基于加权熵的访问控制策略安全性分析研究

王 超,陈性元
(解放军信息工程大学,河南郑州 450004)

摘 要: 为解决访问控制策略的安全性分析问题,提出了一种基于信息熵的策略量化分析理论.首先,根据信息论中加权熵的知识定义了策略安全熵,提出了非授权访问行为的最大不确定性计算方法.然后,分别给出了典型访问控制策略的一维安全熵和 N 维安全熵,并对结果进行了证明.最后,依据安全熵分析了典型访问控制策略的安全性.

关键词: 加权熵;策略安全熵;安全策略;自主访问控制策略;强制访问控制策略
中图分类号: TN911.23 **文献标识码:** A **文章编号:** 0372-2112 (2013)01-0047-05
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2013.01.009

An Approach for Security Analysis to Access Control Policy Based on Entropy-Weigh

WANG Chao, CHEN Xing-yuan
(*The PLA Information Engineering University, Zhengzhou, Henan* 450004, *China*)

Abstract: To resolve the problem of security analysis to access control policy, a quantitative analysis theory is proposed based on information entropy. Firstly, the policy security entropy is defined according to Entropy-Weigh, and computing way for uncertainty of unauthorized access behaviors is bring forward. Then, the security entropy of one dimension and N dimensions of typical policies are given, and the results are proved. Finally, the lackage and change trend of typical policies' security are analyzed.

Key words: entropy-weigh; policy security entropy; access control policy; DAC policy ; MAC policy

1 引言

访问控制策略是用户如何使用信息系统中信息资源的规则、指南和定义,是进一步制定控制规则、安全程序的必要基础[1].典型的访问控制策略主要有自主访问控制策略和强制访问控制策略,但是无论自主访问控制策略还是强制访问控制策略,都只是对如何控制用户的访问行为进行了定义和说明,而访问控制策略对非法行为的控制力度如何?能解决哪些安全问题?安全性与用户和资源数量、访问次数的关系是什么?如何比较不同访问控制策略的优略?目前,由于缺少针对访问控制策略的安全性量化分析和度量方法,这些问题并未有明确和科学的答案.这给信息系统的管理者选择和应用合适的访问控制策略或安全机制造成了混乱和困难.本文借鉴信息熵对事物不确定测度的思想,提出安全熵的概念,为访问控制策略的量化分析提供了科学方法.

2 安全熵

2.1 信息熵的相关知识

熵这一概念最初用于热力学,美国数学家香农将其引入信息论[2,3],提出了信息熵的概念.信息熵用于信息无序程度的度量.信息熵理论可以在工程科学和社会科学的诸多领域得到应用[4~6],已有学者将熵引入到对信息安全风险和事件不确定性的量化分析上[7~10].也有学者[11]基于信息熵理论对经典的 BLP 访问控制模型进行了量化分析,使用条件熵设置安全门限来度量模型的安全性,给出了"下向信息流"安全的条件,并证明了该条件下系统仍保持其保密性,但该方法仅限于对"下向信息流"方面.

(1)离散随机信源的信息熵

对于离散型随机信源(变量)X,其符号集为 $A:a_i(i=1\ 2,\cdots,q)$,q 为符号集的个数,事件 a_i 发生的概率记为 $P(a_i)$,其概率空间$[X,p(x)]$如下:

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_q \\ p(a_1) & p(a_2) & & p(a_q) \end{bmatrix}$$

该离散型随机信源的信息熵为:

$$H(X) = -\sum_{i=1}^{q} p(a_i)\log p(a_i)$$

其中,$p(a_i) \geqslant 0(i=1,2,\cdots,q)$且 $\sum_{i=1}^{q} p(a_i) = 1$

收稿日期:2012-02-29;修回日期:2012-07-30
基金项目:国家 973 重点基础研究发展计划(No.2011CB311801);河南省科技创新人才计划(No.114200510001)

(2)N 维扩展随机信源的信息熵

一系列离散随机变量构成随机矢量,即由多个离散变量构成的随机变量序列 $\boldsymbol{X}=(X_1\cdots X_i\cdots X_N)$,其中每个随机变量 $X_i(i=1,2,\cdots,N)$.若多维随机序列中各维离散随机变量的概率分布都相同,且相互之间是无依赖的、统计独立的,则 N 维随机矢量的联合概率分布满足

$$\begin{aligned}P(\boldsymbol{X})&=P(X_1X_2\cdots X_N)\\&=P_1(X_1)P_2(X_2)\cdots P_N(X_N)\end{aligned}$$

若不同时刻的离散随机变量取值于同一符号集 A:$\{a_1,\cdots,a_q\}$则有

$$P(x=a_i)=P(a_{i_1},a_{i_2},\cdots,a_{i_N})=\prod_{i_k=1}^{q}P(a_{i_k})$$

式中,a_i 是 N 维随机矢量的一个取值,即 $a_i=(a_{i_1},a_{i_2},\cdots,a_{i_N})$,而 $P(a_{i_k})$ 是符号集 A 的一维概率分布.

根据信息熵的定义,N 次扩展信源的熵

$$\begin{aligned}H(\boldsymbol{X})=H(X^N)&=-\sum_{X^N}p(\boldsymbol{X})\log p(\boldsymbol{X})\\&=-\sum_{X^N}p(a_i)\log p(a_i)\end{aligned}$$

其中 $\sum_{X^N}$ 为对所有 N 维随机矢量求和的简写.

(3)加权熵

香农定义的信息熵是撇开人的主观因素的,它只是概率的函数.但在实际场合中,各随机事件虽以一定的概率发生,而各种事件的发生对人们有着不同的价值和效用.为了把主观价值和主观意义引进信息的测度中,对每一个事件 a_i 指定一个非负实数 $\omega_i\geqslant 0(i=1,2,\cdots,q)$,这组实数称为事件的权重.若信源 X 的权重分布$[X,\omega_i]$为

$$\begin{bmatrix}X\\\omega\end{bmatrix}=\begin{bmatrix}a_1&a_2&&a_q\\\omega_1&\omega_2&\cdots&\omega_q\end{bmatrix}$$

则信源 X 的加权熵为

$$H(X)=-\sum_{i=1}^{q}\omega_i p(a_i)\log p(a_i)$$

2.2 安全熵

定义 1(安全熵) 非授权访问行为的最大不确定性测度.

我们将系统中的访问事件作为随机变量看待,则加权熵可用来描述访问事件的不确定性测度.由于安全熵是对非授权事件的整体测度,可令非授权事件的权值为 1,合规事件(符合安全策略的事件)的权值为 0.由此得到的加权熵反映了“非授权访问事件”的不确定性测度,为了与香农的信息熵有所区别,本文称之为安全熵.

众所周知,实际的信息系统中信息失泄往往是由多次访问行为共同作用的结果.因此,为了反映多个访问行为对信息安全的影响,可将访问事件看作由多个离散事件构成的事件序列 $\boldsymbol{X}=(X_1\cdots X_i\cdots X_N)$.按照信息论中定义,将随机矢量中离散变量的个数 N 称为随机矢量的维数.N 维事件序列的安全熵称为 N 维安全熵.

3 策略安全性度量方法

定义 2(策略安全熵) 策略安全熵 $H_P(X)$指违反某策略 P 的非授权访问事件 X 的最大平均不确定性测度. $H_P(X)=-\sum_{i=1}^{q}\omega_i p(a_i)\log p(a_i)$

其中,$p(a_i)\geqslant 0(i=1,2,\cdots,q)$,且 $\sum_{i=1}^{q}p(a_i)=1$,$\omega_i\geqslant 0(i=1,2,\cdots,q)$,$P$ 为安全策略.

策略安全熵描述了某个安全策略下的发生非授权事件的整体概率测度,反映了安全策略对访问行为的控制力度,因此策略的安全熵可以认为是策略的安全强度的一种体现.策略熵越小说明该策略下非授权事件的有序程度越大,混乱程度越小,非授权事件的不确定性越小,策略的安全强度越大.策略熵越大则说明该策略下非授权事件的有序程度越小,混乱程度越大,非授权事件的不确定性越大,策略的安全强度越小.

定义 3(自主策略安全熵) 违反自主访问控制策略的非授权事件平均不确定性,记为 $H_{P_d}(X)$.N 维事件的自主策略安全熵称为 N 维自主策略安全熵,记为 $H_{P_d}(\boldsymbol{X}^N)$.

定义 4(强制策略安全熵) 违反强制访问控制策略的非授权事件平均不确定性,记为 $H_{P_m}(X)$.N 维事件的强制策略安全熵称为 N 维强制策略安全熵,记为 $H_{P_m}(X)$.

4 典型访问控制策略的安全熵

4.1 自主访问控制策略的安全熵

(1)一维自主策略安全熵 $H_{P_d}(X)$

定理 1 一维自主策略安全熵

$$H_{P_d}(X)=\frac{(q+1)\log(q)}{2}$$

其中 $q=2mn$,m,n 分别为用户和资源的个数.

证明 系统的访问操作都可以分解为读 r 和写 w 原子操作,对于 m 个用户和 n 个资源,可能发生的操作方式为 $2mn$ 种.则随机事件 X 的符号集为 A:$a_i(i=1\,2,\cdots,q)$,$q=2mn$.不失一般性,令随机事件等概率发生,$P(a_i)=1/q(i=1\,2,\cdots,q)$.若系统存在 k 条负授权,发生违规操作的可能性就有 $k(k\leqslant q)$种可能性,不妨将非授权事件的事件集合记为$\{a_1,a_2,\cdots,a_k\}$.则 k 条负授权策略时的随机事件 X 的概率分布$[X,P(a_i)]$为

$$\begin{bmatrix} X \\ P(a_i) \end{bmatrix} = \begin{bmatrix} a_1 & \cdots & a_k & a_{k+1} & \cdots & a_q \\ \frac{1}{q} & \frac{1}{q} & \frac{1}{q} & \frac{1}{q} & \frac{1}{q} & \frac{1}{q} \end{bmatrix}$$

由于合法事件对系统的安全性没有影响,可令非授权事件的熵权为 1,合规事件的熵权为 0,得随机事件 X 的熵权分布$[X,\omega_i]$为

$$\begin{bmatrix} X \\ \omega \end{bmatrix} = \begin{bmatrix} a_1 & \cdots & a_k & a_{k+1} & \cdots & a_q \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

根据安全熵的计算公式,k 条负策略时的一维自主策略安全熵 $H_{P_d(k)}(X)$ 为

$$H_{P_d(k)}(X) = -\sum_{i=1}^{q} \omega_i p(a_i)\log p(a_i)$$

$$= -\sum_{i=1}^{k} 1 \times p(a_i)\log p(a_i) - \sum_{i=k+1}^{q} 0 \times p(a_i)\log p(a_i)$$

$$= \frac{k\log(q)}{q}$$

平均一维自主策略安全熵 $H_{P_d}(X)$为

$$H_{P_d}(X) = \frac{\sum_{k=0}^{q} H_{P_d(k)}(X)}{q} = \frac{\sum_{k=1}^{q} k\log(q)}{q}$$

$$= \frac{\log(q)}{q}\sum_{k=1}^{q} k = \frac{(q+1)\log(q)}{2}$$

得证.

(2)N 维自主策略安全熵 $H_{P_d}(X^N)$

定理 2　N 维自主策略安全熵

$$H_{P_d}(X^N) = \frac{\sum_{k=1}^{q}\sum_{i=1}^{[N/2]} kC_N^{2i}(\frac{1}{2})^i q^{(N-i-1)}}{(q+1)q/2}(\frac{\log(q^N)}{q^N})$$

其中 $q=2mn$,m,n 分别为用户和资源的个数.

证明　在主体和客体之间,可能通过多次访问实现间接违规,如对于负授权"$\neg(s_2,o_1,r)$",可以通过 s_1 读 o_1,s_1 写 o_2,s_2 读 o_2 来实现间接的信息传递,也就是用户 s_2 得到了 o_2 中的信息,这显然违反了该负授权,但是由于其他的三个事件可能并未违反安全策略,因此在自主访问控制策略中,这种间接信息流是允许发生的.由于访问事件都是用户对资源的操作,因此这种间接非授权访问都是在奇数次访问时所形成.

对于一条负授权,可以形成 $C_{n-1}^{(j-1)/2}C_{m-1}^{(j-1)/2}$种长度为 j 不同组合形式的间接非授权事件.包含某一长度为 j 的间接非授权事件的 N 维事件序列共有 $q^{(N-j)}$种可能的排列组合方式.因此共有 $C_N^j C_{n-1}^{(j-1)/2}C_{m-1}^{(j-1)/2}q^{(N-j)}$种 j 次访问违规的情况.对于 k 条负授权,最多有 $kC_N^j C_{n-1}^{(j-1)/2}C_{m-1}^{(j-1)/2}q^{(N-j)}$种可能性.

实际的信息系统中 mn 通常很大,因此非授权事件可能数的计算可以简化

$$kC_N^j C_{n-1}^{(j-1)/2}C_{m-1}^{(j-1)/2} \approx kn^{(j-1)/2}m^{(j-1)/2}$$

$$\approx k(mn)^{(j-1)/2} = kC_N^j(\frac{q}{2})^{(j-1)/2}.$$

由此 N 维事件序列在 k 条负授权时,可能形成的非授权事件数量最大为

$$\sum_{j=3,5,\ldots}^{N} kC_N^j(\frac{q}{2})^{(j-1)/2}q^{(N-j)}$$

$$= \sum_{j=3,5,\ldots}^{N} kC_N^j(\frac{1}{2})^{(j-1)/2}q^{(N-j)+(j-1)/2}$$

$$= \sum_{j=3,5,\ldots}^{N} kC_N^j(\frac{1}{2})^{(j-1)/2}q^{(N-(j-1)/2+1)}$$

$$\approx \sum_{i=2}^{[N/2]} kC_N^{2i}(\frac{1}{2})^i q^{(N-i-1)}$$

对 k 取均值得 $\dfrac{\sum_{k=1}^{q}\sum_{i=1}^{[N/2]} kC_N^{2i}(\frac{1}{2})^i q^{(N-i-1)}}{(q+1)q/2}$

在信息论中事件等概率时的信息熵的最大,因此令随机事件等概率发生,即 $P(a_i)=(1/q)^N$,此时计算得到的 N 维自主安全熵最大.

$$H_{P_d}(X^N) \mid HRU$$

$$= -\frac{\sum_{k=1}^{q}\sum_{i=1}^{[N/2]} kC_N^{2i}(\frac{1}{2})^i q^{(N-i-1)}}{(q+1)q/2}(\frac{1}{q^N})\log(\frac{1}{q^N})$$

$$= \frac{\sum_{k=1}^{q}\sum_{i=1}^{[N/2]} kC_N^{2i}(\frac{1}{2})^i q^{(N-i-1)}}{(q+1)q/2}(\frac{\log(q^N)}{q^N})$$

得证.

4.2　强制访问控制策略的安全熵

(1) 一维强制策略安全熵 $H_{P_m}(X)$

定理 3　一维强制策略安全熵

$$H_{P_m}(X) = \frac{\log(6)}{3}$$

证明　强制访问控制策略与具体某个用户或资源无关,仅与用户和资源安全级之间的关系有关,如(机密性策略):信息不能从高安全级流向低安全级.用户和资源之间可能的事件为{信息从高安全级用户流向低安全级资源,信息从低安全级用户流向高安全级资源,信息从同安全级用户流向同安全级资源,信息从低安全级资源流向高安全级用户,信息从高安全级资源流向低安全级用户,信息从同安全级资源流向同安全级用户},分别记为$\{a_1,a_2,\cdots,a_6\}$.

我们将用户和资源之间的信息流向作为随机变量 X,则 X 的符号集为 A:$a_i(i=1,2,\cdots,6)$.不失一般性,令随机事件等概率发生,$P(a_i)=1/6$,根据强制访问控制策略,a_1 和 a_5 为非授权事件,得随机事件 X 的熵权分布

$$\begin{bmatrix} X \\ \omega \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

根据安全熵的计算公式,一维强制策略安全熵 $H_{P_m}(X)$ 为

$$H_{P_m}(X) = -\sum_{i=1}^{6}\omega_i p(a_i)\log p(a_i)$$
$$= -1\times p(a_1)\log p(a_1) - 1\times p(a_5)\log p(a_5)$$
$$= \frac{\log 6}{3}$$

得证.

(2) N 维强制策略安全熵 $H_{P_m}(\boldsymbol{X})$

定理 4 N 维强制策略安全熵

$$H_{P_m}(\boldsymbol{X}) = \frac{\log(6)}{3}$$

证明 强制访问控制是一种流策略,这种流策略是符合偏序关系的格,而格具有传递性,Denning[13]已对此进行了定义和证明.因此强制访问控制策略无法形成类似自主访问控制策略那样的间接违规信息流,多次合规事件不会形成一次非授权事件.对于随机事件序列 $\boldsymbol{X}=(X_1X_2,\cdots,X_N)$,某个符号(维)是否为非授权事件与之前和之后的事件无关,只与本次事件是否为非授权事件有关.这就相当于计算信息论中无记忆平稳信源 N 维扩展信源的信息熵,有

$$H(\boldsymbol{X}) = NH(\boldsymbol{X})$$

因此可得

$$H_{P_m}(\boldsymbol{X}) = NH_{P_m}(\boldsymbol{X}) = N\frac{\log(6)}{3}$$

得证.

5 访问控制策略的安全性分析

基于访问控制策略的安全熵,我们对其安全性进行分析.

(1)自主访问控制策略的安全性分析

自主访问控制策略的一维安全熵仅与 q 相关,因此对于单次访问事件,非授权访问行为的发生只与信息系统中的用户数和资源数量相关,并且非授权事件的整体测度不会随着访问次数的增加而改变.由于随着 q 的增大,安全熵越来越大,因此自主访问控制策略在用户数和资源数增加时,非授权访问行为的混乱程度增大,说明对非授权访问行为的控制力度越来越弱,当 q 非常大时,自主访问控制策略对非授权访问行为的控制将非常困难.

相比一维安全熵,其 N 维安全熵除了与 q 相关外,还与事件序列的维数 N 相关,这说明访问次数影响着非授权行为发生的可能性,N 越大,则非授权访问行为的数量越多,混乱程度越大.N 维安全熵实质上反映的是由于间接信息流所导致的非授权访问行为的不确定性测度.

(2)强制访问控制策略的安全性分析

强制访问控制策略的一维安全熵是一个常数,说明强制访问控制策略中发生非授权事件的可能性与系统中的资源和用户个数无关,因此对任何强制访问控制策略保护下的信息系统,非授权事件行为的预期是一致的,即使当 q 非常大时,强制访问控制策略对非授权访问行为的控制也是不变的.并且由于 $H_{P_d}(X) >> H_{P_m}(X)$,说明强制访问控制策略对非授权访问行为的控制力度要远远大于自主访问控制,因此不难得出结论:强制访问控制策略的安全性高于自主访问控制策略.

强制访问控制策略的 N 维安全熵不再是个常数,与事件序列的维数 N 相关.但是与自主访问控制策略的 N 维安全熵相比,强制访问控制策略的 N 维安全熵随着 N 的增加而线性增大,只是对单次访问非授权访问行为的简单求和,而不像自主访问控制的 N 维安全熵一样,随着 N 的增加访问非授权访问行为可能性呈指数级增长,从这方面来说,强制访问控制策略的安全性也要高于自主访问控制策略.

6 结束语

本文为访问控制策略的安全性分析提供了一种科学的度量方法.通过对自主访问控制策略和强制访问控制策略的安全性分析可知,该方法所得出的结论与我们对这两个访问控制策略的预先认识是一致的,并且结论更全面和翔实.该方法具有很强的实用性和科学性.

参考文献

[1] 张红旗,王鲁.信息安全技术.北京:高等教育出版社,2008.
ZHANG Hong-qi, WANG Lu. Information Security Technology [M]. Beijing: Higher Education Press, 2008. (in Chinese)

[2] Shannon C E. A mathematical theory of communication [J]. Bell System Technical Journal, 1948, 26(3): 379 – 423, 623 – 656.

[3] 傅祖芸.信息论—基础理论与应用[M].北京:电子工业出版社,2007.

[4] 丁晓青,吴佑寿.模式识别统一熵理论[J].电子学报,1993,21(8):1 – 8.
DING Xiao-qing, WU You-shou. Unify entropy theory for pattern recognition [J]. Acta Electronica Sinica, 1993, 21(8): 1 – 8. (in Chinese)

[5] 扬明.决策表中基于条件信息熵的近似约简[J].电子学报,2007,35(11):2156 – 2160.
YANG Ming. Approximate reduction based on conditional information entropy in decision table [J]. Acta Electronica Sinica, 2007, 35(11): 2156 – 2160. (in Chinese)

[6] Morio J, et al. A characterization of Shannon entropy and Bhattacharyya measure of contrast in polarimetric and interferometric SAR image [J]. Proceedings of the IEEE, 2009, 97

(6): 1097 – 1108.

[7] 付钰,吴晓平,叶清等.基于模糊集与熵权理论的信息系统安全风险评估研究[J].电子学报,2010,38(7): 1489 – 1494.
FU Yu, WU Xiao-ping, YE Qing et al. An Approach for information systems security risk assessment on fuzzy set and entropy-weigh[J]. Acta Electronica Sinica, 2010, 38(7): 1489 – 1494. (in Chinese)

[8] 赵冬梅,马建峰,王跃生.信息系统的模糊风险评估模型[J].通信学报,2007,28(4): 51 – 56.
ZHAO Dong-mei, MA Jian-feng, WANG Yue-sheng. Model of fuzzy risk assessment of the information system [J]. Journal on Communication, 2007, 28(4): 51 – 56. (in Chinese)

[9] 张义荣,鲜明,王国玉.一种基于网络熵的计算机网络攻击效果定量评估方法[J].通信学报,2004,25(11): 158 – 165.
ZHANG Yi-rong, XIAN Ming, WANG Guo-yu. A quantitative evaluation technique of attack effect of computer network based on network entropy [J]. Journal on Communications, 2004, 25(11): 158 – 165. (in Chinese)

[10] 王栋,潘少明,吴吉春 等.确定风险分析先验概率分布的最大熵方法[J].应用基础与工程科学学报,2006,14(Sup): 318 – 325.
WANG Dong, PAN Shao-ming, et al. Deriving the prior probability distribution of risk analysis with the use of the principle of maximum entropy [J]. Journal of Basic Science and Engineering, 2006, 14(Sup): 318 – 325. (in Chinese)

[11] 胡 俊,沈昌祥,张 兴.一种 BLP 模型的量化分析方法[J].小型微型计算机系统,2009,30(8): 1605 – 1610.
HU Jun, SHEN Chang-xiang, ZHANG Xing. Quantitative analysis method to BLP model [J]. Journal of Chinese Computer Systems, 2009, 30(8): 1605 – 1610. (in Chinese)

[12] D Denning. A lattice model of secure information flow [J]. Communications of the ACM, 1976, 19(5): 236 – 243.

作者简介

王　超　男.1975 年 3 月出生,河南长垣人.解放军信息工程大学四院讲师,计算机应用技术专业在读博士生,主要从事网络安全、多级安全等方面的研究工作.
E-mail: wangchao302@sina.com

陈性元　男.1963 年 11 月出生,安徽无为人,博士.现为解放军信息工程大学三院院长、教授、博士导师,主要从事网络与信息安全方面的研究工作.
E-mail: chxy302@vip.sina.com