

基于朴素贝叶斯分类器的网络安全态势评估方法

文志诚*, 曹春丽, 周 浩

(湖南工业大学 计算机与通信学院, 湖南 株洲 412007)

(* 通信作者电子邮箱 zcwen@mail.shu.edu.cn)

摘 要: 针对目前网络安全态势评估范围局限、信息来源单一、时空复杂度较高且准确性偏差较大等问题, 提出了一个朴素贝叶斯分类器的网络安全态势评估方法, 充分考虑了多信息源与多层次异构信息融合, 具有快速高效性, 从整体动态上展示出网络当前安全状况, 准确地反映了网络当前安全态势。最后利用网络实例数据, 对所提出的朴素贝叶斯分类器的网络安全态势评估模型和算法进行了验证, 实验结果表明了所提方法的正确性。

关键词: 朴素贝叶斯; 网络安全态势; 态势评估; 评估方法; 分类器

中图分类号: TP393.08 **文献标志码:** A

Network security situation assessment method based on Naive Bayes classifier

WEN Zhicheng*, CAO Chunli, ZHOU Hao

(College of Computer and Communication, Hunan University of Technology, Zhuzhou Hunan 412007, China)

Abstract: Concerning the problem that the current network security situation assessment has characteristics of limited scope, single information source, high time and space complexity and big deviation in accuracy, a new network security situation assessment method based on Naive Bayes classifier was proposed. It fully considered multi-information sources and fusion of multi-level heterogeneous information, and had the features of rapidity and high efficiency. It dynamically demonstrated the whole security state of the network, which could precisely reflect the network security situation. Finally, the proposed method was verified using the reality data from network and its validity was proved.

Key words: Naive Bayes; network security situation; situation assessment; assessment method; classifier

0 引言

Bass^[1]于1999年首次提出了网络态势感知(Cyberspace Situation Awareness, CSA)的概念, 并指出“基于信息融合的网络态势感知”将成为网络安全与管理的发展方向。网络安全态势感知是网络态势感知的一种, 从整体动态上把握网络当前的安全状况、预测未来发展趋势。网管人员根据宏观分析和预测结果, 及时作出决策, 将网络损失和风险降到最低。

网络安全态势评估主要研究整体上从网络中的实体赋予获取、理解和预测网络安全要素的能力, 并依此生成应对网络安全中的威胁策略, 为实现异构、泛化网络中各种安全实体的协同工作与信息融合, 构建无缝的网络安全体系提供一种新的思路^[2]。网络安全态势评估结果的合理性与真实性非常关键, 对于安全策略的制定具有深远的影响, 因为安全策略的制定与实施主要依赖于评估的可信程度。一般从底层决策指标开始, 逐层进行可信度评估, 直到最高层, 从而得到一个整体网络安全态势。

本文针对传统安全态势评估的范围局限、信息来源单一、时空复杂度较高且准确性偏差较大等问题, 将朴素贝叶斯分类器引入态势评估之中, 在深入研究评估方法的基础上, 提出基于朴素贝叶斯分类器的网络安全态势推理方法, 并结合网络三级分层的基础运行性、脆弱性与威胁性指数的推理进行逐层融合, 能快速高效地融合多层异构数据源, 给网管人员展现出一个宏观整体的网络安全状况。

1 相关工作

近年来, 态势评估已成为网络安全领域的研究热点, 国内外学者已提出了诸多理论与方法, 建立了有效的评估模型与方法, 如贝叶斯网络、模糊推理^[3]、博弈论、图模型等。文献[4]综合安全评估和大规模网络的研究成果, 提出了安全态势分层指标评估模型和25个候选指标, 并建立了评估的指标体系, 有机组织了候选指标并进一步抽象。文献[5]在基于隐马尔可夫模型(Hidden Markov Model, HMM)的网络安全态势评估中, 提出了以随机方式来获取所改进观测序列与转移矩阵的方法, 能有效表征网络的安全性。文献[6]提出了一种基于径向基函数(Radical Basis Function, RBF)神经网络的安全态势预测方法, 能在神经网络节点规模较小的环境下达到较为精确的预测效果, 非常适合大规模网络环境下的安全态势预测。文献[7]建立了一个基于反向传播(Back Propagation, BP)神经网络的安全评估模型, 并在此基础上利用RBF网络预测当前安全态势, 利用遗传算法对网络参数进行优化, 能较准确地预测安全态势。文献[8]提出了从大量安全报警信息中提取有效的威胁并识别出当前状态的方法, 并提出了一种时空关联分析的网络实时动态威胁识别与量化评估方法。文献[9]基于专家经验意见融合的推理过程, 实现了一种基于D-S(Dempster-Shafer)证据融合推理的网络安全态势评估方法, 并通过脆弱性指数实验对其进行了验证, 二者可以相互印证。文献[10]针对网络安全态势评估问题, 对多种已有态势评估方法进行了比较

收稿日期: 2015-03-04; 修回日期: 2015-04-13。 基金项目: 国家自然科学基金资助项目(61073186)。

作者简介: 文志诚(1972-), 男, 湖南东安人, 副教授, 博士, 主要研究方向: 软件工程、网络安全; 曹春丽(1980-), 女, 湖南常宁人, 讲师, 硕士研究生, 主要研究方向: 项目管理; 周浩(1981-), 男, 湖南湘潭人, 讲师, 硕士研究生, 主要研究方向: 软件工程。

和分析,提出了一种基于神经网络的评估方法,设计出一种基于BP神经网络的安全态势评估方法。

由于攻击信息具有不确定性和多变性等特点,在态势评估的方法论上仍存在诸多争论点,评估结果的正确性和合理性更饱受质疑。传统的评估方法主要问题在于信息来源单一、感知范围局限、时空复杂度较高及准确性不高且不易操作。

2 网络安全态势

网络安全态势 从网络基础运行性 (*Runnability*)、网络脆弱性 (*Vulnerability*) 和网络威胁性 (*Threat*) 三个方面通过评估函数融合而成,即存在评估函数 h , 有: $SA = h(Runnability_{net}, Vulnerability_{net}, Threat_{net})$, 从三个不同角度向网管人员展示当前网络安全整体状况。

网络的基础运行 由网络上所有组件的基础运行性评估函数融合而成,即存在评估函数 g_1 , 有: $Runnability_{net} = g_1(Runnability_{com,1}, Runnability_{com,2}, \dots, Runnability_{com,m})$, 其他两个维度如网络脆弱性与网络威胁性情形类似,都由组件相应的评估函数 g_2 和 g_3 融合而成。

组件的基础运行性 由与运行信息相关的决策变量 X 通过评估函数融合而成,即存在评估函数 f_1 , 有: $Runnability_{com} = f_1(X_1, X_2, \dots, X_n)$, 其他两个维度如组件脆弱性与组件威胁性形成类似,由相应的评估函数 f_2 和 f_3 融合而成。

计算机网络结构中存在大量的主机、服务器、路由器、防火墙和入侵检测系统 (Intrusion Detection System, IDS) 等各种网络硬件,称之为组件。每个维度都有组件和网络之分,如基础运行性,有组件基础运行性和网络基础运行性,而网络基础运行性则由 N 个组件基础运行性评估融合生成,为了区别术语网络 (network) 与组件 (component), 相应的标识符以下标 net 和 com 作为区别。

本文主要确定三个评估函数 f, g, h , 一旦确定了此三个评估函数,当采集到决策变量 X 值时,容易通过相应的评估函数逐层融合,最后获得整个网络安全态势 SA 。其中,评估函数 f 分为 f_1, f_2 和 f_3 , 评估函数 g 分为 g_1, g_2 和 g_3 。评估函数 g 和 f 通过朴素贝叶斯分类器来实现,而评估函数 h 则由各项指标经验加权而成。

3 朴素贝叶斯分类器构建

3.1 朴素贝叶斯分类器

在朴素贝叶斯分类模型中,用一个 n 维特征向量 X 来表示训练样本数据,设类集合 C 有 m 个不同的取值,则时间复杂度为 $O(m * n)$ 。输入到朴素贝叶斯分类器是一个 n 维向量 $X \in \mathbf{R}^n$, 而 X 分类器的输出是一个类别标签集合 $Y = \{c_1, c_2, \dots, c_k\}$ 。当给定一个输入 n 维向量 $x \in X$, 则分类器给出其所属于的类别标签 $y \in Y$ 。这里, x, y 分别是集合 X 和 Y 上的随机变量,分类器样本训练集为 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, $P(X, Y)$ 表示输入变量 X 与输出变量 Y 的概率联合分布。

朴素贝叶斯分类器对 $P(X = x | Y = c_k)$ 作了较强的假设,也即条件独立性假设,各个决策变量独立同分布。有:

$$P(X = x | Y = c_k) = P(X^{(1)} = x_1, X^{(2)} = x_2, \dots, X^{(n)} = x_n | Y = c_k) =$$

$$\prod_{j=1}^n P(X^{(j)} = x_j | Y = c_k)$$

由贝叶斯定理,可计算 $P(Y = c_k | X = x)$ 的值,有:

$$P(Y = c_k | X = x) = \frac{P(X = x | Y = c_k)P(Y = c_k)}{\sum_k P(X = x | Y = c_k)P(Y = c_k)} = \frac{P(Y = c_k) \prod_j P(X^{(j)} = x_j | Y = c_k)}{\sum_k P(Y = c_k) \prod_j P(X^{(j)} = x_j | Y = c_k)}; k = 1, 2, \dots, K$$

朴素贝叶斯分类器具有简单和有效的分类模型^[11],假设各决策变量独立,参数易于获取且推理结果比较近似等特点,在网络安全态势评估上具有先天优势。

3.2 决策变量离散化

决策变量 X 可取离散和连续型两种观测值,而朴素贝叶斯分类器中的节点都使用离散值,为了便于应用,需把连续型离散化。根据实际意义,连续型决策变量 X 可离散化为“高、中、高、中、中低、低”或“2、1、0、-1、-2”五等值。若决策变量本来就是离散型取值,则按实际情况取这五等值。

引理1 设连续型 X 服从高斯分布,即 $X \sim N(\mu, \sigma^2)$, 则 $Z = (X - \mu)/\sigma \sim N(0, 1)$, μ 表示 X 的数学期望, σ^2 表示方差。

根据概率论知识,把决策变量 X 的历史大样本观测值划分为五个互不相交的区间 $SS_i: (-\infty, \mu - 3\sigma) \cup (\mu + 3\sigma, +\infty), (\mu - 3\sigma, \mu - 2.5\sigma) \cup (\mu + 2.5\sigma, \mu + 3\sigma), (\mu - 2.5\sigma, \mu - 2\sigma) \cup (\mu + 2\sigma, \mu + 2.5\sigma), (\mu - 2\sigma, \mu - \sigma) \cup (\mu + \sigma, \mu + 2\sigma)$ 和 $[\mu - \sigma, \mu + \sigma]$ 。

经计算,五个区间 $SS_i (i = 1 \sim 5)$ 对应的概率 $PS_i (i = 1 \sim 5)$ 分别为 0.26%、0.98%、3.32%、27.18% 和 68.26%,也就是连续型决策变量 X 取“-2、-1、0、1、2”时对应的概率。

在实际应用中,当监测到决策变量 X 值时,由引理1高斯分布标准化后,观察 Z 值落入五个区间 SS_i 的情况,确定决策变量 X 离散化为“-2、-1、0、1、2”中的某个相应值。

3.3 决策变量的遴选

在实际应用中,有必要遴选出一些具有典型代表性的指标,剔除一些与安全态势评估不相关的、冗余的指标,形成网络安全态势评估所需的决策变量。

计算两个决策变量 x_i 和 x_j 的相关系数:

$$\rho_{x_{ij}} = Cov(x_i, x_j) / \sqrt{D(x_i) * D(x_j)}$$

$Cov(x_i, x_j)$ 为 x_i 和 x_j 的协方差,其中:

$$Cov(x_i, x_j) = E\{[x_i - E(x_i)][x_j - E(x_j)]\} = E(x_i x_j) - E(x_i)E(x_j)$$

根据第3.2节指标离散化的方法,每个连续型观测指标可以离散化为五等。在某一个时间段监测若干个数据,以出现的频率近似它们的概率,代入其相关概率公式中计算。给定一个任意实数 $0 < \varepsilon < 1$, 若相关系数 $|\rho_{x_{ij}}| < \varepsilon$, 认为 x_i 和 x_j 不相关,两个决策变量都可保留;否则,说明相关,产生冗余,可剔除一个影响不大的决策变量。

3.4 构建朴素贝叶斯分类器

决策变量 X 是一个向量,每个分量对应于朴素贝叶斯分类器一个具体的叶子节点 x_i , 取离散或连续型两种观测值;本文需要构建两类朴素贝叶斯分类器,一类是组件级的朴素贝叶斯分类器,如图1所示,由三个子分类器构成,分别代表三

个评估函数 f_1 、 f_2 和 f_3 ; 另一类是网络级的朴素贝叶斯分类器, 如图2所示, 也由三个子分类器构成, 分别代表三个评估函数 g_1 、 g_2 和 g_3 。

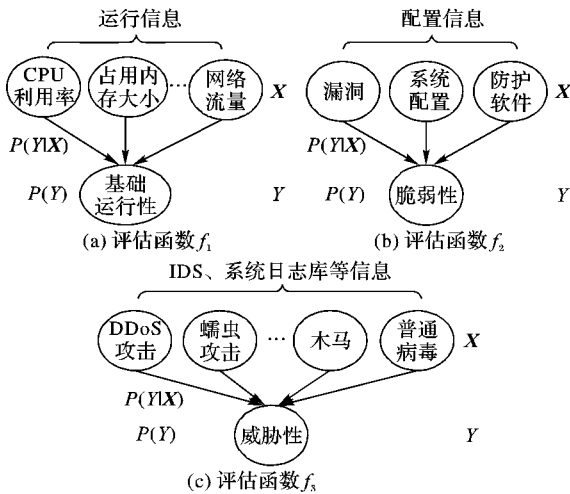


图1 组件级朴素贝叶斯分类器 f

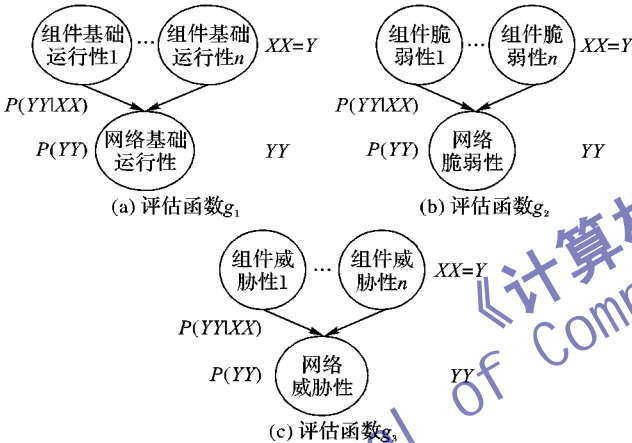


图2 网络级朴素贝叶斯分类器 g

在图1的组件级朴素贝叶斯分类器中, 三类相关指标看成决策变量 X , 而三个类别看成 Y , 其中 X 和 Y 都取五等离散值, 也就是说决策变量 X 的分量 x_i 可以指CPU利用率、占用内存大小、网络流量等, 可取五等离散值, 而类别 Y 的分量 y_i 可以指基础运行性、脆弱性、威胁性, 也取五等离散值。

图2的网络级朴素贝叶斯分类器中, 存在 n 个组件, 任一个组件的一维作为决策变量 XX , 而网络的三个类别看成 YY , 它们共同构成一个朴素贝叶斯分类器。注意, 图2中决策变量的 XX 就是图1的类属 Y , 也即图1的评估函数 f 是图2评估函数 g 的基础。

在组件级朴素贝叶斯分类器 f 中, 当采集到决策变量 X 的值时, 经过离散化预处理, 通过训练好的朴素贝叶斯分类器 f , 把目前状态推理分类给适当的类 Y , 具有一定的概率 $P(Y)$, Y 取五等离散值, 五个概率之和为1; 再由网络级朴素贝叶斯分类器 g , 把目前状态分类给适当的类 YY , 也具有一定的概率 $P(YY)$, YY 取五等离散值, 五个概率之和为1。

3.5 参数确定

经上述方法, 构建两类朴素贝叶斯分类器, 若要能在实际上应用, 必须要获取相应条件概率 $P(Y|X)$ 和 $P(YY|XX)$, 一般通过大样本的参数学习得到。以图1的朴素贝叶斯分类器 f 为例, 当采集到决策变量 X 连续型值后, 经离散化预处理, 取

相应的五等化值 $X^{(j)} = x_j$; 对于类别 Y 的采集一般通过专门软件如360安全防护软件等, 获得其推荐值, 再通过五等离散化类别 c_k ; 通过参数学习确定 $P(Y|X)$ 。如图1所示的朴素贝叶斯分类器 f , 通过大样本参数学习, 只需要训练估计 $P(Y = c_k)$ 与 $P(X^{(j)} = x_j | Y = c_k)$ ($1 \leq i \leq n, 1 \leq k \leq m$) 的值即可, 从而可对决策变量 X 分类为 Y :

$$P(Y = c_k | X = x) = \frac{P(X = x | Y = c_k)P(Y = c_k)}{\sum_k P(X = x | Y = c_k)P(Y = c_k)}$$

这里, 经过大样本观察, 有:

$$P(Y = c_k) = s_k/s$$

$$P(X^{(j)} = x_j | Y = c_k) = s_{kj}/s_k$$

其中: s_k 为样本训练集中类别为 c_k 的样本数, s 为样本总数, s_{kj} 为样本训练集中类别为 c_k 且属性取值 x_j 的样本数。

4 安全态势评估

4.1 组件级态势评估

组件级态势评估函数 f , 通过如图1所示的朴素贝叶斯分类器来实现的。当采集到一组决策变量的值 X , 经过分类器 f 得到它们所属类别 Y , 各类别具有一定的概率, 表示为:

$$P(Y) = P(Y = c_k | X = x)P(X = x) = \frac{P(X = x | Y = c_k)P(Y = c_k)}{\sum_k P(X = x | Y = c_k)P(Y = c_k)}; \\ c_k = 2, 1, 0, -1, -2 \quad (1)$$

式(1)表示, 决策变量 X 取定值时, 经朴素贝叶斯分类器推理, 类属 $Y = c_k$ 具有一定的概率 $P(Y)$ 。也就是说, 图1的三个朴素贝叶斯分类器 f , 每一个类别都具有五个 c_k 对应的概率 $P(Y = c_k)$, 它们是图2所示的朴素贝叶斯分类器的基础 (因为 $XX = Y$)。

4.2 网络级态势评估

网络级态势评估函数 g 通过如图2所示的朴素贝叶斯分类器来实现, 以评估函数 f 为基础。在图1中, 当采集到决策变量 X 值经朴素贝叶斯分类器 f , 网络上每个组件上基础运行性、脆弱性和威胁性都具有五个类别及相应的概率, 以组件基础运行性为例, 令:

$$P(XX) = P(Y) = P(Y = c_k | X = x); \\ c_k = 2, 1, 0, -1, -2 \quad (2)$$

式(2)中, X 可取“CPU利用率、占用内存大小、子网流量变化率、子网数据流总量、子网内不同大小数据包的分布等”, Y 为“基础运行性”。

在图2的朴素贝叶斯分类器评估函数 g 中, 有:

$$P(YY) = P(XX = xx)P(YY = c_k | XX = xx) = \left[P(YY = c_k) \prod_j P(XX^{(j)} = xx_j | YY = c_k) \right. \\ \left. P(XX^{(j)} = xx_j) \right] / \left[\sum_k P(YY = c_k) \prod_j P(XX^{(j)} = xx_j | YY = c_k) \right. \\ \left. P(XX^{(j)} = xx_j) \right]; c_k = 2, 1, 0, -1, -2 \quad (3)$$

从式(3)中, 可得出网络基础运行性、网络脆弱性与网络威胁性三维中每维取五等离散化值的概率 $P(YY = c_k)$, 再作为4.3节图3网络安全态势评估函数 h 的基础。

4.3 网络安全态势评估

如图3所示,网络安全态势 SA 由网络基础运行性、网络脆弱性与网络威胁性三维通过评估函数 h 向上融合生成。

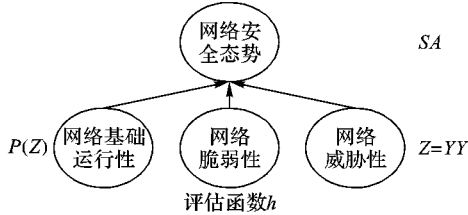


图3中的决策变量 Z 其实就是图2中的类属 YY ，为了便于叙述,用 Z 表示决策变量 YY 。经过图2的朴素贝叶斯分类器推理,可得每个维度都有五种离散型概率取值,令:

$$P(Z = c_k) = P(YY = c_k); c_k = 2, 1, 0, -1, -2 \quad (4)$$

由于网络安全态势 SA 由三个维度通过评估 h 融合生成,而每个维度由五等加权生成,以网络基础运行性 $Runnability_{net}$ 为例,根据经验,它的实值可以定义如下:

$$Runnability_{net} = 100 * [P(Z = 2) + 0.5 * P(Z = 1) - 0.5 * P(Z = -1) - 5 * P(Z = -2)] \quad (5)$$

由于网络安全态势值需取0~100的实值,所以式(5)中乘上了100。按此方法计算网络基础运行性接近实际,因为评估网络安全态势,主要看位于“高”时的概率,也要突出位于“低”和“中低”时的情况,而当位于“中”时的概率可以忽略不计。

本节从网络的基础运行性、网络的脆弱性与网络的威胁性再向上通过评估函数 h 最终生成网络的安全态势 SA 。有:

$$SA = h(Runnability_{net}, Vulnerability_{net}, Threat_{net}) = \eta_1 Runnability_{net} + \eta_2 Vulnerability_{net} + \eta_3 Threat_{net} \quad (6)$$

可根据经验确定式(6)中权值参数 η_i 的值。网络安全态势中,基础运行性表征网络正常运行,居主导地位,所占比重应该最大,可取值为0.5;而其他两项也有可能网络导致网络安全态势降低,因此可各占比重0.25,即可取: $\eta_1 = 0.5, \eta_2 = 0.25, \eta_3 = 0.25$,这三个权值 η 的取定具有经验性,可参考专家的经验意见。 SA 结果取0~100的实值,为当前网络安全态势,从底层逐步通过评估函数 f, g 和 h 生成。

4.4 评估算法

4.4.1 朴素贝叶斯分类器参数学习算法

输入 决策变量 X 大样本观察数据;

输出 朴素贝叶斯分类器。

$s \leftarrow$ 决策变量 X 样本总数

let $s_k = 0, s_{kj} = 0,$

for every s

if $Y = c_k$ then $s_k = s_k + 1$

if $X^{(j)} = x_j$ then $s_{kj} = s_{kj} + 1$

endfor

compute every $P(Y = c_k) = s_k / s$

compute every $P(X^{(j)} = x_j | Y = c_k) = s_{kj} / s_k$

output parameter $P(Y)$ and $P(X|Y)$

4.4.2 网络安全态势评估算法

输入 决策变量 X 一次观察数据;

输出 网络安全态势 SA 。

采集一组决策变量 X 实时观测值,并离散化五等

for every Y in $\{Runnability, Threat, Vulnerable\}$ and c_k in $\{2, 1, 0, -1, -2\}$

$-1, -2\}$

$P(Y) \leftarrow P(Y = c_k | X = x) * P(X = x)$

endfor

let $XX = Y$

for every XX in $\{Runnability, Threat, Vulnerable\}$ and c_k in $\{2, 1, 0, -1, -2\}$

$P(YY) \leftarrow P(XX = xx) * P(YY = c_k | XX = xx)$

endfor

for RVT in $\{Runnability_{net}, Vulnerability_{net}, Threat_{net}\}$

$RVT_n \leftarrow 100 * [P(Z = 2) + 0.5 * P(Z = 1) - 0.5 * P(Z = -1) - 5 * P(Z = -2)]$

endfor

compute $SA \leftarrow h(Runnability_{net}, Vulnerability_{net}, Threat_{net}) =$

$0.5 * Runnability_{net} + 0.25 * Vulnerability_{net} + 0.25 * Threat_{net}$

output SA

5 仿真实验

本章采用 Matlab 7.0 进行仿真实验,实验数据主要来源于:一类是通过开发一个安装在各个网络组件上的软件监测得到的实时数据;一类来源于 Snort 入侵检测系统中的观测数据,并将各类恶意网络流量的数据按照预先规则注入到正常流量中,来获得实验中所需要的异常数据。

在一个设定的10s时间内,动态采集2000个大样本作为离散化的历史数据,当所采集的每个决策变量为大样本数据时(样本量足够大),计算其样本的数学期望 μ 与方差 σ^2 ,按照引理1,为每个连续型决策变量 x_i 划分为五个离散取值区域 SS_k ,每个区域有相应的概率 $PS_k (k = 1, 2, 3, 4, 5)$ 。

经过组件2000个大样本数据参数学习,获得朴素贝叶斯分类器 f 的参数 $P(Y|X)$ 近似值,以决策变量 X 为 CPU 利用率及类属 Y 为基础运行性为例,得到表1的参数。对于图1来说,有多少个决策变量 X ,就有多少个这样的参数表1。

表1 朴素贝叶斯分类器的参数表

$P(Y X)$		组件维度基础运行性 $Y/\%$				
		高	中高	中	中低	低
CPU 利用率	高	76.23	18.77	3.62	0.98	0.40
	中高	10.18	73.27	10.72	4.32	1.51
	中	4.32	9.62	73.17	9.77	3.12
	中低	0.58	7.32	11.22	69.46	11.42
	低	0.26	1.08	3.62	23.71	71.33

在异常情况下,组件不安装任何防病毒软件,且对此组件施实木马和蠕虫等病毒攻击,会对各类决策变量产生影响,CPU 利用率、内存使用情况及网络流量等明显增加。经异常数据不断流入,网络中存在一定数量的异常情况组件,通过决策变量采集、五等离散化后,组件经遴选后的三类决策变量值如表2所示,表示某个时刻该组件上所有决策变量取值。网络上多少个组件,在某个时刻 t 时就有多个这样的参数表2。

决策变量 X 取值如表2所示,经图1所示的三个评估函数 f_1, f_2 和 f_3 融合后,得到如表3所示的一个组件三个维度的概率。网络中有多少个组件,则就有多少个参数表3。

当网络上 N 个组件各自经评估函数 f 融合后,再经图2所示的三个评估函数 g_1, g_2 和 g_3 融合,得到如表4网络级三维的概率。一个网络上只有一个参数表4。

根据表4的取值,网络级三维如网络基础运行性、网络脆弱性与网络威胁性由式(5)计算,可得三维数值为(28.010, 46.625, 0),再经融合函数 h 加权,得 $SA = 25.66$ 。

表2 经遴选后的决策变量所取离散值

组件基础运行性决策变量五等值		组件脆弱性决策变量五等值		组件威胁性决策变量五等值	
决策变量	取值	决策变量	取值	决策变量	取值
CPU 利用率	1	网络漏洞数目及等级	1	报警数目	-1
内存使用情况	1	系统配置	1	DDoS	-1
子网关键设备平均存活时间	2	防护软件是否安装	1	蠕虫攻击	-1
子网流量变化率	0	关键设备漏洞数目及等级	1	木马和普通病毒数目	-2
子网数据流总量	0	子网内安全设备数目	1	子网带宽使用率	-1
子网不同大小数据包的分布	1	子网内各关键设备开放端口	2	子网数据流入量	-1
子网内存存活关键设备数目	1			子网流入量增长率	0
子网平均无故障时间	2				

表3 朴素贝叶斯分类器 f 融合组件情况

CPU 利用率	组件基础运行性维度的概率/%	组件脆弱性维度的概率/%	组件威胁性维度的概率/%
高	20.56	20.84	1.83
中高	43.08	63.06	4.11
中	28.32	11.13	15.11
中低	5.88	3.82	60.18
低	2.16	1.15	18.77

表4 朴素贝叶斯分类器 g 融合网络三维情况

CPU 利用率	网络基础运行性维度的概率/%	网络脆弱性维度的概率/%	网络威胁性维度的概率/%
高	21.16	22.55	1.81
中高	41.22	62.57	6.18
中	29.36	10.18	20.03
中低	6.12	3.62	62.22
低	2.14	1.08	10.76

经过多次决策变量 X 数据观测,根据上述三级评估函数 f, g, h 数据融合,绘出如图4所示的网络安全态势图,反映出本时间段内的安全态势波动情况,给网络管理员一个整体宏观的展现,以便及时调整相应的安全策略。

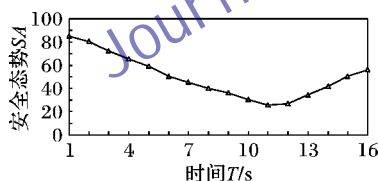


图4 网络安全态势图

6 结语

本文提出了一个基于朴素贝叶斯分类器的网络安全态势评估方法,给出了解决网络安全与管理的一个尝试方案,充分考虑了多信息源与多层次异构信息融合,从整体动态上生成网络当前安全态势,准确地反映了网络当前安全状况,能提高网管员对整个网络运行状况的全局认知与理解,当发现安全态势异常时,辅助指挥员及时准确地作出高层决策,弥补当前网管的不足。

本文的难点在于朴素贝叶斯网的构建以及数据的获取,今后的研究工作包括完善网络安全态势评估方法,进一步提高算法的效率,研究更全面的安全态势因子及其表示方法。

参考文献:

- [1] BASS T. Intrusion detection systems and multisensor data fusion [J]. Communications of the ACM, 2000, 43(4): 99-105.
- [2] JAKOBSON G. Mission cyber security situation assessment using

impact dependency graphs [C]// Proceedings of the 2011 14th International Conference on Information Fusion. Piscataway: IEEE, 2011: 1-8.

- [3] ZHAO J, ZHOU Y, SHUO L. A situation awareness model of system survivability based on variable fuzzy set [J]. Telkomnika: Indonesian Journal of Electrical Engineering, 2012, 10(8): 2239-2246.
- [4] WANG J, ZHANG F, FU C, et al. Study on index system in network situation awareness [J]. Journal of Computer Applications, 2007, 27(8): 1907-1909. (王娟, 张凤荔, 傅骅, 等. 网络态势感知中的指标体系研究[J]. 计算机应用, 2007, 27(8): 1907-1909.)
- [5] XI R, YUN X, ZHANG Y, et al. An improved quantitative evaluation method for network security [J]. Chinese Journal of Computers, 2015, 38(4): 749-758. (席荣荣, 云晓春, 张永铮, 等. 一种改进的网络安全态势量化评估方法[J]. 计算机学报, 2015, 38(4): 749-758.)
- [6] LI F, ZHENG B, ZHU J, et al. A method of network security situation prediction based on AC-RBF neural network [J]. Journal of Chongqing University of Posts and Telecommunications: Natural Science Edition, 2014, 26(5): 576-581. (李方伟, 郑波, 朱江, 等. 一种基于 AC-RBF 神经网络的网络安全态势预测方法[J]. 重庆邮电大学学报: 自然科学版, 2014, 26(5): 576-581.)
- [7] XIE L, WANG Y. New method of network security situation awareness [J]. Journal of Beijing University of Posts and Telecommunications, 2014, 37(5): 31-35. (谢丽霞, 王亚超. 网络安全态势感知新方法[J]. 北京邮电大学学报, 2014, 37(5): 31-35.)
- [8] LYU H, PENG W, WANG R, et al. A real-time network threat recognition and assessment method base on association analysis of time and space [J]. Journal of Computer Research and Development, 2014, 51(5): 1039-1049. (吕慧颖, 彭武, 王瑞梅, 等. 基于时空关联分析的网络实时威胁识别与评估[J]. 计算机研究与发展, 2014, 51(5): 1039-1049.)
- [9] TANG C, TANG S, QIANG B. Assessment and validation of network security situation based on DS and knowledge fusion [J]. Computer Science, 2014, 41(4): 107-110. (唐成华, 汤申生, 强保华. DS 融合知识的网络安全态势评估及验证[J]. 计算机科学, 2014, 41(4): 107-110.)
- [10] XIE L, WANG Y, YU J. Network security situation awareness based on neural network [J]. Journal of Tsinghua University: Science and Technology, 2013, 53(12): 1750-1760. (谢丽霞, 王亚超, 于巾博. 基于神经网络的网络安全态势感知[J]. 清华大学学报: 自然科学版, 2013, 53(12): 1750-1760.)
- [11] YAO P. Integration algorithm research based on Naive Bayes [D]. Guangzhou: South China University of Technology, 2013: 13. (姚沛津. 基于朴素贝叶斯的集成算法研究[D]. 广州: 华南理工大学, 2013: 13.)