

图像质量评价研究综述^{*}

周景超 戴汝为 肖柏华

(中国科学院自动化研究所复杂系统与智能科学重点实验室 北京 100080)

摘要 图像质量评价是一个经典的研究课题,其目标是设计算法,给出和人的主观感受相一致的评价值。在许多图像处理的应用领域中,图像质量评价算法都具有重要意义。目前已提出多种评价算法,本文对这些算法进行综述。首先描述图像质量的主观评价方法,它给出科学、稳定的主观评价;其次对各种客观评价算法进行分类介绍,重点探讨了各类算法的思路及特点;然后详细描述了衡量算法性能的定量指标;最后总结了图像质量评价算法研究的发展趋势。

关键词 图像质量评价,性能衡量,图像处理

Overview of Image Quality Assessment Research

ZHOU Jing-chao DAI Ru-wei XIAO Bai-hua

(Key Lab of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China)

Abstract Image quality assessment is a classical research issue. It aims at an evaluation consistent to human's subject perception and is very important to numerous image processing applications. Currently, many image quality assessment algorithms have been proposed. In this paper, we give an overview. Firstly, we present the subjective evaluating method which offers scientific and stable appraise value. Secondly, objective image quality assessment algorithms are classified and introduced with emphasis on their nature and characteristic. Then, the quantitative criterions for evaluating the performance of these algorithms are described in details. At last, we summarize the trends of future research.

Keywords Image quality metric, Performance evaluation, Image processing

1 引言

图像是人类感知和机器模式识别的重要信息源,其质量对所获取信息的充分性和准确性起着决定性的作用,然而在图像的采集、压缩、处理、传输及显示等过程中产生了图像降质问题。如何在视频传输、字符识别、安全监控、医学等领域建立有效的图像质量评价机制具有重大意义。

目前国内的一些高校和科研机构,如复旦大学^[1]、上海交通大学^[2]、南京航空航天大学^[3]、中国空间技术研究院^[4]、微软亚洲研究院^[5]等(限于篇幅,未全部列举),都对其进行研究。国外对图像质量评价算法的研究自上世纪80年代开始,如VQEG(国际电信联盟的视频质量专家小组)^[6]、LIVE(美国得克萨斯大学奥斯丁分校图像与视频工程实验室)^[7]、美国Sarnoff公司^[8]等研究机构都有长期的研究。

图像质量评价的主体是人,因此人眼视觉特性(HVS,如多通道结构、掩盖效应、对比度敏感度和视觉非线性定律等)对评价算法的研究具有重要意义。但是,对人眼视觉特性的研究涉及生理、心理等方面,迄今对其仍没有充分理解,特别是对视觉的心理特性还难以找出定量的描述方法,所以目前还缺乏一套全面的、统一的图像质量评价体系。

图像质量评价的方法有两种:(1)主观评价方法:设计实验,由观测者对图像质量进行评价;(2)客观评价方法:采用算法对图像质量进行评价。其中主观评价方法天然地和人的主观感受相符,但它费时、复杂,还会受到观测者专业背景、心理

和动机等主观因素的影响,并且不能结合到其他算法中使用;客观评价方法方便、快捷,容易实现并能结合到应用系统中,但它和人的主观感受有出入。通常提到的图像质量评价算法是指客观评价算法,其目标是获得与主观评价结果相一致的客观评价。

2 主观评价方法

为得到定量的主观评价结果,需要设计主观评价实验,其中包括三个步骤:(1)准备样本集;(2)由观测者对样本集的图像质量进行评价;(3)对主观评价结果进行加工。

2.1 准备样本集

准备样本集时需注意如下两个方面:(1)样本集中所包含的失真类型要全面,常见的失真类型有:有损压缩引起的失真、噪声、图像传输过程中受信道衰减影响引起的失真等;(2)图像内容要丰富,包含背景复杂的自然场景、侧重于细节表现的人物造型、背景简洁、醒目的标语牌等。设计算法时如果采用的样本集比较小,就容易出现“过训练”的问题。

目前,有一些公开的数据集可供使用,如VQEG^[6]的Phase-I和LIVE^[7]。其中LIVE数据集由29幅高分辨率、高质量的彩色图像组成,考虑了五种失真类型,并且给出了图像的DMOS(Difference Mean Opinion Score)评分,便于研究者使用。

2.2 观测者对图像质量进行评价

观测者对图像质量进行评价时,需考虑显示设备的配置、

^{*} 中国科学院自动化研究所青年科技创新基金资助,国家自然科学基金(60602031)资助。周景超 博士研究生,研究方向为视频文字提取、图像质量检测;戴汝为 研究员,中国科学院院士,研究方向为自动控制、模式识别、人工智能、智能控制及思维科学;肖柏华 研究员,研究方向为模式识别与图像处理、智能系统、信息检索与智能信息处理。

观测距离、观测者的来源、对观测者的培训、每一个观测小组的人员个数和观测图像数目、每次观测的时间以及如何对图像进行评价等方面,文献[6]中对图像主观评价过程有详细的描述。

2.3 对主观评价结果进行加工

观测者进行评价时,其情绪、动机、知识背景等因素均会对评价结果产生影响,使评价结果变得不准确。为此,需要对实验中得到的原始评价结果进行处理,以减少这些主观因素的影响。在下面的描述中, $i=1, \dots, M$ 代表不同的观测者, $j=1, \dots, N$ 代表不同的图像, r 代表观测者对一幅图像的原始评分。

(1) 由于观测者对图像的主观评价是一个松散的分布,为了减轻主观因素的影响,得到一个比较稳定的结果,需要对原始数据进行取舍。

a) 评分出界的检测。对于同一测试组中同一幅图像的主观评价值,其均值、标准差为:

$$U = \frac{1}{M} \sum_{i=1}^M r_{i,j}$$

$$S = \sqrt{\frac{\sum_{i=1}^M (U - r_{i,j})^2}{M-1}}$$

$$\epsilon = 1.96 \cdot \frac{S}{\sqrt{M}}$$

其中 $r_{i,j}$ 为一个主观评分, $[U-\epsilon, U+\epsilon]$ 为 95% 置信区间,在此区间之外的评分被视为出界评分,舍弃。评分出界检测做 1~2 次。

b) 对观测者的淘汰。如果同一测试组内某一个观测者的评分出界次数 $N \geq N_0$, 则认为该观测者不具备评价的资格,舍弃他的全部评价值。其中 N_0 为经验值,参考值为 16。

(2) 图像内容的复杂程度、观测者对图像内容的喜好等因素都有可能影响其对图像质量的评分。为了减轻这种影响,可采用“双刺激法”,即同时给出失真图像和参考图像,要求观测者对二者都进行评分,以评分之差作为评价结果。

$$d_{i,j} = r_{i,ref(j)} - r_{i,j}$$

$$d'_{i,j} = \frac{d_{i,j} - \text{Min}_i}{\text{Max}_i - \text{Min}_i}$$

其中 $r_{i,ref(j)}$ 是第 i 个观测者对第 j 幅图像的参考图像的评分, $d'_{i,j}$ 是对每一个观测者的评分进行归一化后的结果。

最后,得到一幅图像的 DMOS 值:

$$DMOS_j = \frac{1}{M} \sum_{i=1}^M d'_{i,j}$$

3 客观评价方法

客观评价算法根据其对参考图像的依赖程度,可分成三类。(1)全参考:需要和参考图像上的像素点做一一对应的比较;(2)半参考:只需要和参考图像上的部分统计特征做比较;(3)无参考:不需要具体的参考图像。其中全参考算法是研究时间最长、发展最成熟的部分。

3.1 全参考算法

全参考算法种类繁多。根据算法采用的技术路线,可分为基于误差统计量的算法和基于 HVS 模型的算法。这类方法的特点是数学含义清晰,便于实现,能够捕获像素层面的微小失真。

基于误差统计量的算法的思路为:通过设计特征来比较失真图像和参考图像的局部差异,然后在整幅图像上求出一个总的平均统计量,并把这个统计量与图像质量关联起来。

其中有一部分算法来源于通信理论,比较信号之间的保真度,如峰值信噪比(PSNR)^[6]、目标信息密度^[4]等;另外一部分算法来源于图像处理技术,如考虑邻域差异^[9,10]、颜色差异^[1,11]、内容相关性^[12]等。这类算法的缺点是:仅注重衡量失真图像和参考图像之间的差异,并没有考虑这些差异对人的视觉感知的影响,因而和主观评价之间有较大的出入。目前这类算法有两个发展趋势:(1)研究已有特征之间的关系,从中选出互补的特征,组成新的特征空间,采用分类或者回归的方法对图像质量进行评价;(2)利用人眼视觉特性对误差统计量进行加权。Ismail Avcibas 采用 Kohonen 自组织映射(SOM)的方法以直观的形式表现出各种特征之间的相似或者互补关系,供设计新的评价算法时参考^[11]。汪孔桥等结合人眼感兴趣区域(ROI)理论对峰值信噪比标准进行修正,根据 ROI 区域的划分,调整不同区域的加权值^[13];Chiu 等考虑到人眼对边缘失真的敏感远大于平坦区域,在算法中对边缘失真予以重视^[14];Makoto Miyahara 等提出的 PQS 算法把图像失真的原因分为亮度差异、空间频率失真、结构上的干扰和随机误差,首先通过设计特征来统计各类失真的程度,然后用 PCA(Principle Component Analysis)对特征进行选择,最后在样本集上训练以确定不同特征在质量评价时的权重^[15]。

基于 HVS 模型的算法通过对 HVS 的某些底层特性进行建模,将失真图像和参考图像之间的绝对误差映射为能被人眼觉察的 JND(Just noticeable difference)单位。如果误差高于视觉的敏感门限,则表明其能够被人眼觉察,否则可以被忽略。通过上述处理可以获得和观测者的主观感受相一致的评价结果。算法处理过程为:(1)分块:将整幅图像分成规则的图像块,便于计算;(2)分解:将图像块根据不同的空间频率和方向进行分解,每一个空间频率和方向角度的组合被称为一个“通道”; (3)比较:对失真图像和参考图像进行对比,得到一幅“差异图像”; (4)掩盖:在每个通道上加入时-空滤波器,模拟 HVS 的底层特性,计算差异在参考图像的背景下显露出来的程度;(5)建模:计算整幅图像上的差异显露的平均值,建立模型,将其与图像的主观评价值联系起来,并在训练集上估计模型参数^[16,17]。这类方法多用于视频图像质量检测,由于其综合模拟了 HVS 的各项特性,能够较好地保证主客观评价的一致性,但算法复杂、运算量大,并且受制于当前对 HVS 机制的认识水平,仍需进一步的研究。

3.2 半参考算法

半参考算法可以分为两种:基于图像特征统计量的算法和基于数字水印的算法。这类算法的特点是只需从参考图像中提取部分统计量用于比较,无需原始的像素级别的信息。

基于图像特征统计量的算法的思路是:首先将图像分块,然后以图像块为单位分别统计失真图像和参考图像的相关特征,最后比较这些统计特征之间的差异,从而对图像质量做出评价。Zhou Wang 等提出的 SSIM 算法通过在失真图像和参考图像中剥离亮度、对比度和线性相关强度等结构信息进行比较,从而对图像的失真程度进行衡量^[18]。

NSS(Natural Scene Statistics)是对大量自然场景图像进行小波分解后,对不同尺度、方向和位置上的小波系数进行统计,从而得到统计特征^[19]。根据这些统计特征,采用 GSM(Gaussian Scale Mixture)模型图像进行建模^[20]。Hamid 等提出的 VIF(Visual Information Fidelity)算法即是建立在 GSM 模型上,算法的前提假设是:无论是哪种失真类型,均会使图像的统计特征发生变化,如果得出失真图像和参考图像

在统计特征上的差异,就可以对图像质量做出合理的评价^[21]。

数字水印是按照一定的策略向原始数据中添加的某些数字信息。基于数字水印的算法的思路是向图像数据中嵌入水印信息,当水印信息和图像数据经过相同的系统处理后,从输出图像中提取水印信息,然后比较原始水印和提取出来的水印,以水印信息的质量状况作为图像质量评价的依据^[2,22]。

半参考算法介于全参考算法和无参考算法之间,具备二者各自的优点:一方面,经过大量样本的训练后能够取得很好的性能(如 NSS 和 SSIM 算法的性能衡量^[23]);另一方面,用于参考的统计量与原始的参考图像相比数据量很小,这在对图像或者视频传输系统的质量监控中具有优势。

3.3 无参考算法

无参考算法可以分为两种:针对失真类型的算法和基于机器学习的算法。这类方法的特点无需参考图像,灵活性强。

无参考算法中研究较多的失真类型是 jpeg 和 jpeg2000 (即: H. 264 视频压缩标准) 压缩失真。在对图像进行 jpeg 压缩时,首先将整幅图像分成许多统一大小的图像块,在每个图像块上采用离散余弦变换(DCT),对 DCT 系数进行压缩。由于不可避免地存在量化和截断误差,因此压缩后的图像会出现模糊和块状失真。无参考算法初期主要根据这两种失真现象对压缩后的图像进行质量评价。Zhou Wang 等提出一种频域评价算法,根据失真图像在某些特征频率上的峰值和能量从高频到低频的转移来衡量其失真程度^[24]; Yuukou Horita 等和 R. Muijs 等根据 DCT 压缩后图像块边界上梯度值变化的剧烈程度设计相应的统计量,来衡量块状失真的程度^[25,26]; P. Marziliano 等和 E. P. Ong 等提出通过计算图像中边缘向其邻域内的扩散程度来衡量模糊失真程度的算法^[27,28]。jpeg2000 采用小波变换,通过舍弃较大尺度上的小波系数达到压缩的目的,这造成在图像细节上的失真,主要表现为振铃和模糊。与 jpeg 失真不同, jpeg2000 失真没有固定的结构信息,而是和图像内容有关,如振铃现象出现在图像中比较明显的边缘上。H. R. Sheikh 等根据图像在不同尺度上的小波系数的衰减规律,提出一种衡量 jpeg2000 失真的算法^[29]。

基于机器学习的算法基于如下假设:质量相近的图像在底层特征上具有相同的规律,而这些规律可以通过机器学习的方法得到。与前面提到的无参考算法不同,它并不需要分析失真的原因以及如何设计特征来衡量失真的程度,而是将学习得到的“特征”直接作为图像质量评价的标准。这类算法首先将图像分块,以每块区域内的像素值作为原始特征向量,然后采用不同的机器学习策略从中提取出分类能力强的新特征向量,最后结合主观评价利用新特征向量进行分类或者回归。Tong Hanghang 等采用 MMD^[30] 策略提取特征向量,用 MDH(Multi-dimensional Histogram)方法估计类条件概率密度函数,然后设计 Bayes 分类器对不同质量的图像进行分类、评价^[5]; Gastaldo 等采用循环反向传播(CBP)神经网络进行特征提取和质量评估^[31]。这类算法的优点在于能够借助机器学习领域内的研究成果,提高算法性能,但由于目前算法直接以图像中的像素点作为原始特征而没有进行相关处理,在性能上处于劣势。

无参考算法的难点在于如何使评价结果尽量不受图像内容的影响。算法进行质量评价的依据是从图像中提取的某些特征统计量,而这些统计量受图像内容的影响。如在内容复

杂的图像上得到的统计量往往大于同质量的内容简单的图像上的统计量,由于没有具体的参考图像作为标准,容易造成对复杂图像质量的“过估计”(评价价值大于实际值)和对简单图像的“欠估计”(评价价值小于实际值)。算法应该在整个样本集上得到相对一致的评价价值。

4 衡量算法性能的定量指标

图像质量评价算法应该具备以下三个特性:(1)准确性:主、客观评价价值之间的差异较小;(2)单调性:客观评价价值应随主观评价价值的增减而增减;(3)一致性:算法在测试集上表现出的性能与其在训练集上表现的性能相近似。

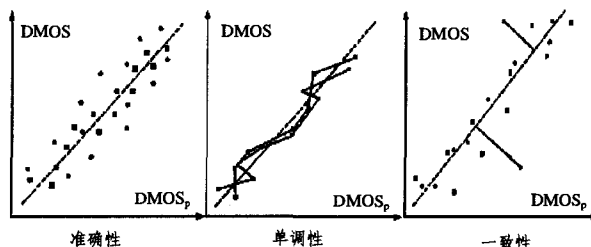


图1 评价算法性能的三个特性

为便于理解,结合图1进行说明。图中所示为算法A(方形点)和B(圆形点)的评价结果在 DMOS-DMOS_p 上的分布情况,其中 DMOS 反映观测者给出的主观评价价值,DMOS_p 反映评价算法给出的客观评价价值。左图表明,A 的偏差要比 B 小一些,即 A 的准确性更高。从中间图看出,A 给出的评价价值随着 DMOS 值的增加而增加,而 B 给出的评价价值中有两个点随着 DMOS 值的增加而减小,所以 A 的单调性更好一些。右图表明,B 的评价价值虽然在多数情况下和 DMOS 值之间的偏差都比 A 小,但在两个点上偏差特别大,所以 B 的一致性没有 A 好。

上述性质可以通过定量标准进行衡量。衡量之前需要对 DMOS 值进行非线性回归^[6]。在理想的情况下,DMOS_p = DMOS,二者为简单的线性关系,但是,主观评价过程会引入一些非线性因素,如果直接对 DMOS_p 和 DMOS 求相关的话,二者的差异会比较大,同时衡量算法性能的重点不在于客观评价价值和主观评价价值之间的线性关系,而在于二者对应关系的稳定性以及在样本集上误差的大小。所以,在计算定量标准之前需要根据训练集上 DMOS_p 和 DMOS 的分布情况进行非线性回归,以减弱非线性因素的影响。有两种非线性回归可供参考:多项式拟合和 Logistic 回归,根据实际情况选择误差较小的一个。

衡量算法性能的定量指标主要有如下三个:

(1) Linear Correlation Coefficient

$$r = \frac{N \sum y \hat{y} - \sum y \sum \hat{y}}{\sqrt{N \sum y^2 - (\sum y)^2} \cdot \sqrt{N \sum \hat{y}^2 - (\sum \hat{y})^2}}$$

其中 y 和 $\hat{y}$ 分别为 DMOS 和 DMOS_p, N 为样本的个数。

(2) Root Mean Squared Error

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y - \hat{y})^2}$$

变量含义同上。

(3) Spearman Rank Order Correlation Coefficient

$$r_s = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

其中 D 为等级间差异,在此处为 DMOS_p 和 DMOS 之间差值的绝对值。

5 图像质量评价算法的研究趋势

图像质量评价算法的研究呈现以下三个趋势:

(1)从单纯的客观评价算法转化为主客观相结合的评价算法;(2)无参考算法开始成为研究的热点;(3)面向具体任务的评价算法逐渐增多。

图像质量评价算法的目标是得到与人的主观评价相一致的评价结果,因此一致性是衡量算法性能的最主要方面。近来提出的算法都在不同程度上采用了 HVS 特性,或者是图像中提取人眼感兴趣的边缘、色彩等结构化信息,或者是在评价模型中保留参数,根据训练集上的主观评价价值来确定这些参数,或者是把 HVS 系统看成一个黑箱,通过机器学习的方法挑选出人眼感兴趣的特征。所有这些都是为了让算法给出的客观评价结果尽可能地与主观评价价值相一致。同时,众多研究结果表明,加入 HVS 特性的算法要优于单纯的客观评价算法。因此,主客观相结合成为评价算法未来发展的一个趋势。

在许多实际应用中,经常得不到参考图像或者获得参考图像付出代价太大,因而要求算法降低对参考图像的依赖程度。同时,观测者往往并不需要参考图像就能够对图像质量做出合理的评价。这表明观测者在进行评价时抓住了反映图像质量的最本质的特征,也说明对无参考算法的研究更有可能揭示人类视觉感知的原理。因而,无参考算法开始引起众多研究者的关注,相关的研究成果也不断增多,成为未来发展的趋势之一。

在不同的应用场合下对图像质量进行评价的侧重点也有所不同,如在数码摄影时注重图像在细节上的刻画,而在医学图像处理时则注重图像在结构上的清晰度。同时大量的研究表明,建立一套通用的图像质量评价体系是比较困难的。Is-mail 采用方差分析(ANOVA)的方法来考察各种评价算法对失真类型(JPEG 压缩、SPIHT 压缩、模糊、加性噪声)和图像内容(编织物、人脸、自然场景)等外部因素的敏感程度^[11]。结果表明,一种评价算法只是对一种或者少数的几种失真比较敏感,而对其他类型的失真检测效果并不理想。事实上,评价图像质量的问题从一开始就是面向任务需求的,比如图像压缩评价、图像滤波评价等。目前国内的科研人员陆续开展了第二代身份证数码照片质量评价、面向退化文字识别的图像质量判别^[32]以及面向微装配机器人自动调焦跟踪的图像清晰度判别等许多面向任务需求的图像质量评价研究,这也是评价算法未来的一个发展方向。

结束语 在图像处理的实际应用中,会遇到质量不一的各种图像,如何准确地对其质量进行评价并结合评价价值做进一步的处理,具有重要的实用价值。同时,图像质量评价算法的研究能够促进对人类视觉感知的研究,具有重要的理论意义。鉴于此,本文系统地介绍了设计图像质量评价算法所必需的三个环节:(1)设计主观评价实验,在样本集上给出相对准确、稳定的主观评价价值;(2)根据实际情况,设计客观评价算法对图像质量做出评价;(3)对照主观评价价值,采用科学的定量标准来衡量设计出来的客观评价算法的性能。

在设计评价算法时应把握好以下两个方面:

(1)图像特征提取与特征选择问题。提取既能准确刻画图像的状态,又能符合人类视觉特性的特征,是设计算法的

关键。

(2)客观评价价值与主观评价价值相一致的问题。为使评价算法具有较好的性能,在设计时应尽量使其得出的评价价值和主观评价价值相一致。

随着图像质量评价算法的不断成熟,将有更多的软件和硬件产品投入实用,评价算法的研究具有广阔的应用前景。

参考文献

- [1] 叶佳,张建秋,胡波. 客观评估彩色图像质量的超复数奇异值分解法. 电子学报, 2006, 35: 28-33
- [2] 殷晓丽,方向忠,翟广涛. 一种 JPEG 图片的无参考图像质量评价方法. 计算机工程与应用, 2006, 18: 79-81
- [3] 蒋森,朱剑英. 基于奇异值分解的图像质量评价. 东南大学学报(自然科学版), 2006, 36: 643-646
- [4] 时红伟,陈世平. 一种基于现代通信理论的图像质量评价技术. 中国空间科学技术, 2005, 1: 1-6
- [5] Tong Hanghang, Li Mingjing, Zhang Hongjiang, et al. Learning no-reference quality metric by examples // Multimedia Modeling Conference. 2005: 247-254
- [6] VQEG. Final report from the video quality experts group on the validation of objective models of video quality assessment. <http://www.vqeg.org>, 2000
- [7] LIVE. <http://live.ece.utexas.edu/research/quality>
- [8] Sarnoff. <http://www.sarnoff.com/products-services>
- [9] DiGesu V, Staravotov V V. Distance-based functions for image comparison. Pattern Recognition Letters, 1999, 20(2): 207-213
- [10] Starovitev V V, Kose C, Sankur B. Generalized distance based matching of non-binary images // IEEE International Conference on Image Processing. 1998
- [11] Avcibas I, Sankur B, Sayood K. Statistical Evaluation of Image Quality Measures. Journal of Electronic Imaging, 2002: 206-223
- [12] Eskicioglu A M, Fisher P S. Image Quality Measures and Their Performance. IEEE Transactions on Communication, 1995, 43(12): 2959-2965
- [13] 汪孔桥,沈兰荪,荆昕. 一种基于视觉兴趣性的图像质量评价方法. 中国图象图形学报, 2000, 5: 300-303
- [14] Chiu Yi-Jen, Berger T. A Software-Only Video Codec Using Pixel Wise Conditional Differential Replenishment and Perceptual Enhancements. IEEE Transactions on Circuits and Systems for Video Technology, 1999(9)
- [15] Miyahara M, Kotani K, Algazi V R. Objective Picture Quality Scale for Image Coding. IEEE Transactions on communications, 1998, 46
- [16] van den Branden Lambrecht C J, Verscheure O. Perceptual ality Measure Using a Spatio-Temporal Model of the Human Visual System // Proceedings of SPIE. 1996(2668): 450-461
- [17] Wang Zhou, Bovik A C. A Human Visual System Based Objective Video Distortion Measurement System // International Conference on Multimedia Processing and Systems. 2001
- [18] Wang Zhou, Bovik A C, Sheikh H R, et al. Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Transactions on Image Processing, 2004, 13: 600-612
- [19] Srivastava A, Lee A B, Simoncelli E P, et al. On Advances in Statistical Modeling of Natural Images. Journal of Mathematical Imaging and Vision, 2003
- [20] Wainwright M J, Simoncelli E P. Scale Mixtures of Gaussians and the Statistics of Natural Images. Advances in Neural Information Processing System, 2000, 12: 855-861
- [21] Sheikh H R, Bovik A C. Image Information and Visual Quality. IEEE Transactions on Image Processing, 2006, 15: 430-444
- [22] Yang Fu-zheng, Wang Xin-dai, Chang Yi-lin, et al. A No-reference Video Quality Assessment Method Based on Digital Watermark // IEEE International Symposium on Personal, Indoor and Mobile Radio Communication Proceedings. 2003: 2707-2710

(下转第 8 页)

3)比较内存中内核表项的值与 System.map 中系统调用表项与已保存的中断描述符表项的值,可以发现被劫持的系统调用。

KSTAT 有多个选项来帮助用户检测 LKM 后门的存在,最常用的就是-s 选项,使用这个选项,KSTAT 将导出核心系统调用的所有地址。如果探测出的地址和原来的地址不同,就会给出一个警告。另一个常用的选项是-p,这个选项列出当前在系统中运行的所有进程,包括被 LKM 后门隐藏的进程,并给出提示信息。

KSTAT 属于传统方法,它不可避免地带有传统方法的缺陷。另外,研究人员在 2001 年提出了逃避 KSTAT 检测的 LKM 后门实现方法。其原理是对原始系统调用加一个跳转指针指向新的调用,这样并不需要添加新的系统调用到系统调用列表中,也不需要修改原始系统调用的入口地址,从而可以逃避 KSTAT 的检测^[15]。

4.3 EPA 方法

EPA(Execution Path Analysis,可执行路径分析)方法的原理是:假如有人修改了内核函数,内核的指令执行路径将改变,在系统调用过程中指令的数目将和原始的内核不同。恶意代码在返回到用户空间时肯定有一些附加行为。这意味着将有许多与未被感染系统不同的代码被执行。可以通过测试系统调用的可执行指令的数目变化来确定系统是否被侵入^[16,17]。EPA 方法不属于传统方法,因为系统被入侵后它也能正确分析,它的应用成为传统 LKM 后门检测方法的有力补充。不过 EPA 类检测工具自身也容易受到攻击和欺骗,目前尚在研究之中。

4.4 其它

另外也有针对 LKM 某个方面进行的检测。例如 chkrootkit 对 ps 的输出与 /proc 下的数字目录做比较,如果发现有不符,则报警。但是假如入侵者有针对性地将 chkrootkit 使用的系统命令也做修改,chkrootkit 将无法监测到 LKM 后门的存在。文献[18]在 chkrootkit 的基础上,利用隐藏的文件可以被访问以及进程数量有限两个特点,提出了基于穷举和对比的 Linux 隐藏进程查找方法。这两个方法的特点都是在系统被入侵后也能正确运行,但是检测结果的完备性存在一定的不足。例如文献[18]只能检测到隐藏的进程名和进程文件信息,至于被加载的 LKM 后门信息则无法确定。

结束语 本文分析了 LKM 后门的原理以及威胁,并在此基础上对其检测方法进行了讨论。在目前的技术条件下,还没有通用的、精确的方法能检测到 LKM 后门,这就需要管理员未雨绸缪、加强管理,防止任何能让别人加载 LKM 的机

会。文献[2,19]提到了内核固化的思想,就是让系统不支持 LKM,虽然杜绝了 LKM 后门的存在,但是对系统上的开发却不利。在将来的检测上,更多地趋向于混合利用上述几种检测方法,以达到互相补充、互相弥补、提高检测精度的目的。

参考文献

- [1] Pragmatic / THC. Complete Linux Loadable Kernel Modules (v1.0). <http://packetstormsecurity.nl/docs/hack/LKMHACKING.html>, 1999
- [2] 袁源,戴冠中,罗红,等.基于 LKM 的 Linux 安全检测器的设计与实现.计算机应用研究,2005,22(7):131-133
- [3] 李海刚,崔杜武,姚全珠,等. Linux 模块技术分析及应用.计算机工程,2003,29(1):120-122
- [4] Rubini A, Corbet J. Linux Device Drivers[M]. O'REILLY & Associates, Inc, 2001
- [5] Plaguez. Weakening the Linux Kernel[J]. Phrack Magazine, 1998, 52
- [6] 王永杰,刘京菊,孙乐昌. Linux 可装载模块的开发和应用.计算机应用研究,2002,19(7):143-146
- [7] 毛德操,等. Linux 内核源代码情景分析. 浙江:浙江大学出版社,2001:191-256
- [8] 时金桥,方滨兴,胡铭曾,等. Linux 系统调用劫持:技术原理、应用及检测.计算机工程与应用,2003,39(32):167-170
- [9] 王斌,须文波,冯斌.利用 LKM 的 Linux 审计功能实现.计算机工程,2004,30(3):136-138
- [10] Xpiloveryou. Linux 2.4.18 内核下基于 LKM 的系统调用劫持. <http://dev.csdn.net/Develop/article/28/28152.shtm>
- [11] Palmers. Advances in Kernel Hacking[J]. Phrack Magazine, 2002, 59
- [12] 兰薇薇,贾卓生.利用 LKM 记录键盘事件的研究与实现.网络安全技术与应用,2005,3:24-26
- [13] Rd. Writing linux kernel keylogger[J]. Phrack Magazine, 2002, 59
- [14] 孙淑华,马恒太,张楠,等.后门植入、隐藏与检测技术研究.计算机应用研究,2004,7:78-81
- [15] E4gle. 逃避 kstat 的检测的 lkm 程序的实现方法(linux). <http://www.xfocus.net/articles/200108/258.html>
- [16] 王松涛,吴灏. Linux 下基于可执行路径分析的内核 rootkit 检测技术研究.计算机工程与应用,2005,41(11):121-123
- [17] Rutkowski J K. Execution path analysis; finding kernel rks[J]. Phrack Magazine, 2002, 59
- [18] 袁源,戴冠中,罗红.利用 Perl 实现 Linux 下隐藏进程的查找.计算机工程与应用,2007,43(3):102-105
- [19] Stealth. Kernel Rootkit Experiences [J]. Phrack Magazine, 2003, 61
- [20] Ong E P, Lin W S, Lu Z K, et al. No-reference JPEG2000 Image Quality Metric // International Conference on Multimedia and Expo. 2003, 1: 6-9
- [21] Sheikh HR, Bovik AC, Cormack L. No-Reference Quality Assessment Using Natural Scene Statistics; JPEG2000. IEEE Transactions on Image Processing, 2005, 14: 1918-1927
- [22] Vasconcelos N. Feature Selection by Maximum Marginal Diversity; Optimality and Implications for Visual Recognition. Computer Vision and Pattern Recognition, 2003(1): 762-769
- [23] Gastaldo P, Zunino R. No-Reference Quality Assessment of JPEG Images by Using CB PNeural Networks // IEEE International Symposium on Circuits and Systems. 2004, 5: 772-775
- [24] Liu Chunmei, Wang Chunheng, Dai Ruwei. Low Resolution Character Recognition by Image Quality Evaluation // IEEE International Conference of Pattern Recognition, 2006, 1: 864-867
- [25] Sheikh H R, Sabir M F, Bovik A C. A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. IEEE Transactions on Image Processing, 2006, 15: 3441-3451
- [26] Wang Zhou, Bovik A C, Evans B L. Blind Measurement of Blocking Artifacts in Images // IEEE International Conference on Image Processing. 2000, 3: 981-984
- [27] Horita Y, Arata S, Murai T. No Reference Image Quality Assessment for JPEG/JPEG2000 Coding // The 12th European Signal Processing Conference. 2004
- [28] Muijs R, Kirenko I. A No-Reference Blocking Artifact Measure for Adaptive Video Processing // The 13th European Signal Processing Conference. 2005
- [29] Marziliano P, Dufaux F, Winkler S, et al. A No-reference Perceptual Blur Metric // IEEE International Conference on Image Processing. 2002, 3: 57-60

(上接第4页)

- [23] Sheikh H R, Sabir M F, Bovik A C. A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. IEEE Transactions on Image Processing, 2006, 15: 3441-3451
- [24] Wang Zhou, Bovik A C, Evans B L. Blind Measurement of Blocking Artifacts in Images // IEEE International Conference on Image Processing. 2000, 3: 981-984
- [25] Horita Y, Arata S, Murai T. No Reference Image Quality Assessment for JPEG/JPEG2000 Coding // The 12th European Signal Processing Conference. 2004
- [26] Muijs R, Kirenko I. A No-Reference Blocking Artifact Measure for Adaptive Video Processing // The 13th European Signal Processing Conference. 2005
- [27] Marziliano P, Dufaux F, Winkler S, et al. A No-reference Perceptual Blur Metric // IEEE International Conference on Image Processing. 2002, 3: 57-60