同，a 分别表示样本变形病毒样本体的大小值，根据 a 值的进行 4 个区间的划分。

一维变形病毒，样本病毒体大小的变化范围：［0，a）∪（a，$a+g$］，a 与 g 均自然数，单位为 K。

二维变形时病毒，样本病毒体大小的变化范围：[0，a) ∪ (a，$a+g$]，a 与 g 均自然数，单位为 K。

三维变形时病毒，样本病毒体大小的变化范围：[0，a) ∪ (a，$a+g$]，a 与 g 均自然数，单位为 K。

四维变形时病毒，样本病毒体大小的变化范围：[0，a) ∪ (a，$a+g$]，a 与 g 均自然数，单位为 K。

根据夹角余弦计算公式得出的相似性系数计算，可以找到较为相似的两种病毒，根据变形病毒再选择一个最优区间，作为未知病毒的识别区间。

3）通过余弦夹角公式的计算可以得出以下的结论：

根据余弦夹角公式可以知道，当两个病毒的相似性系数值越大，则这两个病毒就越相似，

如果被检测的未知病毒与一维变形样本病毒的相似性系数的值最大，并且这两种病毒的功能差不多或完全一样，同时这两种病毒的相似功能又落在了同一变化区间内，那么该病毒为一维病毒的可能性就很大。

如果被检测的未知病毒与二维变形样本病毒的相似性系数的值最大，并且这两种病毒的功能差不多或完全一样，同时这两种病毒的相似功能又落在了同一变化区间内，那么该病毒为二维病毒可能性就很大。

如果被检测的未知病毒与三维变形样本病毒的相似性系数的值最大，并且这两种病毒的功能差不多或完全一样，这两种病毒的相似功能又落在了同一变化区间内，那么该病毒为三维病毒的可能性就很大

如果被检测的未知病毒与四维变形样本病毒的相似性系数的值最大，并且这两种病毒的功能差不多或完全一样，同时这两种病毒的相似功能又落在了同一变化区间内，那么该病毒为四维病毒的可能性就很大。

5 变形计算机病毒在虚拟机中的检测步骤

5.1 变形病毒的特征提取原理

在虚拟机中运行，根据文献[24]对比参照感染文件和感染前参照文件[24-25]，可以得出病毒片段的位置。根据病毒片段的插入位置计算相对位置，统计多个样本相对位置均值、标准差以及标准差的标准差，其中相对位置标准差的标准差体现了变形病毒变形的特征；进一步收集系统文件感染的变形特征，和不同系统时间下的变形特征，这些统计值组合成为特征值集合[24]。

一维变形样本病毒特征提取如图 1 所示。

二维变形样本病毒特征提取如图 2 所示。

三维变形样本病毒体增加系统文件感染特征，采用类似的统计方法，只是统计文件范围扩大[24]。

四维变形样本病毒体增加时间变化特征，采用类似的统计方法，只是设置虚拟机的初始条件，造成时间上差异[24]。

5.2 变形病毒检测算法的 3 个阶段

5.2.1 第一阶段：检测对象是否含有疑似病毒

1）选定某两到 3 个原始文件，计算这些原始程序的校验和值，程序的大小值。

再将校验和法检测出的结果，比较法得出值都保存在

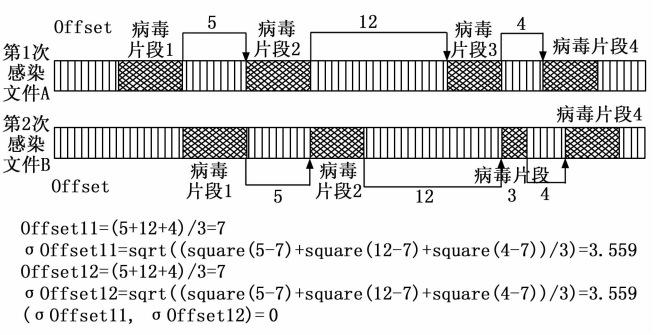

图 1 特征提取示意图

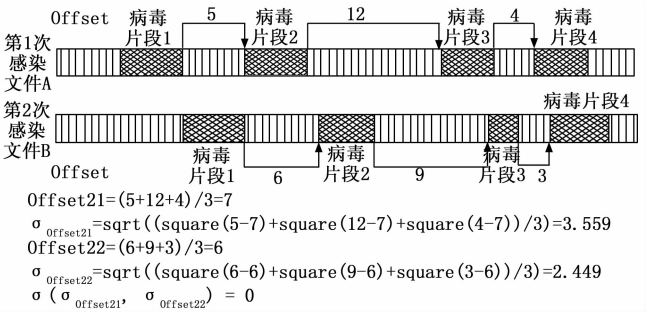

图 2 特征提取示意图

正常文件中保存作为今后比较的依据。

2）在一个星期后再将这些前面提到的过的文件，程序提出，计算这些文件，程序的校验和值。

3）如果原始序的校验和值，比较法的值发生了改变，那么该程序就可能被病毒感染了，则此再计算该程序的大小值。

4）如果程序的大小与原始程序的大小出现了差异，则下一步采用计算机行为权值计公式进行计算，判断该程序是否是病毒。

检测流程如图 3 所示。

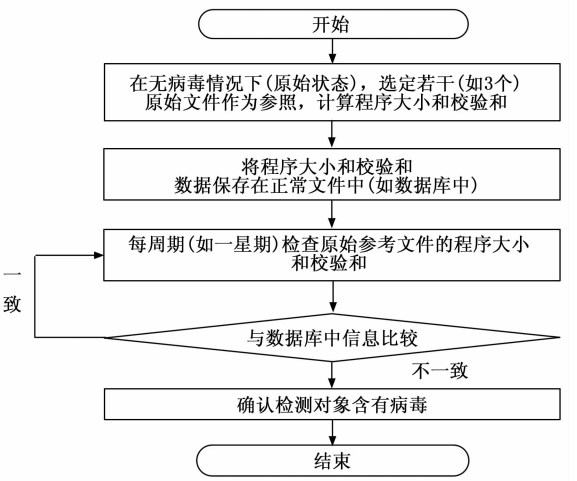

图 3 检测是否含有疑似病毒流程图

5.2.2 第二阶段：确认检测对象含有病毒：

1）在采用计算机病毒代码权值计算公式计算时，首先判断该程序是否具有复制性。

2）如果未知程序具有复制的功能，不在人工干预条件下自动产生了一个完全相同的程序，则再进一步判断该程序是否能够与其他善意的程序相互融合，如果该程序具有感染性，能够与别地程序融为一体，那么该程序就病毒。

3）最后如果个某个程序具有以上两个行为属性，又具有寄生性，就能判断该这个程序就是病毒。

5.2.3　第三阶段：变形病毒在虚拟机种的检测和运行

1）取出有两个或多个不同的已知病毒的特征值作为夹角余弦公式计算相似性系数的依据。

2）在算法中需要建立一个衡量，相似性系数强度的标准函数，在算法中采用夹角余弦公式该函数对未知病毒与已知病毒进行相似性计算。

3）虚拟机和沙箱只是一个运行环境，对被怀疑的程序进行检测需要算法。

4）先通过夹角余玄的公式，寻找到与未知病毒体大小十分相似的已知病毒体。

5）从未知病毒行为和功能上的观察和分析，确定未知病毒与已知同一病毒是否具有相同的功能，病毒机理，变形机制。

6）变形病毒能产生 4 种形式的变形，但 4 种变形机制不一样，能产生不同种类的病毒，通过未知病毒在虚拟机中的运行，观察 4 种变形病毒的各种功能，变形机制，变形机理再对这 4 种变形病毒进行功能上的判断，就判断该未知病毒是一维病毒，二维病毒，三维病毒和四维病毒。

7）根据上述几个步骤，采用夹角余玄的公式计算对未知病毒与已知病毒进行的相似性计算，就能得出未知程序和已知程序的检测结果。

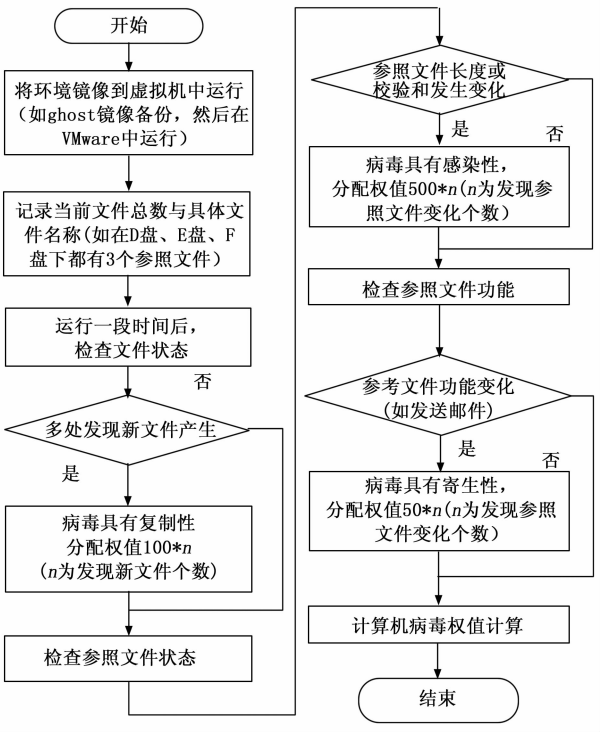

图 4　确认检测对象含有病毒流程图

8）如果两个病毒体的大小值相似性很强，十分相似，并且在病毒体现的各种功能又落在同一个区间内，那么这两种病毒，就可以是一维病毒，二维病毒，三维病毒或四维病毒。确认流程如图 5 所示。

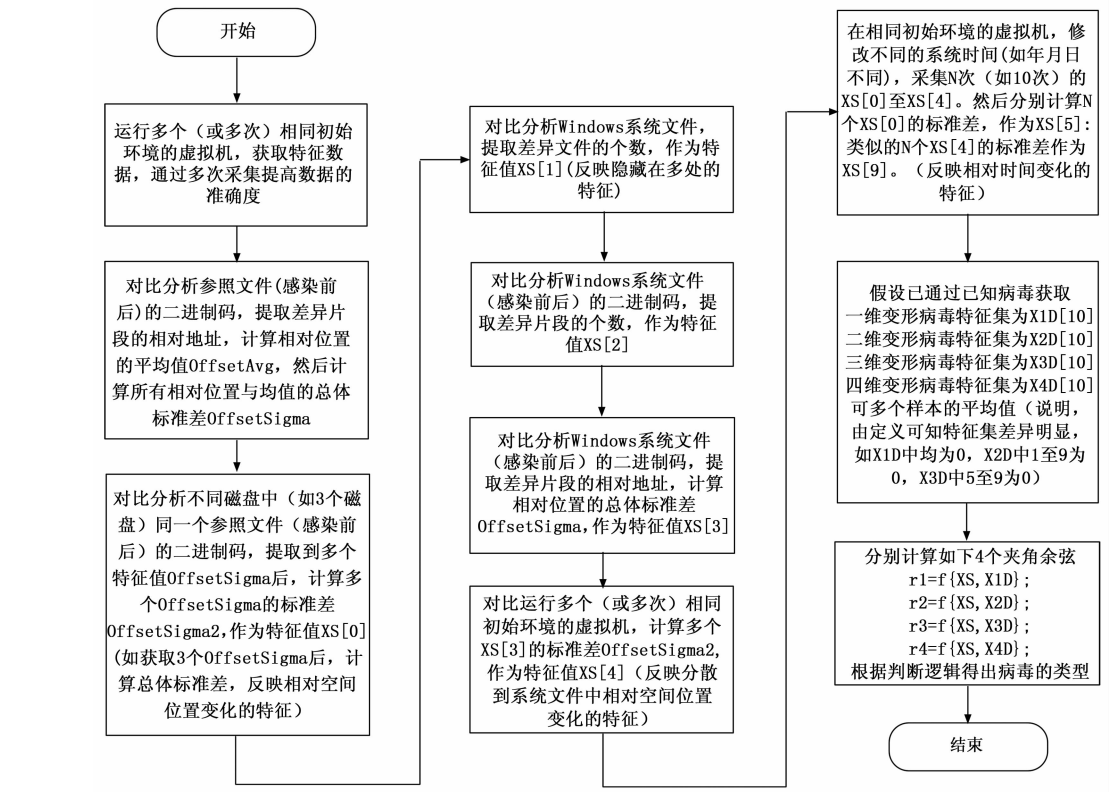

图 5　变形病毒种类确认流程图

表 1 检测结果

<table>
<tr><th>测试样本</th><th>第 1 阶段检测结果</th><th>第 2 阶段检测结果</th><th>第 3 阶段检测结果</th></tr>
<tr><td>myVirus1</td><td>QQ.exe：
感染前文件长度:76400
感染前校验和:0x5E1E26
感染后文件长度:76400
感染后校验和:0x5E206B
WeChat.exe：
感染前文件长度:568880
感染前校验和:0x3784CA7
感染后文件长度:568880
感染后校验和:0x3784EEF</td><td><table><tr><th></th><th>权重</th><th>个数</th></tr><tr><td>复制性</td><td>100.0</td><td>0</td></tr><tr><td>感染性</td><td>500.0</td><td>2</td></tr><tr><td>寄生性</td><td>50.0</td><td>2</td></tr></table>感染权值:1100</td><td><table><tr><th></th><th>相似度</th></tr><tr><td>一维</td><td>1.000</td></tr><tr><td>二维</td><td>0.000</td></tr><tr><td>三维</td><td>0.000</td></tr><tr><td>四维</td><td>0.000</td></tr></table>相似度落在一维变形病毒区间</td></tr>
<tr><td>myVirus2</td><td>QQ.exe：
感染前文件长度:76400
感染前校验和:0x5E1E26
感染后文件长度:76400
感染后校验和:0x5E2068
WeChat.exe：
感染前文件长度:568880
感染前校验和:0x3784CA7
感染后文件长度:568880
感染后校验和:0x3784EF1</td><td><table><tr><th></th><th>权重</th><th>个数</th></tr><tr><td>复制性</td><td>100.0</td><td>0</td></tr><tr><td>感染性</td><td>500.0</td><td>2</td></tr><tr><td>寄生性</td><td>50.0</td><td>2</td></tr></table>感染权值:1100</td><td><table><tr><th></th><th>相似度</th></tr><tr><td>一维</td><td>0.000</td></tr><tr><td>二维</td><td>1.000</td></tr><tr><td>三维</td><td>0.267</td></tr><tr><td>四维</td><td>0.182</td></tr></table>相似度落在二维变形病毒区间</td></tr>
<tr><td>myVirus3</td><td>QQ.exe：
感染前文件长度:76400
感染前校验和:0x5E1E26
感染后文件长度:76400
感染后校验和:0x5E2068
微信.exe：
感染前文件长度:568880
感染前校验和:0x3784CA7
感染后文件长度:568880
感染后校验和:0x3784EF4
svchost.exe
感染前文件长度:57360
感染前校验和:0x40DF41
感染后文件长度:57360
感染后校验和:0x40E29C</td><td><table><tr><th></th><th>权重</th><th>个数</th></tr><tr><td>复制性</td><td>100.0</td><td>0</td></tr><tr><td>感染性</td><td>500.0</td><td>3</td></tr><tr><td>寄生性</td><td>50.0</td><td>3</td></tr></table>感染权值:1650</td><td><table><tr><th></th><th>相似度</th></tr><tr><td>一维</td><td>0.000</td></tr><tr><td>二维</td><td>0.098</td></tr><tr><td>三维</td><td>0.551</td></tr><tr><td>四维</td><td>0.375</td></tr></table>相似度落在三维变形病毒区间</td></tr>
</table>

6 试验数据

因没有真实病毒，模拟了 3 种病毒进行试验，如下：

1）一维变形病毒 myVirus1：感染 QQ.exe 和微信.exe，具有 3 个病毒片段，片段中内容变化。

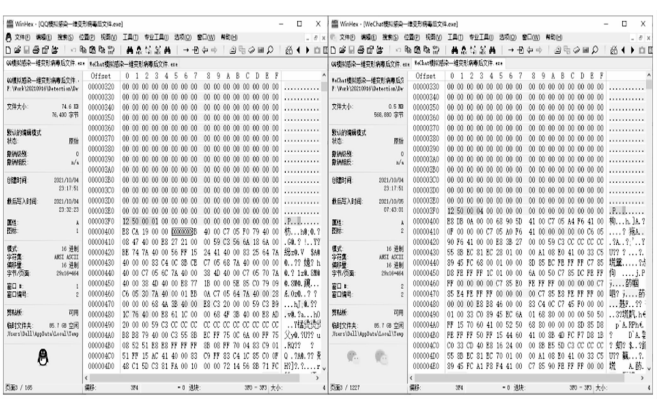

图 6 模拟 QQ 和微信感染一维病毒

2）二维变形病毒 myVirus2：感染 QQ.exe 和微信.exe，具有 3 个病毒片段，片段中内容变化，相对位置变化。

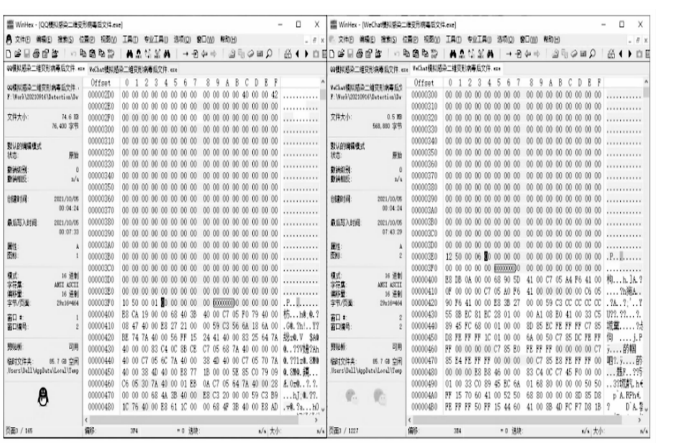

图 7 模拟 QQ 和微信感染二维病毒

3）三维变形病毒 myVirus3：感染 QQ. exe、微信 . exe 和 svchost. exe，具有 3 个病毒片段，片段中内容变化，相对位置变化，同时感染系统文件，需系统文件中的片段构成完整的病毒。

四维变形病毒原理相似，引入时间变化即可，此处不展开。

采用本文中检测算法结果如图 8～图 10 和表 1 所示。

图 8　myVirus1 相似度计算

图 9　myVirus2 相似度计算

图 10　myVirus3 相似度计算

7　结束语

在本文中研究了已有的病毒检测算法，得知当前病毒检测算法存在不足之处。目前变形病毒对互联网的危害最大，因此本人结合已有的病毒检测算法，从病毒的行为，校验和法，比较法的角度提出了一种基于行为，校验和值与比较法的新病毒检测算法。

参考文献：

[1] 刘巍伟，石　勇，郭　韩，等. 一种基于综合行为特征的恶意代码识别 [J]. 电子学报，2009（9）：696－700.

[2] 胡传裴. 计算机病毒智能检测及清除方法探究科 [J]. 科技资讯，2019（2）：24－25.

[3] 何　敏. 计算机病毒检测技术探究 [J]. 电脑知识与技术，2016（8）：39－40.

[4] 张　衡. 计算机病毒检测与查杀技术相关思考 [J]. 数字通信世界，2019（1）：119－120.

[5] 黄河夫. 计算机病毒智能检测及清除方法探究 [J]. 信息科技，2018，20（10）：159－160

[6] 李延香，袁　辉. 计算机病毒及其防治策略研究 [J]. 自动化与仪器仪表，2016（4）：209－210.

[7] 宋紫华，郭　春，蒋朝惠. 一种基于网络流量分析的快速木马检测方法 [J]. 计算机与现代化，2019(6)：9－15.

[8] 闫小妹. 网页木马的防御与检测技术研究 [J]. 电脑迷，2018(3)：75－76.

[9] 刘正宏. 变形病毒的分析与检测 [J]. 网络安全技术与应用，2009（5）：19－20.

[10] 苗　强，赵　琳，杜　明. 变形病毒技术研究及反病毒策略设计 [J]. 科技资讯，2008（8）：107.

[11] 王珊珊，孔韦伟，张　捷. 基于计算机变形病毒及其防治现状的探讨 [J]. 计算机与数字工程，2007（8）：78－80.

[12] 慈庆玉. 计算机变形病毒技术探讨 [J]. 中国数据通信，2005（1）：37－39.

[13] 祝　恩，殷建平，蔡志平，等. 计算机病毒自动变形机理的分析 [J]. 计算机工程与科学，2002（6）：14－17.

[14] 汤雲茜. 计算机病毒与防范 [J]. 信息与电脑，2019（9）：95－96.

[15] 杨　磊. 计算机病毒的检测与防御技术分析 [J]. 数字技术与应用，2017（4）：215.

[16] 沈继涛. 计算机病毒检测技术的现状与发展 [J]. 电子技术与软件工程，2017（4）：220.

[17] 陈桂生，张　哲. 计算机病毒检测方法的分析 [J]. 商丘职业技术学院学报，2007（2）：39－41.

[19] 陈菊红，顾林晴. 诊治计算机病毒的比较法 [J]. 现代电子技术，1996（1）：45－48.

[20] 卢　峰. 计算机网络应用病毒防护技术 [J]. 信息与电脑，2019（9）：22－24.

[21] 吴永娜. 黄苗苗. 多态变形技术原理分析及对策 [J]. 吉林工程技术师范学院院报，2011（10）：64－66.

[22] 朱俚治. 一种基于 k-近邻算法的最优解算法 [J]. 计算机与数字工程，2018（1）：35－38.

[23] 朱俚治. 程序属性的检测与程序属性的分类 [J]. 计算机测量与控制，2018，26（3）：103－106.

[24] KRIS K，谭明金，伍红兵，等. 黑客反汇编揭秘（第二版）[M]. 北京：电子工业出版社，2010.