2015 南京理工大学大学生数学建模竞赛

承　诺　书

我们仔细阅读了中国大学生数学建模竞赛的竞赛规则.

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的, 如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们授权全国大学生数学建模竞赛组委会，可将我们的论文以任何形式进行公开展示（包括进行网上公示，在书籍、期刊和其他媒体进行正式或非正式发表等）。

我们参赛选择的题号是（从 A/B 中选择一项填写）：　A　

我们的参赛报名号为（报名编号）：　54　

所属学院（请填写完整的全名）：　教育实验学院　

参赛队员 (打印并签名) ：1.　丁天扬　

2.　陈　彤　

3.　闫诗晗　

日期：　2016　年　5　月　24　日

评阅编号（由组委会评阅前进行编号）：

信息时代的系统安全性度量

摘　要

针对问题一，基于“2015 年信息安全事件汇总报告”以及其它网络数据，通过归纳与分析近年来多发的信息安全事件的原因与类别，可以从中抽象出信息安全度量的五项重要指标：系统预警能力、安全防护能力、隐患发现能力、反应处置能力以及信息安全费用。对系统进行信息安全程度的评估，最主要的难点就是定性指标的定量处理。一般的安全报告大多只是将事件做分类或是对某个具体的系统的某些安全事件发生的数量做了统计，没有数据直接对信息安全指标做分类及统计，所以建立一个具有一般性的计算系统安全程度的数学模型具有一定的困难。在已有数据的基础上，我们采用了层次分析的方法，先计算出影响信息安全程度有关因素的相对权重值，对各因素权重值排序，做横向比较，为之后在具体的实例中分析权重打下数学基础；接着用模糊综合评判，根据现今已有的信息系统安全性的评价标准，综合得到信息系统的安全等级，从而可以完成信息安全的度量。

针对问题二，首先确定将上述提到的五项指标作为评价信息系统安全的核心指标，搜集能够反映这些指标的相关数据，基于主成分分析法分析其对系统安全程度的贡献率，选出累积贡献率达 90%及以上的的几项指标，对其做综合评价，给出一个综合指标，对这个综合指标进行分析，并给出计算指标的方法。

针对问题三，需要我们对现有的手机系统的安全程度作出评估与计算。首先，整个手机信息系统涉及物理环境及保障、硬件设施、软件设施和管理者四个方面,其中软件设施的安全性评估涉及五类风险因素,分别为：计算机操作系统 M1、网络操作系统 M2、网络通信协议 M3、通用应用平台 M4 和网络管理软件 M5。更具体化的指标一般为恶意程序、垃圾短信、诈骗短信、骚扰电话、诈骗电话。通过第一问的模型，以及对网络数据的处理，可以初步评价出手机信息系统的安全性。接着通过利用第二问的模型，可以在肯定了手机系统总体安全性的基础上对四大手机操作系统进行安全性比较，并得出分析结果。

关键词

层次分析；模糊综合；主成分分析；手机系统安全

1　问题重述

斯诺登事件为我们敲响了信息安全的警钟，也让我们更进一步认识到当前网络信息安全所面临形势的严峻性。保障我国网络信息安全，是当前面临的重要问题。信息安全度量是业界公认的一个难题，信息安全度量一般需要回答两个问题：信息系统安全不安全?信息系统的安全程度是多少? 通过查阅相关资料。请你们小组解决以下问题：

问题1：基于“2015年信息安全事件汇总报告”（http://mt.sohu.com/20160113/n434399073.shtml）以及其它网络数据，建立一个计算信息系统（孤立隔离，或广泛互联的系统）的安全程度的数学模型。

问题2：选取一个重要的信息安全度量指标，说明选取的理由，并给出计算该指标的数学模型。

问题3：利用上述两个模型，具体对你们小组成员所持有的手机信息系统进行研究，给出计算结果，并进行简单分析与比较。

2　问题分析

对于问题 1，首先考虑如何衡量一个系统的安全程度。通常情况下，常常通过分析某个信息系统在一段时期内发生安全事件的多少来分析一个系统是否安全。基于“2015年信息安全事件汇总报告”以及相关的其他网络数据，发现系统多发的信息安全事件往往有以下几类：信息泄漏、电信诈骗、网络漏洞、网络谣言、手机流量疯跑、网站被黑、非法集资、网络暴力、网络暴力以及其他安全事件。根据这些信息安全事件发生的不同原因，我们可以从中抽象出信息安全度量的几个重要指标：系统预警能力、安全防护能力、隐患发现能力、反应处置能力以及信息安全费用。但是，一段时期内信息安全事件的发生会由于多方面的原因而发生不同程度的变化，一种可能是系统的安全措施的增强或削弱，第二种可能是外部攻击的增多或减少，第三种可能是信息系统所承载的信息的吸引力降低了。诸多不确定因素的存在增加了信息安全度量的复杂性,也决定了对信息安全的度量是多维的，而且各个维之间还存在着非线性的复杂关联。基于以上分析可知，评估至少要对以下因素进行分析[1]：

①信息和信息系统自身价值判断；

②外部安全环境变化,即对手的攻击强度；

③时间因素的分析；

④保障措施的强度和费用等。

对系统进行信息安全程度的评估，最主要的难点就是定性指标的定量处理。基于层次分析法及模糊综合评判的信息安全程度评估方法，有效地解决了这一难点。层次分析法可以计算影响信息安全程度有关因素的相对权重值，对各因素权重值排序，做横向比较，为采取相关措施提供有力依据；模糊综合评判根据现今已有的对信息系统安全性的评价，综合得到信息系统的安全等级，从而可以完成信息安全的度量。

对于问题二，首先确定将系统预警能力、安全防护能力、隐患发现能力、反应处置能力以及信息安全费用[1]这五项指标作为评价信息系统安全的核心指标，搜集能够反映这些指标的相关数据，基于主成分分析法分析各项指标对系统安全程度的贡献率，选出累积贡献率达 90%及以上的的几项指标，对其做综合评价，给出一个综合指标，对这个综合指标进行分析，并给出计算指标的方法。

对于问题三，题目选择了我们身边的手机作为案例。首先，我们可以分析出整个手机信息系统认为涉及物理环境及保障、硬件设施、软件设施和管理者四个方面,其中软

件设施的安全性评估涉及五类风险因素,分别为:计算机操作系统 M1、网络操作系统 M2、网络通信协议 M3、通用应用平台 M4 和网络管理软件 M5。然二从更日常化的指标中可以找到恶意程序、垃圾短信、诈骗短信、骚扰电话、诈骗电话这五项评价系统安全性的指标。通过第一问的模型，以及对网络数据的处理，可以初步评价出手机信息系统的安全性。接着通过利用第二问的模型，可以在肯定了手机系统总体安全性的基础上对手机操作系统进行安全性比较，并得出分析结果。

3 模型假设

1.数据来源于2014年中国互联网网络安全报告，假设样本选择的数据具有较强的代表性，能够反映相应指标的特征变化。

2.在分析过程中，假设在2014年这一段时间内，不会出现影响相应指标结果的政府政策或者特殊活动，数据具有稳定性。

3.假设指标彼此独立，之间没有相互过多的影响，便于进行分析。

4 重要参数说明

参数符号	含义
U	网络信息安全事件分类指标
V	系统信息安全等级
TR	从 U 到 V 的模糊变换
X1	通信行业安全信息发布数量
X2	CNVD 发布补丁数量
X3	CNVD 收录漏洞数量
X4	网络安全事件处理数量
X5	网络安全事件接受数量
y	网络安全性能评估重要指标
Z	网络安全性能评估综合指标
b	网络安全指标对综合指标的贡献率
a	网络安全指标对综合指标的累积贡献率
S	手机信息安全事件分类指标

5 模型的建立与求解

5.1 计算系统安全程度的数学模型

5.1.1 层次分析法[5]

层次分析法是一种简便、灵活又实用的多准则决策方法。它的算法步骤如下：

步骤 1 建立层次结构模型

步骤 2 构造判断矩阵对同一层次的指标两两比较其相对重要性得出相对重要性,得

出相对权值的比重的比值ω_i/ω_j，以此来构造判断矩阵，评判标准如下：

模糊数	模糊语义变量	隶属函数	描述
1	同样重要	（1，1，2）	根据经验和主观判断，准则 i 与 j 同样重要
3	稍微重要	（2，3，4）	根据经验和主观判断，准则 i 与 j 稍微重要
5	比较重要	（4，5，6）	根据经验和主观判断，准则 i 与 j 更加重要
7	明显重要	（6，7，8）	根据经验和主观判断，准则 i 与 j 相对明显重要
9	极其重要	（8，9，9）	根据经验和主观判断，准则 i 与 j 相比极其重要

由此得到的矩阵如下：

$$A=\begin{bmatrix} 1 & \frac{\omega_1}{\omega_2} & \cdots & \frac{\omega_1}{\omega_n} \\ \frac{\omega_2}{\omega_1} & 1 & \cdots & \frac{\omega_2}{\omega_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\omega_n}{\omega_1} & \frac{\omega_n}{\omega_2} & \cdots & 1 \end{bmatrix}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

上式判断矩阵 A 为$n\times n$的方阵，其主对角线为 1，ω_m/ω_n为相应指标的权重比例，通过a_{ij}表示，a_{ij}为i与j两因素相对权值的比值，满足$a_{ij}=1/a_{ji}$，$i\neq j$，$i,j=1,2,3...n$，$a_{ij}>0$，可按比例标度法对重要性程度赋值。

如果结果不满足$\sum_{i=1}^{n}a_n=1$，则将其归一化。

5.1.2 模糊综合评判法

模糊综合评判，是对具有多种属性的事物，或者说其总体优劣受多种因素影响的事物，做出一个能合理地综合这些属性或因素的总体评判。层次结构模型不论是多层的还是单层的，都要有两个关键的步骤：确定模糊关系 R，R 是从因素集 X 到评语集 Y 的一个模糊映射；计算模糊评判子集 B=A·R。

在指标个数较少的情况下，运用一级模糊综合评判，而在复杂的系统中，由于考虑的因素较多，又存在一定的层次性，就必须采用分层逐级评判的方法进行，即模糊多层次综合评判法。

一级模糊综合评判模型的建立，主要包括以下步骤：

● 确定因素集。对信息安全程度从多个方面进行综合评判，如信息安全事件发生的次数等，所有这些因素构成了评价指标体系集合，即因素集，记为

$$U=\{u_1,u_2,...u_n\}$$

用来表示某问题需要考虑的因素，

● 确定评语集。由于每个指标的评价值的不同，往往会形成不同的等级，如对信息安全程度的评价有很好（一级），较好（二级），一般（三级），较差（四级）等。有各种不同的决断构成的集合称为评语集，记为

$$V=\{v_1,v_2,...v_n\}$$

● 确定各因素的权重。一般情况下，因素集的各因素在综合评价中所起的作用是不相同的。由于对U中各因素有不同的侧重，需要对每个因素赋予不同的权重，它是U上的一个模糊子集，记为

$$A=\{a_1,a_2,...a_n\}$$

并且规定$\sum_{i=1}^{n}a_n=1$。

● 确定模糊综合判断矩阵。对于指标 u_i 来说，对各个评语的隶属度为 V 上的模糊子集。对指标 u_i 的评判记为

$$R=\{r_1,r_2,...r_n\}$$

各指标的综合模糊判断矩阵为

$$R=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}$$

它是一个从 U 到 V 的模糊关系矩阵。

● 综合评判。如果有一个从 U 到 V 的模糊关系 $R=(r_{ij})_{n\times m}$，那么利用 R 就可以得到一个模糊变换：

$$T_R:F(U)\to F(V)$$

由此变换，就可以得到综合评判结果$B=A\circ R$，记为

$$B=\{b_1,b_2,...b_n\}$$

如果评判结果$\sum_{i=1}^{n}b_n=1$将它归一化。

在模糊综合评判的上述步骤中，建立单因素评判矩阵 R 和确定权重分配 A，是两项关键性的工作，没有统一的格式可以遵循，一般采用统计实验或专家评分等方法求出。

5.1.3 案例分析

通过查找网上的信息安全的相关报告，现以 2014 年互联网网络安全报告[2]上的数据为例进行分析。

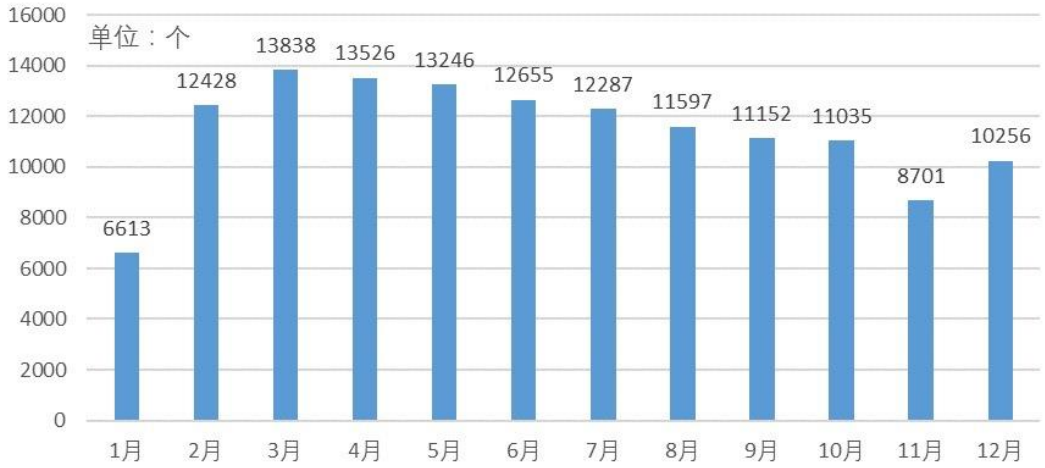

图 5.1　2014 年我国境内被篡改网站数量月度统计

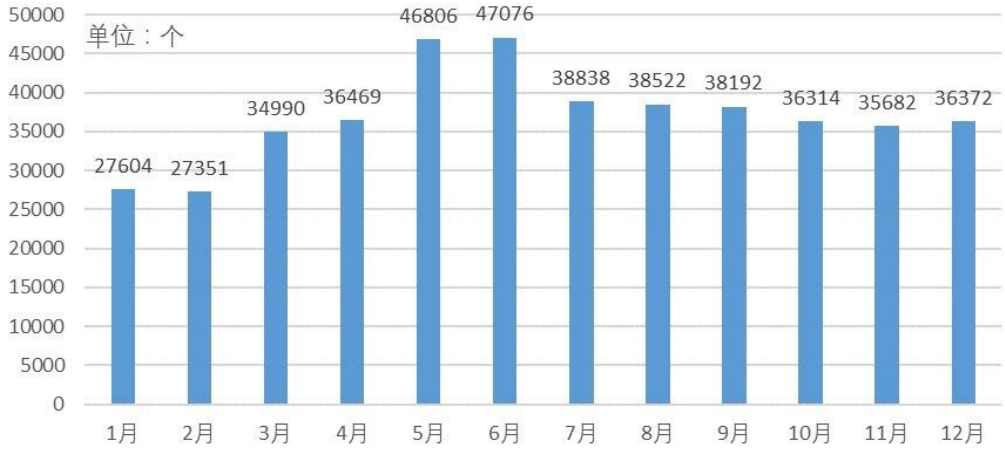

图 5.2　2014 年截获挂马网站数量月度统计

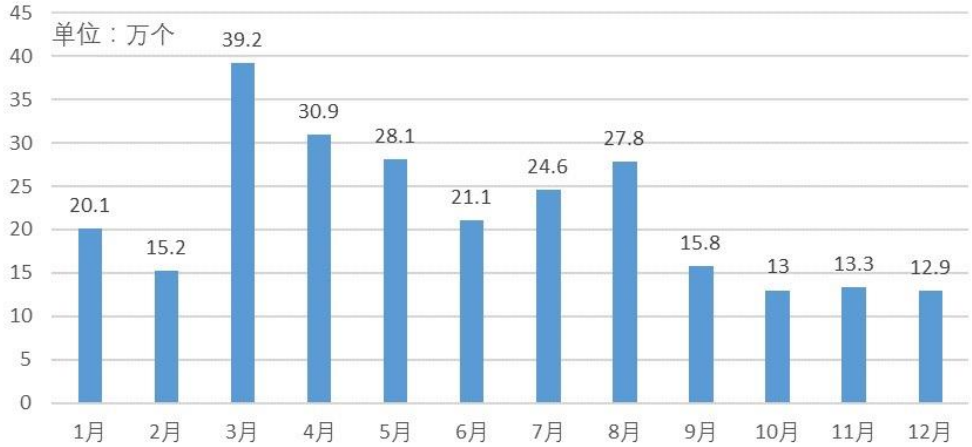

图 5.3　2014 年每月新增钓鱼网站数量月度统计

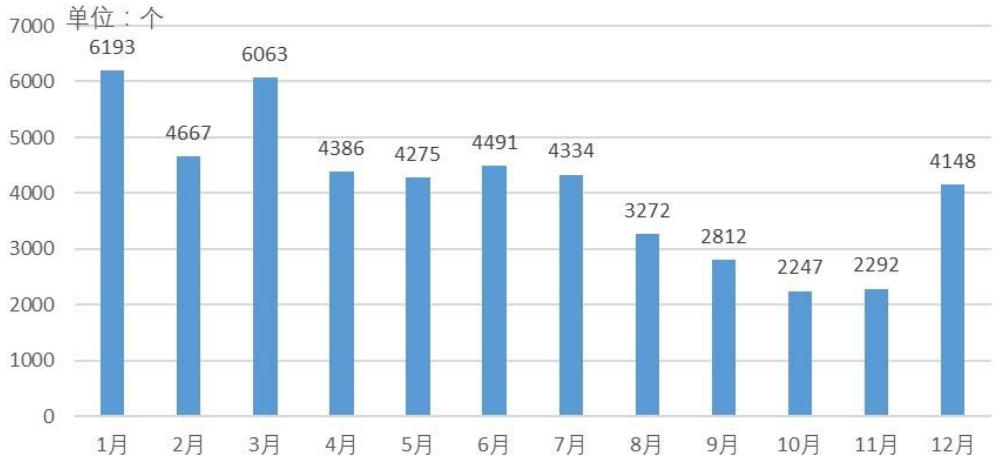

图 5.4　2014 年我国境内被植入后门网站数量月度统计

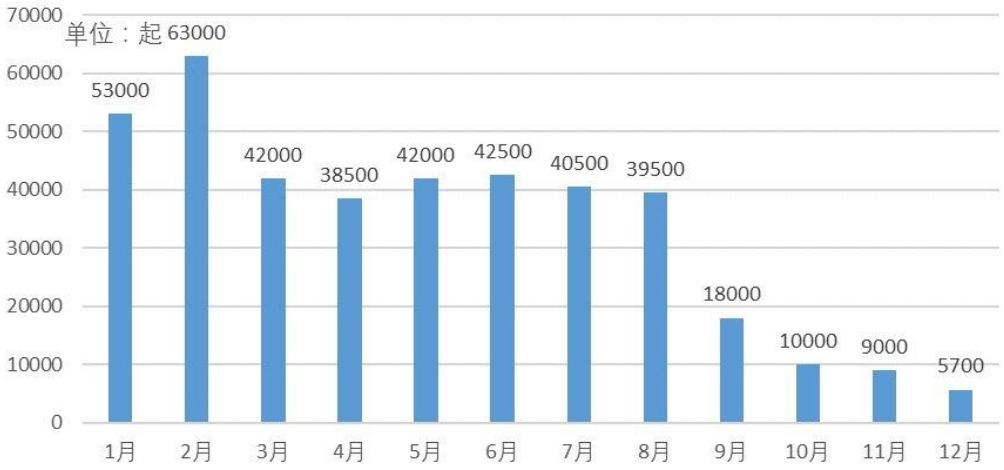

图 5.5　2014 年攻击流量在 1Gbit/s 以上 DDos 攻击事件数量月度统计[3]

下表是对反映 5 个安全指标的网站信息安全事件发生数量的数据统计表。

	2014 年我国境内网站被篡改数量月度统计 / 个	2014 年截获挂马网站数量月度统计 / 个	2014 每月新增钓鱼网站数量（网页仿冒）/ 万个	2014 年我国境内被植入后门网站数量月度统计 / 个	2014 年攻击流量在 1Gbit/s 以上 DDoS 攻击事件数量月度统计
1 月	6613	27604	20.1	6193	53000
2 月	12428	27351	15.2	4667	63000
3 月	13838	34990	39.2	6063	42000
4 月	13526	36469	30.9	4386	38500
5 月	13246	46806	28.1	4275	42000
6 月	12655	47076	21.1	4491	42500
7 月	12287	38838	24.6	4334	40500
8 月	11597	38522	27.8	3272	39500

9 月	11152	38192	15.8	2812	18000
10 月	11035	36314	13	2247	10000
11 月	8701	35682	13.3	2292	9000
12 月	10256	36372	12.9	4148	5700

表 5.1　2014 年网站信息安全事件发生数量

首先运用 MatlabR2015b 对上述的五项事件对应的指标进行层次分析。根据专家已有的评估标准两两对比确定其相对重要性，得到相对权重的比值，构造判断矩阵。计算出各个因素的权重，结果如下：

```
>> a=[1 1/3 1 1/3 7
3 1 3 3 7
1 1/3 1 1/3 5
3 1/3 3 1 9
1/7 1/7 1/5 1/9 1];
>> [v,d]=eig(a);
>> d(1,1)
ans =
    5.2966
>> v(:,1)
ans =
    0.2486
    0.7863
    0.2266
    0.5149
    0.0583
```

由于 $\sum_{i=1}^{n} a_n \neq 1$，将其归一化，得到五项指标权重分别为：

0.1355　0.4286　0.1235　0.2806　0.0318.

则权重的模糊向量确定为 $A=\{0.1355,0.4286,0.1235,,0.2806,0.0318\}$

接下来确定因素集为

$$U=\{\text{网站篡改},\ \text{网站挂马},\ \text{网页仿冒}\ ,\ \text{网站后门},\ \text{流量攻击}\}$$

评语集为

$$V=\{\text{很强},\ \text{较强},\ \text{中等}\ ,\text{较弱},\ \text{很弱}\}$$

其中定义安全程度的Ⅰ级为很强，Ⅱ级为较强，Ⅲ级为中等，Ⅳ级为较弱，Ⅴ级为很弱，因此安全程度被划分为Ⅰ~Ⅴ级。

然后根据信息安全事件发生的次数来划分各指标的等级。信息安全发生的次数越少，则系统的安全程度越高。

经过划分，得到模糊综合判断矩阵 R：

$$R=\begin{bmatrix} 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{2} & 0 \\ \frac{1}{6} & \frac{1}{12} & \frac{7}{12} & 0 & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{6} & \frac{1}{4} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{5}{12} & \frac{1}{12} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{12} & 0 & \frac{1}{2} & \frac{1}{6} \end{bmatrix}$$

结合上一步求得的 A 矩阵，带入 Matlab 求解，结果如下：

```
>> a=[0.1355 0.4286 0.1235 0.2806 0.0318]
a =
    0.1355    0.4286    0.1235    0.2806    0.0318
>> r=[0 1/6 1/3 1/2 0
1/6 1/12 7/12 0 1/6
1/4 1/6 1/4 1/6 1/6
1/6 1/6 5/12 1/12 1/6
1/4 1/12 0 1/2 1/6]
r =
         0    0.1667    0.3333    0.5000         0
    0.1667    0.0833    0.5833         0    0.1667
    0.2500    0.1667    0.2500    0.1667    0.1667
    0.1667    0.1667    0.4167    0.0833    0.1667
    0.2500    0.0833         0    0.5000    0.1667
>> b=a*r
b =
    0.1570    0.1283    0.4430    0.1276    0.1441
>>
```

最终得到模糊综合评判结果：

$B=(0.1570,\ 0.1283,\ 0.4430,\ 0.1276,\ 0.1441)$

取数值最大的评语作为综合评判结果，则该网络系统的安全程度评判结果为“中等”——III级。

5.2 信息安全度量指标的选取与计算

5.2.1 主成分分析法

主成分分析是一种数学变换方法。它把给定的一组变量 $x_1,x_2,...,x_k$，通过线性变换，转换成一组不相关的变量 $y_1,y_2,...,y_k$（两两相关系数为 0 的随机变量，或样本向量彼此相互垂直的随机变量）。在这种变换中，保持变量的总方差（$x_1,x_2,...,x_k$ 的方差之和）不变，同时，使 y_1 具有大方差，称为第 1 主成分；y_2 具有次大方差，称为第 2 主成分。依次类推，原来有 k 个变量，就可以转换出 k 个主成分。最后一个主成分具有的方差最

小，并且和前面的主成分都不相关。但在实际应用中，为了简化问题，通常不是找出 k 个主成分，而是找出 q(q<k)个主成分就够了，只要这 q 个主成分反映出原来 k 个变量的绝大部分的方差即可。

对于问题二：将衡量系统安全程度的指标定为以下五项，基于主成分分析法分析各项指标对系统安全程度的贡献率，从而选出贡献率最高的一项指标，计算出其值。

1) 系统预警能力

主要由预警时间和预警空间两个指标进行度量。预警时间指标可以从漏洞信息发布到利用此漏洞实施攻击实际发生的时间差,以及某种攻击方式的出现到某大规模蔓延之间的时间差统计得到。预警空间指标可以用预警信息发布时，被 度量对象已经遭受该类型攻击或脆弱性被利用以实施攻击的 空间范围来做出评价,为统计值。

2) 安全防护能力

指使用技术和管理的手段来保证信息和信息系统的保密性、完整性、真实性、可用性、可控性和不可否认性、可靠性等。该指标可以从研究对象单位时间内(如一年)基于某一个(多个)安全属性(如: 保密性)的安全事件的数量、事件起因分析、风险级别、造成的损失、潜在威胁估算等实施度量。同时，可运用渗透测试等手段进行实验分析。

3) 隐患发现能力

隐患发现能力指数主要由检测技术的有效性如漏报率和误报率)，检测制度的完备性和检测的实时性等指标合成。

4) 反应处置能力

反应处置是对发生的安全事件、破坏行为和过程，能及时做出处理,限制潜在的损失和破坏。可从不同规模(可分为大、中、小三类)安全事件的反应时间、技术手段、管理措施、处理效果等方面进行考评。

5) 信息安全费用

信息安全费用包括人力资源、时间资源和财务资源等内 容。人力资源主要包括人员培训、专职人员等; 时间资源包括信息系统的安全建设、安全运营和安全维护等所占用的时间因素; 财务资源包括购置一切信息安全软、硬件产业和服务的财务支出。

以下五项数据[2]可分别表征以上五项指标。

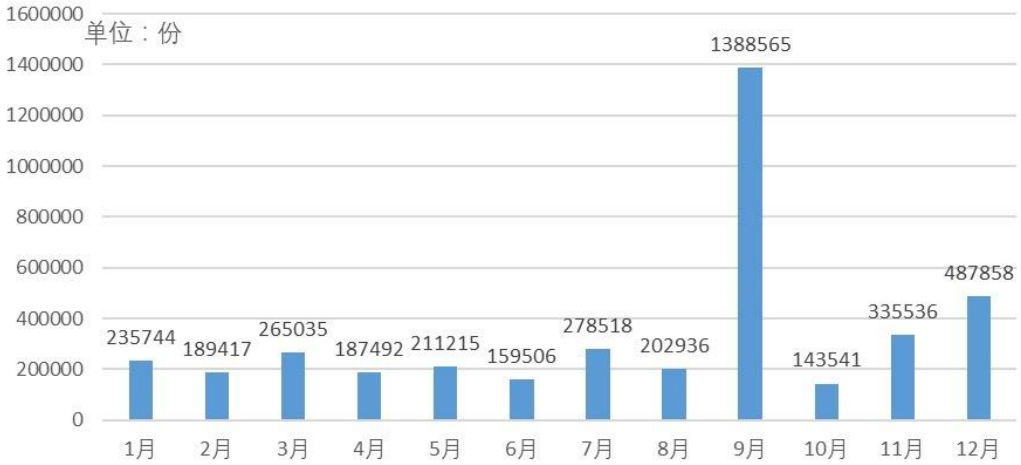

图 5.6　2014 年通信行业事件月度报送数量统计图

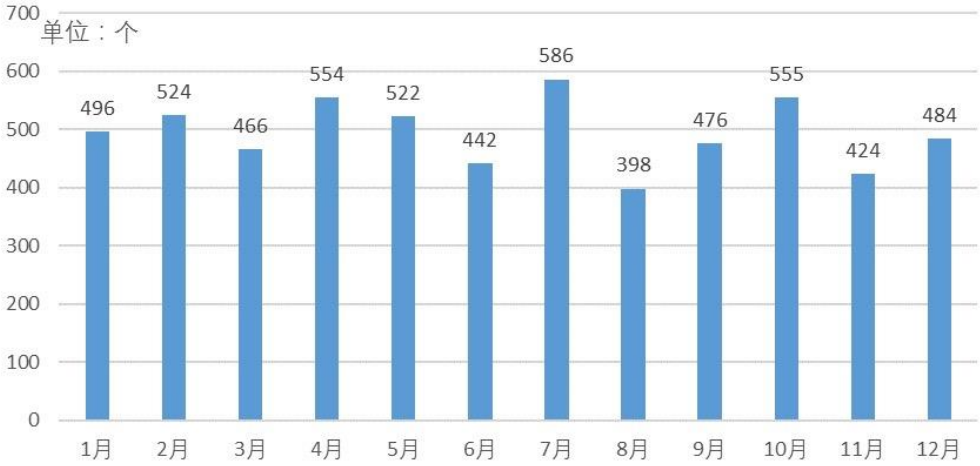

图 5.7　2014 年 CNVD 收录漏洞补丁数量统计图

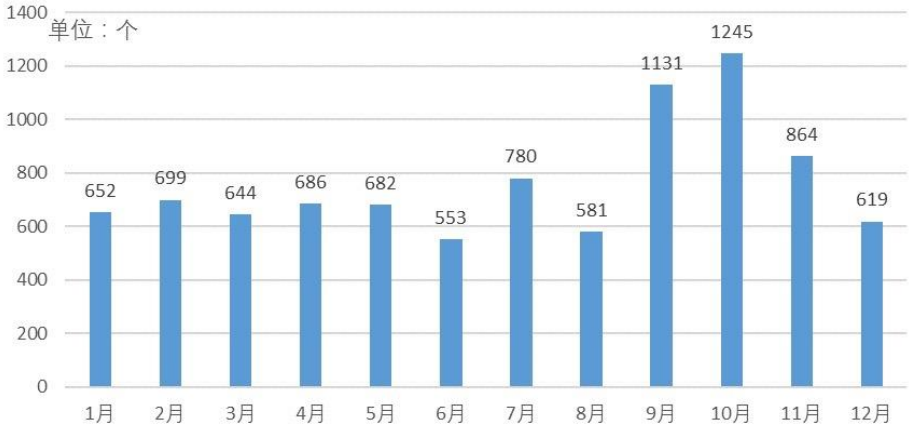

图 5.8　2014 年 CNVD 收录漏洞数量统计图

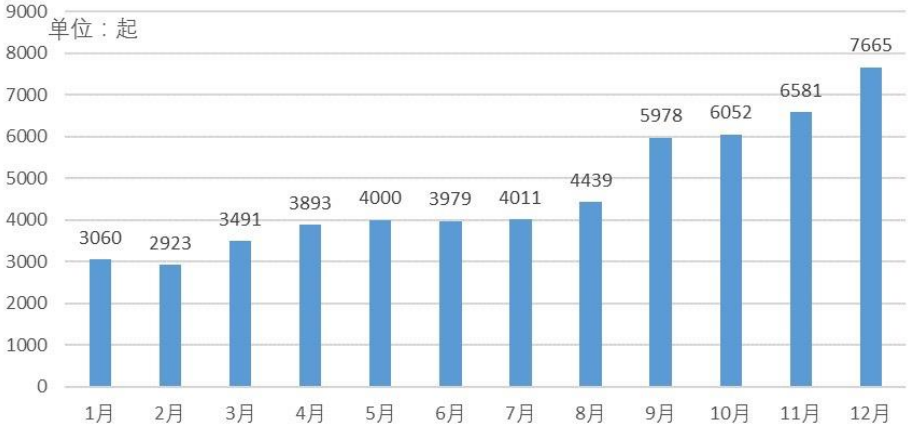

图 5.9　2014 年 CNCCRT/CC 网络安全事件处置数量统计图

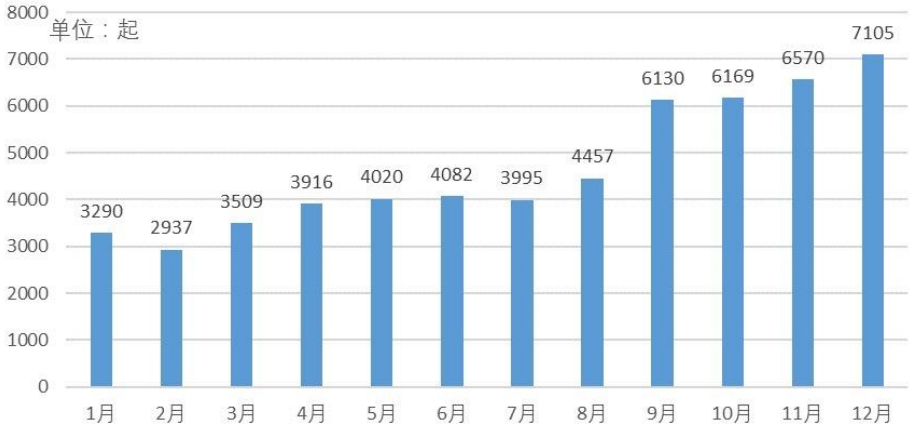

图 5.10　2014 年 CNCCRT/CC 网络安全事件接受数量统计图

下表是综合了反映以上 5 个指标的信息安全情况的数据统计表。

	通信行业安全信息发布情况/份	发布补丁数量/个	2014cnvd 收录漏洞数量/个	事件处理情况/起	安全事件接受情况/起
1 月	235744	496	652	3060	3290
2 月	189417	524	699	2923	2937
3 月	265035	466	644	3491	3509
4 月	187492	554	686	3893	3916
5 月	211215	522	682	4000	4020
6 月	159506	442	553	3979	4082
7 月	278518	586	780	4011	3995
8 月	202936	398	581	4439	4457
9 月	1388565	476	1131	5978	6130
10 月	143541	555	1245	6052	6169
11 月	335536	424	864	6581	6570
12 月	487858	484	619	7665	7105

表 5.2　2014 年信息行业安全情况统计表

用 X_1，X_2，X_3，X_4，X_5分别表示通信行业安全信息发布情况、发布补丁数量、2014cnvd 收录漏洞数量、事件处理情况、安全事件接受情况。用 i=1，2，...，12 分别表示 1 月、2 月、...、12 月，第 i 年 x_1，x_2，...，x_5的取值分别记作$[a_{i1}, a_{i2}, ..., a_{i5}]$，构成矩阵 $A=(a_{ij})_{17\times5}$。

基于主成分分析法[5]的评价步骤如下：

（1）对原始数据进行标准化处理。将各指标值a_{ij}转换成标准化指标$\tilde{a}_{ij}$，有

$$\tilde{a}_{ij}=\frac{a_{ij}-\mu_j}{s_j}, i=1,2,...,12, j=1,2,...,5,$$

其中：$a_{ij}=\frac{1}{12}\sum_{i=1}^{12}a_{ij}, s_j=\sqrt{\frac{1}{12-1}\sum_{i=1}^{12}(a_{ij}-\mu_j)^2}$，$j=1,2,...,5,$即$\mu_j, s_j$为第 j 个指标的样

本均值和样本标准差。对应地，称

$$\tilde{x}_j = \frac{x_j - \mu_j}{s_j}, j = 1,2,...,5$$

为标准化指标变量。

（2）计算相关系数矩阵 R。相关系数矩阵 $R = (r_{ij})_{5\times5}$，有

$$r_{ij} = \frac{\sum_{i=1}^{12} \tilde{a}_{ki} \cdot \tilde{a}_{kj}}{12-1}, i, j = 1,2,...,5$$

其中：$r_{ii} = 1, r_{ij} = r_{ji}, r_{ji}$ 是第 i 个指标与第 j 个指标的相关系数。

（3）计算特征值和特征向量。计算相关系数矩阵 R 的特征值 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_5 \geq 0$，及对应的标准化特征向量 $u_1, u_2, ..., u_5$，其中 $u_j = [u_{1j},\ u_{2j},\ ...,\ u_{3j}]^T$，由特征向量组成 5 个新的指标变量

$$\begin{aligned} y_1 &= u_{11}\tilde{x}_1 + u_{12}\tilde{x}_2 + ... + u_{15}\tilde{x}_5 \\ y_2 &= u_{12}\tilde{x}_1 + u_{22}\tilde{x}_2 + ... + u_{25}\tilde{x}_5 \\ &\vdots \\ y_5 &= u_{15}\tilde{x}_1 + u_{25}\tilde{x}_2 + ... + u_{55}\tilde{x}_5 \end{aligned}$$

其中：y_1 是第 1 主成分，y_2 是第 2 主成分，...，y_5 是第 5 主成分。

（4）选择个主成分，$p(p \leq 5)$ 计算综合评价值：

①计算特征值 $\lambda_j (j = 1,2,...,5)$ 的信息贡献率和累积贡献率。称

$$b_j = \frac{\lambda_j}{\sum_{k=1}^{5} \lambda_k}, j = 1,2,...,5$$

为主成分 y_j 的信息贡献率：而且称

$$\alpha_p = \frac{\sum_{k=1}^{p} \lambda_k}{\sum_{k=1}^{5} \lambda_k}$$

为主成分 $y_1, y_2, ..., y_p$ 的累积贡献率。当 α_p 接近于 1（α_p=0.85, 0.90, 0.95）时，则选择前 p 个指标变量 $y_1, y_2, ..., y_p$ 作为 p 个主成分，代替原来 5 个指标变量，从而可对 p 个主成分进行综合分析。

②计算综合得分：

$$Z = \sum_{j=1}^{p} b_j y_j$$

其中：b_j 为第 j 个主成分的信息贡献率，根据综合得分值就可进行评价。

根据公式，将上表所给的各项数据代入 Matlab 软件：

```
>> a=[235744 496 652 3060 3290
189417 524 699 2923 2937
```

```
265035 466 644 3491 3509
187492 554 686 3893 3916
211215 522 682 4000 4020
159506 442 553 3979 4082
278518 586 780 4011 3995
202936 398 581 4439 4457
1388565 476 1131 5978 6130
143541 555 1245 6052 6169
335536 424 864 6581 6570
487858 484 619 7665 7105]
a =
      235744         496         652        3060        3290
      189417         524         699        2923        2937
      265035         466         644        3491        3509
      187492         554         686        3893        3916
      211215         522         682        4000        4020
      159506         442         553        3979        4082
      278518         586         780        4011        3995
      202936         398         581        4439        4457
     1388565         476        1131        5978        6130
      143541         555        1245        6052        6169
      335536         424         864        6581        6570
      487858         484         619        7665        7105
>> a=zscore(a);
r=corrcoef(a);
[x,y,z]=pcacov(r)
x =
    0.4162    0.0353    0.8245   -0.3817    0.0051
   -0.0985    0.8255   -0.2210   -0.5078    0.0472
    0.4177    0.5290    0.1030    0.7255   -0.0936
    0.5586   -0.1489   -0.3852   -0.2456   -0.6760
    0.5749   -0.1238   -0.3353   -0.0990    0.7293
y =
    2.7146
    1.2582
    0.7090
    0.3168
    0.0014
z =
   54.2918
   25.1650
   14.1802
    6.3361
    0.0270
```

```
>> f=repmat(sign(sum(x)),size(x,1),1);
x=x.*f
x =
    0.4162    0.0353   -0.8245    0.3817    0.0051
   -0.0985    0.8255    0.2210    0.5078    0.0472
    0.4177    0.5290   -0.1030   -0.7255   -0.0936
    0.5586   -0.1489    0.3852    0.2456   -0.6760
    0.5749   -0.1238    0.3353    0.0990    0.7293
>> num=3;
df=a*x(:,[1:num]);
tf=df*z(1:num)/100;
[stf,ind]=sort(tf,'descend');
stf=stf',ind=ind'
stf =
    1.5374    1.5016    0.9872    0.7748    0.0099   -0.3186   -0.3746
-0.7488   -0.8257   -0.8276   -0.8488   -0.8669
ind =
    10     9    12    11     7     4     5     8     3     2     6     1
```

求得相关系数矩阵的前 5 个特征根及其贡献率如下表 5-3 所列：

序号	特征根	贡献率	累计贡献率
1	2.7146	54.2918	54.2918
2	1.2582	25.165	79.4568
3	0.709	14.1802	93.637
4	0.3168	6.3361	99.9731
5	0.0014	0.027	100

表 5.3　主成分分析结果

可以看出，前三个特征根的累计贡献率就达到 93%以上，主成分分析效果很好。下面选取前三个主成分进行综合评价。前三个特征根对应的特征向量见下表 5-4。

	X_1	X_2	X_3	X_4	X_5
第1特征向量	0.4162	-0.0985	0.4177	0.5586	0.5749
第2特征向量	0.0353	0.8255	0.5290	-0.1489	-0.1238
第3特征向量	-0.8245	0.2210	-0.1030	0.3852	0.3353

表 5.4　标准化变量的前 3 个主成分对应的特征向量

由此可得三个主成分分别为

$Y_1=0.4162X_1-0.0985X_2+0.4177X_3+0.5586X_4+0.5749X_5$

$Y_2=0.0353X_1+0.8255X_2+0.5290X_3-0.1489X_4-0.1238X_5$

$Y_3=-0.8245X_1+0.2210X_2-0.1030X_3+0.3852X_4+0.3353X_5$

分别以三个主成分的贡献率为权重，构建主成分综合评价模型

$Z=54.2918Y_1+25.1650Y_2+14.1802Y_3$

因此选取 Y_1 为指标，且计算 $Y_1=0.4162X_1-0.0985X_2+0.4177X_3+0.5586X_4+0.5749X_5$

5.3 手机系统安全性分析

在现代社会中，手机似乎已经成为了人们生活中必不可少的组成元素。手机作为网络终端的组成部分，其系统的安全与整个网络的系统安全息息相关。而且，由于随着手机支付功能的不断完善，手机安全从个人资料的保密问题上升到与个人资金有关的财产问题。因而，对手机信息系统进行安全性的分析有着必要性及重要性。

关于手机信息系统的安全风险一般可认为涉及物理环境及保障、硬件设施、软件设施和管理者四个方面,而软件设施的安全性评估也涉及五类风险因素,分别为:手机操作系统 M1、网络操作系统 M2、网络通信协议 M3、通用应用平台 M4 和网络管理软件 M5。

然而，对于我们手机用户而言，即作为移动终端的使用者，我们对手机安全性进行评估可利用更加直观的指标进行分析。根据《2014 年中国移动互联网安全报告》及《2014 年中国互联网网络安全报告》，我们可得如下数据：

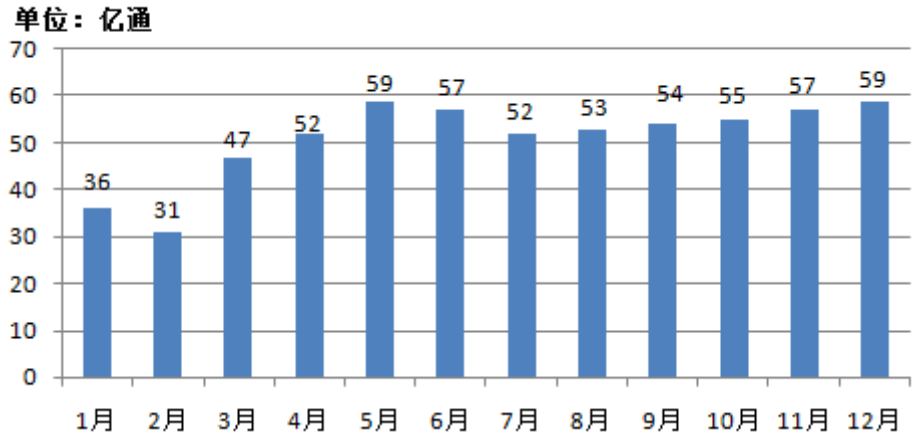

图 5.11　2014 年捕获新增恶意程序样本数月度统计

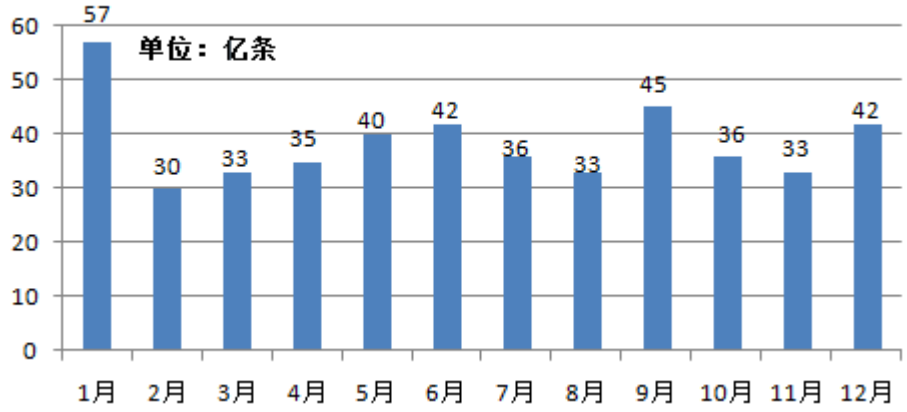

图 5.12　2014 年垃圾短信数月度分布

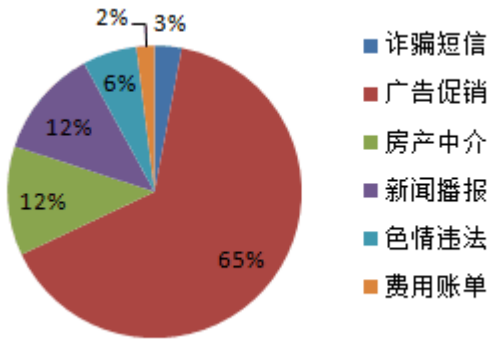

图 5.13 垃圾短信分类

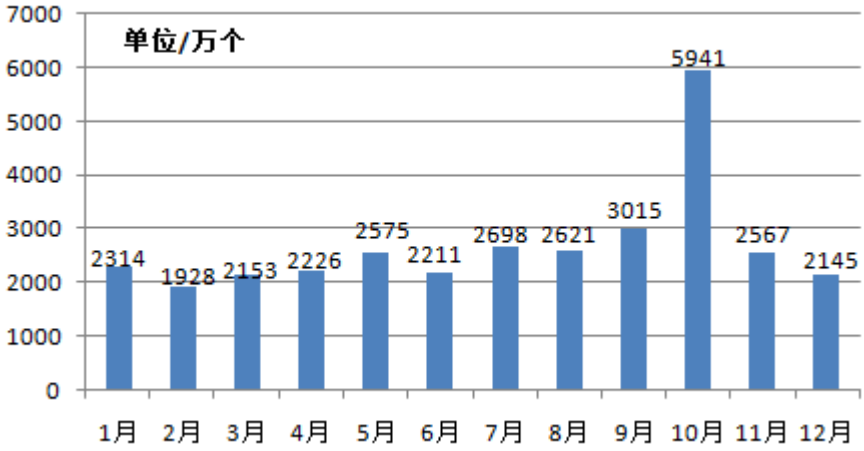

图 5.14 2014 年骚扰电话数月度分布

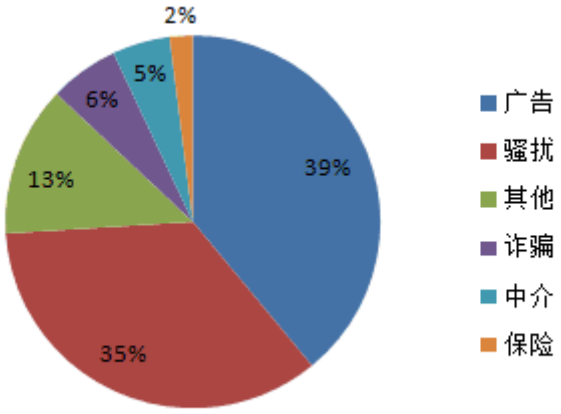

图 5.15 骚扰电话分类

将数据进行整理可得下表：

	恶意程序/万	非诈骗类垃圾短信/亿	诈骗短信/亿	骚扰电话/亿	诈骗电话/亿
1月	2314	55.29	1.71	12.60	2.16
2月	1928	29.10	0.90	10.85	1.86
3月	2153	32.01	0.99	16.45	2.82
4月	2226	33.95	1.05	18.20	3.12
5月	2575	38.80	1.20	20.65	3.54
6月	2211	40.74	1.26	19.95	3.42
7月	2698	34.92	1.08	18.20	3.12
8月	2621	32.01	0.99	18.55	3.18
9月	3015	43.65	1.35	18.90	3.24
10月	5941	34.92	1.08	19.25	3.30
11月	2567	32.01	0.99	19.95	3.42
12月	2145	40.74	1.26	20.65	3.54

表5.5　手机安全事件汇总表

首先可见，对于评价手机系统安全性的指标可定为恶意程序、垃圾短信、诈骗短信、骚扰电话、诈骗电话这五项。

运用 MatlabR2015b 对上述的对应的指标进行层次分析。根据专家已有的评估标准两两对比确定其相对重要性，得到相对权重比值，构造判断矩阵。计算出各个因素的权重，结果如下：

```
>>   a=[1 1/7 1/3 1/5 1
7 1 3 1 5
3 1/3 1 1/3 2
5 1 3 1 3
1 1/5 1/2 1/3 1];
[v,d]=eig(a);
d(1,1)
ans =
    5.0740
>> v(:,1)
ans =
    0.1128
    0.7185
    0.2639
    0.6165
    0.1458
```

由于上式计算得出的5个分量之和不为1，故将其归一化得

a=[0.0607,0.3868,0.1421,0.3319,0.0785]

接下来确定因素集为

S={恶意程序、垃圾短信、诈骗短信、骚扰电话、诈骗电话}

评语集为

$$V=\{\text{很强，}\ \text{较强，}\ \text{中等}\ \text{，较弱，很弱}\}$$

其中定义安全程度的Ⅰ级为很强，Ⅱ级为较强，Ⅲ级为中等，Ⅳ级为较弱，Ⅴ级为很弱，因此安全程度被划分为Ⅰ~Ⅴ级。

然后根据信息安全事件发生的次数来划分各指标的等级。信息安全发生的次数越少，则系统的安全程度越高。

经过划分，得到模糊综合判断矩阵 R：

$$R=\begin{bmatrix} \frac{1}{12} & \frac{9}{12} & \frac{1}{12} & 0 & \frac{1}{12} \\ \frac{1}{12} & \frac{6}{12} & \frac{1}{12} & \frac{3}{12} & \frac{1}{12} \\ \frac{4}{12} & \frac{4}{12} & \frac{3}{12} & 0 & \frac{1}{12} \\ 0 & \frac{2}{12} & \frac{5}{12} & \frac{5}{12} & 0 \\ \frac{1}{12} & \frac{1}{12} & 0 & \frac{4}{12} & \frac{6}{12} \end{bmatrix}$$

结合上一步求得的 A 矩阵，带入 Matlab 求解，结果如下

```
>> a=[0.0607 0.3868 0.1421 0.3319 0.0785]
a =
    0.0607    0.3868    0.1421    0.3319    0.0785
>> r=[1/12 9/12 1/12 0 1/12
1/12 6/12 1/12 3/12 1/12
4/12 4/12 3/12 0 1/12
0 2/12 0 5/12 5/12
1/12 1/12 0 4/12 6/12]
r =
    0.0833    0.7500    0.0833         0    0.0833
    0.0833    0.5000    0.0833    0.2500    0.0833
    0.3333    0.3333    0.2500         0    0.0833
         0    0.1667         0    0.4167    0.4167
    0.0833    0.0833         0    0.3333    0.5000
>> b=a*r
b =
    0.0912    0.3481    0.0728    0.2612    0.2267
```

最终得到模糊综合评判结果： b =[0.0912 0.3481 0.0728 0.2612 0.226]

取数值最大的评语作为综合评判结果，则该手机系统的安全程度评判结果为较强。

由上述结果可见，用互联网大数据结合模糊层次分析理论可以得出我们一般使用的手机系统都是较安全的。这是与一般事实相符合的。也就是说虽然生活中人们因为手机

系统的不安全性受到精神或财产损失的事情时有发生，但是，将这些事情放到整个社会环境下，在整个基数很大的情况下，对于整个手机系统安全性的评价还是很安全的。这也是体现了存在即是合理的原理。

既然已经肯定了手机系统整体的安全性，那么我们可以对手机单个的组成部分进行安全性评估。对于手机用户而言，手机整体的物理环境及保障、硬件设施、管理者三个方面可以视为恒定的参数，整个手机信息系统的安全性所起作用可以视为恒值 a。若设系统的安全性评估结果为 S，则 S 与 S1 呈正相关（其中 S1 为软件设施的安全性评估结果）。故对手机系统安全性的评估结果可仅考虑软件设施的安全性评估结果。尽管对于软件设施的评价存在手机操作系统 M1、网络操作系统 M2、网络通信协议 M3、通用应用平台 M4 和网络管理软件 M5，然而事实上，对于广大的手机用户本身，网络操作系统 M2、网络通信协议 M3、通用应用平台 M4 和网络管理软件 M5 这四点并不是可以有直观差异性感受的因素，因而，我们将选择手机操作系统 M1 作分析。

对于手机操作系统而言，存在着设备防火墙（device firewall）、内建功能安全性（built-in security）、数据保护（data protection）、授权机制（authentication）、虚拟化（virtualization）、同步支持性（support for ActiveSync）、设备擦除（device wipe）、手机设备管理（mobile device management）等安全项目。按照第二题中所涉及的基本变量：系统预警能力、安全防护能力、信息安全费用、反应处置能力、隐患发现能力，可将上述 8 个项目各自分类到所属变量类别下。

系统预警能力	设备防火墙
隐患发现能力	虚拟化
安全防护能力	内建功能安全性、数据保护、授权机制
信息安全费用	手机设备管理
反应处置能力	同步支持性、设备擦除

表 5.6　手机系统安全项目分类

由网络数据可得如下的数据表格

	系统预警能力	安全防护能力			隐患发现能力	反应处置能力		信息安全费用
	设备防火墙	功能安全性	数据保护	授权机制	虚拟化	同步支持性	设备擦除	手机设备管理
Blackberry OS	4	3	4	5	0	0	5	4
Apple OS	0	4	2	2	1	2	2	3
Windows Phone	0	4	3	3	0	3	3	1
Android	0	3	2	2	2	2	1	2

表 5.7　四大手机系统安全性对比表

对于上述数据采用题目二中所描述的模型，可得

```
a=[4 3 4 5 0 0 5 4
0 4 2 2 1 2 2 3
0 4 3 3 0 3 3 1
0 3 2 2 2 2 1 2 ]
a =
     4     3     4     5     0     0     5     4
     0     4     2     2     1     2     2     3
     0     4     3     3     0     3     3     1
     0     3     2     2     2     2     1     2
a=zscore(a);
r=corrcoef(a);
[x,y,z]=pcacov(r)
x =
    0.4189    0.1107   -0.0088    0.7300   -0.2542   -0.4021   -0.1203
0.1963
   -0.1976   -0.5537    0.6085    0.0352    0.1710   -0.1448   -0.3250
0.3564
    0.3924   -0.2264   -0.2869   -0.2221   -0.3580    0.4801   -0.4041
0.3757
    0.4138   -0.1141   -0.1973    0.0073    0.6918   -0.0666   -0.4147
-0.3490
   -0.2732    0.5603   -0.0801    0.1635    0.4206    0.2317   -0.1909
0.5578
   -0.3673   -0.3483   -0.2043    0.6140    0.0121    0.5069   -0.0046
-0.2606
    0.3985   -0.2493    0.0397    0.0737    0.3419    0.2108    0.7097
0.3272
    0.3032    0.3456    0.6772    0.0861   -0.0737    0.4768   -0.0672
-0.2921
y =
    5.5682
    1.8491
    0.5827
    0.0000
    0.0000
    0.0000
    0.0000
    0.0000
z =
   69.6021
   23.1139
    7.2840
```

```
    0.0000
    0.0000
    0.0000
    0.0000
    0.0000
f=repmat(sign(sum(x)),size(x,1),1);
x=x.*f
x =
    0.4189   -0.1107   -0.0088    0.7300   -0.2542   -0.4021    0.1203
0.1963
   -0.1976    0.5537    0.6085    0.0352    0.1710   -0.1448    0.3250
0.3564
    0.3924    0.2264   -0.2869   -0.2221   -0.3580    0.4801    0.4041
0.3757
    0.4138    0.1141   -0.1973    0.0073    0.6918   -0.0666    0.4147
-0.3490
   -0.2732   -0.5603   -0.0801    0.1635    0.4206    0.2317    0.1909
0.5578
   -0.3673    0.3483   -0.2043    0.6140    0.0121    0.5069    0.0046
-0.2606
    0.3985    0.2493    0.0397    0.0737    0.3419    0.2108   -0.7097
0.3272
    0.3032   -0.3456    0.6772    0.0861   -0.0737    0.4768    0.0672
-0.2921

num=3;
df=a*x(:,[1:num]);
tf=df*z(1:num)/100;
[stf,ind]=sort(tf,'descend');
stf=stf',ind=ind'
stf =
    2.3639   -0.1171   -0.7545   -1.4924
ind =
     1     3     2     4
```

由上述结果可见，黑莓系统的安全性能较好，WindowsPhone 次之，安卓系统最差。

对于此种现象，我们通过查阅资料可知：

黑莓系统由于企业级安全性和易管理性使该平台成为企业用户最可靠的选择，其企业服务器(BES)提供的许多功能和增强保护不能应用于使用黑莓互联网服务(BIS)的用户自备的手机。苹果的 iOS 的应用机制提供给用户很多保护，因为所有应用都是在一个公用的内存环境中沙盒化的。与此同时 iPhone 和 iPad 的硬件特性也提供了更好的安全性，它们不能加入存储卡。iOS 的易管理性也可与黑莓媲美，黑莓 IT 管理员可以完全控制他们的设备，而在 iOS 的用户只要提供权限，IT 部门也可以对他们手机中的项目进行设置。

Windows Phone 则是微软吸取过去的经验，用权限和隔离技术创建了沙盒进程，并基于系统中的策略系统定义了该腔室中的进程是否有访问权限，从而将 Windows Phone 构建为强健和安全的智能手机操作系统。

Android 存在的安全问题主要由于其没有统一的系统升级机制,即许多用户在长期使用缺乏保护的有漏洞的系统。尽管从一方面而言，它是一个特权分隔的操作系统，应用不经用户同意不能访问网络。每个应用运行在独立的沙盒中，它们的权限是用户分别授予的。但不幸的是，用户在急于使用应用的时候经常不看提示就同意授权，那些提示经常也是很不清晰的，当应用被授权以后，就可以为所欲为了。

对于我们日常生活中常用的 IOS 和安卓系统，IOS 由于其系统的密封性，在安全性上还是超过了安卓系统的。

6　模型的分析、缺陷与推广

1、对于系统安全程度的评价，我们我们主要利用的是一种层次分析法以及基于层次分析法的模糊层次分析法的评价方式。这两种方式在系统性的评价，即考虑多因素对结果影响的评价方面，具有着清晰、明确的自身优势，也很适合对多目标、多准则、多时期的系统模型进行评价。而且这两种方式对于定量数据的要求较少，因而对于实际问题的解决有着独特优势。这也从理论上说明我们选择这两种方式的结合进行系统安全程度的评价具有合理性。

2、在实际运用中，只需对于不同的被评价系统选择合乎该系统特色的指标便可以对此系统进行评价。但该评价模型的缺陷在于，只能进行定性的评价而无法进行定量评价。而且，整个评价模型的评价结果和选择的评价指标及人为确定的权重比率，使整个评价模型受人为影响较大。

例如在最后对手机系统进行安全性评估时，本文在多套指标中，选取了与生活联系较紧密的一套评价指标。这样的好处是对于模糊判断矩阵有更多的数据支持，但缺陷是评价过于单一化，也较为浅显。

3、为补充层次分析法及模糊层次分析法缺少的定量评价，我们引入了主成分分析法。该方法可对于原始的各单一指标进行权重的线性组合，得出多种的综合指标，并以总权重得出最重要的一项综合指标。该方法的优势明显在于其可通过定量计算，对于不同的系统安全性进行比较分析。但是，由于我们所建模型只选择了一项综合指标作为最终评价方式，因而对系统差异性较小的进行比较时可能出现较大误差。

例如在评价手机操作系统的安全性时，我们选择了 8 项代表性项目分为 5 大类进行主成分分析，最终可得出苹果系统安全性最高，安卓系统安全性最低的结论，但是对于塞班和微软手机系统则无法做出准确衡量。

总之，该模型简单合理，虽然在某些方面无法做到完全精准，但是其广泛的可行性对于系统安全程度的评价与比较还是有着重要意义。

7　参考文献

[1] 吕欣. 信息安全度量理论和方法研究〔A〕全国计算机安全学术交流会论文集(第二十二卷)[C]. 合肥：中国科学技术大学出版社，2007. 56-60

[2] 国家计算机网络应急技术处理协调中心.2014 年中国互联网网络安全报告[R].北京：人民邮电出版社, 2015.6

[3] 国家计算机网络应急技术处理协调中心.2015 年我国互联网网络安全态势综述， http://www.cert.org.cn/publish/main/upload/File/2015%20Situation.pdf，2016.5.23

[4] 姜启源，谢金星，叶俊.数学模型（第三版）[M].北京：高等教育出版社，2003：224.

[5] 司守奎，孙兆亮.数学建模算法与应用（第二版）[M].北京：国防工业出版社，2015：387-390.

[6] 百度手机卫士.2014 年中国移动互联网安全报告，http://wk.baidu.com/view/97f4609ee009581b6ad9eb22?pcf=2#modile.qq.com，2016.5.23

附录

本次建模使用了以下软件：MATLAB R2015b、Grapher、Excel、Numbers
其中，MATLAB 代码如下：

```
>> a=[1 1/3 1 1/3 7
3 1 3 3 7
1 1/3 1 1/3 5
3 1/3 3 1 9
1/7 1/7 1/5 1/9 1];
>> [v,d]=eig(a);
>> d(1,1)
>> v(:,1)·····1-1


>> a=[0.1355 0.4286 0.1235 0.2806 0.0318]
>> r=[0 1/6 1/3 1/2 0
1/6 1/12 7/12 0 1/6
1/4 1/6 1/4 1/6 1/6
1/6 1/6 5/12 1/12 1/6
1/4 1/12 0 1/2 1/6]
>> b=a*r·····1-2


>> a=[235744 496 652 3060 3290
189417 524 699 2923 2937
265035 466 644 3491 3509
187492 554 686 3893 3916
211215 522 682 4000 4020
159506 442 553 3979 4082
278518 586 780 4011 3995
202936 398 581 4439 4457
1388565 476 1131 5978 6130
143541 555 1245 6052 6169
335536 424 864 6581 6570
487858 484 619 7665 7105]
>> a=zscore(a);
r=corrcoef(a);
[x,y,z]=pcacov(r)
>> f=repmat(sign(sum(x)),size(x,1),1);
x=x.*f
>> num=3;
df=a*x(:,[1:num]);
tf=df*z(1:num)/100;
[stf,ind]=sort(tf,'descend');
```

```
stf=stf',ind=ind'……2


>>   a=[1 1/7 1/3 1/5 1
7 1 3 1 5
3 1/3 1 1/3 2
5 1 3 1 3
1 1/5 1/2 1/3 1];
[v,d]=eig(a);
d(1,1)
>> v(:,1)……3-1


>> a=[0.0607 0.3868 0.1421 0.3319 0.0785]
>> r=[1/12 9/12 1/12 0 1/12
1/12 6/12 1/12 3/12 1/12
4/12 4/12 3/12 0 1/12
0 2/12 0 5/12 5/12
1/12 1/12 0 4/12 6/12]
>> b=a*r……3-2


a=[4 3 4 5 0 0 5 4
0 4 2 2 1 2 2 3
0 4 3 3 0 3 3 1
0 3 2 2 2 2 1 2 ]
a=zscore(a);
r=corrcoef(a);
[x,y,z]=pcacov(r)
f=repmat(sign(sum(x)),size(x,1),1);
x=x.*f
num=3;
df=a*x(:,[1:num]);
tf=df*z(1:num)/100;
[stf,ind]=sort(tf,'descend');
stf=stf',ind=ind'……3-3
```