Journal of Computer Applications
计算机应用,2013,33(12):3486-3489,3498
ISSN 1001-9081
CODEN JYIIDU
2013-12-01
http://www.joca.cn

文章编号:1001-9081(2013)12-3486-04
doi:10.11772/j.issn.1001-9081.2013.12.3486

基于状态标注的协议状态机逆向方法

黄笑言*,陈性元,祝　宁,唐慧林
(信息工程大学,郑州 450004)
(*通信作者电子邮箱 eileen0908@qq.com)

摘　要:协议状态机可以描述一个协议的行为,帮助理解协议的行为逻辑。面向文本类协议,首先利用统计学方法提取表示报文类型的语义关键字;然后利用邻接矩阵描述报文类型之间的时序关系,基于时序关系进行协议状态标注,构建出协议的状态转换图。实验表明,该方法可以正确地描述出报文类型的时序关系,抽象出准确的状态机模型。

关键词:协议逆向;协议语义;协议会话;协议状态机;邻接矩阵
中图分类号:TP393　　**文献标志码**:A

Protocol state machine reverse method based on labeling state

HUANG Xiaoyan*, CHEN Xingyuan, ZHU Ning, TANG Huilin
(*Information Engineering University, Zhengzhou Henan* 450004, *China*)

Abstract: Protocol state machine can describe the behavior of a protocol, which can help to understand the behavior logic of protocol. Oriented towards text protocols, a statistical method was firstly used to extract the semantic keyword of representative message type, and an adjacency matrix was used to describe the sequential relationship between the message types, based on which the protocol states were labeled and a state transition diagram was built. The experimental results show that the method can accurately describe the sequential relationship between the message types and abstract state machine model accurately.

Key words: protocol reverse; protocol semantic; protocol session; protocol state machine; adjacency matrix

0　引言

在网络安全中,很多基于协议的安全技术都以协议规范说明为基础。比如为了提高防御粒度,新一代的防火墙[1]和入侵检测系统[2]利用协议规范进行深度包检测和状态行为检测,从而能高效、精确地识别出恶意传输;高交互型蜜罐系统[3]基于协议规范可以生成各种脚本以监听各种远程请求,实现多种服务的仿真。

但是很多网络私有协议没有公开自己的规范说明,比如MicroSoft的网络文件共享SMB(Server Message Block)协议[4]、Oracle数据库访问的TNS协议[5]、各种IPTV和及时通信软件使用的协议[6]等。另外,即使对于公开规范的协议,不同厂商在软件的具体实现时并没有严格按照规范说明去设计,因此越来越多的研究人员通过协议逆向的方法自动获取协议规范说明,以支撑其他网络安全技术的实施。

协议规范定义了协议报文的格式和协议状态机[7]:前者规范了协议报文由哪些字段组成,每个字段的位置、类型和取值含义等[8];后者规定了协议报文的时序关系,体现出协议的行为逻辑。目前大多研究集中于反向推断协议的格式,较少研究协议状态机的逆向。事实上,逆向出协议状态机可以描述一个协议的行为,帮助理解协议的行为逻辑,进一步应用到入侵检测或蜜罐系统中,因此本文对协议状态机逆向方法进行研究。

1　相关研究

目前关于协议状态机逆向的研究分为两类[9]。一类是指令级的分析方法。这种方法需要在指令级监控协议实体对报文的解析过程,实现起来比较复杂,在实际应用中很难获得对协议实体的控制权,加之很多软件为了防止其代码被逆向,加强了软件的模糊性。另一类网流级的分析方法是以嗅探得到的网络数据流为分析对象,它的可行性在于同一报文格式对应的多个报文样本具有相似性,会话内报文的时序关系体现了协议状态转换的信息。由于实现起来简单,因此近年来很多研究者开始研究基于网流级的状态机逆向方法。

2007年,Shevertalov等[10]提出协议状态机逆向的工具PEXT,将协议的运行过程划分为多个阶段(子会话),每个子会话完成不同的功能,被定义成一个状态。PEXT以最长公共子序列长度为相似度指标,对报文样本进行聚类,将协议流转化成一系列的聚类ID,在协议流之间提取相同的聚类ID序列标注成一个协议的状态,根据状态转换序列生成状态机,通过合并算法得到涵盖所有协议会话实例的最小确定状态机。

2009年,Trifilo等[11]将会话中的每条报文(包括不同的方向)定义为一个状态,认为协议报文中通常存在一些报文状态域标志了当前的状态逻辑,通过分析二进制协议报文中各字节的变化分布来识别状态域并构建状态机;并考虑通过

收稿日期:2013-06-13;**修回日期**:2013-08-19。
基金项目:国家973计划项目(2011CB311801);河南省科技创新人才计划项目(114200510001)。
作者简介:黄笑言(1989-),女,福建福州人,硕士研究生,主要研究方向:信息安全、协议逆向;　陈性元(1963-),男,安徽无为人,教授,博士,主要研究方向:信息安全、分布式操作系统;　祝宁(1981-),男,辽宁抚顺人,讲师,博士,主要研究方向:网络对抗;　唐慧林(1980-),男,安徽枞阳人,讲师,硕士,主要研究方向:信息安全。

检测状态的前一状态和随后的状态来区分由同一特征值表示的不等状态,避免构建出的状态机产生错误的报文序列。

2011 年,Wang 等[12] 提出了协议的概率状态机(Probabilistic Protocol State Machine,PPSM)模型构建单方向网流的状态机。PPSM 首先利用统计学的方法找到网流中最频繁的字符串;而后利用围绕中心点的划分(Partitioning Around Medoids,PAM)聚类方法获取协议的状态关键字,根据关键字为每一个数据包分配状态类型;最后以概率的形式描述状态之间的转换,构建概率协议状态机。

2009 年,Comparetti 等[13] 综合利用网流级和指令级的信息提出了完整的协议逆向方案 Prospex,为客户端的输出报文逆向状态机,旨在识别表示相似应用情景的状态。首先利用指令执行序列分析技术,抽取每个报文的格式,继而结合格式特征和执行特征对报文进行聚类,抽取更普遍的报文格式,识别会话中的每个报文类型;然后构建增广前缀树(Augmented Prefix Tree Acceptor,APTA)接受网络会话中的所有报文类型序列,继而从观察到的会话中抽取报文类型之间的顺序特征,以正则表达式表示,称为先决条件;接着对 APTA 的每个状态用那个状态允许输入的报文类型集合进行标注,表示其符合先决条件的可接受的报文类型集;最后使用 Exbar 算法将 APTA 最小化。

以上依据网流级状态机逆向的方法有以下不足:Trifilo 等[11] 的方法依赖于各种报文类型字段在报文格式的同字节偏移位置上出现,不适于报文类型字段的字节位置不固定的文本协议;Wang 等[12] 的方法适于文本协议,但是只依据频繁字符串标识报文状态,没有区分字符串之间的层次,无法准确提取出报文的语义字段。

另外,目前构建状态机的方法都是先构建一棵状态前缀树,再利用启发式方法进行状态机的合并和简化,这会导致初始构建的状态机过于庞大,并且在状态机简化过程中容易导致路径的错误合并,无法准确地描述报文类型之间的时序关系。

为解决以上问题,本文依据网络流的分析提出一种面向文本类协议的状态机逆向方法。首先利用语义关键字的分布特征和偏移属性提取语义关键字,识别出会话中的报文类型;然后利用有向图的邻接矩阵描述报文类型之间的时序关系,进行状态标注;最后实现协议状态机的逆向,体现协议行为逻辑。其流程如图 1 所示。

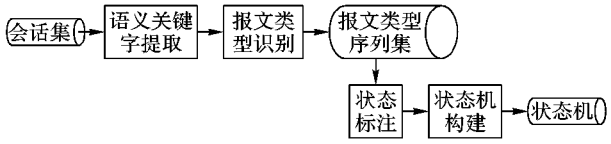

图 1　状态机重构流程

2　状态机逆向方法

2.1　语义关键字提取方法

定义 1　语义关键字。报文中表示报文类型的字段称为语义关键字。

为了识别文本语义关键字,第一步使用处理文本协议的通用方法,将连续可打印 ASCII 码(不少于 3 字节)标记成一个文本块,根据分隔符将文本块划分为一系列的文本 token;然后利用启发式规则过滤掉特征明显的参数,包括 IP 地址、版本号、URL 等[14],过滤后的文本 token 作为候选语义关键字,并记录它在原文本 token 序列中的位置。第二步通过考察关键字在会话集中的分布特征,来识别常用语义关键字。第三步发现关键字在报文中的偏移特征,一方面作为第二步语义关键字识别规则的补充,另一方面提高报文类型的识别速度。

2.1.1　语义关键字的会话分布特征

语义关键字的分布特征指的是语义关键字通常频繁地出现在会话中,而不是出现在会话的每个报文中。本文综合关键字在会话中的频数和报文的频数来刻画语义关键字的分布特征,定义两个识别规则,同时满足两种识别规则的文本 token 即为语义关键字。

给定协议的会话集 S 和报文集 M,$|S|$ 表示会话数,$|M|$ 表示报文数。对于语义关键字 t,$|M_t|$ 表示包含 t 的报文数,即 t 在报文中的频数;$|S_t|$ 表示包含 t 的会话数,即 t 在会话中的频数。

规则 1　定义 δ_t 度量语义关键字 t 在报文中的频率:$\delta_t = |M_t| / |M|$,要求 $0 < \delta_t < 1$ 且与 0 和 1 有显著差异。

当 $\delta_t \geqslant 1$ 或 $\delta_t \approx 0$ 时均不符合语义关键字的分布特征:

$\delta_t \geqslant 1$ 即 $|M_t| \geqslant |M|$ 时,t 为协议标识级的关键字;

$\delta_t \approx 0$ 即 $|M_t| \ll |M|$ 时,t 为变量 token,而非协议的关键字。

规则 2　定义 φ_t 度量语义关键字 t 在会话中的分布与报文中的分布的一致性:$\varphi_t = |M_t| / |S_t|$,要求 φ_t 与 1 没有显著差异。

当 $\varphi_t = 1$ 时表示 t 在会话中出现的次数与报文中出现的次数一致,符合语义关键字的分布特征。

2.1.2　语义关键字的报文偏移特征

上述识别规则容易忽略很少出现在网流中的语义关键字,由于文本协议有一定的格式规范,语义关键字往往出现在报文的固定位置上,因此本文进一步发现语义关键字的偏移属性,并基于偏移属性对 2.1.1 节的识别规则进行修正。

定义 2　$pos\_permanent(t)$ 表示 t 的固定偏移率。令 $PF_m(t)$ 表示 t 在报文第 m 个位置出现的频率,$PF_m(t) = N_m(t) / |M_t|$。其中 $N_m(t)$ 表示 t 在每个报文文本 token 序列中第 m 个位置出现的次数,则 $pos\_permanent(t) = \max(PF_m(t))$。

规则 3　若 $pos\_permanent(t) = \max(PF_m(t)) \geqslant \vartheta$,则 t 具有固定偏移属性,m 是其固定偏移,此时 m 偏移处的关键字都是语义关键字。

本文通过发现语义关键字的报文偏移特征,一方面对 2.1.1节的识别规则进行修正,补充进出现频率较少的语义关键字;另外,基于语义关键字的偏移特征可以进行语义关键字的迅速定位,实现报文类型的快速识别,将协议的每个会话转化成报文类型序列。

2.2　状态标注

定义 3　协议状态是协议的一个逻辑概念,特定状态下协议实体可以接受特定事件和执行相应动作。

由于服务器发送的报文多是一些表示服务器服务能力的命令码,无法反映协议的行为逻辑,因此本文同先前的工作一样,从客户端发送的报文类型序列中构建状态机。

不同的报文类型可能会引起状态的转换,因此通常利用状态转换图表现报文类型之间的时序关系,描述协议的行为逻辑。先前的文献首先构建状态前缀树,接受所有的报文类型序列再进行状态的合并和化简,导致初始构建的状态前缀树过于庞大,且需要大量的比较操作。而本文利用邻接矩阵描述报文类型之间的关系,基于报文类型之间的关系对状态进行标注,可以省去状态前缀树的构造,直接构建状态转

换图。

主要思路是首先将具有强顺序约束的报文类型序列标注为一个状态,然后找到会话的必经报文类型序列,每个必经报文类型序列标注为一个状态,最后基于报文类型之间的弱顺序约束对两两必经报文类型中间的可选报文类型集进行状态标注。

2.2.1 构造邻接矩阵

本文利用邻接矩阵表示各报文类型之间的邻接顺序关系。假设会话集共有 n 个报文类型,一个 $n \times n$ 的矩阵 $\boldsymbol{V}$ 表示各报文类型的相邻关系,矩阵的元素 v_{ij} 取值为:

1)1,当报文类型 a_i 和报文类型 a_j 存在先后邻接关系;

2)0,当报文类型 a_i 和报文类型 a_j 不存在先后邻接关系。

若一个会话中出现重复的报文类型,表示所处的状态相同,则当作新的会话重新开始对报文类型的邻接关系进行学习。

2.2.2 基于强顺序约束关系的状态标注

定义 4 若报文类型 A 与报文类型 B 具有强顺序约束关系,则 A 与 B 有严格的邻接顺序(中间不能有其他报文类型)且总是绑定出现。

定义 5 若 A 与 B 有强顺序约束关系,B 与 C 有强顺序约束关系,则报文类型序列 A,B,C 具有强顺序约束关系。

本文将具有强顺序约束关系的报文类型序列定义成合并一个状态转换,对邻接矩阵进行合并。

规则 4 若元素 v_{ij} 不为 0,且它是它所在的行和列中唯一不为空的元素,且 v_{ji} 为 0(说明 a_i 和 a_j 总是绑定出现的),则 a_i 和 a_j 之间具有强顺序约束关系,合并成一个状态转换。具体方法为删除 a_i 所在的行和 a_j 所在的列,用 $\langle a_i,a_j \rangle$ 替换行元素的 a_j 和列元素中的 a_i。

2.2.3 基于会话必经路径的状态标注

定义 6 会话必经路径是指所有会话必须出现的报文类型序列,这些报文类型之间有严格的顺序关系。

本文首先引入形式概念的表示方法表示报文类型与会话的隶属关系,然后利用这种隶属关系和有向图的最短路径设计会话必经路径搜索算法,找到会话开始和结束之间的必经路径,将必经路径中的每个报文类型标识成一个状态转换。

1)报文类型与会话的隶属关系。

$K = (G,A,I)$,其中:G 为所有会话的集合,A 为所有报文类型的集合,$I \subseteq G \times A$ 为 G 和 A 中元素之间的关系集合。对于 $g \in G, a \in A, (g,a) \in I$ 表示会话 g 包含报文类型 a。

2) 矩阵的幂。

$\boldsymbol{V}^r$ 表示 r 个矩阵 $\boldsymbol{V}$ 相乘,记 $\boldsymbol{V}^r = (v_{ij}^{(r)})$。$v_{ij}^{(r)}$ 表示节点 a_i 到节点 a_j 长度等于 r 的通路数。

3)会话必经路径搜索算法。

```
nec_load (a1,am,S,r)          // a1 和 am 表示两个报文类型,
                              // S 表示会话的集合,r 表示 a1 到 am 通路的长度
利用 Strassen 矩阵相乘算法计算 V^r
if (v1m^r = 0):r ++; nec_load (a1,am,S,r)
else
  load = shortest_load (a1,am,r)    // 搜索 a1 到 am 的最短路径
  M = {g ∈ G | (g,a) ∈ I, ∀a ∈ load}
                              //M 表示最短路径中每个节点的会话集的交集
if (M = S)                    //若为全集(即每个会话都有)
  return nec_load
else
  S = S - M; r ++; nec_load(a1,am,S,r);
shortest_load (a1,ay,r):      // 搜索 a1 到 ay 的最短路径
H = {ai | v1i^(r-1) * viy ≠ 0}    //ai 为 a1 到 ay 必经的节点
For each h in H
  Add h to load [n]           //add h to a1 与 ay 中间
  r --;
if (r > 1) shortest_load (a1,ay,h,r)
else return load              //v1i^n * vim = 0,n < r - 1
```

4)会话开始到结束的必经路径。

令 a_1 为会话中第 1 个报文类型,表示会话的开始;a_m 为会话中最后一个报文类型,表示会话的结束。本文假设会话的开始报文类型和结束报文类型分别只有一个,若为多个则分开讨论,分别构建状态机。

```
S = U,U 表示所有的会话集合
r = 1
nec_load (a1,am,S,r)          //搜索 a1 到 am 的必经路径
```

2.2.4 对可选出现的报文类型集进行状态标注

若 a_t 与 a_g 是必经路径上的两个报文类型,其先后顺序关系是 $a_t \rightarrow a_g$;T 为 a_t 与 a_g 之间可选出现的报文类型集。本文在 T 中查找报文类型集之间的强顺序约束关系,每个报文类型集所包含的报文类型导致一个状态转换,导致相同状态转换的报文类型具有等价关系。

定义 7 若报文类型 a 与报文类型 b 具有弱顺序约束关系,则 a 与 b 有严格的邻接顺序(中间不能有其他报文类型),但不是绑定出现。

定义 8 若两个报文类型集 A 与 B 之间具有强顺序约束关系,则对于 $\forall a \in A, \forall b \in B$,$a$ 与 b 之间具有弱顺序约束关系。

定义 9 若报文类型集 A 与 B 之间具有强顺序约束关系,报文类型集 B 与 C 之间具有强顺序约束关系,则报文类型集 A、B、C 之间具有强顺序约束关系。

本文首先利用邻接矩阵查找可选报文类型之间的弱顺序关系,然后进行合并和连接,得到报文类型集之间的强顺序约束关系。

规则 5 设邻接矩阵 $\boldsymbol{V}$ 是经简化后的邻接矩阵,即报文类型之间不存在强顺序约束关系,若邻接矩阵元素 v_{ij} 不为 0,v_{ji} 为 0,则 a_i 与 a_j 具有弱顺序关系。

比如报文类型 a、b、c、d 之间的弱顺序关系为 $a \rightarrow c$,$a \rightarrow b$,$b \rightarrow d$,$c \rightarrow d$,通过合并和连接后,得到报文类型集之间的强顺序关系,即一个报文类型集序列 $a \rightarrow (b,c) \rightarrow d$。

2.3 状态转换图的构建

利用 2.2 节的状态标注方法对报文类型引起的状态转换进行状态标注,构造协议的状态转换图。首先将具有强顺序约束关系的报文类型序列定义为一个状态转换;然后依据 2.2.3节得到会话必经报文类型画出基本的状态转换图;最后依据 2.2.4 节方法将具有等价关系的报文类型定义成一个状态转换,构造出完成的状态转换图。

3 实验验证分析

本文在 Windows XP 环境下利用 Python2.7 编写代码实现状态机逆向的方法,为了对本文的方法进行验证,选取广泛应用于网络中的两种文本类协议 SMTP 和 FTP 进行状态机逆向。通过在 Windows XP 下利用 WebMail 搭建 SMTP 服务器,利用 Serv-U 搭建 FTP 服务器,获取训练数据,然后将某校园网的真实网络流量作为测试数据。

由于基于网络轨迹的协议逆向一个普遍存在的缺陷是无法学习到训练集中未出现的行为,因此本文要求训练集覆盖每个协议的主要功能,使得逆向出的状态机能够反映协议的主要行为逻辑。

3.1　关键字提取

表 1 为利用 50 个 SMTP 会话(432 条报文)提取的 SMTP 协议关键字;表 2 为利用 100 个 FTP 会话(2500 条报文)提取的 FTP 协议关键字。

表 1　SMTP 语义关键字

类型序号	语义关键字	类型序号	语义关键字
1	EHLO	6	RESET
2	HELO	7	EMPTY CONTENT
3	MAIL FROM	8	CONTENT
4	RCPT TO	9	QUIT
5	DATA		

表 2　FTP 语义关键字

类型序号	语义关键字	类型序号	语义关键字
1	USER	8	PORT
2	PASS	9	LIST
3	QUIT	10	XMKD
4	CMD	11	TYPE
5	CDUP	12	STOR
6	RNFR	13	RETR
7	RNTO	14	DELE

3.2　状态机逆向

本文利用在入侵检测和信息检索中广泛使用的两个评估指标——召回率和准确率对逆向的状态机的质量进行评估。

3.2.1　召回率

状态机的召回率用来衡量状态机对测试集的覆盖程度,表示状态机可以接受多少个协议会话。令 N 表示规范测试集会话 I 总数,TP 表示 I 被状态机接受的会话数,则 $Recall_{FSM} = TP/N$。

为了测试召回率,本文利用真实网络流量,对状态机的召回率进行评估。真实网络流量的 SMTP 会话数为 855,FTP 会话数为 642。

图 2 为利用不同大小的 SMTP 协议训练集训练出的状态机的召回率。实验结果表明训练集的 SMTP 报文个数达到 300 时,召回率稳定在 94.1%,即逆向出的 SMTP 状态机可以接受 803 个 FTP 会话,余下的 52 个会话使用了 SSL 加密传输,没有被状态机识别出来。图 3 为 SMTP 训练集的报文个数达到 200 时逆向出的状态机,由于 SMTP 第一条报文有两种报文类型,HELO 和 EHLO,图 3 显示的是第一种报文类型开头的状态机。

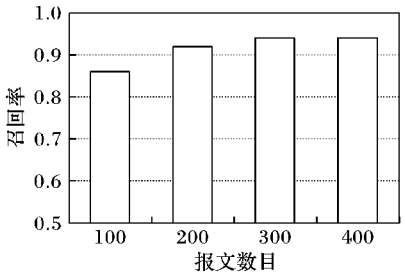

图 2　SMTP 协议状态机召回率

图 4 为利用不同大小的 FTP 协议训练集训练出的状态机的召回率。实验结果表明训练集的 FTP 报文个数达到 2 000 条时,召回率稳定在 91.7%,即逆向出的 FTP 状态机可以接受 589 个 FTP 会话,在余下的 53 个会话中,有 12 个会话出现了训练集中未出现的命令,另外 41 个会话使用了 SSL 加密传输。图 5 为 FTP 训练集的报文个数达到 1 500 条时逆向出的 FTP 状态机。

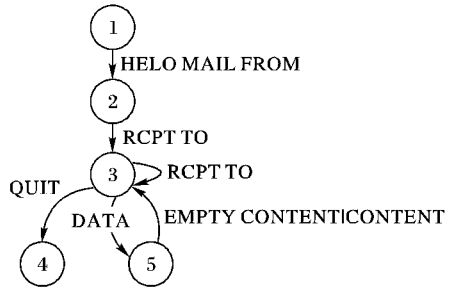

图 3　SMTP 协议状态机

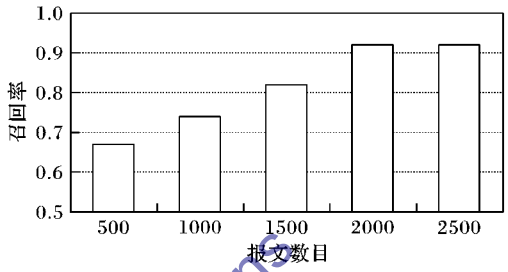

图 4　FTP 协议状态机召回率

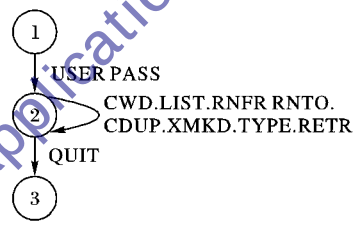

图 5　FTP 协议状态机

3.2.2　准确率

状态机的准确率用来衡量状态机的可靠性,表示被状态机接受的会话中,有多少个会话是符合协议规范的,令 M 表示被状态机接受的会话数,CP 表示符合协议规范的会话数,则 $Precision_{FSM} = CP/M$。

为了测试准确率,本文对测试集中的会话以 0.1 的概率进行随机修改,创建不符合协议规范的会话,包括关键报文丢失和报文之间的乱序。通过这种方式,在已经被 SMTP 协议状态机接受的 786 个 SMTP 协议会话中,创建无效会话 78 个。在已经被 FTP 协议状态机接受的 489 个 FTP 协议会话中,创建无效会话 64 个。

对于修改后的会话集,SMTP 状态机接受 SMTP 会话为 708 个,未被接受的会话均是无效会话;而 FTP 状态机接受 FTP 会话为 473 个。经检查发现,由于程序对 FTP 会话进行随机修改,而 FTP 规范中的很多报文类型是不需要严格顺序的,随机修改后的会话很多仍然是符合协议规范的,经人工过滤掉这些有效会话后,其状态机的准确率达到了 100%。

4　结语

现有逆向协议状态机的方法需要根据协议会话中的报文类型顺序构建初始状态前缀树,这会出现大量的冗余状态。本文利用邻接矩阵描述报文类型之间的时序关系,基于时序关系进行协议状态的标注,构建出协议的状态转换图,并对文本类协议进行了实验验证。结果表明,该方法可以正确地描述出报文类型的时序关系,抽象出准确的协议状态机模型。下一步,本文准备对二进制类的协议进行逆向实验,以验证本文方法的有效性。

(下转第 3498 页)

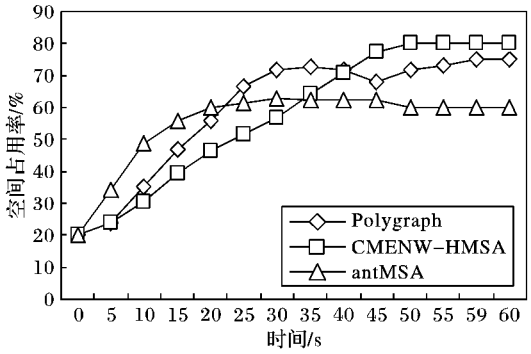

图 3 Code-RedII 特征提取空间占用比较

实验结果证明了本文方法可以提高特征提取的时空效率,并且可以得到更加精确的泛化程度更高的特征序列,误报率更低。

6 结语

本文借鉴蚁群算法的觅食特性,将改进后的蚁群算法应用于多态蠕虫的特征提取当中,提出了一种基于蚁群算法进行多序列比对的多态蠕虫的特征提取方法 antMSA。antMSA 方法可以有效克服蚁群算法的停滞现象,扩大搜索空间,大大减少了多态蠕虫特征提取的时空开销,同时能够得到精确度更高的特征片段,降低误报率。现有的特征提取方法仍依赖于纯净的数据源,因此如何提高方法的抗噪声能力是未来需要进一步解决的问题。

参考文献:

[1] KIM H-A, KARP B. Autograph : toward automated , distributed worm signature detection [C]// SSYM'04: Proceedings of the 13th Conference on USENIX Security Symposium. Berkeley: USENIX Association, 2004, 13: 271 – 286.

[2] NEWSOME J, SONG D. Dynamic taint analysis for automatic detection, analysis, and signature generation of exploits on commodity software [C]// NDSS 2005: Proceedings of the 12th Annual Network and Distributed System Security Symposium. San Diego: BibSonomy, 2005: 1 – 38.

[3] NEWSOME J, KARP B, SONG D. Polygraph: automatically generating signatures for polymorphic worms [C]// SP '05: Proceedings of the 2005 IEEE Symposium on Security and Privacy. Washington, DC: IEEE Computer Society, 2005: 226 – 241.

[4] 唐勇,卢锡城,胡华平,等.基于多序列联配的攻击特征自动提取技术研究[J].计算机学报,2006,29(9):1533 – 1541.

[5] TANG Y, XIAO B, LU X C. Using a bioinformatics approach to generate accurate exploit-based signatures for polymorphic worms [J]. Computers & Security, 2009, 28(8): 827 – 842.

[6] TANG Y, CHEN S. Defending against Internet worms: a signature-based approach [C]// Proceedings of the INFOCOM 2005. Piscataway: IEEE, 2005: 1384 – 1394.

[7] CRANDALL J R, WU S F, CHONG F T. Experiences using minos as a tool for capturing and analyzing novel worms for unknown vulnerabilities [C]// DIMVA'05: Proceedings of the Second International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment, LNCS 3548. Berlin: Springer-Verlag, 2005: 32 – 50.

[8] NEEDLEMAN S B, WUNSCH C D. A general method applicable to the search for similarities in the amino acid sequence of two proteins [J]. Journal of Molecular Biology, 1970, 48(3): 443 – 453.

[9] DORIGO M, MANIEZZO V, COLOMI A. Ant system optimization by a colony of coorperating Agents[J]. IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics, 1996, 26(1): 29 – 41.

[10] 梁栋,霍红卫.自适应蚁群算法在序列比对中的应用[J].计算机仿真,2005,22(1):100 – 106.

[11] 郑松,候迪波,周泽魁.动态调整选择策略的改进蚁群算法[J].控制与决策,2008,23(2):225 – 228.

[12] 卢家兴,郭帆,余敏.静态检测多态溢出攻击代码的方法[J].计算机应用,2010,30(12):3349 – 3353.

[13] OLUSOLA A A, OLADELE A S, ABOSEDE D O. Analysis of KDD'99 intrusion detection dataset for selection of relev-vance features [C]// WCECS 2010: Proceedings of the World Congress on Engineering and Computer Science. San Francisco: [s. n.], 2010.

[14] UCI knowledge discovery in databases archive [EB/OL]. [2013 – 03 – 20]. http://kdd.ics.uci.edu/databases/kddcup99/kddcup.data.gz.

(上接第 3489 页)

参考文献:

[1] KRUEGER T, KRÄMER N, RIECK K. ASAP: automatic semantics-aware analysis of network payloads [C]// Proceedings of the 2011 International ECML/PKDD Conference on Privacy and Security Issues in Data Mining and Machine Learning, LNCS 6549. Berlin: Springer-Verlag, 2011: 50 – 63.

[2] 郝耀辉,郭渊博,刘伟.基于有限自动机的密码协议入侵检测方法[J].计算机应用研究,2008,25(1):230 – 234.

[3] KRUEGER T, GASCON H, KRÄMER N, *et al.* Learning stateful models for network honeypots [C]// AISec '12: Proceedings of the 5th ACM Workshop on Security and Artificial Intelligence. New York: ACM, 2012: 37 – 48.

[4] How samba was written [EB/OL]. [2013 – 01 – 16]. http://samba.org/ftp/tridge/misc/french_cafe.txt.

[5] 李伟明,张爱芳,刘建财,等.网络协议的自动化模糊测试漏洞挖掘方法[J].计算机学报,2011,34(2):242 – 255.

[6] ANTUNES J, NENVES N, VERSSIMO P. Reverse engineering of protocols from network traces [C]// Proceedings of the 18th Working Conference on Reverse Engineering. Piscataway: IEEE, 2011: 169 – 178.

[7] 田园,李建斌,张振.一种逆向分析协议状态机模型的有效方法[J].计算机工程与应用,2011,47(19):63 – 67.

[8] 黎敏,余顺争.抗噪的未知应用层协议报文格式最佳分段方法[J].软件学报,2013,24(3):604 – 617.

[9] 潘璠,吴礼发,杜有翔.协议逆向工程研究进展[J].计算机应用研究,2011,28(8):2801 – 2806.

[10] SHEVERTALOV M, MANCORIDIS S. A reverse engineering tool for extracting protocols of networked applications [C]// WCRE 2007: Proceedings of the 14th Working Conference on Reverse Engineering. Piscataway: IEEE, 2007: 229 – 238.

[11] TRIFILIO A, BURSCHKA S, BIERSACK E. Traffic to protocol reverse engineering [C]// CISDA 2009: Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications. Piscataway: IEEE, 2009: 1 – 8.

[12] WANG Y P, ZHANG Z B, YAO D F, *et al.* Inferring protocol state machine from network traces: a probabilistic approach [C]// ACNS '11: Proceedings of the 2011 Applied Cryptography and Network Security, LNCS 6715. Berlin: Springer-Verlag, 2011: 1 – 18.

[13] COMPARETTI P M, WONDRACEK G, KRUEGEL C, *et al.* Prospex: protocol specification extraction[C]// Proceedings of the 30th IEEE Symposium on Security and Privacy. Piscataway: IEEE, 2009: 110 – 125.

[14] 赵咏,姚秋林,张志斌,等.TPCAD:一种文本类多协议特征自动发现方法[J].通信学报,2009,32(S1):28 – 35.