

针对空间协议的模板分析方法

李 华, 王 韬, 郑天明, 赵新杰

(军械工程学院 计算机工程系, 河北 石家庄 050003)

摘 要: 通过距离判别算法和模板分析思想的结合, 提出一种新的基于多级模板分析的空间传输协议识别算法. 根据空间协议特征构建一级模板, 利用一级模板间距离建立二级模板, 然后基于距离判断算法进行模板分析, 实现协议识别. 实验结果表明: 多级模板分析可提高空间协议识别效率, 该方法为协议识别技术研究提供新的思路.

关键词: 卫星网络协议; 模板分析; 协议识别; 多级模板分析

中图分类号: TP309

文献标识码: A

文章编号: 1000-7180(2013)01-0008-05

Template Analysis Method for Space Protocol

LI Hua, WANG Tao, ZHENG Tian-ming, ZHAO Xin-jie

(Department of Computer Engineering, Ordnance Engineering College, Shijiazhuang 050003, China)

Abstract: A new space transportation protocol identification algorithm based on multilevel template analysis is proposed by combining the distance judge algorithm and template analysis idea. The first level template can be built by space protocol characteristics, the second level template can be built by the distances between the first level templates. Then, the distance judge algorithm is used to make the template analysis and realize protocol identification. Experiment results demonstrate that multilevel template analysis improves the efficiency of space protocol identification and provides new ideas for the research of protocol identification techniques.

Key words: satellite network protocol; template analysis; protocol identification; multilevel template analysis

1 引言

随着卫星通信技术的发展, 卫星通信协议族在不断扩大、完善^[1], 出于安全性、适用性等方面的考虑, 研究卫星网络协议识别技术对于空间协议安全性分析具有重要的意义.

空间数据是根据卫星网络协议所产生的规则集合, 符合一定的规则和结构^[2], 为模板分析思想在协议识别研究中的应用奠定了理论基础. 目前, 已提出一些新型的空间传输协议. 但原有模板分析算法^[3]存在识别效率较低问题, 为此, 文中提出一种基于多级模板分析的协议识别算法, 通过分析协议结构构建协议特征的一级模板, 判断一级模板间各模板的距离, 在此基础上建立二级模板, 并基于距离判断算法实现协议的快速识别.

2 相关理论

2.1 空间协议结构

空间数据在传输过程中具有如下特点: 传输时延长、信道误码率高、非对称性、适合同地面终端进行协议转换.

空间通信协议体系结构^[4-6]自下而上包括: 物理层、数据链路层、网络层、运输层和应用层. 其中, 每一层又包括若干个可供组合的协议. 网络层协议识别是进行上层协议识别并获取截获数据真实信息的基础, 所以本文重点研究网络层协议识别. 空间通信传输协议标准 SCPS 结构如图 1 所示. 在空间传输协议识别过程中, 数据经常以比特流形式存在, 使得特征位难以提取, 分析难度增大.

收稿日期: 2012-05-30; 修回日期: 2012-06-28

基金项目: 国家自然科学基金(60772082)

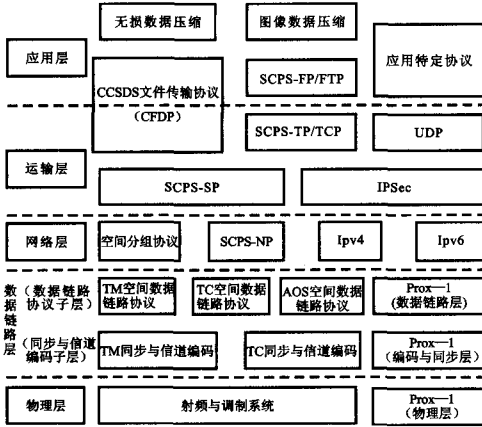


图 1 SCPS 协议结构

2.2 距离判别

距离判别主要通过定义样本指标 X 的观测值 $x(p$ 维)到各总体的距离,根据其大小判定样本属于那个总体.下面首先给出有关距离的概念.

2.2.1 统计距离

设 D 是某个 p 维空间的一个点集,在 D 中的任何两点有一个非负函数 d ,对任意 $x_i, x_j \in D$,如果 $d_{ij} = d(x_i, x_j)$,且满足:

- ① $d_{ij} \geq 0$, 当且仅当 $x_i = x_j$ 时, $d_{ij} = 0$,
- ② $d_{ij} = d_{ji}$,
- ③ $d_{ij} \leq d_{ik} + d_{kj}$,

则称 d_{ij} 为点 x_i 和 x_j 间的距离.例如,对任意 $x_i = (x_{i1}, \dots, x_{ip})'$, $x_j = (x_{j1}, \dots, x_{jp})' \in R_p$, 有

(1) Euclidean(欧几里德)距离:

$$d(x_i, x_j) = \left(\sum_{a=1}^p (x_{ia} - x_{ja})^2 \right)^{\frac{1}{2}}$$

(2) Minkovski(明考夫斯基)距离:

$$d_{ij}^{(m)} = \left(\sum_{a=1}^p |x_{ia} - x_{ja}|^m \right)^{\frac{1}{m}}, \text{ 其中 } m > 0.$$

Minkovski 距离是欧氏距离的推广,当 $m=2$ 时为欧氏距离,当 $m=1$ 时为绝对距离,当 $m=\infty$ 时称为 Chebyshev(切比雪夫)距离,此时有

$$d_{ij}^{(\infty)} = \max_{1 \leq a \leq p} |x_{ia} - x_{ja}|$$

(3) B 模距离. 设 B 为正定阵,定义

$$d_{ij} = [(x_i - x_j)' B (x_i - x_j)]^{\frac{1}{2}}$$

是 x_i 和 x_j 的 B 模距离,如果 x_i 和 x_j 是来自总体 (μ, Σ) ,其中 μ 为总体均值, Σ 为总体协方差,取 $B = \Sigma^{-1}$,则称

$$d_{ij}^2 = (x_i - x_j)' \Sigma^{-1} (x_i - x_j)$$

是 x_i 和 x_j 的 Mahalanobis(马哈拉诺比斯)距离,以下简称马氏距离,这种距离对平移和非异线性变换不变.

以上的各种距离度量在实际应用中以及计算复

杂性方面,效果各不相同.马氏距离考虑了样本的各输入变量分量相关性影响,而上述其它距离没有考虑各输入变量分量相关性影响.同时,在马氏距离在许多情况下,便于分析和计算,因此,常被用作判别分析的准则.

2.2.2 多总体的马氏距离判别

在应用马氏距离判别时,对于一个新样本 X ,要判断 X 来自哪个总体,应首先计算 X 到各总体的马氏距离,并判定距离 X 最短的总体.在对空间协议类型进行判别时,各协议总体样本的协方差矩阵不相等,下面仅以各协方差矩阵不相等的情况予以详细讨论.

设有 k 个总体 $G_1, G_2, \dots, G_k$, 对应均值和协方差矩阵分别是 $\mu_1, \mu_2, \dots, \mu_k$ 和 $\Sigma_1, \Sigma_2, \dots, \Sigma_k$, 它们均不相等.首先,计算样本 X 到每个总体的距离:

$$D^2(X, G_a) = (X - \mu_a)' \Sigma_a^{-1} (X - \mu_a)$$

$$a = 1, 2, \dots, k$$

则相应的判别规则为 $X \in G_i$. 如果

$$D^2(X, G_i) = \min_{1 \leq a \leq k} D^2(X, G_a)$$

且 $\mu_1, \mu_2, \dots, \mu_k$ 和 $\Sigma_1, \Sigma_2, \dots, \Sigma_k$ 均未知时,可以通过相应的样本值来替代.设 $X_1(a), \dots, X_n(a)$ 是来自总体 $G_a(a=1, 2, \dots, k)$ 中的样本,则 $\mu_a(a=1, 2, \dots, k)$ 和 Σ 可估计为:

$$\hat{\mu}_a = \bar{X}^{(a)} = \frac{1}{n_a} \sum_{i=1}^{n_a} X_i^{(a)},$$

$$\hat{\Sigma}_a = \frac{1}{n_a - 1} \sum_{i=1}^{n_a} (X_i^{(a)} - \bar{X}^{(a)}) (X_i^{(a)} - \bar{X}^{(a)})',$$

$$a = 1, 2, \dots, k$$

3 多级模板分析算法

空间数据根据协议产生,而协议一般具备一定的规则和结构.针对这一特性,提出了一种基于多级模板分析的协议识别算法,算法根据协议特征构建一级模板 $P(X_i)$,为提高识别效率,在通过判断各个一级模板间距离的基础上,建立二级模板 $P(X_i)'$,基于距离判别算法进行识别,通过二级模板的转换机制,可实现协议的快速识别.

常见的空间协议结构特征归纳为四类,不同类对应的数据在分析过程中要采用不同的处理方式.

(1) 确定的标识位:在分析中,某些标识位的位置和值是固定的,如果分析的数据均满足标识位特征,则称数据和标识位是完全匹配的.它可作为分类的前提,并为分类提供依据.

(2) 数据区间:是一个多位的集合,取值是一个

范围区间, 对其的判断应该采用范围控制方式, 如根据数据长度等。

(3) 离散值: 标识根据协议的可能结果列出的数据项, 如通过本层协议确定的上层协议格式等。

(4) 随机值: 此类数据常给分析带来较大困难, 在分析过程中, 如何跳过其所属区域, 使其在搜索过程中不对确定数据、数据区间、离散值产生影响, 对提高协议分析准确度十分重要。

依据上述数据类型, 构造的模板也应具有四种类型:

(1) 特征模板 stamp: 一组数据的模板单元。

(2) 上下限模板 boundary: 具有区间容错能力的数据区间。

(3) 点阵模板 concourse: 具有离散值的模板集合。

(4) 长度模板 length: 该模板可有效去除数据中不确定数据对数据分析的影响, 且具有占位能力, 其内容没有特殊规定。

通过协议特征构建模板 $P(X_i)$, 并存储 N 个模板串, 通过组合的方式形成原型样本串 $\text{sty}_i[P(X_j), P(X_k), \dots, P(X_l)]$, 待测试串 $\text{pending}[D(X_1), D(X_2), \dots, D(X_N)]$ 依次与存储的原型串 $\text{sty}_i[P(X_j), P(X_k), \dots, P(X_l)]$ 进行比较, 如果符合, 则判断该数据使用这一模板, 反之更换模板, 直到选择出和数据串一致的模板。该方式可以准确匹配出被测数据所使用的模板和协议。

改进一: 创建二级模板

由于协议种类繁多, i 值较大, 依次比较各模板 (最差情况循环执行比较 i 次) 必然会造成识别效率低下, 为了解决这一问题, 采用二级模板机制, 所谓二级模板就是对模板采用判断距离的方式, 分析各个串的距离差, 分析过程中对协议转换采用距离差, 通过一级模板间距离差或相似性, 按照最近邻的类别来标记。

$$\left. \begin{array}{l} \text{sty}_1[P(X_j), P(X_k), \dots, P(X_l)] \\ \text{sty}_2[P(X_j), P(X_k), \dots, P(X_l)] \\ \vdots \\ \text{sty}_n[P(X_j), P(X_k), \dots, P(X_l)] \end{array} \right\} \Rightarrow \text{sty}'[P(X_j), P(X_k), \dots, P(X_l)]$$

将各种协议所构成的统一模板 $\text{sty}'[P(X_1), P(X_2), \dots, P(X_l)]$ 作为数据协议识别的新模板。

改进二: 采用阶段记录

将各种模板通过步数变换, 形成统一的模板, 并将差距得分进行阶段记录, 采用阶段记录可以有效

根据差异位定位的方式, 记录距离产生位置, 为应用快速转换模板提供了保证。处理测试数据时, 将待测数据同二级模板的初始态 (初始为该数据包使用最频繁的协议组合 $\text{sty}_i[P(X_j), P(X_k), \dots, P(X_l)]$) 进行比较, 如果不匹配, 判断不匹配位置, 依据待分析数据特征, 通过最小步数变换方式, 将后续模板块中差距最小的模板组合作为当前的“活跃”模板, 使得模板分析具有快速模板转换功能。反复执行上述过程, 直到找到最匹配待分析数据的模板, 完成协议识别。表 1 给出多级模板分析中的模板转换算法。

表 1 模板转换算法

输入: 任一模板 P_x , 模板 P_i

输出: 模板。

1: $m = \text{length}[P_x]$, $n = \text{length}[P_i]$

2: $R = \min[m, n]$

3: $c[0, 0] \leftarrow 0$

4: $i \leftarrow 0$

5: do $i \leftarrow i + 1$

6: until $i = R$

7: $j \leftarrow 0$

8: do $j \leftarrow j + 1$

9: $c[0, j] \leftarrow j$

10: until $j = R$

11: $i \leftarrow 0$; $j \leftarrow 0$

12: do $i \leftarrow i + 1 \parallel j \leftarrow j + 1$

13: if $(P_x = \text{find}(P_i))$

14: $U(\text{Res}) = t$

15: else

16: goto 12

17: return $U(\text{Res})$

18: End

根据模板转换算法, 利用模板间关联关系, 可完成不同模板间转换, 快速选择出最合适的模板, 用于空间数据识别分析。模板更换过程如图 2 所示。

由图 2(a) 可知, 数据 D5, D6 与当前模板不匹配, 将模式 5、6 带入到模板集中采用最小转换机制, 转换为模板 7, 进而最终转换为模板 9, 通过图 2(d) 可知该数据采用的是模板 9, 从而分析出当前数据所使用的协议。

多级模板分析算法可以通过模板相似度关系,

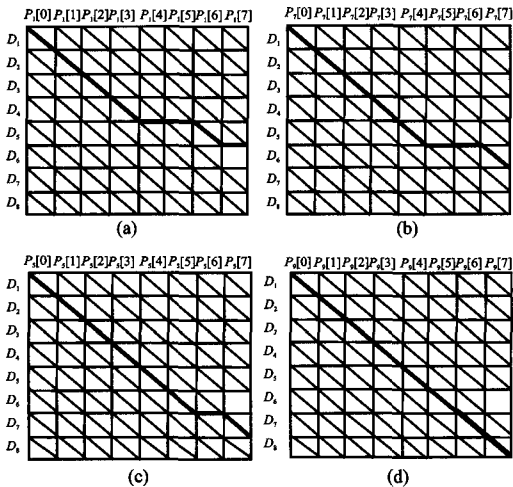


图 2 模板转换图

选取和待分析数据最接近的模板,有效解决空间数据模板选择问题.依据数据特征构建一级模板,并通过一级模板间的关系构建二级模板,在识别过程中通过最小转换机制,可达到最高的识别效率.多级模板分析算法执行流程如图 3 所示.

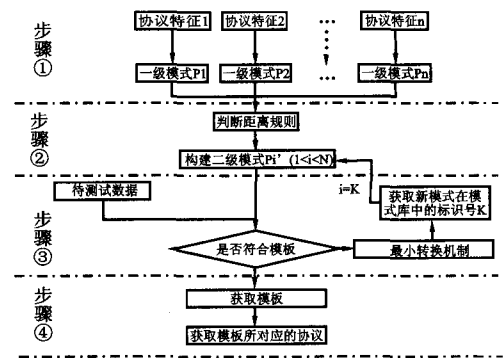


图 3 数据分析过程

以空间传输协议 SCPS 网络层中的 SCPS-NP 协议分析为例,对其构建一级模板,根据协议中是否包含扩展位,可分别搭建模板,如表 2 和 3 所示.

通过分析表 2 和表 3,两种协议的差异主要是在第四个属性.文中采用模板转换、插入或转换、删除的方式,可以快速的实现模板间的互换,提高识别效率.

表 2 无扩展位

序号	模板名称	模板类型	模板范围/取值	所占 Bit
1.	标示符	stamp	确定值“001”	3
2.	数据长度	boundary	>104	13
3.	传输层标示符	concourse	属于值域 ts	4
4.	控制	concourse	属于值域 con1	4
5.	头校验和	long	——	16

表 3 含扩展位

序号	模板名称	模板类型	模板范围/取值	所占 Bit
1.	标示符	stamp	确定值“001”	3
2.	数据长度	boundary	>104	13
3.	传输层标示符	concourse	属于值域 ts	4
4.	控制	concourse	属于值域 con2	12
5.	源地址	long	——	8
6.	基本质量服务	long	——	8
7.	头校验和	long	——	16

注:ts={0001,0100,0101,0110,1000,1010,1011}
con1={0000,0001,0011}
con2={101 * * 01 * * 011,1011101 * * 010},
* * 表示连续两位为 0、1 任意值

4 实验结果比较分析

4.1 实验结果

在 Windows XP 环境下用 Visual C++ 6.0 语言实现了本文测试程序与算法,算法针对空间协议 SCPS,根据几种当前常用的网络层、传输层协议生成数据,并将数据进行分类,结果如图 4 所示.

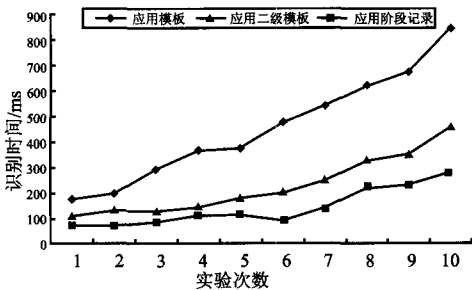


图 4 实验对比结果

4.2 实验结果分析

多级模板分析算法通过创建二级模板及采用阶段记录两种方法对模板分析算法进行了改进,通过图 4 中实验结果可以发现,随着实验中数据量的增大,传统的应用模板分析算法识别时间显著增加,而当采用了二级模板后识别时间明显减少,此后再使用了阶段记录的方法后,识别效率也有所提高.

通过这种改进算法,有效的屏蔽了各种应用模式串匹配算法改进时可能遇到的模式串长度不足等问题,通过建立模板的方法对协议的各种特征进行了规范,实验证明了这种算法有较高的识别准确率和识别效率.

5 结束语

本文提出的改进多级模板分析算法可以较好的

解决原有模板分析算法在模板数量较多时匹配效率较低这一问题. 通过模板间距离建立二级模板, 在匹配过程中同时应用一级模板和二级模板进行, 应用于协议识别领域, 提高了协议识别效率.

参考文献:

- [1] Ercetin O, Ball M O, Tassiulas L. Modeling study for evaluation of aeronautical broadband data requirements over satellite networks[J]. IEEE Transactions on Aerospace and Electronic Systems, 2005, 41(1):361-370.
- [2] Gupta B, Ghosh K, Dutta D, et al. Broadcasting in complete and incomplete star interconnection networks [J]. International Journal of Computer System Science and Engineering, 2001, 4(2):205-213.
- [3] Suresh C, Josyula R R, Pankaj R. Template attacks [C]// Proceedings of Canadian Healthcare Engineering Society, Redwood Shores. CA, USA, 2003:13-28.
- [4] CCSDS 130. 0-G-2. Overview of space communications protocols, report concerning space data system standards[S]. Green Book, Washington, DC, USA, 2007.

- [5] CCSDS 720. 1-G-3. CCSDS file delivery protocol (CFDP)—part 1: introduction and overview, report concerning space data system standards [S]. Green Book, Washington, DC, USA, 2007.
- [6] CCSDS 720. 2-G-3. CCSDS file delivery protocol (CFDP)—part 2: implementers guide, report concerning space data system standards [S]. Green Book, Washington, DC, USA, 2007.
- [7] CCSDS 720. 3-G-1. CCSDS file delivery protocol (CFDP)—part 3: interoperability testing final, report concerning space data system standards [S]. Green Book, Washington, DC, USA, 2007.

作者简介:

- 李 华 男, (1981—), 博士研究生. 研究方向为网络安全.
- 王 韬 男, (1964—), 教授, 博士生导师. 研究方向为网络安全.
- 郑天明 男, (1985—), 硕士. 研究方向为网络安全.
- 赵新杰 男, (1986—), 博士研究生. 研究方向为网络安全.

(上接第 7 页)

- [3] Becker V D. Epidemic routing for partially connected ad hoc networks[R]. Technique Report, CS-2000-06, Department of Computer Science, Duke University, Durham, NC, 2000.
- [4] Spyropoulos T, Psounis K, Raghavendra C S. Spray and wait: An efficient routing scheme for intermittently connected mobile networks[C]//Proc. of the 2005 ACM SIGCOMM Workshop on Delay-Tolerant Networking. Philadelphia: ACM, 2005:252-259.
- [5] Keränen A, Ott J, Kärkkäinen T. The ONE Simulator for DTN Protocol Evaluation[C]//Proc. of the 2nd Int'l Conf. on Simulation Tools and Techniques, Brussels, Belgium: ACM Press, 2009.
- [6] Chen L J, Chen Y C, Sun T, et al. Finding self-similarities in opportunistic people networks [C]//Proc. IEEE INFOCOM Mini-Symposium. Anchorage, Alaska, 2007.

- [7] Zhang X, Kurose J, Levine BN, et al. Study of a bus-based disruption tolerant network: mobility modeling and impact on routing [C]//Proc. ACM Mobicom Canda; Quebec, 2007:195-206.

作者简介:



钱景辉 男, (1978—), 讲师, 硕士. 研究方向为传感器网络与移动自组网.