工业网络流量异常检测的概率主成分分析法

侯重远，江汉红，芮万智，刘亮
（海军工程大学电气与信息工程学院，430033，武汉）

摘要：针对主成分分析(PCA)法用于工业测控网络流量异常检测时存在的误报率高的问题，提出了一种基于概率主成分分析(PPCA)的检测算法.首先通过分析误报成因，建立了工业测控网络流量矩阵的PPCA模型，然后使用迭代变分贝叶斯算法辨识该模型的参数，再利用模型参数估计值求解流量矩阵的秩的分布函数并得到秩的极大似然估计值，最后以秩的跃变状况为判据进行异常流量检测.模拟攻击实验表明，该方法使漏报率平均下降了32%，从而有效降低了PCA方法的误报率.

关键词：工业网络；流量异常检测；主成分分析；误报率；变分贝叶斯
中图分类号：TP301 **文献标志码**：A **文章编号**：0253-987X(2012)02-0070-06

A Probabilistic Principal Component Analysis Approach for Detecting Traffic Anomaly in Industrial Networks

HOU Chongyuan, JIANG Hanhong, RUI Wanzhi, LIU Liang
(College of Electric and Information Engineering, Naval University of Engineering, Wuhan 430033, China)

Abstract: An algorithm using probabilistic principal component analysis (PPCA) is proposed to reduce the false alarm rate of anomaly detection of industrial networks using traditional principal component analysis (PCA). A PPCA model of industrial network traffic matrix is established by analyzing the causes of false alarm. Parameters in the model are identified by using the iterative variational Bayesian algorithm, and then are used to infer the rank of the PPCA model. Traffic anomalyis finally detected by making judgement on the rank. Simulated attack experiments show that the proposed method decreases false alarm rate by 32% in average, and effectively reduces the false alarm rate of PCA method.

Keywords: industrial networks; traffic anomaly detection; principal component analysis; false alarm rate; variational Bayesian method

随着工业网络在各领域的普及，近年开始出现以物理设施为攻击目标的网络攻击.这类攻击的特点是不以信息的窃取或软件功能的破坏为目标，而是以工业网络为媒介，通过恶意操纵物理设施的控制系统，达到破坏物理设施的目的.这类攻击的后果往往是灾难性的，因此研制面向工业网络的入侵检测系统十分迫切.

工业网络根据应用场合(如化工、电力、船舶等)不同差异很大，这种差异主要表现在通信协议、帧格式、传输介质等方面.然而，流量异常检测问题的本质是发现网络中流量规律的异常变化情况，只与工业网络的业务类型有关，而各类场合下的业务类型彼此相似，因而流量异常检测算法与具体应用场合无关.具体表现在绝大多数应用场合下，工业网络流量都包含着轮询、远程IO、组态、诊断、逢变即报(change of state，COS)、故障告警、周期刷新、同步

收稿日期：2011-06-14. 作者简介：侯重远(1982—)，男，博士生；江汉红(通信作者)，男，教授，博士生导师. 基金项目：国家自然科学基金创新研究群体科学基金资助项目(50421703).
网络出版时间：2011-10-27 网络出版地址：www.cnki.net/kcms/detail/61.1069.T.20111027.0957.008.html

测量等业务类型，各种应用场合都不超出这些类型，而网络攻击的流量特征一般与这些正常业务类型差异较大，流量异常检测算法以此原理检测攻击. 因此，本文研究的流量异常检测算法，与具体应用场合无关，研究具有普遍性.

针对工业网络的典型攻击是 Stuxnet 蠕虫，它攻击了 ProfiBus 工业测控网络，恶意操纵了可编程逻辑控制器(Programmable Logic Controller，PLC)所控制的机电设备. 除了 Stuxnet 外，外网入侵的途径还有利用工业无线设备的口令破解途径入侵、利用外网穿透的分布式拒绝服务攻击以及利用间谍芯片的硬件方式入侵等. 然而，在工业网络所有可能的攻击形式中，仅有 Stuxnet 在工程实际中被成功运用，因而代表了工业网络入侵技术的主要发展方向，因此本文以 Stuxnet 攻击方式作为研究载体，其他潜在的但尚未在工程实际中实现的攻击形式将在下一步工作中开展研究.

流量异常检测是网络入侵检测的主要方法之一[1]. 基于主成分分析(principal component analysis，PCA)的流量异常检测方法因其高检测率和抗变种能力而日益得到广泛重视和应用[2-5]. 这类方法以流量源节点和目的节点(OD)矩阵的滑动窗口为输入，通过 PCA 降维和对残差的 Q 统计量阈值判断来检测流量异常. 如文献[6]提出针对 PCA 的毒害流量机制，文献[7]提出面向防御毒害流量的鲁棒 PCA 算法，文献[8]提出 PCA 方法的多尺度异常检测算法，文献[9]指出 PCA 在反映时域相关性上的缺陷并提出了解决方案. 然而，这些工作仍以大时间尺度下的高速网络为研究对象，流量的多分形效应不足以影响 PCA 异常检测的性能. 文献[10-11]将 PCA 方法用于无线 mesh 网络的异常检测，能够实时检测拒绝服务和端口扫描等多种攻击，但这种方法是面向无线 mesh 网络的异构性拓扑而设计，并不适用于工业网络.

与互联网在大时间尺度上的流量特点十分相似，工业网络流量具有周期、平稳的特点，本文将用于互联网大尺度异常检测的 PCA 方法用于 ProfiBus 低速工业网络的异常检测时，发现该方法检测率很高，但存在误报率高的缺点，而流量波形显示随机突发流量和攻击流量无法区分是导致误报的原因. 这种低频的随机突发流量在能量和时间尺度上都与攻击流量接近，造成了 PCA 将随机突发流量和攻击流量都变换到了异常子空间，导致了误报.

为解决这一问题，本文引入概率 PCA 模型来描述随机突发流量对 PCA 的影响，然后利用变分贝叶斯(variational Bayesian，VB)理论精确的推断能力对概率 PCA 模型的秩进行推断，通过检测秩的变化判断异常流量，从而抑制随机突发流量对异常检测的干扰.

1　PCA 异常检测高误报率的原因分析

1.1　PCA 检测异常流量的原理

PCA 异常检测方法是通过分析流量 OD 矩阵的异常变化来实现异常检测[2-4].

定义 1　流量 OD 矩阵由网络中各个源节点和目的节点(简称 OD 对)之间流量需求(traffic demand)的时间序列构成[7].

大多数网络攻击在单条 OD 链路上的流量特征不明显，但各单条 OD 链路上的攻击流量之间存在相关性. 因此，用 PCA 对流量 OD 矩阵降维，可以得到流量 OD 矩阵的主成分，从而使原本分散在多条链路上的攻击流量被集中. PCA 检测算法用周期性和平稳性较强的流量主成分建立正常子空间，用突发性较强的流量主成分建立异常子空间，从而得到对流量 OD 矩阵的划分. 异常流量会使流量 OD 矩阵在异常子空间上的投影系数的 Q 统计量超出阈值，因而这一特性被 PCA 检测算法作为异常流量的判据.

1.2　随机突发流量是导致 PCA 高误报的原因

我们将 PCA 异常检测方法用于波特率较低的 Profibus 工业网络，研究工业网络的经典分布式攻击 Stuxnet 所引起的流量异常现象. 在研究中发现，Stuxnet 引起的异常流量主成分“淹没”在随机突发的噪声中，使得 PCA 检测算法将突发噪声也误报为异常流量. 分析导致这种误报现象的原因，可以归纳为以下两点.

(1)工业网络的波特率较低，使得分形现象产生的随机突发流量在大时间尺度上已不可忽略.

ProfiBus 工业网络一般工作在 1 Mb/s 以下的总线速率上，经虚拟令牌总线协议分时后，每个主站的实际波特率往往不超过 100 kb/s 量级，而互联网链路的信息传输速率远高于此. 互联网在较小时间尺度上的流量特性与工业网络较大时间尺度上的流量特性相似，网络在较小时间尺度上会表现出分形和多分形现象，前者是后者在 Mandelbrot 矩阵为定常时的情形. 网络流量分形现象的原因复杂，并且仍在争论之中，但以大数据块的传输、通信子网设备的

非线性、以及协议栈的分层特点等为主因. 因而,工业网络在亚秒尺度上就存在分形导致的随机突发流量.

(2)工业网络的攻击持续时间较短,与随机突发流量的时间尺度接近.

针对工业网络的攻击往往在很短时间采用猝发方式向全网大量执行器发送恶意指令. 与针对互联网的攻击(如 DDos 和蠕虫端口扫描等)相比,工业网络攻击的时间尺度小得多,造成攻击流量主成分与随机突发流量在持续时间上处于同一数量级,因此 PCA 无法从异常子空间投影系数上分辨攻击流和随机突发流.

2 基于概率 PCA 模型及其 VB 推断的异常检测方法

2.1 流量的概率 PCA 模型

概率 PCA(probabilistic principal component analysis, PPCA)模型被用来提高 PCA 处理随机数据源的性能[12].

设流量 OD 对的个数为 p,检测时间窗内的流量采样点数为 n,则网络流量信号的观测模型为

$$\boldsymbol{D}_{p\times n}=\boldsymbol{M}_{(r)p\times n}+\boldsymbol{E}_{p\times n} \tag{1}$$

式中:$\boldsymbol{M}_{(r)p\times n}$是流量的观测信号矩阵,$r$ 为秩,流量的量值须经归一化和零中心化后输入此矩阵;$\boldsymbol{E}_{p\times n}$是随机突发引起的观测噪声矩阵,为使计算复杂度满足实时检测的需求,采用均值为 0、方差为 ω^{-1} 的高斯分布近似随机突发流量. 因此,网络流量信号的 PPCA 模型为

$$f(\boldsymbol{D}\mid\boldsymbol{A},\boldsymbol{L},\boldsymbol{X},\omega,r)=N(\boldsymbol{ALX}',\omega^{-1}\boldsymbol{I}_p\otimes\boldsymbol{I}_n) \tag{2}$$

式中:$\boldsymbol{ALX}'$是 $\boldsymbol{M}_{(r)p\times n}$的经济奇异值分解,$\boldsymbol{A}\in\mathbf{R}^{p\times r}$和 $\boldsymbol{X}\in\mathbf{R}^{n\times r}$都是酉矩阵;$\boldsymbol{I}_p$ 和 $\boldsymbol{I}_n$ 为单位矩阵;$\boldsymbol{L}$ 是非零奇异值组成的对角阵,即

$$\boldsymbol{L}=\mathrm{diag}(\boldsymbol{l});\quad \boldsymbol{l}=[l_1,\cdots,l_r]^{\mathrm{T}};\quad l_1>l_2>\cdots>l_r>0 \tag{3}$$

2.2 使用 VB 理论推断概率 PCA 模型的秩

PPCA 模型的秩反映了主成分的总量. 网络上的攻击流量有使秩增大的趋势,而随机突发流量则有使秩分布趋于平滑的趋势. 这一结论的依据在于以下两点:①流量的全部主成分由周期恒稳的若干个正常主成分与突然出现的若干个异常主成分组成[2],由于攻击流量的出现会导致异常主成分个数的增多,而正常主成分并未改变,因而总的主成分个数将增加,又由于秩反映了主成分的总量,因此攻击流量有使秩增大的趋势;②根据式(2),随机突发流量可以看做高斯部分与非高斯部分之和,其中高斯部分由于已计入 PPCA 模型,不会影响秩的分布,而非高斯部分对秩的分布的影响由于没有任何先验信息因而会对秩的分布函数产生扁平化影响. 综合以上两点原因,随机突发流量会使秩的分布平滑化. 根据这一原理设计检测算法,就需要对 PPCA 模型的秩进行推断. 在贝叶斯推断体系中,基于极大似然原理的 EM 和 MAP 等方法在解决 PPCA 模型的推断问题时,利用模型的秩作为先验信息,因而不具有对秩的推断能力;近年发展起来的变分贝叶斯(VB)理论尽管计算量较大,但不仅可以推断模型的秩,而且能够得到秩的分布函数. 因此,本文采用 VB 算法作为 PPCA 模型的秩的推断算法[13].

根据贝叶斯公式,r 的后验分布可以表示为

$$f(r\mid\boldsymbol{D})\propto f(\boldsymbol{D}\mid r)f(r) \tag{4}$$

用结森不等式进行近似,则有

$$\begin{aligned}\ln f(\boldsymbol{D}\mid r)\approx{}&\ln f(\boldsymbol{D}\mid r)-\mathrm{KL}(\tilde{f}(\boldsymbol{A},\boldsymbol{X},\boldsymbol{L},\omega\mid\boldsymbol{D},r)\,\|\\&f(\boldsymbol{A},\boldsymbol{X},\boldsymbol{L},\omega\mid\boldsymbol{D},r))=\\&\int_{\Theta^*}\tilde{f}(\boldsymbol{A},\boldsymbol{X},\boldsymbol{L},\omega\mid\boldsymbol{D},r)(\ln f(\boldsymbol{D},\boldsymbol{A},\boldsymbol{X},\boldsymbol{L},\omega\mid r)-\\&\ln(\tilde{f}(\boldsymbol{A},\boldsymbol{X},\boldsymbol{L},\omega\mid\boldsymbol{D},r))\mathrm{d}\boldsymbol{A}\mathrm{d}\boldsymbol{X}\mathrm{d}\boldsymbol{L}\mathrm{d}\omega\end{aligned} \tag{5}$$

式中:KL(·|·)是 2 个分布的 Kullback-Leibler (KL)偏差;$\tilde{f}$ 是分布 f 的最优近似.

求解式(5)的难度在于矩阵高维积分带来的复杂运算问题,而文献[13]则使用变分近似和变量分离技巧克服了这一问题,得到精度很高的估计. 本文使用下面算法实现对秩的估计.

(1)用 IVB 算法[13]求解 VB 方程

$$\left.\begin{aligned}&\boldsymbol{k}_{\boldsymbol{A}}=G(p,\hat{\omega}\boldsymbol{l}_{:r}\circ\boldsymbol{k}_{\boldsymbol{X}}\circ\hat{\boldsymbol{l}})\\&\boldsymbol{k}_{\boldsymbol{X}}=G(n,\hat{\omega}\boldsymbol{l}_{:r}\circ\boldsymbol{k}_{\boldsymbol{A}}\circ\hat{\boldsymbol{l}})\\&\boldsymbol{\mu}_{\boldsymbol{l}}=\boldsymbol{k}_{\boldsymbol{X}}\circ\boldsymbol{l}_{:r}\circ k_{\boldsymbol{A}}\\&\phi=\hat{\omega}^{-1}\\&\hat{\boldsymbol{l}}=\boldsymbol{\mu}_{l+}(\phi)^{1/2}\varphi(\boldsymbol{\mu}_{\boldsymbol{l}},\phi)\\&\hat{\boldsymbol{l}}^{\mathrm{T}}\hat{\boldsymbol{l}}=r\phi+\boldsymbol{\mu}_{\boldsymbol{l}}^{\mathrm{T}}\hat{\boldsymbol{l}}-\phi^{1/2}\kappa(\boldsymbol{\mu}_{\boldsymbol{l}},\phi)^{\mathrm{T}}\mathbf{1}_{\mathrm{r},1}\\&\hat{\omega}=pn[\boldsymbol{l}^{\mathrm{T}}\hat{\boldsymbol{l}}-2(\boldsymbol{k}_{\boldsymbol{X}}\circ\hat{\boldsymbol{l}}\circ\boldsymbol{k}_{\boldsymbol{A}})^{\mathrm{T}}\boldsymbol{l}_{:r}+\hat{\boldsymbol{l}}^{\mathrm{T}}\hat{\boldsymbol{l}}]\end{aligned}\right\} \tag{6}$$

式中:符号∘是 Hadamard 积;$\kappa(\cdot,\cdot)$和 $G(\cdot,\cdot)$是截断高斯分布的辅助函数;$\boldsymbol{l}_{:r}$指 $\boldsymbol{l}$ 的前 r 行. 上标^指估计值;$G(\cdot,\cdot)$是矩阵超几何分布的各阶偏导的简化式

$$G(p,\boldsymbol{l}_F)=\frac{\partial}{\partial \boldsymbol{l}_{F,i}}\ln{}_0\boldsymbol{F}_1(p/2,\boldsymbol{l}_F^2/4) \tag{7}$$

其中，$\boldsymbol{F}$ 是 Mises-Fisher 分布的参数矩阵，${}_0\boldsymbol{F}_1(\cdot,\cdot)$ 是 $\boldsymbol{F}\boldsymbol{F}^{\mathrm{T}}$ 的超几何分布函数，$\boldsymbol{l}_F$ 是 $\boldsymbol{F}$ 的经济奇异值向量.

(2)将 IVB 算法收敛后得到的各估计数值代入 PPCA 模型的秩的分布函数(8)[13]，得到秩的变分近似估计

$$\left.\begin{aligned}
&\tilde{f}(r\,|\,\boldsymbol{D})\propto\exp\{R_1+R_2+R_3+R_4+R_5+R_6\}\\
&R_1=-\frac{r}{2}\ln\pi+r\ln 2+\ln\Gamma\left(\frac{r}{2}+1\right)+\ln(r!)\\
&R_2=\frac{1}{2}\phi^{-1}(\boldsymbol{\mu}_l^{\mathrm{T}}\boldsymbol{\mu}_l-\hat{\boldsymbol{l}}^{\mathrm{T}}\boldsymbol{\mu}_l-\boldsymbol{\mu}_l^{\mathrm{T}}\hat{\boldsymbol{l}}+\hat{\boldsymbol{l}}^{\mathrm{T}})\\
&R_3=\ln{}_0\boldsymbol{F}_1\left(\frac{1}{2}p,\frac{1}{4}\boldsymbol{F}_A\boldsymbol{F}_A^{\mathrm{T}}\right)-\hat{\omega}(\boldsymbol{k}_X\circ\hat{\boldsymbol{l}}\circ\boldsymbol{k}_A)^{\mathrm{T}}\boldsymbol{l}_{;r}\\
&R_4=\ln{}_0\boldsymbol{F}_1\left(\frac{1}{2}n,\frac{1}{4}\boldsymbol{F}_X\boldsymbol{F}_X^{\mathrm{T}}\right)-\hat{\omega}(\boldsymbol{k}_X\circ\hat{\boldsymbol{l}}\circ\boldsymbol{k}_A)^{\mathrm{T}}\boldsymbol{l}_{;r}\\
&R_5=\sum_{j=1}^{r}\ln\left[\mathrm{erf}\left(\frac{\bar{\boldsymbol{l}}_j-\boldsymbol{\mu}_{j,l}}{(\phi)^{1/2}}\right)+\mathrm{erf}\left(\frac{\boldsymbol{\mu}_{j,l}}{(\phi)^{1/2}}\right)\right]\\
&R_6=r\ln((\pi\phi/2)^{1/2}-(\theta+1)\ln\rho)
\end{aligned}\right\} \tag{8}$$

式中：$\boldsymbol{F}_A$ 和 $\boldsymbol{F}_X$ 是 $\boldsymbol{F}$ 奇异值分解后的酉矩阵；θ 和 ρ 是 Gamma 分布的参数；$\bar{\boldsymbol{l}}_j$ 是 $\boldsymbol{l}_j$ 的上界.

(3)PPCA 模型的秩的后验信息已完全体现在式(8)中，为简化检测算法，本文使用 r 的极大似然估计作为最终的估计值

$$\hat{r}=\arg\max_r\tilde{f}(r\,|\,\boldsymbol{D}) \tag{9}$$

2.3　使用秩的变化建立异常判据

上节得到了秩的点估计值 $\hat{r}$. 将估计值 $\hat{r}$ 与其历史数据比较，可以通过 $\hat{r}$ 的跃变情况设计检测算法. 本文将 $\Delta\hat{r}$ 的阈值设定为 1，即当主成分个数的增加量超过 1 个时则报警.

综上，本文流量异常检测算法如图 1 所示.

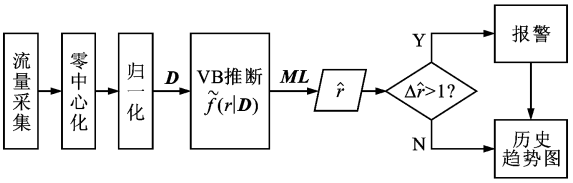


图 1　流量异常检测算法流程图

3　实验与结果分析

3.1　异常检测实验系统搭建

实验环境按图 2 搭建，使用网络分析仪采集 ProfiBus 流量. 流量数据通过 OPC Tools 工具箱导入至 Simulink，并按照图 1 算法流程进行异常检测.

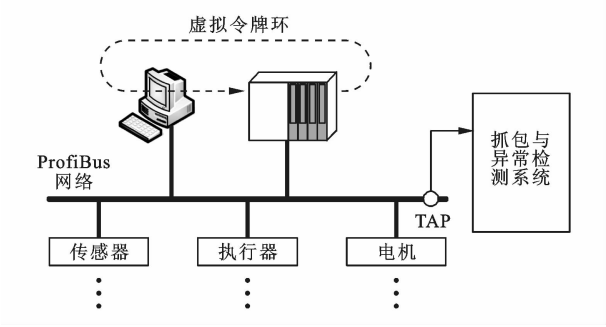
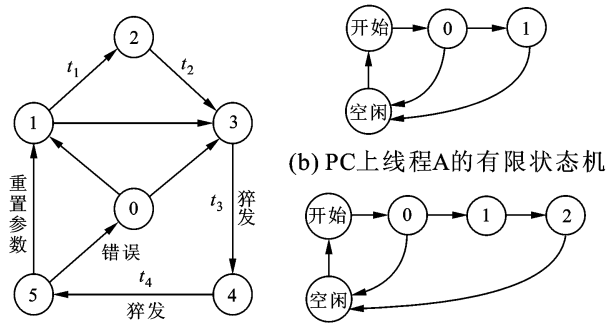


图 2　异常检测实验系统搭建

3.2　Stuxnet 的攻击流量的模拟

Stuxnet 是工业网络的典型攻击，但它只在遇到特定型号的可编程逻辑控制器(PLC)时才会自我激活，目前实验室不能激活. 为此，本文根据铁门塞克和 ESET 的 Stuxnet 代码逆向工程报告[14-15]，模拟了 Stuxnet 在传播和攻击时的网络通信行为，从而实现了对攻击流量的模拟.

Stuxnet 的网络通信行为的有限状态机如图 3 所示，各状态描述如表 1～表 3 所示.



(a) PLC上的有限状态机　(b) PC上线程A的有限状态机　(c) PC上线程B的有限状态机

图 3　Stuxnet 的网络通信行为有限状态机

表 1　Stuxnet 在 PLC 上的有限状态机状态描述

状态	状态描述
1	攻击初始化. 询问受控设备的运行状态，判断是否进行攻击.
2	攻击等待状态. 等待时间 t_1 为数周至数月不等.
3	攻击准备状态. t_2 为 2 h. 准备完成后，第 1 次向底层设备猝发恶意指令.
4	第 2 次向底层设备猝发恶意指令. t_3 为 15～50 min.
5	重置所有参数.
0	重置失败，判断继续攻击或初始化.

表2 Stuxnet在PC上线程A的有限状态机状态描述

状态	状态描述
开始	数据收集,完成同步前准备
0	询问被控PLC是否满足同步条件
1	发出PLC同步攻击命令
空闲	空闲等待. t_5 为5 min

表3 Stuxnet在PC上线程B的有限状态机状态描述

状态	状态描述
开始	数据收集,完成攻击PLC前准备
0	询问PLC是否满足攻击条件
1	向PLC发出代码块移动指令
2	向PLC传输Stuxnet样本
空闲	空闲等待. t_6 为15 min

3.3 检测效果的对比分析

本文检测算法基于PPCA模型及其VB秩推断,而文献[2-4]算法基于PCA模型及异常子空间的Q统计量阈值检测.本文在多种攻击强度下对这2种算法的误报率、检测率和接收者操作特性(receiver operating characteristic, ROC)曲线进行了比较分析.

攻击强度是平均攻击流量发送速率与链路最大带宽的归一化比值,误报率是指误报样本占报警总次数的比率.2种算法的误报率随攻击强度变化的曲线如图4所示.由图4可见,基于PPCA模型的算法有效降低了误报率,尽管在攻击强度较小时误报率没有明显改善,但随着攻击强度的增大,误报率明显下降,平均下降率达到32%,可见本文算法可以有效抑制误报现象.在攻击强度较小时,3种算法的误报率都达到100%,没有检测攻击的能力;随着攻击强度增大,基于PPCA模型的误报率大幅降低,且逐渐趋于饱和.

检测率是指正确检测到的攻击样本数占总攻击次数的比率.2种算法的检测率随攻击强度变化的曲线如图5所示.从图5可以看到,随着攻击强度的增大,基于PCA模型的检测算法和本文基于PPCA模型的算法在检测率上都有显著提高,二者提高幅度的差异不显著,因此性能接近.本文算法在检测率方面改善不明显,原因主要有以下2方面.

(1)攻击的分布式特性.由于Stuxnet是分布式攻击,而对分布式攻击的检测取决于检测算法对各

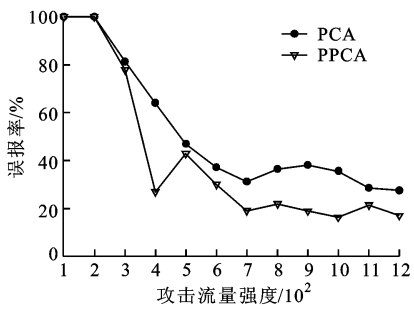


图4 基于2种模型算法的误报率比较

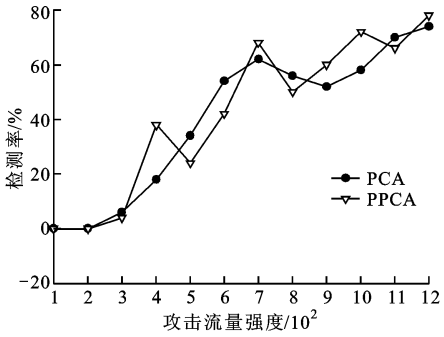


图5 基于2种模型算法的检测率比较

分支攻击流量的聚合程度,因此聚合程度高的算法其检测率才能提高,而本文算法在对分支流量的聚合能力上,本质上仍属于主成分分析,因此检测率也与主成分分析异常检测算法的检测率相近.可见,本文算法在显著降低误报率的同时,尚不能有效提高检测率.

(2)工业网络业务的冲击性特性.工业网络常常出现冲击性背景流量,即在很短时间段内,流量由各分支节点向控制中心节点或数据中心节点汇聚,此时背景流量的主成分会急剧增大而后急剧减小,呈现冲击性特点.此时,Stuxnet引起的异常流量主成分会因强度的相对不显著而"淹没"在冲击性背景流量中而不易被检测.检测算法若能在冲击性背景流量产生期间将攻击流量从冲击性背景中分离,则也能提高检测率,但图5中2种算法都不具有这种分离能力,因而这也是检测率不能改善的另一个原因.

ROC曲线综合反映了一种检测算法的性能,曲线下方覆盖面积大的算法更优.2种算法的ROC曲线如图6所示.由于设备性能限制,只能得到灵敏度在0%~80%附近的测试样本,由这部分样本的ROC曲线可以看出,基于PPCA模型的算法下方面积更大,因此其性能明显优于基于PCA模型的算法.

综合以上结果,本文基于PPCA模型及其秩推断的检测算法有效降低了传统PCA算法的误报率.

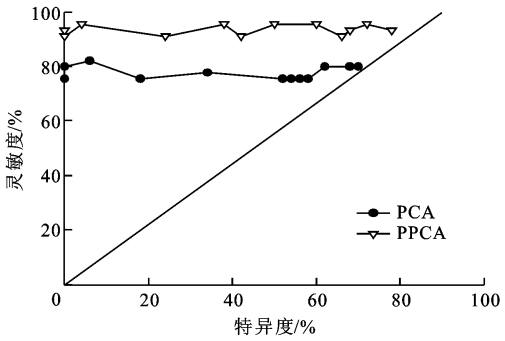


图 6　基于 2 种模型算法的 ROC 曲线对比

4　结束语

主成分分析法是成熟的异常检测方法，本文针对主成分分析法在工业网络异常检测时所出现的误报问题，首先分析了误报原因，进而提出使用改进的概率主成分分析方法解决误报问题的思路，通过建立基于概率主成分分析模型的变分贝叶斯秩推断的异常判据，有效降低了异常检测方法的误报率. 在实验中，我们也注意到本文方法在检测率方面还有改进空间，因此下一步的研究重点是进一步提高检测率. 工业测控网络入侵和攻击的潜在途径还有很多，未来的工作中将加强对其他潜在攻击途径防范措施的研究.

参考文献：

[1] 颜若愚，郑庆华. 使用交叉熵检测和分类网络异常流量[J]. 西安交通大学学报，2010，44(6)：10-15.
YAN Ruoyu，ZHENG Qinhua. Using cross entropy to detect and classify network anomalous traffic [J]. Journal of Xi′an Jiaotong University，2010，44(6)：10-15.

[2] LAKHINA A，CROVELLA M，DIOT C. Diagnosing network-wide traffic anomalies[C]// Proceedings of ACM SIGCOMM 2004：Conference on Computer Communications. New York，USA：ACM，2004：219-230.

[3] LAKHINA A，CROVELLA M，DIOT C. Characterization of network-wide anomalies in traffic flows[C]// Proceedings of the 2004 ACM SIGCOMM Internet Measurement Conference. New York，USA：ACM，2004：201-206.

[4] LAKHINA A，CROVELLA M，DIOT C. Mining anomalies using traffic feature distributions [J]. Computer Communication Review，2005，35(4)：217-228.

[5] 张文铸，刘佳，袁坚，等. 基于 PCA 的对等网络流量时空特性监测[J]. 清华大学学报：自然科学版，2010，50(4)：561-564.
ZHANG Wenzhu，LIU Jia，YUAN Jian，et al. PCA based approach for monitoring the spatial-and-temporal characteristics of P2P traffic [J]. Journal of Tsinghua University：Science and Technology，2010，50(4)：561-564.

[6] RUBINSTEIN B，NELSON B，HUANG L，et al. Compromising PCA-based anomaly detectors for network-wide traffic，UCB/EECS-2008-73 [R]. Berkeley，USA：UCB，2009.

[7] 钱叶魁，陈鸣. 面向 PCA 异常检测器的毒害攻击和防御机制[J]. 电子学报，2011，39(3)：543-548.
QIAN Yekui，CHEN Ming. Poison attack and defense strategies on PCP based anomaly detector[J]. Acta Electrionica Sinica，2011，39(3)：543-548.

[8] CHATZIGIANNAKIS V，PAPAVASSILIOU S，ANDROULIDAKIS G. Improving network anomaly detection effectiveness via an integrated multi-metric-multi-link（M^3L）PCA-based approach[J]. Security and Communication Networks，2009，2(3)：289-304.

[9] BRAUCKHOFF D，SALAMATIAN K，MAY M. Applying PCA for traffic anomaly detection：problems and solutions[C]// Proceedings of IEEE INFOCOM 2009. Piscataway，NJ，USA：IEEE，2009：2866-2870.

[10] ZAIDI Z，HAKAMI S，MOORS T，et al. Detection and identification of anomalies in wireless mesh networks using principal component analysis [J]. Journal of Interconnection Networks，2009，10(4)：517-534.

[11] ZAIDI Z R，HAKAMI S，LANDFELDT B，et al. Real-time detection of traffic anomalies in wireless mesh networks [J]. Wireless Networks，2010，16(6)：1675-1689.

[12] BISHOP M，TIPPING E. Probabilistic principal component analysis [J]. Journal of the Royal Statistical Society，1999，61(3)：611-622.

[13] VÁCLAVÁ，ANTHONY Q. The variational Bayes method in signal processing[M]. Berlin，Germany：Springer，2006：57-88.

[14] NICOLAS F，LIAM O M，ERIC C. W32. stuxnet dossier [EB/OL]. [2010-9-30]. http：//www. symantec. com/connect/blogs/w32stuxnet-dossier. html.

[15] ALEKSANDR M，EUGENE R，DAVID H，et al. Stuxnet under the microscope [EB/OL]. [2010-7-19]. http：//www. eset. com/resources/white-papers/Stuxnet_Under_the_Microscope. pdf.

（编辑　刘杨）