DOI:10.11721/cqnuj20150121

基于粒子群优化的检测网络攻击方法*

赵 攀,邱 玲,陈 超
(四川理工学院 计算机系,四川 自贡 643000)

摘要:为了有效判断网络数据包是否存在被攻击的可能性,在以往的研究基础上提出了一种新的检测算法 DMPS(Detection method based of particle susarm)。首先该算法根据数据包属性的离散度定义了状态检测指标,并利用粒子群优化方法给出了标准差分布的计算流程,以此判断数据包的异常状况。最后,通过 OPNET 和 Matlab 进行仿真实验,深入研究了影响该算法的关键因素,同时对比了与其他算法之间的性能状况,结果表明 DMPS 具有较好的适应性。

关键词:攻击;检测;异常;粒子群优化
中图分类号:TP393 **文献标志码**:A **文章编号**:1672-6693(2015)01-0106-05

随着 Internet 的迅速发展,网络攻击给用户带来了极大的安全隐患,如何有效地检测和防御攻击成为研究的热点和重点[1-6]。目前,根据攻击意图可以将攻击事件划分为 4 大类:Denial of service attacks(DoS),Remote to local attacks(R2L),User to root attacks(U2R),Probes。其中,DoS 针对网络服务有效性进行破坏,使主机或网络不能及时接收并处理外界请求,从而不能提供给合法用户正常的服务。而根据网络攻击提出的检测方法可以划分为:1) 隐马尔可夫模型,将每个攻击意图视为有限状态机的一个状态,通过算法计算每个状态的概率;2) 贝叶斯博弈,预测攻击者和防御者在下一个博弈阶段将会选择的攻击和防御措施的概率;3) 有色时间 Petri 网,重点检测攻击事件;4) 攻击响应事件相关性,通过分析攻击的前后关系,预测同一攻击者对同一个对象可能的攻击概率。同时,由于实际流量存在分形和突发特性,而网络检测流量统计模型 FARIMA(p,d,q)能较好地刻画流量分形和突发特性[7-8],因此可结合该模型对到达数据包进行分组拟合,以此分析流量的异常性。对此,国内外学者做了大量工作。文献[9]针对异常网络行为会偏离正常语法规则的特点,利用改进的隐马尔可夫模型建立了一种网络攻击检测方法,其原理建立在正常行文样本的学习基础上。文献[10]结合 4 种不同的检测方法,建立了基于隐马尔可夫的分布式拒绝服务攻击的检测模型,但该模型只能针对网络层协议,对于应用层缺乏有效检测机制。文献[11]根据节点行为特征的多元属性,并结合精确检测攻击行为和对传感数据的有效融合,建立了一种高效的网络攻击检测方法,以达到降低网络开销和误报率的目的。为了采用轻量级协议交互方式来获取客户端和服务器双方的实时状态信息,文献[12]在对等网络中建立了一种抵御 DoS 攻击的自适应安全框架。文献[13]基于进程异常场景自动分析攻击载荷句法,建立了一种代码注入攻击自动分析和响应方法,能够有效识别和阻断基于同一未知漏洞的各种代码注入攻击,显著地降低响应虚警概率和系统对外服务的响应时间。文献[14]利用蝴蝶突变模型刻画数据包异常行为,建立了一种突变级数的异常流量状态检测方法,但攻击检测的误报率有待进一步提高。

在上述研究的基础上,本文基于粒子群优化算法提出了一种新的检测方法,并结合数据包属性的离散度给出了检测指标,同时通过获得数据包属性的标准差分布来判断是否存在被攻击的可能性。最后,以 OPNET 和 MATLAB 进行仿真实验,对比研究了该算法与其他算法之间的性能状况。论文结构如下所述:第 1 节给出了网络攻击评价指标,第 2 节结合粒子群优化算法建立了攻击检测方法,第 3 节对该检测算法进行仿真实验,第 4 节对论文进行了总结。

* **收稿日期**:2013-09-11 **修回日期**:2013-10-30 **网络出版时间**:2015-1-7 16:04
资助项目:国家自然科学基金(No. 60372013);四川省教育厅重点项目(No. 13ZA0118);人工智能四川省重点实验室开放基金项目(No. 2012RYY02);四川理工学院培育项目(No. 2012PY13);企业信息化与物联网测控技术四川省高校重点实验室项目(No. 2013WYJ01)
作者简介:赵攀,男,副教授,研究方向为计算机网络,E-mail:598256382@qq.com
网络出版地址:http://www.cnki.net/kcms/detail/50.1165.N.20150107.1604.021.html

1 网络攻击评价指标

由于 HTTP 协议中数据包可以看作遵循一定标准的字符串，具有 k 个通用属性的数据域组成，令数据包 $\boldsymbol{Y}=[y_1, y_2, \cdots, y_k]$，其中 y_k 表示 Data 头部，Host 头部，源端 IP 地址，目的端 IP 地址等。对于 n 个数据包的样本集合 $\boldsymbol{Z}=[\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_n]$，则可以表示为

$$\boldsymbol{Z}=\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1k} \\ y_{21} & y_{22} & \cdots & y_{2k} \\ \vdots & \vdots & & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nk} \end{bmatrix}。\quad (1)$$

那么这 n 个数据包第 k 个属性可采用序列 $\boldsymbol{Z}_k=[y_{1k}, y_{2k}, \cdots, y_{nk}]$ 来表示。令第 k 个属性的离散度为 $\beta(\boldsymbol{Z}_k)$，则整个样本集合的离散度 $\beta(\boldsymbol{Z})$ 为

$$\beta(\boldsymbol{Z})=\frac{1}{k}\sum_{i=1}^{k}\beta(\boldsymbol{Z}_i)。\quad (2)$$

同时，假设某属性 k 的平均离散度为 $\overline{\beta(Z_k)}$，这里采用标准差 λ 来刻画数据包属性与整体平均离散度之间的偏差：

$$\lambda=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\beta(Z_{ik})-\overline{\beta(Z_k)})^2}，\quad (3)$$

其中，λ 越大意味着该数据包与标准样本偏离越远，越有可能被攻击篡改信息。

针对上述评价指标，本文曾经结合云模型建立了网络攻击检测方法（Detection method based of cloud model，DMCM），其算法流程如图 1 所示。在云模型中，将待检测某数据包 $Y=[y_1, y_2, \cdots, y_k]$ 视作云滴，云团是由一系列云滴构成，某个云滴代表定性对象的一次实现，并且云滴之间的顺序是无关的。云滴根据定义的规则发生器，在域内产生一个随机确定度，通过这一随机确定度来激活后件云发生器，以此产生新的云滴。但是由于云模型的不确定度量较大，使得计算量偏大。因此，本文利用粒子群优化（Particle swarm optimization，PSO）算法来建立新的网络攻击检测方法，以此减小计算量，同时避免陷入局部最优。

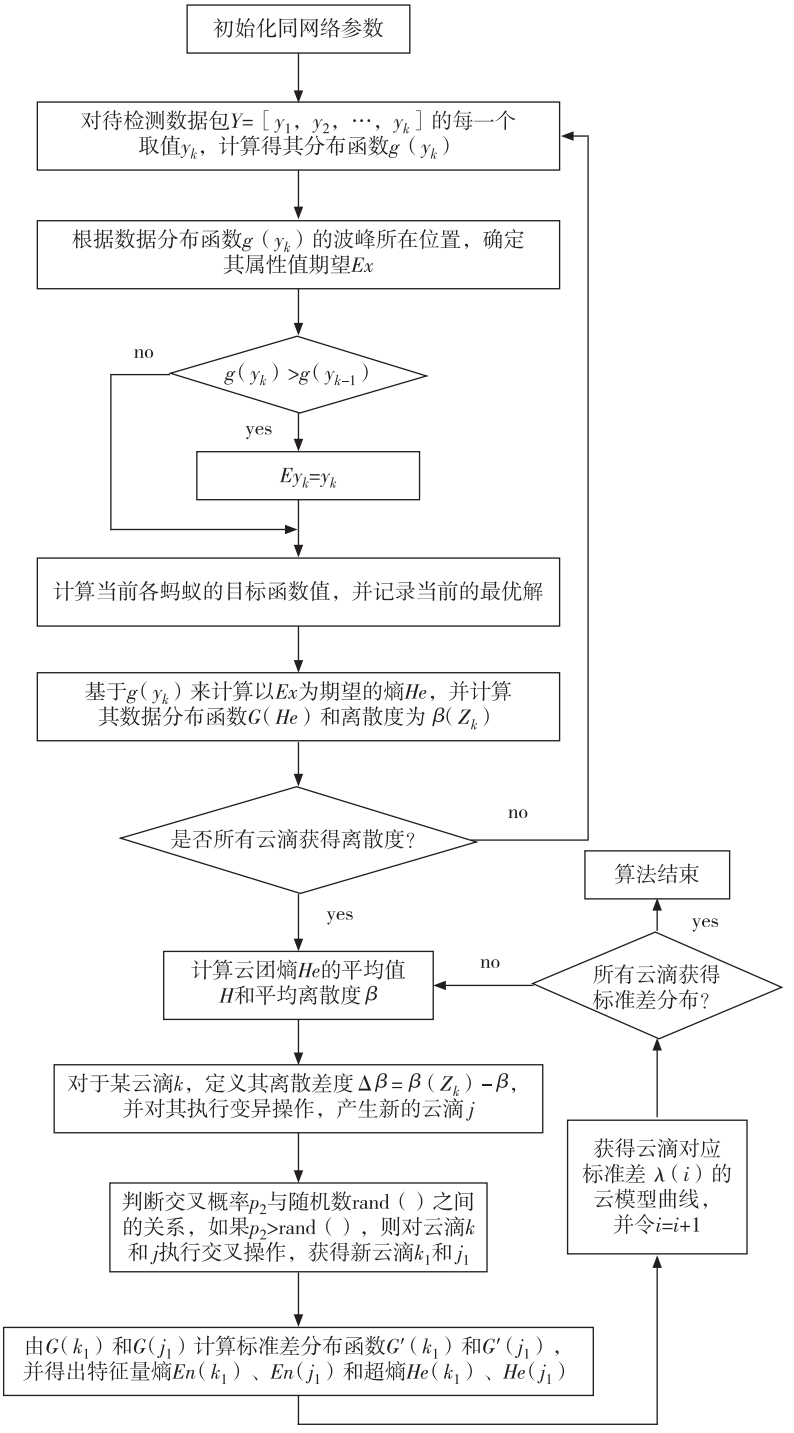

图 1　DMCM 算法流程图

2 基于粒子群的攻击检测方法

粒子群优化算法[15-16]是通过模拟鸟群觅食行为而发展起来的一种基于群体协作的随机搜索算法。粒子群优化算法中，每个优化问题的解都是搜索空间中的一个粒子，所有的粒子都有一个由被优化的函数决定的适应值，每个粒子还有一个速度决定他们优化的方向和距离，其他粒子就追随当前最优粒子在解空间中搜索，通过迭代找到最优解。在每一次迭代中，粒子通过两个极值来更新自身状态：个体极值和种群全局极值，个体极值是粒子本身所找到的最优解，种群全局极值是整个种群目前找到的最优解。

但是粒子群优化算法容易导致局部最优,所以本文结合变异算子来改进粒子群优化算法的缺陷,以此提高检测网络攻击的成功率。具体算法DMPS如下所述:

1) 在开始时刻初始化网络参数,并确定粒子群规模 n,产生粒子群的初始位置 $s(i,0)$和初始速度 $v(i,0)$;

2) 将待检测某数据包 $Y=[y_1,y_2,\cdots,y_k]$视作粒子 i,判断当前粒子 i 的标准差 λ 是否小于阈值 $\lambda_{\max}$,如果不满足则根据(4)式和(5)式所示的变异算子替换 $s(i,0)$,否则保持不变;

$$s(i,t+1)=s(i,t)+(s(max,t)-s(min,t))(1-\mathrm{rand}())^{\frac{1}{1+\varphi}}, \tag{4}$$

其中,$s(max,t)$和 $s(min,t)$分别为粒子位置 $s(i,t)$的边界,φ 为变异分布指数,rand()为(0,1)之间的随机数;

3) 按照(5)式所示的适应度函数计算粒子 i 的适应值 $f(\lambda)$

$$f(\lambda)=\frac{1}{1+\mathrm{e}^{-\lambda}} \tag{5}$$

并将粒子 i 的最佳位置 s_{opt} 确定为当前位置,同时暂时令 s_{opt} 为种群的最佳位置 s;

4) 根据当前最佳位置 s_{opt},结合(6)式和(7)式更新粒子的位置和速度

$$v(i,t)=v(i-1,t)+\eta|s_{opt}-s(i,t)|(1-\mathrm{rand}()) \tag{6}$$

$$s(i,t)=s(i-1,t)+v(i-1,t)\Delta t \tag{7}$$

其中,η 为状态更新参数,$\eta>0$;

5) 如果粒子 i 的标准差 λ 小于阈值 $\lambda_{\max}$则跳转到步骤6),否则以 $s(i,t)$作为初始点,采用变异算子计算的结果替换 $s(i,t)$,并重新计算其适应度;

6) 判断粒子 i 的适应度是否优于 s_{opt} 的适应度,如果是则令 s_{opt} 为粒子 i 的适应值;

7) 判断粒子 i 的适应度是否优于 s 的适应度,如果是则令 s 为粒子 i 的适应值;

8) 输出当前粒子的标准差 $\lambda(i)$,并令 $i=i+1$,跳转到步骤2)重复计算,直至获得所有粒子标准差分布 $\lambda=[\lambda(1),\lambda(2),\cdots,\lambda(k)]$,同时判断每个 $\lambda(i)$是否超出规定阈值,如果超出则存在被攻击的可能性;

9) 算法结束。

3 数学仿真

针对上述改进的DMPS算法,本文利用OPNET和Matlab进行仿真实验来验证其有效性。这里基于OPNET建立如图2所示的仿真拓扑结构,其网络参数设置为:每个数据包大小为512 b,链路带宽为20 M,每个网络节点缓冲区为1024 Kb,延时10 ms。其中,节点 S 作为数据源端(IP地址设为192.168.1.1),节点 D 作为数据接收端(IP地址设为192.168.1.100),节点 f 为攻击源(IP地址设为192.168.1.2),不定期向网络中发动攻击(包括DoS、Probe、R2L和U2R等),其余为中转节点(从节点 a 到节点 g 的IP地址分别设为192.168.1.3到192.168.1.9)。设置粒子群规模 $n=100$,变异分布指数 $\varphi=0.5$,状态更新参数 $\eta=2$。令节点 S 处每秒发送1 000个数据包到节点 D,每个数据包 $Y=[y_1,y_2,y_3]$,y_1 表示数据包长度,y_2 表示数据包时间戳,y_3 表示数据包源端地址。在节点 D 处设置监听,收集从节点 S 发送的数据包并进行状态分析,令节点 f 发动DoS攻击。这里将DMPS算法、DMCM算法、文献[9]提出的IHMM(Improved HMM model based method for detecting attacks)算法以及节点 D 处实际监听的结果进行对比,图3中显示了其数据包长度检测的情况。从图3可以看出,DMPS算法检测的数据包长度状态 y_1 与节点 D 处实际监听结合比较接近,其次是DMCM算法。对检测数据进行数据分析,DMPS、DMCM、IHMM与实际结果的误差分别为:3.62%、5.05%和8.28%。

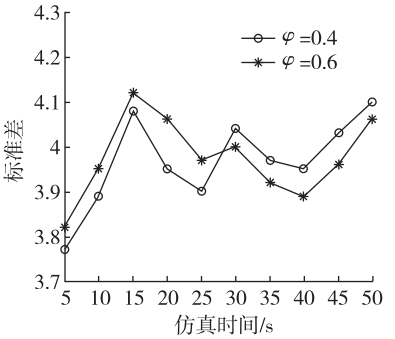

图2 不同变异分布指数 φ 下数据包长度标准差比较

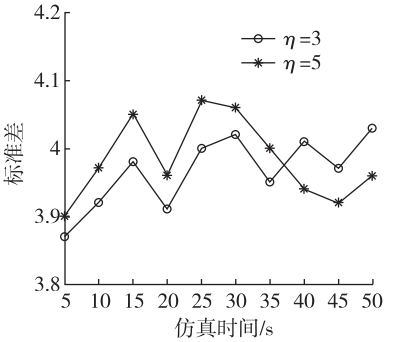

图3 不同状态更新参数 η 下数据包长度标准差比较

其次,在图4中给出了在上述仿真环境下,DMPS、DMCM和IHMM这三种算法的数据包长度标准差 λ 变化情况。从图4可以看出,在实验进入平稳过程后(仿真时间15 s后),DMPS所对应曲线的标准差小于其余两种算法。并且从标准差的抖动情况来看,DMPS也趋于稳定,而IHMM对应曲线

的抖动较大。

在表 1 中显示了当节点 f 分别发动 DoS、Probe、R2L 和 U2R 攻击下，DMPS、DMCM 和 IHMM 算法的检测成功率、漏报率以及误报率对比情况。从表 1 可以看出，本文改进的 DMPS 算法性能较 DMCM 和 RETMMAD 算法有了明显提高。

表 1　不同攻击种类下检测结果比较

攻击种类	DMPS 算法/%			DMCM 算法/%			IHMM 算法/%		
	成功率	漏报率	误报率	成功率	漏报率	误报率	成功率	漏报率	误报率
DoS	95.23	3.77	0	95.47	4.53	0	91.85	9.15	0
Probe	91.15	8.75	0	89.21	10.79	0	86.41	13.59	0
R2L	88.03	8.09	2.98	86.75	9.54	3.71	83.68	11.09	5.23
U2R	84.17	15.83	0	82.69	17.31	0	80.12	19.88	0

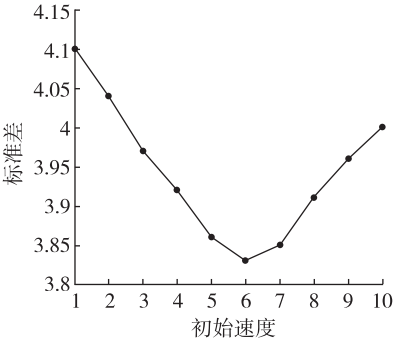

图 4　数据包长度标准差 λ 与粒子群初始速度 v 之间的变化关系

同时，为了进一步研究 DMPS 算法性能，这里对关键参数进行分析。假设令节点 f 仍然发动 DoS 攻击。图 2 给出了不同变异分布指数 φ 下数据包长度标准差 λ 的变化情况。从图 2 可以看出，在仿真初期，变异分布指数 φ 越小对应曲线的标准差越小，而在仿真后期，变异分布指数 φ 越大对应曲线的标准差越小。由于前期在节点 D 处聚集的数据包较少，标准差 λ 小于阈值 λ_{max} 的粒子数量较少，此时较小的变异分布指数 φ 就能获得最佳位置 $s(i,0)$ 和较优的适应值 $f(\lambda)$，从而使得数据包长度标准差 λ 较小；而随着节点 D 处聚集的数据包增多，标准差 λ 小于阈值 λ_{max} 的粒子数量较多，需要较大的变异分布指数 φ 才能获得最佳位置 $s(i,0)$ 和较优的适应值 $f(\lambda)$，因此此时变异分布指数 φ 越大对应曲线的标准差越小。而图 3 中给出了不同状态更新参数 η 下数据包长度标准差 λ 的变化情况。类似于图 2 中的现象，曲线出现了突变。在仿真初期状态更新参数 η 越小对应曲线的标准差越小，而在仿真后期状态更新参数 η 越大对应曲线的标准差越小。其原因在于，仿真初期节点 D 处聚集数据包较少，并且粒子群整体性能较低，稍微调整状态更新参数 η 将对粒子的位置和速度起到非常明显的作用，从而比较容易获得全局最优，而此时状态更新参数 η 越大反而会加大系统开销，不利于全局寻优操作；而仿真后期随着聚集数据包增加以及粒子群整体性能的提高，如果要明显改善粒子群性能，需要相对较大的状态更新参数 η，此时表现出状态更新参数 η 越大对应曲线的标准差越小。

最后图 4 给出了数据包长度标准差 λ 与粒子群初始速度 v 之间的变化关系。从图 4 可以看出，随着粒子群初始速度 v 的增加，标准差呈现出先降低后递增的趋势。这一点容易理解，在初始阶段粒子群整体性能偏低，此时加大粒子寻优的初始速度有利于更快收敛并获得全局最优，此时标准差呈现降低趋势；但是达到极值后，粒子群整体性能处于较优程度，进一步增加初始速度 v 只能使系统开销增加，不能达到提高性能的目的，从而标准差呈现递增趋势。

4 结束语

为了有效判断网络是否存在被攻击现象，本文在以往研究的基础上利用粒子群优化算法提出了一种新的检测算法 DMPS。首先该算法结合数据包属性的离散度和标准差定义了数据包样本判别指标，同时基于粒子群优化算法给出了其标准差分布的计算流程。最后，以 OPNET 和 Matlab 进行仿真实验，对比研究了该算法与 DMCM 算法、IHMM 算法在不同攻击种类下的性能状况，结果发现 DMPS 具有较好的适应性。在后续研究中，可以考虑结合多种网络环境来建立诸如 DoS、Probe、R2L 和 U2R 等各类攻击的检测方法。

参考文献：

[1] Liu P, Zang W, Yu M. Incentive-based modeling and inference of attacker intent, objectives, and strategies[J]. ACM Transactions on Information and System Security, 2005, 8(1): 78-118.

[2] Wang W, Daniels E. A graph based approach toward network forensics analysis[J]. ACM Transactions on Information and System Security, 2008, 12(1): 1-33.

[3] Qiu X Q, Paterson R. An innovative network security vulnerability modeling method and tool[J]. IEEE Communications Magazine, 2010, 48(1): 104-108.

[4] 严芬，殷新春，黄皓. 基于 MLL-AT 的网络攻击建模方法研究[J]. 通信学报，2011，32(3)：115-124.
Yan F, Yin X C, Huang H. Research on establishing network intrusion modeling based on MLL-AT[J]. Journal on

Communications,2011,32(3):115-124.

[5] 徐艳群,张斌,秦小铁.一种基于FGHI的集群入侵检测模型[J].重庆师范大学学报:自然科学版,2013,30(1):81-83.
Xu Y Q,Zhang B,Qin X T. Clustering intrusion detection model based on grey fuzzy *K*-mean clustering[J]. Journal of Chongqing Normal University:Natural Science,2013,30(1):81-83.

[6] 何保荣,李建荣,王超智.基于分类算法的校园网络入侵检测系统研究[J].电子工程设计,2013,21(12):12-13,17.
He B R,Li J R,Wang C Z. Study of the intrusion detection system in campus network based on the classification algorithm[J]. Electronic Design Engineering,2013,21(12):12-13,17.

[7] 杨双懋,郭伟,唐伟.基于FARIMA-GARCH模型的网络业务预测算法[J].通信学报,2013,34(3):23-31.
Yang S M,Guo W,Tang W. Network traffic prediction based on FARIMA-GARCH model[J]. Journal on Communications,2013,34(3):23-31.

[8] 陈晓天,刘静娴.改进的基于小波变换和FARIMA模型的网络流量预测算法[J].通信学报,2011,32(4):153-157.
Chen X T,Liu J X. Network traffic prediction based on wavelet transformation and FARIIMA[J]. Journal on Communications,2011,32(4):153-157.

[9] 杨晓峰,孙明明,胡雪蕾,等.基于改进隐马尔可夫模型的网络攻击检测方法[J].通信学报,2010,31(3):95-101.
Yang X F,Sun M M,Hu X L,et al. Improved HMM model based method for detecting cyber attacks[J]. Journal on Communications,2010,31(3):95-101.

[10] 赵攀,江宇波,邱玲,等.一种新的网络攻击检测方法[J].四川理工学院学报:自然科学版,2014,27(4):21-23,28.
Zhao P,Jiang Y B,Qiu L,et al. A etection method of network attacks[J]. Journal of Sichuan University of Science & Engineering:Natural Science Edition,2014,27(4):21-23,28.

[11] 程宏兵,容淳铭,黄晓,等.高效的攻击检测与数据融合算法[J].通信学报,2012,33(9):85-94.
Cheng H B,Rong C M,Huang X,et al. Efficient attack detection and data aggregation algorithm[J]. Journal on Communications,2012,33(9):85-94.

[12] 赵攀,江宇波,邱玲,等.基于元胞退火算法的僵尸网络传播特征研究[J].四川理工学院学报:自然科学版,2014,27(2):32-36.
Zhao P,Jiang Y B,Qiu L,et al. Study of botnet spread characteristic based on Cellular Annealing Algorithm[J]. Journal of Sichuan University of Science & Engineering:Natural Science Edition,2014,27(2):32-36.

[13] 袁恩隆,李飞,唐籍涛,等.改进蚁群算法的云存储任务调度算法研究[J].四川理工学院学报:自然科学版,2014,27(1):41-44.
Yuan E N,Li F,Tang J T,et al. Research on task schedule algorithm of cloud storage based on improved Ant Colony Algorithm[J]. Journal of Sichuan University of Science & Engineering:Natural Science Edition,2014,27(1):41-44.

[14] 熊伟,胡汉平,王祖喜,等.基于突变级数的网络流量异常检测[J].华中科技大学学报:自然科学版,2011,39(1):28-31.
Xiong W,Hu H P,Wang Z J,et al. Network traffic anomaly detection method based on catastrophe progression [J]. Journal of Huazhong University of Science and Technology:Natural Science Edition,2011,39(1):28-31.

[15] 罗金炎.连续型粒子群优化算法的均方收敛性分析[J].电子学报,2012,40(7):1364-1367.
Luo J Y. The analysis of continuous Particle Swarm Optimization Algorithm's mean square convergence[J]. Acta Electronica Sinica,2012,40(7):1364-1367.

[16] 李华,陈超,卢令.无线传感器网络拓扑控制算法的改进[J].四川理工学院学报:自然科学版,2014,27(3):65-69.
Li H,Chen C,Lu L,et al. Improvement of Topology Control Algorithm for wireless sensor networks[J]. Journal of Sichuan University of Science & Engineering:Natural Science Edition,2014,27(3):65-69.

The Method of Detection Network Attacks Based on Particle Swarm Optimization

ZHAO Pan, QIU Ling, CHEN Chao

(Faculty of Computer Science, Sichuan University of Science & Engineering, Zigong Sichuan 643000, China)

Abstract: In order to effectively determine the possibility of attacks for network packets, a new detection algorithm DMPS is proposed by previous studies. At first, the state indicators is defined with the discreteness of packet characteristic in this algorithm, and the calculation process of standard deviation distribution is presented to judge the anomaly of packet by Particle Swarm Optimization. Finally, a simulation with OPNET and Matlab was conducted to study the key factors of DMPS. Compared to the performance of other algorithm, the results shows that it has better adaptability.

Key words: attack; detection; anomaly; particle swarm optimization

(责任编辑 游中胜)