doi:10.11830/ISSN.1000-5013.201701019

卡尔曼熵值模型的网络安全态势估计

朱闻亚[1,2]

(1. 武汉大学 经济与管理学院,湖北 武汉 430000;
2. 义乌工商职业技术学院 机电信息学院,浙江 义乌 322000)

摘要: 针对网络安全态势估计问题,提出一种基于卡尔曼熵值模型的估计方法.根据熵值关联度,筛选出影响网络安全的关键因素,构建回归方程;构建网络安全评估的状态模型和测量模型.采用卡尔曼滤波对网络安全态势进行估计.结果表明:文中方法可以对网络安全态势作出精确估计.

关键词: 关联度;回归方程;安全估计;测量模型

中图分类号: TP 393.4 **文献标志码:** A **文章编号:** 1000-5013(2017)01-0101-04

Network Security Situation Assessment Based on Kalman Entropy Model

ZHU Wenya[1,2]

(1. School of Economics and Management, Wuhan University, Wuhan 430000, China;
2. School of Mechanical and Information, Yiwu Industrial and Commercial College, Yiwu 322000, China)

Abstract: Aiming at the problem of network security situation assessment, an estimation method is proposed based on Kalman entropy model. Key factors influencing the network security is selected according to the entropy correlation in order to construct regression equation, which help establish the state model and the measurement model of network security assessment. Then Kalman filter is used to estimate the network security situation. Results show that this method can accurately estimate the network security situation.

Keywords: correlation degree; regression equation; safety estimation; measurement model

为了提升网络使用安全水平,国家提出了网络信息安全战略[1].在网络安全问题应对早期阶段,3种最典型的方法[2-3]是脆弱性评估法、防火墙监控法和入侵检测法.随后,各种网络安全的监控和评估方法陆续出现.韦勇等[4]根据网络登陆访问和信息处理的相关日志文件,提出一种网络性能修正算法,进而构建一个适用于网络安全态势评估的全新模型.姜伟等[5]采用攻防博弈模型对网络安全程度进行评价,并提出最优主动防御策略以增加网络安全.黄同庆等[6]构建了一种实时的网络安全态势预测方法.刘玉岭等[7]构建了一种时空维度分析方法,以实现对网络安全态势的预测.任江伟等[8]提出一种信息融合的网络安全态势评估模型,从多个角度提取网络安全的表征信息,并将这些信息融合在一起作为网络安全水平的判断依据.赵颖等[9]提出网络安全的未来发展趋势在于数据的可视化、网络状态的可视化,让监控人员可以更加直观地发现当前网络所处的状态.本文根据已有研究成果,结合灰熵关联度和卡尔曼滤波,构建一种新的网络安全态势评估方法,以实现更加准确的网络安全态势估计.

收稿日期: 2016-11-25

通信作者: 朱闻亚(1980-),男,副教授,博士研究生,主要从事计算机软件理论、网络安全的研究.E-mail:zhuwenya2002@163.com.

基金项目: 浙江省教育厅高等教育教学改革项目(JG2015343)

1 网络安全态势估计方法

1.1 卡尔曼滤波模型

卡尔曼滤波模型的状态方程表示为

$$\boldsymbol{Y}(k+1)=\boldsymbol{G}(k+1,k)\cdot\boldsymbol{Y}(k)+\boldsymbol{U}_1(k). \tag{1}$$

式(1)中:$\boldsymbol{Y}(k)$用于表达网络安全系统在第 k 时刻所表现出来的状态;$\boldsymbol{G}(k+1,k)$用于表达网络安全系统从第 k 时刻所表现出的状态到第 $k+1$ 时刻状态的转移,也称为状态转移矩阵;$\boldsymbol{U}_1(k)$用于表达整个网络安全运行过程中的噪声或可能出现的误差[10].

卡尔曼滤波模型的量测方程表示为

$$\boldsymbol{z}(k+1)=\boldsymbol{B}(k)\cdot\boldsymbol{Y}(k+1)+\boldsymbol{U}_1(k). \tag{2}$$

式(2)中:$\boldsymbol{z}(k+1)$表达网络安全系统在第 $k+1$ 时刻所能观测到的信息;网络安全的状态向量 $\boldsymbol{Y}(k+1)$也是可以量测的.

对于 $k\geqslant 1$ 后的各个状态向量 $\boldsymbol{y}(i)$,如果 $i>n$,通过卡尔曼滤波模型的状态方程可以得到 n 时刻后的状态,再通过量测方程的观察,就可以估计出 n 时刻之后的网络安全状态信息.

1.2 网络安全态势的新信息估计

用 $\boldsymbol{M}(k)$表示第 k 个时刻的网络安全新信息,那么,对于网络安全态势 $\boldsymbol{z}(k)$,其新信息的表达式为

$$\boldsymbol{M}(k)=\boldsymbol{z}(k)-\boldsymbol{z}_1(k),\qquad k=1,2,\cdots. \tag{3}$$

式(3)中:$\boldsymbol{z}_1(k)$的计算需要借助最小二乘法,它是网络安全态势的最小二乘估计结果.

根据新信息理论,按照上述过程计算出网络安全态势新信息.

1.3 卡尔曼熵值模型网络安全态势估计方法

步骤 1 用 $\boldsymbol{z}(k)$表达第 k 个时刻的网络安全态势数据信息,那么,$\boldsymbol{z}(k+1)$就表达了第 $k+1$ 时刻的网络安全态势数据信息.采用灰度熵值的计算方法,通过比较关联度的大小确定影响网络安全态势的 m 个关键参数,这时 $\boldsymbol{y}_i(k)$就表达了第 k 个时刻关键参数 i 的信息,$\boldsymbol{y}_i(k+1)$就表达了第 $k+1$ 个时刻关键参数 i 的信息,进而可以建立一个网络安全态势和关键参数之间的回归方程,即

$$\left.\begin{aligned}
\boldsymbol{z}(k+1)&=b_{0,0}\boldsymbol{z}(k)+b_{0,1}\boldsymbol{y}_1(k)+\cdots+b_{0,m}\boldsymbol{y}_m(k)+\delta_0,\\
\boldsymbol{y}_1(k+1)&=b_{1,0}\boldsymbol{z}(k)+b_{1,1}\boldsymbol{y}_1(k)+\cdots+b_{1,m}\boldsymbol{y}_m(k)+\delta_1,\\
&\vdots\\
\boldsymbol{y}_m(k+1)&=b_{m,0}\boldsymbol{z}(k)+b_{m,1}\boldsymbol{y}_1(k)+\cdots+b_{m,m}\boldsymbol{y}_m(k)+\delta_m.
\end{aligned}\right\} \tag{4}$$

式(4)中:参数 $b_{0,0},b_{0,1},\cdots,b_{m,m}$和 $\delta_0,\delta_1,\cdots,\delta_m$ 均为回归系数,可以通过最小二乘法求得.

步骤 2 在网络安全态势和关键参数之间回归方程的基础上,根据卡尔曼滤波模型的基本原理,构建网络安全态势的卡尔曼状态方程和量测方程,分别为

$$\left.\begin{aligned}
\boldsymbol{Y}(k+1)&=\boldsymbol{G}(k)\cdot\boldsymbol{Y}(k)+\boldsymbol{U}_1(k),\\
\boldsymbol{z}(k+1)&=\boldsymbol{B}(k)\cdot\boldsymbol{Y}(k)+\boldsymbol{U}_2(k).
\end{aligned}\right\} \tag{5}$$

步骤 3 对 $\boldsymbol{Y}(0)$等向量信息进行数据初始化.

步骤 4 求取网络安全态势的新信息 $\boldsymbol{z}(k)$,表示为

$$\boldsymbol{M}(k)=\boldsymbol{z}(k)-\boldsymbol{B}(k)\boldsymbol{G}(k)\boldsymbol{Y}_i(k-1). \tag{6}$$

式(6)中:$\boldsymbol{B}(k)\boldsymbol{G}(k)\boldsymbol{Y}_i(k-1)$用于表达 $\boldsymbol{z}(k)$的一个估计值.

2 结果与分析

用 m_i 表达计算机 i 上的抗体数量,$m_{i,j}$表达计算机 i 上遭受到第 j 类攻击的抗体数量,θ_j 表达第 j 类攻击的可能造成的危险程度,ϑ_i 表达计算机 i 的重要程度,y_i 表达网络安全状态下计算机 i 上的抗体数量,那么,需要计算以下 3 种网络安全风险.

第 1 种风险:单台计算机遭受攻击的风险,其数学表达式描述为

$$S_{\mathrm{h}}=1-\frac{1}{1+\ln(\vartheta_i\mid m_i-y_i\mid+1)}. \tag{7}$$

第 2 种风险：整个网络遭受第 j 类攻击的风险，其数学表达式描述为

$$S_j^{\text{sys}} = 1 - \frac{1}{1 + \ln(\theta_j \sum \vartheta_i \mid m_{i,j} - y_i \mid + 1)}. \tag{8}$$

第 3 种风险：整个网络遭受所有可能攻击的风险，其数学表达式描述为

$$S_{\text{sys}} = 1 - \frac{1}{1 + \ln(\sqrt{\sum_i (\vartheta_i \mid m_i - y_i \mid)^2} + 1)}. \tag{9}$$

通过判断上述 3 类风险和对应的抗体浓度，就可以形成对计算机网络安全态势的大致判断. 上述 3 类风险在数值表达上可能存在较大的差异，为此，对其进行归一化处理，使这三类风险的数值均分布在 0 到 1 的数值区间上，其数学公式为

$$y_1 = \frac{y - y_{\min}}{y_{\max} - y_{\min}}. \tag{10}$$

式(10)中：$y_{\max}$为网络安全态势的极大值；$y_{\min}$为网络安全态势的极小值；y 为网络安全态势的当前值.

在上述处理的基础上，为灰度熵值关联度的计算选择 3 个常见的网络安全影响参数，即网络遭受的攻击强度、计算机网络流量和计算机网络流量的变化率，如表 1 所示.

表 1　网络安全的 3 个影响参数

Tab. 1　Three influencing parameters on network security

组别	网络安全态势	攻击强度	网络流量	流量变化率
第 1 组	0.19	0.102 28	0.143 93	0.169 3
第 2 组	0.20	0.104 59	0.692 57	0.038 8
第 3 组	0.22	0.115 46	0.212 84	0.621 7
⋮	⋮	⋮	⋮	⋮
第 25 组	0.44	0.171 92	0.152 99	0.522 9
第 26 组	0.41	0.148 33	0.177 50	0.168 3
第 27 组	0.42	0.142 71	0.156 53	0.226 3
⋮	⋮	⋮	⋮	⋮

结合表 1 数据，采用式(4)，计算网络安全态势影响因素中前 10，20，30 组的灰熵关联度，结果如表 2 所示. 由表 2 可知：攻击强度与网络安全态势的灰熵关联度最大.

表 2　各组数据的灰熵关联度

Tab. 2　Grey entropy correlation degree of group data

参数	前 10 组			前 20 组			前 30 组		
	攻击强度	网络流量	流量变化率	攻击强度	网络流量	流量变化率	攻击强度	网络流量	流量变化率
灰熵	2.301 4	2.285 1	2.292 7	2.335 8	2.291 1	2.302 4	3.336 9	3.285 3	3.310 6
最大差异熵	—	2.423 3	—	—	2.915 7	—	—	3.392 8	—
灰熵关联度	0.892 5	0.753 6	0.768 8	0.888 7	0.743 1	0.776 2	0.835 1	0.752 6	0.749 8

由表 2 的分析结果可知：强度攻击对于网络安全态势的灰熵关联最大，因此，选择它作为卡尔曼熵值模型的网络安全态势估计值.

以表 1 中 30 组攻击强度数据作为网络安全态势的表征数据，借助提出的基于卡尔曼熵值模型网络安全预态势估计方法进行数据拟合，拟合结果如图 1 所示. 图 1 中：e 为预测值与真实值的偏差；n 为数据个数；T 代表网络安全态势的真实值；P 代表基于卡尔曼熵值模型网络安全预态势估计方法得到的估计值. 由图 1 可知：两条曲线在第 6 个值后实现了比较好的拟合，这种状态一直持续到第 25 个值；随后，因为曲线 T 的剧烈变化，使得曲线 P 和曲线 T 有了一定的偏差，但趋势上一直拟合较好.

在拟合处理的基础上，进一步对后续 20 组数据进行预测，结果如图 2 所示.

由图 2 可知：基于卡尔曼熵值模型网络安全预态势估计方法准确地估计了未来 20 个时刻上的网络安全态势，与网络安全态势的真值以较小的偏差吻合在一起. 结果表明：经过 30 个数据的训练，基于卡尔曼熵值模型网络安全预态势估计方法已经适合本实验条件下的估计；基于卡尔曼熵值模型网络安全预态势估计方法具有理想的估计性能和估计精度.

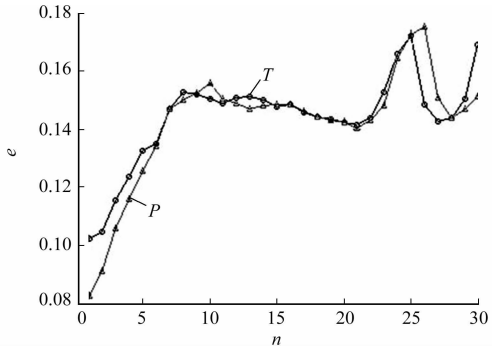

图1 前30组数据的拟合曲线

Fig. 1 Fitting curves of first 30 sets of data

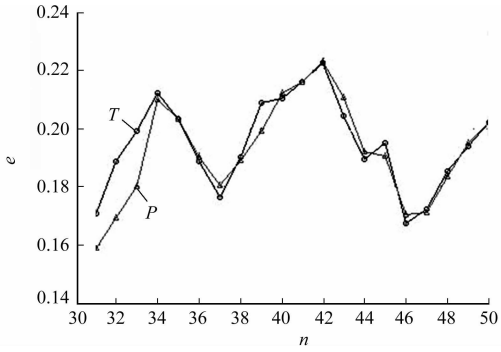

图2 后20组数据的预测曲线

Fig. 2 Fitting curves of last 20 groups of data

3 结束语

网络安全态势估计对于计算机网络安全具有重要意义.文中在卡尔曼滤波模型的基础上,构建一种卡尔曼熵值网络安全估计方法.首先,借助灰熵关联理论分析网络安全态势的影响因素;其次,利用关键影响因素构建网络安全态势估计的卡尔曼状态方程和卡尔曼量测方程;最后,采用卡尔曼滤波分析的过程执行对网络安全态势的估计.实验结果以攻击强度为网络安全态势的表征指标进行估计,估计结果显示:文中提出的基于卡尔曼熵值模型的网络安全态势估计方法,具有准确的估计精度和良好的估计性能,对于网络安全态势的预判具有较好的适用性.

参考文献:

[1] 韩文智.计算机文本信息挖掘技术在网络安全中的应用[J].华侨大学学报(自然科学版),2016,37(1):67-70.

[2] 林闯,汪洋,李泉林.网络安全的随机模型方法与评价技术[J].计算机学报,2005,28(12):1943-1956.

[3] 李方伟,张新跃,朱江,等.基于信息融合的网络安全态势评估模型[J].计算机应用,2015,35(7):1882-1887.

[4] 韦勇,连一峰.基于日志审计与性能修正算法的网络安全态势评估模型[J].计算机学报,2009,32(4):763-772.

[5] 姜伟,方滨兴,田志宏,等.基于攻防博弈模型的网络安全测评和最优主动防御[J].计算机学报,2009,32(4):817-827.

[6] 黄同庆,庄毅.一种实时网络安全态势预测方法[J].小型微型计算机系统,2014,35(2):303-306.

[7] 刘玉岭,冯登国,连一峰.基于时空维度分析的网络安全态势预测方法[J].计算机研究与发展,2014,51(8):1681-1694.

[8] 任江伟,韩跃龙.基于信息融合的网络安全态势评估模型[J].黑龙江科技信息,2015,46(9):353-362.

[9] 赵颖,樊晓平,周芳芳.网络安全数据可视化综述[J].计算机辅助设计与图形学学报,2014,26(5):687-697.

[10] MARSA-MAESTRE I,HOZ E D L,GIMENEZ-GUZMAN J M,*et al*. Design and evaluation of a learning environment to effectively provide network secureity skills[J]. Computers and Education,2013,69(4):225-236.

(**责任编辑**:黄晓楠 **英文审校**:吴逢铁)