基于 SVM 的 Telnet 隐蔽信道检测

吴传伟，孙 瑞，罗 敏
(中国电子科技集团公司第三十研究所，四川 成都 610041)

[摘　要] 基于 Telnet 的隐蔽信道将隐匿的消息直接附加在 Telnet 的网络数据中，并发送至远程 “ 服务器 ” 。由于键盘操作具有任意性，检测这种信道比较困难。通过分析 Telnet 隐蔽信道技术，提出针对该隐蔽信道的检测方法。检测方法使用了一分类支持向量机 (SVM)，抓取用户正常操作的网络数据包作为检测样本，并利用样本数据间的时间间隔构造检测向量。试验表明，利用这种方法对基于 Telnet 的隐蔽信道进行检测，检测率达到 100%，且虚警率较低。

[关键词] 隐蔽信道；Telnet；一分类支持向量机；训练集
[中图分类号] TN918.91　　[文献标识码] A　　[文章编号] 1009-8054(2012)09-097-02

Detection of Telnet Covert Channel based on SVM

WU Chuan-wei，SUN Rui，LUO Min
(No.30 Institute，CETC，Chengdu Sichuan 610041，China)

[Abstract] The covert channel based on Telnet attaches the hidden message to the telnet packet and sends it to the “ remote server ” ，and this channel is very difficult to be detected because of the arbitrary keyboard operation. This paper analyzes the covert channel technologies based on Telnet，and proposes a detection method for the covert channel. The detection method uses one-class support vector machine(SVM)，and by capturing the user's normal operation of the network packets as the template and making use of the time interval between samples，constructs the detection vector. Experiment indicates that，with this method，the detection rate for the covert channel based on telnet could reach 100%，while the false detection rate is fairly low.

[Keywords] covert channel；Telnet；one-class support vector machine(SVM)；training aggregate

0 引言

隐蔽信道有别于一般的通信方式，它将要传送的秘密信息进行伪装和隐藏，通过传统的通信方式或系统漏洞将秘密信息传送出去，而使对手觉察不到秘密通信的存在。随着互联网技术的不断发展，基于网络协议的隐蔽信道技术 [1-2] 成为人们关注的焦点。站在网络安全的角度，隐蔽信道的存在给计算机网络带来了极大的安全隐患。

Telnet 在工作时，计算机终端将用户的键盘操作信息传输到服务器端，而用户的键盘操作是不可预知的，因此很难依据 Telnet 传输的内容来判断是否存在基于 Telnet 的隐蔽信道，但是计算机用户在使用 Telnet 时对键盘的操作与隐蔽信道软件所表现的行为会有所不同。文中采用了一分类支持向量机 (One-Class-SVM) 算法，通过采集正常的 Telnet 数据建立训练集，达到了检测该种隐蔽信道的目的。

1 基于 Telnet 的隐蔽信道

计算机终端用户使用 Telnet 协议，通过本地键盘操作已登录的远程服务器。由于键盘操作具有任意性，构建隐蔽信道的最简单的方法就是将要隐匿的消息直接附加在 Telnet 的网络数据中，并发送至远程 “ 服务器 ” ，为了通信更加安全，可事先将消息进行加密。如果隐蔽信道软件在发送 Telnet 数据时增加一定时延，将使该信道更加难以检测。

2 支持向量机

支持向量机基于统计学习理论和结构风险最小化原则，根据有限的样本信息在模型的复杂性和学习能力之间寻求最佳折衷，以期获得最好的推广能力 [3]。最优分类线就是要求分类面不但能将两类正确分开，而且使分类间隔最大，如图 1 所示。

设线性判别方程 $x \cdot w+b=0$，对其归一化，得到：

$y_i[(w \cdot x_i)+b]-1 \geq 0$，$i=1, 2, \cdots, n$，$(x_i, y_i)$，$i=1, 2, \cdots, n$，$X \in R^d$，$y \in \{+1, -1\}$　　(1)

收稿日期：2012-07-24
作者简介：吴传伟，1981 年生，男，工程师，研究方向：计算机网络安全、移动通信技术等；孙瑞，1982 年生，男，工程师，研究方向：信息网络安全、嵌入式应用等；罗敏，1976 年生，男，高级工程师，研究方向：网络安全、移动通信安全技术等。

分类间隔 $\rho=2/\|W\|$，最优分类面要同时满足式 (1) 及 $\min(\frac{1}{2}\|w\|^2)$。

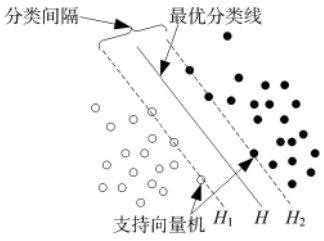

图 1 最优分类面示意

用 Lagrange 乘子方法求式 (2) 最大值：

$$\max(Q(\alpha))=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j(x_i\cdot x_j)\sum_{i=1}^{n}\alpha_i y_i=0$$

和 $\alpha_i\geq 0$，$i=1$，2，…，n (2)

式 (2) 为一个不等式约束下二次函数寻优的问题，存在唯一解 。不为零的解对应的样本就是支持向量，故最优决策函数为：

$$f(x)=\mathrm{sgn}(\sum_{i=1}^{n}\alpha_i^* y_i(x_i\cdot x)+b^*) \quad (3)$$

若样本线性不可分，则在式 (1) 中引入一个松弛项 $\xi_i\geq 0$，得：

$$y_i[(w\cdot x_i)+b]-1-\xi_i\geq 0,\ i=1,2,\cdots,n \quad (4)$$

将目标改为求：

$$\min(w,\ \xi)=\frac{1}{2}\|w\|^2+C(\sum_{i=1}^{n}\xi_i),\ 0\leq\alpha_i\leq C \quad (5)$$

当样本非线性时，需要将输入向量 X 映射到一个更高维的特征空间 H，并构造最优分类超平面。根据 Mercer 条件采用不同的内积函数 $K(x_i,\ x_j)$ 实现非线性分类，$K(x_i,\ x_j)$ 称为核函数，此时最优决策函数变为：

$$f(x)=\mathrm{sgn}(\sum_{i=1}^{n}\alpha_i^* y_i K(x_i,\ x)+b^*)$$

支持向量机将低维空间的非线性问题变换到一个线性高维特征空间，通过内积函数在高维空间求最优分类面。支持向量机分类函数结构上类似一个神经网络，输出是中间节点，每个节点对应一个支持向量，如图 2 所示。

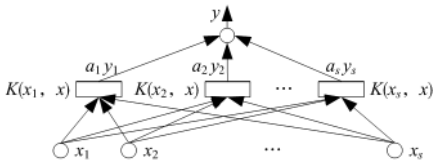

图 2 VM 示意

选择不同的核函数 $K(x_i,\ x_j)$ 可以构造不同的 SVM，目前常用的 $K(x_i,\ x_j)$ 包括：

1) 多项式核函数，$K(x,\ x_i)=(x\cdot x_i+1)^q$。

2) 径向基核函数，$K(x,\ x_i)=\exp(-\frac{1}{2\sigma^2}\|x-x_i\|^2)$。

3) 多层感知器核函数，$K(x,\ x_i)=\tanh(\beta_1(x\cdot x_i)+\beta_2)$。

SVM 一般用来对二分类或多分类问题进行求解，而对于基于 Telnet 的隐蔽通道的检测，由于只能获取正常用户的操作数据，无法得到异常数据，因此不能用一般的 SVM 进行训练，得到最优超平面，这时需要 One-Class-SVM[4]。One-Class-SVM 使用无标记数据训练，以寻找将训练数据集与原点分开的超平面。在一分类支持向量机利用参数 v，$v\in(0,1)$ 代替折中参数 [5]，求：

$$\min(L_D(\alpha))=-\frac{1}{2}\sum_{i,\ j=1}^{l}\alpha_i\alpha_j K(x_i\cdot x_j)$$

其中，$0\leq\alpha_i\leq\frac{1}{vl}$；$\sum_{i=1}^{l}\alpha_i=1$。

判别函数：$f(x)=\mathrm{sgn}(\sum_i\alpha_i k(x_i,\ x)-\rho)$。

3 基于 Telnet 的隐蔽信道检测分析

使用隐蔽信道检测软件的抓包工具获取正常用户的 Telnet 操作信息，将抓包工具的采样时间间隔设置为 0.5 ms，并设置窗口值 w=8。此次试验邀请了 8 名能够熟练操作 Telnet 的人员参与，在数据采集前拟定了操作流程，并要求他们按照该流程进行训练。试验训练集数量为 600。在数据抽样采集过程中发现有些样本数据会明显偏离正常值，这可能是由参与人员长时间操作疲劳导致，不能将这些数据作为训练数据，需要剔除，否则会影响检测效果。由于不同人员习惯不同，使用 Telnet 进行操作的方式也会有所不同，因此剔除数据的比例也不尽相同，比例范围一般在 15%~35%。通过试验表明，如果没有对异常数据进行剔除，利用 One-Class-SVM 进行训练并检测，误检率 $E\approx$35%，若将异常数据进行剔除后进行检测，E<4%。试验表明无论是否对异常数据进行剔除，其检测率都达到 100%，具有很好的检测效果。

Telnet 隐蔽信道检测软件利用了支持向量机算法库 LIBSVM[6](V3.12)，LIBSVM 是一个支持向量机算法的 C 语言函数库，集成了多种支持向量机算法，包括 One-Class-SVM。使用 One-Class-SVM 对 Telnet 隐蔽信道进行检测，选择径向核函数作为支持向量机的核函数，在试验中不断调整径向核函数参数 σ 和折中参数 v，以获取最佳检测效果。通过不断尝试，发现当 v=0.03，σ=2.88~3.52，隐蔽信道检测软件具有较好的检测效果。基于 Telnet 的隐蔽信道检测结果如表 1 所示，其中训练样本 (N) 和测试样本 (T) 都是剔除异常数据后的样本。

表 1 Telnet 隐蔽信道检测结果

NO	W	N	T	E/(%)
1	5	428	367	3.8
2	6	476	406	3.6
3	7	487	352	3.1
4	8	496	386	2.6
5	9	402	421	2.4
6	10	413	236	3.7
7	11	501	389	3.5

(下转第 101 页)

5) 数据压缩模块。为用户数据提供不同压缩率的数据编码压缩支撑。

6) 数据烧录控制模块。将封装好的用户数据和安全控制信息写入光盘 (含擦除、销毁光盘数据)，或从光盘上读出数据进行解析。

7) 加解密控制模块。提供用户身份认证和访问控制、数据加解密、数字签名 / 验证、PIN 码掩盖、完整性校验等功能。

8) 数据快速销毁模块。对软件系统在宿主机内存或硬盘上产生的临时文件进行快速随机擦除。

9) 内存管理控制模块。对宿主机内存或硬盘进行调度管理，满足安全、快速的软件加密要求。

10) 托盘监听程序。监听 USBKey、光盘等接入过程，提供配置系统软件功能的快速操作界面。

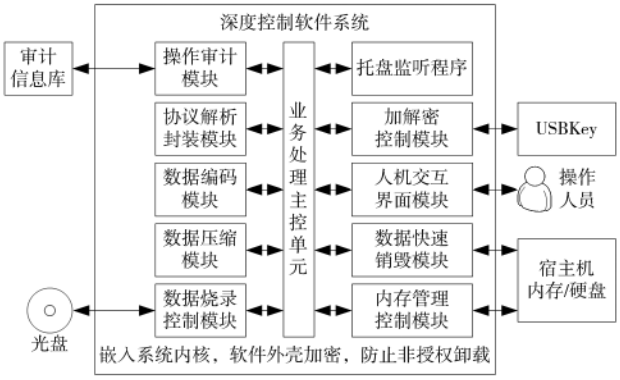

图 4 深度控制软件系统架构

深度控制软件系统采用软件方式实现多个不同强度的密码算法，可采用软件外壳加密技术、USB 加密狗技术和数字签名技术，保证系统软件程序不被反跟踪调试、不被修改和不被静态分析；系统内建多组数据加密工作密钥表和多套算法参数，工作密钥表和算法参数在深度控制软件系统和 USBKey 之间分割存储，并受到本地存储加密和完整性保护；数据加密工作密钥、算法参数可依用户数据类型和密级确定，并由安全控制区域中的系统专用安全控制信息指示；用户所有写入、读取行为均被详细安全审计，审计信息安全存放，仅能授权删除，便于事件追踪。

4 结语

文中针对当前涉密光盘全生命周期安全保密管理需求，设计了一种深度控制系统方案，可实现对涉密光盘数据的机密性、完整性、真实性安全保护，也可对涉密光盘的写入、传递、读取、复制等方面进行策略控制、审计记录和监控管理，可广泛应用于具有安全保密需求的单位内或单位间基于光盘的涉密信息交换过程的管理控制，能有效防止光盘数据失泄密，降低内部人员恶意窃密的攻击风险。

参考文献

[1] 天极群乐 . 八种加密方法保护光盘数据不被盗窃 [J]. 网络与信息，2009(8)：69.

[2] 胡启明 . 简述光盘数据的八种 “ 保护神 ” [J]. 家电检修技术，2009(1)：34-35.

[3] Kew. 探寻光存储创新因子的秘密 (一)——光存储技术解析之光盘加密 [EB/OL]. (2009-07-08)[2012-07-20]. http：//news.mydrivers.com/1/139/139036.htm.

[4] 冯蓓 . 蓝光光盘——{ 驶入蓝色 }[J]. 通信技术，2004(2)：38-39.

[5] 高晓菲，张立 . 光盘刻录技术初探 [J]. 信息记录材料，2005，6(2)：31-35.

[6] 张艳丽，刘嘉勇 . 基于 ECC 数字签名系统的设计与实现 [J]. 信息安全与通信保密，2011(5)：48-49.

[7] 周英红 . 集中管控安全文件系统 [J]. 信息安全与通信保密，2011(12)：42-43.

(上接第 98 页)

4 结语

通过分析基于 Telnet 的隐蔽信道的特征，提出了 One-Class-SVM 的检测方法。通过剔除采集到的异常操作数据构建训练集，建立检测模型，并通过不断调整参数来获取最佳检测效果。试验中发现 One-Class-SVM 分类算法具有较好的分类效果，无论是否对异常数据进行剔除，其检测率都达到 100%，具有很好的检测效果。由于人对 Telnet 的操作习惯可能会随着时间变化，原有的模式将不能准确描述正常的操作行为，这可以通过定期训练或在线训练的方式来解。

参考文献

[1] 吴其祥，李祖猛，马华 . 基于 HTTP 协议的隐蔽信道研究 [J]. 信息安全与通信保密，2009(1)：73-74，77.

[2] 谷传征，王轶，骏薛质 . 基于 DNS 协议的隐蔽信道研究 [J]. 信息安全与通信保密，2011(12)：126-129.

[3] 张学工 . 关于统计学习理论与支持向量机 [J]. 自动化学报，2000(1)：32-42.

[4] JOHN Giffin J，GREENSTADT R. Covert Messaging Through TCP Timestamps[C]. San Francisco，CA：Proceedings of the Second International Privacy Enhancing Technologies Workshop，2002：194-208.

[5] SCHOLKOPF B，PLATT J C，SHAWE J T，et al. Estimating the Support of a High-dimensional Distribution[J]. Neural Computation，2001，13(7)：1443-1471.

[6] CHANG Chih-Chung，LIN Chih-Jen. A Library for Support Vector Machines[EB/OL]. (2005-08-06)[2012-03-26]. http://www.csie.ntu.edu.tw/~cjlin/libsvm/.