

一种基于误用检测的新算法

朱俚治

(南京航空航天大学 信息中心, 江苏 南京 210016)

摘要: 当今攻击网络的手段是多种多样的, 为保护用户在访问网络资源时不受黑客的攻击, 因此需要网络安全设备和网络安全技术。入侵检测技术是一种安全技术, 该技术能够检测出网络中数据包的行为属性, 是正常还是异常。目前入侵检测技术有两种: 误用检测和异常检测。这两种技术都能够阻止网络攻击行为。但要想阻止网络的攻击行为, 必须检测出攻击行为。文中在简述了入侵检测技术、粒子群的某些概念后, 提出了基于粒子群技术在入侵检测中的应用, 最后给出了数据包属性的匹配算法。

关键词: 粒子群; 入侵检测; 误用检测; 异常检测; 数据包

中图分类号: TP301.6

文献标识码: A

文章编号: 1673-629X(2015)02-0135-05

doi: 10.3969/j.issn.1673-629X.2015.02.031

A New Algorithm Based on Misuse Detection

ZHU Li-zhi

(Information Center, Nanjing University of Aeronautics and Astronautics,
Nanjing 210016, China)

Abstract: Today's attack means are varied, in order to protect the user against hackers attack when accessing cyber source, need network security equipment and network security technology. Intrusion detection technology is a security technology, this technology can detect the behavior attribute of data packet in the network, whether is normal or abnormal. The current intrusion detection technology has two types, misuse detection and anomaly detection. The two techniques are able to prevent network from attacks. But to prevent network attacks, must detect attacks. In this paper, the concept of intrusion detection technology, particle swarm is discussed, and then put forward the application of particle swarm optimization technology in intrusion detection, eventually, the matching algorithm of data packet attributes is given.

Key words: particle swarm; intrusion detection; misuse detection; anomaly detection; data packet

0 引言

如今的互联网规模已布满全球, 并且用户的数量达到了数亿。网络是一个资源共享平台, 网络技术的不断更新能够使得更多的用户在访问资源时更快捷、更方便和更丰富。但在访问资源过程中, 用户常常会被黑客攻击, 因此网络安全性日益突出。如果某个网络没有采用安全设备, 那么网络的用户极易受到来自黑客、病毒和蠕虫的攻击; 因此, 如何保护网络用户不受来自网络的攻击变得十分重要。

为了应对来自网络的攻击, 相关的科研单位研发了多种防护技术。依据这些技术, 网络硬件研制单位开发出了多种安全设备, 这些设备主要有防火墙、防毒墙和入侵检测系统。在网络中如果防火墙是第一道防

护墙, 那么入侵检测系统就是第二道防护墙。入侵检测在防护过程中有着不可替代的作用。文中根据已有的入侵检测技术, 提出了一种新的检测方法。

1 入侵检测技术简介

入侵检测技术是一种具有一定智能性的网络安全技术。安全人员开发该系统的目的就是阻止黑客对用户的攻击, 使得用户在访问资源时变得更安全。当今已知的攻击方式有两种: 已知模式的攻击和未知模式的攻击。为了应对来自网络的攻击, 对入侵检测系统开发出了误用检测与异常检测这两种检测方法来发现网络攻击手段^[1-4]。

误用检测是一种基于串匹配模型技术, 当一个行

收稿日期: 2014-03-05

修回日期: 2014-06-11

网络出版时间: 2014-12-27

基金项目: 软件开发环境国家重点实验室开放基金资助项目(SKLSDE-2013KF-02)

作者简介: 朱俚治(1980-), 男, 工程师, 研究方向为网络安全。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20141227.1343.022.html>

为事件进入入侵检测系统,误用检测就会根据行为事件的特征进行模式,来判断这个行为是正常的还是异常的。误用检测的不足之处是在进行模式匹配的时候只能根据已知的行为事件的属性为依据;因此,该技术只能检测出已知模式的网络攻击^[5-7]。

异常检测是基于统计分析的检测技术,异常检测技术对网络行为的检测能力强于误用检测。异常检测最大的优势是能够同时检测出已知的和未知的网络攻击。由于异常检测的检测技术是根据统计学来区分正常行为和异常行为,因此异常检测具有自我学习能力,这使得该技术在检测网络行为的属性时具有一定的智能性。异常检测技术与计算机技术相融合的有以下几种:基于神经网络的入侵检测、基于遗传算法的入侵检测和基于数据挖掘的入侵检测^[5-7]。

2 粒子群的几个重要基本概念

(1)种群:种群是群体的概念,最常见的种群有鸟类、鱼群。

(2)粒子:粒子是种群中的个体,鸟群和鱼群中的个体就是种群中的粒子。

(3)适应度函数:适应度函数能够为种群的粒子的搜索方向和寻找最优解提供依据,加快种群的收敛速度^[8-9]。

(4)目标函数:目标函数的目的是使得粒子在搜索空间寻找最优解,该函数的期望值就是粒子的最优解^[8-9]。

(5)粒子间的协作:如果粒子群中某个粒子找到了最优解,那么其他粒子可以通过粒子之间的相互学习来调整粒子飞行的速度和方向,使得粒子找到最优解。粒子在搜索空间中不会重复搜索同一个位置,这将加快种群收敛速度^[4]。

3 粒子群算法在检测网络数据包上的应用

检测函数:

$$y = f(x) = \frac{n}{m} = \frac{\text{某种未知数据包的大小值}}{\text{已知数据包的大小值}}$$

3.1 粒子群算法简介

粒子群算法是寻找最优解的算法。由于粒子在寻找最优解时能够相互协作,因此当种群中的某个粒子找到了最优解,那么与其他粒子通过相互协作同样能够使其他粒子也找到最优解^[10]。

为了使粒子在搜索空间中较快地找到最优解,算法需通过适应函数来控制 and 调整种群中每个粒子自身飞行的方向和速度。如果粒子在搜索空间中找到了十分符合目标函数的期望值,则此时粒子也就找到了最优解,也就加快了种群的收敛速度^[10]。

3.2 粒子群在检测网络数据包上的应用

对某种未知的数据包组成的粒子群,在这个粒子群中每个数据包都具有自身固有的属性,所以文中将所有某种未知属性的数据包属性提取出来,把数据包组成一个粒子群。而这些由某种未知数据包组成的粒子群都要在搜索空间中找到自己的最优解,因此还需构建一个搜索空间,使得每一个粒子在搜索空间中都能找到最优解。为了加快粒子群的收敛速度,这里将某种已知数据包的属性种类按照一定的规模组成若干个搜索空间。这样使得求解空间中的粒子在搜索空间中都能较快地找到最优值。

基于粒子上述的特点,为了将粒子群技术应用在入侵检测技术上,在这里需要建立两个函数:目标函数和适应函数。

这两个函数具体形式如下:

(1)目标函数:

$$y = f(x) = \frac{n}{m} = \frac{\text{某种未知数据包的大小值}}{\text{已知数据包的大小值}} \approx 1$$

目标函数是用来描述粒子寻找最优解的衡量标准。

(2)适应函数:

$$y = h(x) = \lim_{n \rightarrow m} (1 - \frac{\text{某种未知数据包的大小值}}{\text{已知数据包的大小值}}) = 0$$

适应函数是用来调整粒子飞行的方向,使得粒子的飞行逐步接近最优解。

3.3 粒子群参数初始化

粒子群算法在求最优化问题时,每个粒子根据如下公式来更新自己的速度和新的位置^[6]。

$$v_{k+1} = \omega v_k + c_1 r_1 (pbest_k - x_k) + c_2 r_2 (gbest_k - x_k) \quad (1)$$

$$x_{k+1} = x_k + v_{k+1} \quad (2)$$

其中, r_1 和 r_2 为(0,1)之间的随机数; c_1 和 c_2 为学习因子,一般取 $c_1 = c_2 = 2$; ω 为惯性权重, ω 取较大值粒子群算法具有较强的全局搜索能力, ω 取较小值粒子群算法倾向于局部搜索;速度 v_{k+1} 可以设定在 v_{\min} 和 v_{\max} 之间,当超过这个区域时,就取对应的上下限值; x_k 位置也可设定在一定的 $[x_{\min}, x_{\max}]$ 区间范围内。

由式(1)和(2)所组成的粒子群算法一般称为基本粒子算法。在此基础上,如果将惯性权重 ω 的取值根据迭代次数依次递减,此时粒子群算法被称为惯性权重线性递减粒子群算法。一般情况下,将 ω 的初值设置为 0.9,然后按照迭代次数线性递减到 0.4。惯性权重线性递减粒子群算法是粒子群算法研究领域一种比较常用的算法模型,因此文中采用上述惯性权重线性递减粒子群算法对病毒种群构成的粒子群进行研究^[11-13]。

3.4 求解空间粒子群和搜索空间粒子群的划分

(1)求解空间粒子种群的划分。

从求解空间中取出一个粒子,求解空间中剩余粒子个数为: n_1, n_2, \dots, n_n ; 其属性值分别为 n'_1, n'_2, \dots, n'_n 。

将求解空间中的某个粒子的属性值与求解空间中所有粒子的属性值进行比较。

属性判定函数:

$$g(x) = \left| 1 - \frac{\text{某个求解空间中的粒子属性值}}{\text{求解空间中所有粒子的属性值}} \right|;$$

令 $x = j$

如果粒子 j 与求解空间中的粒子属性的比较有以下结论:

$$g(x) = \left| 1 - \frac{j}{n_1} \right| \approx 0, g(x) = \left| 1 - \frac{j}{n_2} \right| \approx 0, \dots,$$

$$g(x) = \left| 1 - \frac{j}{n_n} \right| \approx 0$$

则将求解空间中这 n 个粒子组成一个子种群,并且将这 n 个粒子从求解空间种群分离出去,组成一个独立的求解空间 A_1 。

再在剩余求解空间中依次取出粒子重复以上的过程,直到将求解空间种群中的粒子划分为若干子种群 A_1, A_2, \dots, A_n 。

(2)搜索空间粒子种群的划分。

从搜索空间中取出一个粒子,求搜索空间中剩余粒子个数为: n_1, n_2, \dots, n_n ; 其属性值分别为 n'_1, n'_2, \dots, n'_n 。

将搜索空间中的某个粒子的属性值与搜索空间中所有粒子的属性值进行比较。

属性判定函数:

$$T(x) = \left| 1 - \frac{\text{某个搜索空间中的粒子属性值}}{\text{搜索空间中所有粒子的属性值}} \right|;$$

令 $x = j$

如果粒子 j 与搜索空间中的粒子属性的比较有以下结论:

$$T(x) = \left| 1 - \frac{j}{n_1} \right| \approx 0, T(x) = \left| 1 - \frac{j}{n_2} \right| \approx 0, \dots,$$

$$T(x) = \left| 1 - \frac{j}{n_n} \right| \approx 0$$

则将搜索空间中这 n 个粒子组成一个子种群,并且将这 n 个粒子从搜索空间种群分离出去,组成一个独立的搜索空间 B_1 。

再在剩余搜索空间中依次取出粒子重复以上的过程,直到将搜索空间种群中的粒子划分为若干独立的搜索空间种群 B_1, B_2, \dots, B_n 。

3.5 粒子群技术在匹配网络数据包上的实现

在子种群 A_1, A_2, \dots, A_n 中依次选出一个种群 A_n ,

在种群 A_1 中提取粒子 a , 粒子 a 的属性值为 a' 。

在子种群 B_1, B_2, \dots, B_n 中依次选出一个种群 B_n , 种群 B_n 中有粒子数 b_1, b_2, \dots, b_n , 粒子群的属性值分别为 b'_1, b'_2, \dots, b'_n 。

(1)目标函数:

$$y = f(x) = \frac{n}{m} = \frac{\text{某种未知数据包的大小值}}{\text{已知数据包的大小值}} \approx 1$$

文中令 $y = f(x) = \frac{n}{m} = \frac{a'}{b'_n}$ 。粒子 a 寻找最优解过程如下:

$$y = f(x) = \frac{a'}{b'_1}, y = f(x) = \frac{a'}{b'_2}, \dots, y = f(x) = \frac{a'}{b'_n}$$

如果存在 $y = f(x) = \frac{a'}{b'_n} \approx 1$, 则粒子 a 找到了最优解。

如果粒子群 A 中的粒子 a 找到了最优解,那么粒子群 A 中的其他粒子根据粒子协作和竞争机制,同样能够很快地找到自身的最优解。根据相同的理由种群 A_1, A_2, \dots, A_n , 都能很快找到各自的最优解。

(2)适应函数:

$$y = h(x) = \lim_{n \rightarrow m} (1 - \frac{\text{某种未知数据包的大小值}}{\text{已知数据包的大小值}}) = 0$$

适应函数是用来调整粒子飞行的方向,使得粒子的飞行逐步接近最优解。

①当 $y = h(x) = \lim_{n \rightarrow m} (1 - \frac{\text{某种未知数据包的大小值}}{\text{已知数据包的大小值}}) > 0$, 有 $1 - \frac{n}{m} > 0$ 。

如果 $1 - \frac{n}{m} > 0$ 时,有以下讨论:

由 $1 - \frac{n}{m} > 0$, 可得 $n < m$, 但目标函数中 $y = f(x) = \frac{n}{m} \approx 1$, 需要 $n \approx m$ 。

为了使 n 无限接近于 m , 达到 $n \approx m$ 的目的,因此在搜索空间中需要寻找属性值比粒子 m 属性值大的粒子,这样才能找到粒子的最优解。

②当 $y = h(x) = \lim_{n \rightarrow m} (1 - \frac{\text{某种未知数据包的大小值}}{\text{已知数据包的大小值}}) < 0$, 有 $1 - \frac{n}{m} < 0$ 。

如果 $1 - \frac{n}{m} < 0$ 时,有以下讨论:

由 $1 - \frac{n}{m} < 0$, 可得 $n > m$, 但目标函数中 $y = f(x) = \frac{n}{m} \approx 1$, 需要 $n \approx m$ 。

为了使 n 无限接近于 m , 达到 $n \approx m$ 的目的,因此在搜索空间中需要寻找属性值比粒子 m 属性值小的

粒子,这样才能找到粒子的最优解。

③当 $y = h(x) = \lim_{n \rightarrow m} (1 - \frac{\text{某种未知数据包的大小值}}{\text{已知数据包的大小值}}) = 0$, 有 $1 - \frac{n}{m} = 0$ 。

如果 $1 - \frac{n}{m} = 0$ 时,则此时粒子在搜索空间中找到了最优解。

3.6 粒子群技术的检测算法

- (1) 初始化搜索空间中的粒子群。
- (2) 初始化求解空间中的粒子群。
- (3) 将未知的数据包组成的种群,划分成若干个子种群,并且将子种群中的粒子数目控制在 30 以内。
- (4) 同样将搜索空间中的粒子种群划分为若干个子种群,并且将子种群中的粒子数目控制在 30 以内。
- (5) 某个未知属性数据包组成的粒子群在搜索空间中寻找最优的解。
- (6) 使用适应函数不断调整粒子飞行的方向和速度。
- (7) 如果粒子群中的某些粒子在第一次搜索中未找到最优解,将这些粒子提取出来组成一个新的粒子群。
- (8) 将新组成的粒子群在下一个搜索空间寻找最优解。

3.7 网络数据包属性的判定

判定函数:

$$y_n = f(x) = \frac{\text{未知数据包的属性} + a}{\text{已知数据包的属性} + a}$$

(1) 一个未知数据包有 i 个属性: i_1, i_2, \dots, i_n , 分别计算出属性的值: i'_1, i'_2, \dots, i'_n 。

(2) 某个已知数据包有 j 个属性: j_1, j_2, \dots, j_n , 分别计算出属性的值: j'_1, j'_2, \dots, j'_n 。

$$y_1 = \frac{i'_1 + a}{j'_1 + a}, y_2 = \frac{i'_2 + a}{j'_2 + a}, \dots, y_n = \frac{i'_n + a}{j'_n + a}$$

当 $i'_n = j'_n$ 时,有 $i'_n + a = j'_n + a$; 如果 $i'_n + a = j'_n + a$,

那么 $\frac{i'_n + a}{j'_n + a} = 1$ 。

当 $i'_n > j'_n$ 时,有 $i'_n + a > j'_n + a$; 如果 $i'_n + a > j'_n + a$, 那么 $\frac{i'_n + a}{j'_n + a} > 1$ 。

当 $i'_n < j'_n$ 时,有 $i'_n + a < j'_n + a$; 如果 $i'_n + a < j'_n + a$, 那么 $\frac{i'_n + a}{j'_n + a} < 1$ 。

讨论:

(1) 当未知数据包的属性值 i'_n 都大于已知数据包的某个属性值 j'_n , 这时 i'_n 值越小, 则 $y_n = f(x)$ 的值越接近于 1, 未知属性的数据包的属性越接近于已知数

据包的属性。

(2) 当未知数据包的属性值 i'_n 部分大于已知数据包的属性值 j'_n , 部分小于已知数据包的属性值 j'_n , 则此时需要讨论:

① 当 $y_n = f(x) = \frac{i'_n + a}{j'_n + a} > 1$, 这时 i'_n 值越小, 则 $y_n = f(x)$ 的值越接近于 1;

② 当 $y_n = f(x) = \frac{i'_n + a}{j'_n + a} < 1$, 这时 i'_n 值越大, 则 $y_n = f(x)$ 的值越接近于 1。

(3) 当未知数据包的属性值 i'_n 都小于已知数据包的某个属性值 j'_n , 这时 i'_n 值越大, $y_n = f(x)$ 的值越接近于 1, 未知属性的数据包的属性越接近于已知数据包的属性。

$$\text{如果 } y_1 = \frac{i'_1 + a}{j'_1 + a} \approx 1, y_2 = \frac{i'_2 + a}{j'_2 + a} \approx 1, \dots, y_n =$$

$$\frac{i'_n + a}{j'_n + a} \approx 1$$

$$y_n = g(x) = y_1 + y_2 + \dots + y_n$$

$$\text{则 } y = g(x) = \lim_{t=1}^{t=n} \sum y_n = n。$$

如果 $y = g(x) = \lim_{t=1}^{t=n} \sum y_n \neq n$, 讨论如下:

① $y = g(x) > n$ 时: 如果 $y = g(x) > n$ 的值越大, 则未知数据包的属性偏离已知数据包的属性就越大。

② $y = g(x) < n$ 时: 如果 $y = g(x) < n$ 的值越小, 则未知数据包的属性偏离已知数据包的属性就越大。

4 小 结

如果未知属性的数据包的属性较大程度地偏离已知数据包的属性, 那么未知属性的数据包与已知属性的数据包不能相匹配。如果未知属性的数据包与已知属性的数据包不匹配, 那么重复上一个过程进行下一个数据包的匹配历程。未知属性的数据包在下次寻找最优解时, 当数据包在搜索空间中寻找最优解时, 则需要更换搜索空间中的粒子。如此重复直到未知属性的数据包找到相匹配的数据包。

根据已知数据包的属性以及粒子群技术得出的结论, 可以知道未知数据包的属性是属于攻击性数据包还是正常数据包。

5 误用检测算法

- (1) 使用入侵检测系统监视网络中的流量。
- (2) 检测的数据包中捕获有未知的数据包。
- (3) 在入侵检测系统中保存在该数据包。
- (4) 提取未知数据包的行为属性, 并计算出属性

的值。

(5)通过粒子群算法需找相匹配的数据包。

(6)提取已知数据包的属性,并计算出属性的值。

(7)比较未知数据包的属性与已知数据包的属性的值。

(8)通过属性值的计算和属性值的比较,得出相应的结论。

6 结束语

文中提出的新的检测算法是基于现有的入侵检测算法,并首次将粒子群技术应用到入侵检测的异常检测中。这是文中的创新点。异常检测是一种成熟的检测技术,异常检测技术与其他技术相结合的应用前景十分广泛,随着计算机技术和人工智能技术的发展,将会出现更多的新异常检测算法。随着入侵检测技术不断的发展,这必将遏制网络攻击事件的发生,网络攻击事件的成功率将会有明显的下降。

参考文献:

[1] 阎巧,谢维信.异常检测技术的研究与发展[J].西安电子科技大学学报,2002,29(1):128-132.

[2] 金文进,杨武.异常检测技术研究综述[J].软件导刊,2008,7(1):10-13.

[3] 刘陶,叶君耀,朱永宣.一种基于统计方法的入侵检测模

作监测机制[J].软件学报,2010,21(10):2584-2598.

[7] 刘志辉,孙斌,谷利泽,等.一种防范 BGP 地址前缀劫持的源认证方案[J].软件学报,2012,23(7):1908-1923.

[8] Ballani H,Francis P,Zhang Xinyang. A study of prefix hijacking and interception in the Internet[J]. ACM SIGCOMM Computer Communication Review,2007,37(4):265-276.

[9] Dimitropoulos X A,Krioukov D V,Fomenkov M,et al. AS relationships: inference and validation[J]. ACM SIGCOMM Computer Communication Review,2007,37(1):29-40.

[10] Ramachandran A,Feamster N. Understanding the network-level behavior of spammers[J]. ACM SIGCOMM Computer Communication Review,2006,36(4):291-302.

[11] Adhikari V K,Jain S,Zhang Zhili. YouTube traffic dynamics

型的研究[J].微计算机信息,2007,23(10-3):120-122.

[4] 纪祥敏,宁正元,林大辉.误用检测技术研究[J].福建电脑,2006(2):6-7.

[5] Lane T,Brodley C E. Temporal sequence learning and data reduction for anomaly detection[J]. ACM Transactions on Information and System Security,1999,2(3):295-331.

[6] Warrender C,Forrest S,Pearlmutt B. Detecting intrusions using system calls: alternative data mode[C]//Proc of IEEE symposium on security and privacy. Oakland:IEEE,1999:133-145.

[7] Denning D. An intrusion detection model[J]. IEEE Transactions on Software Engineering,1987,13(2):222-232.

[8] 潘峰,周倩,李位星,等.标准粒子群优化算法的马尔科夫链分析[J].自动化学报,2013,39(4):381-389.

[9] 高海兵,周驰,高亮.广义粒子群优化模型[J].计算机学报,2005,28(12):1980-1987.

[10] 曾建潮,崔志华.微粒群算法的统一模型及分析[J].计算机研究与发展,2006,43(1):96-100.

[11] 刘志雄,梁华.粒子群算法中随机数参数的设置与实验分析[J].控制理论与应用,2010,27(11):1489-1496.

[12] Eberhart R C,Shi Y. Comparing inertia weights and constriction factors in particle swarm optimization[C]//Proceedings of the IEEE congress on evolutionary computation. [s. l.]:[s. n.],2000:84-88.

[13] 高鹰,谢胜利.免疫粒子群优化算法[J].计算机工程与应用,2004,40(6):4-6.

and its interplay with a tier-1 ISP: an ISP perspective[C]//Proceedings of the 10th ACM SIGCOMM conference on Internet measurement. [s. l.]:ACM,2010:431-443.

[12] Lad M,Oliveira R,Zhang B,et al. Understanding resiliency of internet topology against prefix hijack attacks[C]//Proc of 37th annual IEEE/IFIP international conference on dependable systems and networks. Los Angeles: IEEE, 2007: 368-377.

[13] BGP best path selection algorithm[EB/OL]. 2006. <http://www.cisco.com/warp/public/459/25.shtml>.

[14] Zhao Xiaoliang,Pei Dan,Wang Lan,et al. An analysis of BGP multiple origin AS (MOAS) conflicts[C]//Proceedings of ACM Internet measurement workshop. San Francisco, CA, USA:ACM,2001:31-35.

一种基于误用检测的新算法

作者：[朱偲治, ZHU Li-zhi](#)
作者单位：[南京航空航天大学 信息中心, 江苏 南京, 210016](#)
刊名：[计算机技术与发展](#) 
英文刊名：[Computer Technology and Development](#)
年, 卷(期): 2015 (2)

引用本文格式: [朱偲治, ZHU Li-zhi](#) [一种基于误用检测的新算法](#)[期刊论文]-[计算机技术与发展](#) 2015 (2)