基于在线聚类和检测成本的移动自组网异常检测

王雷春 马传香
(湖北大学数学与计算机科学学院 武汉 430062)

摘 要 移动自组网具有无线信道、动态拓扑、缺乏基础设施和节点资源受限等特点,更易受到安全威胁,且无法部署复杂的安全协议和算法。为了有效检测移动自组网中的异常访问行为,提出了一种基于在线聚类和检测成本的异常检测方案TCDC。TCDC先在单个节点内对访问行为进行在线聚类和处理,然后在不同节点间通过基于检测成本的协同检测进一步确认访问行为。仿真实验表明,该异常检测方案能够有效地检测移动自组网中的异常行为,且消耗资源较少。

关键词 移动自组网,在线聚类,检测成本,异常检测
中图法分类号 TP391 **文献标识码** A

Online Clustering and Detective Cost Based Anomaly Detection Scheme for MANET

WANG Lei-chun MA Chuan-xiang
(School of Mathematics & Computer Science, Hubei University, Wuhan 430062, China)

Abstract Mobile Ad hoc networks(MANET) are highly vulnerable to be attacked and difficult to deploy complicated safe protocols and algorithms due to the open medium, dynamically changing network topology, lack of centralized monitoring and management point, and limited resources. To detect efficiently anomaly behaviors in MANET, this paper proposed a online clustering and detective cost based anomaly detection scheme for MANET, TCDC. In this scheme, TCDC firstly analyzes and deals with access behaviors in single node using online clustering based on access behaviors, and then validates farther access behaviors by cooperative detection based on detective cost among different nodes. Simulation results show TCDC can efficiently detect anomaly behaviors in MANET with less resource consumption.

Keywords Mobile ad hoc networks, Online clustering, Detective cost, Anomaly detection

1 引言

MANET具有无线信道、动态拓扑、没有基础设施和分布式协作等特点,与传统的有线和无线网络相比存在更多的安全问题。MANET可以采用加密和认证等入侵措施来减少入侵,但并不能消除入侵,因此需要研究保护网络系统的第二道防线——入侵检测。另一方面,MANET中节点本身的资源非常有限,这包括1)网络使用移动信道,其本身的物理特性决定了相对有线信道低得多的网络带宽;2)网络中的用户终端通常采用电池供电,且内存较小,CPU性能较低,这给应用程序的设计开发带来一定难度。

针对MANET的安全威胁和资源严重受限等特点,本文提出了一种基于在线聚类和检测成本的异常检测方案。通过在单个节点和节点间分别采用基于访问行为在线聚类的本地检测和基于检测成本的协同检测来共同检测网络中的异常行为,以对抗MANET的各种入侵。

本文第2节介绍相关工作;第3节给出了基于MANET的入侵检测方案TCDC;第4节详细阐述了基于访问行为的在线聚类算法;第5节描述了基于检测成本的协同检测机制;第6节通过模拟实验分析TCDC的性能;最后对本文进行了总结。

2 相关工作

文献[1,2]提出了基于agent的分布式入侵检测体系结构。在该体系结构中,每个节点都参与检测,因此整个网络用于入侵检测的能源消耗很大。文献[3,4]提出使用基于簇的分布式检测方案来解决运行时的资源约束问题。每个节点被选择性地作为一个簇的ID agent,以节省节点能量。这类检测方案主要基于MANET中的分布式特点,通过节点之间的协作来提高检测效率,对于降低节点的资源消耗和提高入侵检测性能方面考虑较少。

Sen等[5]建议采用人工智能学习方法来检测已有攻击,如DoS攻击和路由攻击等。Liu等[6]提出了一种基于节点的混合挖掘异常检测技术——关联规则挖掘和特征挖掘,分别挖掘短期和长期的入侵行为。然而,这些入侵检测方案大多只考虑检测的准确率,算法较为复杂,对于在计算、存储等方面资源严重受限的MANET在实现上存在较大困难。

考虑MANET中节点在资源方面的限制,一些文献[7-10]

到稿日期:2009-10-13 返修日期:2010-01-29 本文受国家自然科学基金(60603069)资助。
王雷春(1974-),男,博士,讲师,主要研究方向为传感器网络等,E-mail:wlc2345702@163.com;马传香(1971-),女,博士,副教授,主要研究方向为移动Ad hoc网络等。

提出了节省资源、提高检测效率的入侵检测方案。文献[7,8]建议通过基于节点间信任度的协作检测来提高入侵检测的性能。Zhou 等[9]提供了一种节约资源、提高效率的综合分析方法，力图在增加安全性能的同时，在资源使用上有更好的效率。Vigna 等[10]给出了一种有效的入侵检测工具 AODV-STAT，它能够在消耗较少节点资源的情况下对 MANET 中的入侵行为进行有效的实时检测。

与文献[1,2,7,8]等提出的入侵检测方案相比，本文提出的方案借鉴了其研究成果，同时对其不足提出了改进。在本地检测上，采用基于在线聚类方法对节点访问行为进行聚类、分析，以提高检测效率和准确率。在联合检测方面，通过考虑信任度、能量和距离等综合检测成本选择协同检测节点和只检测本地不确定访问行为等方法，来减少通信量和增加效率。

3 基于在线聚类和检测成本的异常检测方案

MANET 存储能力较弱，无法保存大量入侵行为的访问模式。为此，本文提出了一种基于访问行为在线聚类和检测成本的异常检测方案(如图 1 所示)，包括两个层次：1)基于访问行为在线聚类的本地检测；2)基于检测成本的协同检测。

图 1 基于在线聚类和检测成本的异常检测方案

基于访问行为在线聚类的本地检测由 MANET 的各个节点独立完成。首先，本地节点收集来自各种信息源的实时网络数据流(包括 MAC 层、网络层、传输层和应用层)并进行预处理；然后，通过基于访问行为在线聚类的方法对本地收集数据进行分析；最后，根据本地检测结果进行响应。

当本地检测有不能确定的访问行为时，则要求邻近的节点参与全局协同检测。通过考虑基于检测成本的方法，本地节点选择部分协同检测节点，将不能确定的访问行为发送给这些节点，共同完成检测任务，并作出相应响应。

在本方案中，基于访问行为在线聚类分析和基于检测成本的协同检测分别是本地检测和协同检测的核心，下面详细讨论这两部分。

4 访问行为在线聚类分析

4.1 访问行为特征

定义 1(访问行为特征) MANET 中节点访问行为的属性集合，包括访问连接使用的 IP 地址、协议、访问系统敏感文件种类、登录次数、流量变化情况等。

定义 2(访问行为特征值) 假定节点访问行为共有 n 个属性，分别用 $a_1,a_2,\cdots,a_n$ 表示，$w_1,w_2,\cdots,w_n$ 是与不同属性对应的权重，则该访问行为特征值为

$$C=\sum_{i=1}^{n}(a_i * w_i) \tag{1}$$

访问行为特征包括多个不同方面，属性和权重取值各有不同要求。如用户登录成功取值为 0，不成功为 1；其权重则随着登录失败次数的增加而增加。系统敏感文件则可分为不同级别，级别越高，取值越大；其权重则随着访问次数的增加而增加。

根据上述定义，访问行为特征值越小，正常访问行为的可能性就越大；反之，异常访问行为的可能性较大。

4.2 访问行为时间序列划分

对 MANET 中节点的访问行为可看作一个以访问时间为自变量、访问行为特征值为因变量的时间序列，即访问行为时间序列，简称访问序列。

定义 3(访问序列) MANET 中对节点按时间顺序访问行为的一系列特征值的集合，记作 $S_C=((t_1,C_1),(t_2,C_2),\cdots,(t_n,C_n))$。其中，$(t_i,C_i)$表示在 t_i 时刻访问行为的特征值 C_i，n 为访问序列的长度，即访问行为个数。

定义 4(访问序列) $S_C=\{(t_i,C_i)\}(i=1,2,\cdots,n)$的划分中心和划分半径。

访问行为特征值划分中心和划分半径定义为

$$C_0=\frac{\sum_{i=1}^{n}C_i}{n} \tag{2}$$

$$r_C=\left[\frac{\sum_{i=1}^{n}(C_i-C_0)^2}{n}\right]^{\frac{1}{2}} \tag{3}$$

访问行为的访问时间划分中心和划分半径定义为

$$t_0=\frac{\sum_{i=1}^{n}t_i}{n} \tag{4}$$

$$r_t=\left[\frac{\sum_{i=1}^{n}(t_i-t_0)^2}{n}\right]^{\frac{1}{2}} \tag{5}$$

令 $L1=\sum_{i=1}^{n}C_i$，$L2=\sum_{i=1}^{n}(C_i-C_0)^2$，$L3=\sum_{i=1}^{n}t_i$，$L4=\sum_{i=1}^{n}(t_i-t_0)^2$，称为访问序列的 4 个特征系数，用 $LF=(L1,L2,L3,L4)$表示。

性质 1(访问序列性质) 设访问序列 $S_C=\{(t_i,C_i)\}(i=1,2,\cdots,n)$的特征系数为 $LF=(L1,L2,L3,L4)$，(t_i,C_i)为下一个访问行为，则新访问序列 $S'_C=\{(t_i,C_i)\}(i=1,2,\cdots,n+1)$的特征系数 LF'分别为

$$L1'=L1+C_{i+1} \tag{6}$$

$$L2'=L2+(C_{i+1}-C_0)^2 \tag{7}$$

$$L3'=L3+t_{i+1} \tag{8}$$

$$L4'=L4+(t_{i+1}-t_0)^2 \tag{9}$$

证明：(从略)。

4.3 算法描述

在算法实现过程中，需要定义 3 个访问序列参数：访问序列访问行为特征值半径阈值 r_{C_th}、时间范围半径阈值 r_{t_th}、时间距离阈值 Δt_{th}。r_{C_th}和 r_{t_th}分别规定了访问序列中访问行为特征值与中心值之间的最大半径和访问行为访问时间到中心时间的最大距离；时间距离阈值 Δt_{th}则规定了访问序列中相邻两个访问行为访问时间的最大距离。

判定方法 1：给定一个访问序列 $S_C=\{(t_i,C_i)\}(i=1,2,\cdots,n)$，其访问特征为 $LF=(L1,L2,L3,L4)$，则下一个访问行为为(t_{n+1},C_{n+1})能够加入 S_C 的充要条件是：新访问序列 $S_C'=\{(t_i,C_i)\}(i=1,2,\cdots,n+1)$的参数同时满足条件 $r_C'<r_{C\_th}$，$r_t'<r_{t\_th}$，$\Delta t'<\Delta t_{th}$。

其中，参数 r_C'，r_t'和 $\Delta t'$由式(2)－式(9)计算得到。

算法原理是考查节点的每个访问行为：如果没有访问序列，则将其加入新的访问序列；如果已经存在访问序列，则计算假定这个访问行为加入后新访问序列的参数。如果参数在规定范围内，则将其加入这个访问序列；否则，将其作为一个新访问序列的成员。

算法详细步骤可描述如下：

算法输入：访问行为(t_i, C_i)。

算法输出：访问序列及其特征系数。

算法步骤：

步骤 1　访问序列初始化.

1. initiate $S_C=\Phi$;

步骤 2　访问序列添加成员，即考察每个访问行为(t_i, C_i)

1. if($S_C==\Phi$) goto 8;
   else next;
2. generate $S_{C\_temp}=S_C+\{(t_i, C_i)\}$;//生成临时的访问序列
3. calculate $L1, L2, L3, L4, r_C', r_t'$ and $\Delta t'$ of S_{C_temp};
4. if($\Delta t>\Delta t_{th}$) goto 8;//大于规定时间距离阈值
   else next;
5. if ($r_t'>r_{t\_th}$) goto 8;//大于规定访问行为半径阈值
   else next;
6. if ($r_C'>r_{C\_th}$) goto 8;//大于规定访问行为时间范围阈值
   else next;
7. generate $S_C'=S_{C\_temp}$;//生成新访问序列
   repeat 步骤 2;
8. $S_C'=\{(t_i, C_i)\}$;
   repeat 步骤 2;

5　基于检测成本的协同检测

5.1　协同检测节点选择

5.1.1　检测成本计算

考虑 MANET 中节点在能量、带宽和计算能力等资源上的限制及检测效率，节点对不确定访问行为进行协同检测时需要考虑检测成本(Detective Cost, DC)，这包括与该节点有关联邻居节点的能量状态、与邻居节点的距离以及它们之间的信任度。能量、距离和信任度属于不同属性，需要进行归一化处理。

设 e_i 是节点 N 邻居节点 N_i 的能量，d_i 和 c_i 分别是 N 和 N_i 之间的距离和信任度。在计算检测成本时，能量和信任度越小，检测成本越高。可通过下式计算归一化值：

$$f_i'=\begin{cases}\dfrac{f_{\max}-f_i}{f_{\max}-f_{\min}}, & f_{\max}\neq f_{\min}\\ 1, & f_{\max}=f_{\min}\end{cases} \tag{10}$$

式中，f 代表节点之间的能量 e_i 或信任度 c_i，$f_{\max}=\mathrm{Max}\{f_i\}$，$f_{\min}=\mathrm{Min}\{f_i\}$。

节点之间的距离越大，检测成本越大。归一化计算方法如下：

$$g_i'=\begin{cases}\dfrac{g_i-g_{\min}}{g_{\max}-g_{\min}}, & g_{\max}\neq g_{\min}\\ 1, & g_{\max}=g_{\min}\end{cases} \tag{11}$$

式中，g 代表节点之间的距离能量 d_i，$g_{\max}=\mathrm{Max}\{g_i\}$，$g_{\min}=\mathrm{Min}\{g_i\}$。

节点 N 和邻居节点 N_i 之间的检测成本 DC_i 可通过下式计算：

$$DC_i=w_e*e_i'+w_d*d_i'+w_c*c_i' \tag{12}$$

式中，e_i'，d_i'和c_i'分别是节点 N 邻居节点 N_i 的能量、距离和信任度归一化系数，w_e，w_d 和 w_c 分别是它们的权重，且 $w_e+w_d+w_c=1$。

DC_i 越大，协同检测要求的资源越多，效率越低；反之，要求的检测资源少，效率高。

5.1.2　协同检测节点选择

协同检测节点选择越多，检测的可信度越高，但消耗资源也多；反之，消耗资源少，但检测可信度不足。为此，本文在选择协同节点时考虑了邻居节点数目 n 和选择比例($p\%$)两个方面的综合因素。当 n 较大时，$p\%$可取较低值；相反，可取较高值。n 和 $p\%$的取值规则可预先确定。

规则 1(协同检测节点选择)　假定节点 N 有 n 个邻居节点 $N_i(i=1,2,\cdots,n)$，根据式(12)计算的检测成本 DC_i 得到邻居节点 N_i 从小到大排序的序列 S_{DC}，n 和 $p\%$的取值规则已知，则被选择作为协同检测节点是序列 S_{DC} 中在 $p\%$范围的前若干个节点。

5.2　协同检测结果计算

当节点 N 接收到协同检测节点 $N_j(j=1,2,\cdots,m)$发回的协同检测信息后，根据下式计算协同检测结果：

$$R=\sum_1^m(P_{C\_j}*Q_j) \tag{13}$$

式中，P_{c_j}是节点 N 对协同检测节点 N_j 的信任度所占比重，$P_{c\_j}=c_j/\sum_1^m c_j$，c_j 是节点 N 对协同检测节点 Q_j 的信任度，Q_j 是访问行为被协同检测节点 N_j 认为是异常行为的可能性。

当 R 大于规定的阈值 R_{th}，则认为是异常访问行为；反之，是正常访问行为。

5.3　算法描述

算法原理：节点 N 首先将不确定访问行为(Access Behavior, AB)的有关信息 $Info_{AB}$ 发送给通过基于检测成本的方法选择的部分邻居节点 $N_j(j=1,2,\cdots,m)$；然后协同检测节点 N_j 根据本地信息对该访问行为 AB 进行联合检测，并将结果返回给节点；节点最后通过基于信任度的方法对返回结果进行确认，共同完成入侵检测。

算法具体描述如下：

算法输入：不确定访问行为。

算法输出：联合检测结果。

步骤 1：节点 N 从邻居节点 $N_i(i=1,2,\cdots,n)$中选择协同检测节点 $N_j(j=1,2,\cdots,m)(m\leqslant n)$

//计算检测成本

1. calculate DC_i of N_i by equation(10), (11) and(12);
   //根据检测成本对邻居节点排序
2. sort$S_{neigh}=\{N_i\}(i=1,2,\cdots,n)$ according to DC_i;
3. choose $S'_{neigh}=\{N_j\}(j=1,2,\cdots,m)$ of S_{neigh} according to n and $p\%$;//选择协同检测节点
4. send $Info_{AB}$ to N_j;

步骤 2：协同检测节点 N_j 对访问行为 AB 联合检测

1. receive $Info_{AB}$ from N;
   //选择特征值最相似的访问行为
2. choose AB' with the most similarity in N_j;
   //协同检测节点 N_j 给 N 发送协同检测结果
3. send detective result of AB' to N;

步骤 3：节点 N 根据返回的结果对访问行为 AB 进行确认

1. receive detective results from N_j;
   //计算协同检测结果
2. calculate R according to equation(13);
   //大于规定的阈值，作为异常访问行为
3. if $R>R_{th}$, take it as anomaly behavior;
   else, take it as normal behavior. //作为正常行为访问

6 模拟实现与性能评估

6.1 性能评估标准

采用的性能评估标准主要有以下几个。

检测率：被正确分类的测试样本个数与全体测试样本个数的比值，是入侵检测技术的总体性能评估标准。检测率越高，说明入侵检测方案的检测能力越强。

误报率：正常样本中被认为是攻击样本的个数与全体正常样本个数的比值。如果误警率很高，那么真正有危险的告警可能会被淹没在无用的误警当中，因此误警率越低越好。

异常检测通信开销：节点间协同检测时相互传输的信息量与网络中总的信息传输量的比值。该值越小，说明协同检测的检测效率越高。

CPU 平均负荷：异常检测时 CPU 的平均利用程度，用百分比表示。CPU 平均负荷越低，说明异常检测需要的资源越少，效率越高。

上述性能评估标准一方面考虑了异常检测对性能的要求（高的检测率和低的误报率），另一方面也考虑了异常检测机制对资源的要求（较低的异常检测通信开销和平均 CPU 负荷），因而是合理的性能评估标准。

6.2 模拟实验结果

采用网络模拟器 NS-2 对本文提出的检测方案进行模拟并分析其性能。网络参数包括节点数 Nodes、2 次连续移动之间的停留时间 Pause time、节点移动速度 Speed、无线传输距离 Tx range 以及各算法所需的其他必要参数。模拟参数的设置如表 1 所列。

表 1 实验参数

参 数	取 值	参 数	取 值
Channel	Wireless Channel	Range	200m×200m
Mac	802-11	Nodes	10～70
Antenna	OmniAntenna	Speed(m/s)	2
IFQ length	50	Pause time(s)	2
Route protocol	AODV	Tx range	30～180m

在 MANET 中，安全威胁种类很多，但威胁最大、最常见的是路由攻击，其次是对传输信息的攻击和假冒攻击。本实验对这 3 种安全威胁进行了综合测试，其比例分别为路由攻击 40%、对传输信息攻击 30%、假冒攻击 30%。

本文选择了 AODVSTAT[10] 作为比较对象，因为 AODVSTAT 与本文提出的检测方案在检测方式上有相似之处，且使用相同的路由协议 AODV。实验收集 TCDC 和 AODVSTAT 在不同时间访问行为的检测率、误报率、异常检测通信开销和 CPU 的平均负荷等方面的实验数据并加以分析，实验结果如图 2－图 5 所示。

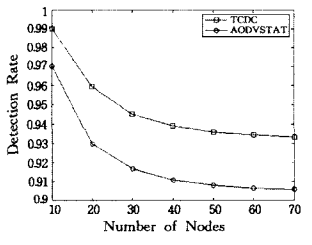

图 2 节点变化情况下检测率

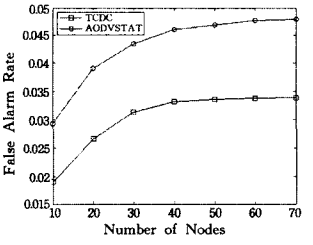

图 3 节点变化情况下误报率

随着节点数的增加，两种检测方案的检测率呈类似对数曲线下降（如图 2 所示）。这是由于网络中节点在刚增加时，每个节点的邻居节点数迅速增加，访问行为数量也快速上升，当节点数增加到一定时，节点的邻居节点数量（特别是一跳范围的邻居节点数量）变化较小，访问行为数量增加也随之减慢。与 AODVSTAT 相比，TCDC 的检测率更高。这是因为 TCDC 通过基于访问行为在线聚类，考虑了对节点前后访问行为的时间相关性，同时对不确定访问行为进行了包括信任度等的联合检测，这些都增加了检测率。

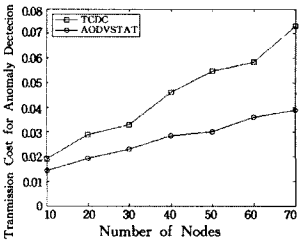

图 4 节点变化情况下异常检测通信开销

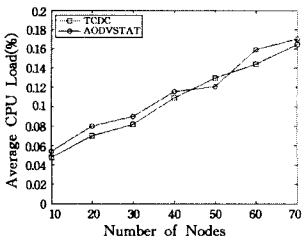

图 5 节点变化情况下平均 CPU 负荷

图 3 显示了两种检测方案在误报率上的变化规律。与图 2 相反，两种检测方案误报率曲线呈类似对数规律上升，但变化原因则相似。在两种检测方案中，TCDC 的误报率更低。一方面，TCDC 通过基于访问行为的在线聚类，考虑了对节点访问行为的前后时间相关性；另一方面，TCDC 采用了考虑协同检测节点能量、距离和信任度等的联合检测方法，有助于降低 TCDC 的误报率。

随着节点数的增加，用于异常检测的通信开销有所增加；与 AODVSTAT 相比，TCDC 用于异常检测传输的通信开销较少（如图 4 所示）。其原因是：1）节点只对不确定访问行为进行协同检测；2）通过基于访问行为的在线聚类，节点对不确定访问行为进行批处理检测；3）节点只选择部分检测成本低的邻居节点进行协同检测。通过以上方法，TCDC 大大减少了节点间通过联合检测而需要的通信开销。

从图 5 可知，随着节点数的增加，两种检测方案的 CPU 平均负荷都在增加，但总体上非常相似。与 AODVSTAT 比较，通过基于访问行为的在线聚类，TCDC 不需要对每个访问进行异常检测，减少了负荷；但聚类也增加了额外的计算负荷。因此，两种检测方案的 CPU 平均负荷相差不大。

结束语 本文从提高异常访问行为检测率和检测性能出发，以访问行为特征值和检测成本为测度，提出了一种新的访问行为在线聚类和检测成本的异常检测方案——TCDC。通过模拟实验表明，TCDC 在节点数变化条件下，在 MANET 中访问行为检测率、误报率和异常检测通信开销等性能方面均优于 AODVSTAT，且有较少的 CPU 资源消耗，从而证明了其作为 MANET 异常检测方案的有效性。

参 考 文 献

[1] Zhang Y G, Lee W. Intrusion detection techniques for mobile MANET[J]. Wireless Networks Journal, 2003, 9(5): 545-556

[2] Yi P, Zou F T, Jiang X H, et al. Multi-agent cooperative intrusion response in mobile ad hoc networks[J]. Journal of Systems Engineering and Electronics, 2007, 18(4)

[3] Huang Y A, Lee W K. A cooperative intrusion detection system for ad hoc networks[C]// Proceedings of the 1st ACM Workshop on Security of Ad Hoc and Sensor Networks. 2003: 35-147

（下转第 167 页）

$$Sim\_Tie_i=\frac{\sum_{a=1}^{m} Sim\_tie_a}{m}$$

②基本义原计算：采用基本义原计算，主要是用来反映多义词 W 与句子语境之间的关系。根据基本义原库可知多义词 W 的基本义原组 $\{BSense_1, BSense_2, \cdots, BSense_k\}$ 和语句除去多义词外的实词组 $\{ConWord_1, ConWord_2, \cdots, ConWord_j\}$。则基本义原与实词组之间的相似度构成一个相似度矩阵：

$$\begin{array}{c} \\ BSense_1 \\ BSense_2 \\ \vdots \\ BSense_k \end{array} \begin{array}{c} \begin{array}{cccc} CWord_1 & CWord_2 & \cdots & CWord_j \end{array} \\ \begin{bmatrix} BSim_{11} & BSim_{12} & \cdots & BSim_{1j} \\ BSim_{21} & BSim_{22} & \cdots & BSim_{2j} \\ \vdots & \vdots & \vdots & \vdots \\ BSim_{k1} & BSim_{k2} & \cdots & BSim_{kj} \end{bmatrix} \end{array}$$

通过该相似度矩阵可得义项 W_i 的基本义原与语句的相似度：

$$Sim\_Base_i=\frac{\sum_{a=1,b=1}^{k,j} BSim_{ab}}{j\times k}$$

③计算义项 W_i 的综合相似度。

$$Sim\_Com_i=\frac{Sim\_Tie_i+Sim\_Base_i}{2}$$

④对多义词的每一个义项执行上述三步计算，假设 Sim_Com_x 为其中最大的值，则 W_x 所代表的意思即为本语句中的含义。

为验证本文提出的消歧策略的可行性和有效性，本文选取人民日报语料库中的文档作为词义消歧的对象。通过与文献[5]中的消歧方法进行比较，可以得出本文消歧策略的可行性和有效性，具体数据如表 1 所列。

表 1　选定文档的词义消歧结果

文档标号	多义词数	文献[5]		本文消歧策略	
		正确数	准确率	正确数	准确率
980101-01-01	164	126	76.83%	131	79.88%
980101-01-02	152	126	82.89%	132	86.84%
980101-01-03	38	30	78.95%	31	81.58%
980101-01-04	88	70	79.55%	73	82.95%
980101-02-01	66	49	74.24%	50	75.76%
980101-02-03	39	23	58.97%	30	76.92%
980101-02-06	41	34	82.93%	35	85.37%
980101-02-07	103	84	81.55%	89	86.41%
980101-01-08	34	26	76.47%	29	85.29%
980101-03-01	65	54	83.08%	58	89.23%

从实验数据来分析，本文所采用的消歧策略相对来说有比较高的准确率，可以达到较好的词义消歧效果，进而保证了语义 Web 服务的名称匹配的正确性。

4　实验结果和性能分析

为了验证上述策略的可行性和有效性，做了相应的测试实验。在对 a_i 的选取为 $a_1=0.25, a_2=0.25, a_3=0.25, a_4=0.25$，并设定各阶段相似度的阈值为 0.60，且最终相似度的阈值为 0.90 的情况下，对测试集合中的 175 个广告服务进行 40 次实验测试，并将 40 次的实验数据平均分为 10 组。表 2 所列为本文的发现框架和文献[6]的发现策略在这 10 组实验中的查全率和查准率[7]。

表 2　两种发现框架的查全率和查准率

实验组序号	查全率		查准率	
	文献[6]	本文	文献[6]	本文
1	88.7%	92.5%	82.6%	89.2%
2	89.3%	93.3%	82.1%	90.0%
3	88.8%	92.6%	81.7%	89.4%
4	89.5%	93.0%	82.0%	90.2%
5	88.1%	93.1%	82.5%	89.3%
6	88.3%	92.9%	81.9%	90.2%
7	87.6%	93.0%	81.8%	89.6%
8	89.5%	91.7%	82.7%	90.0%
9	88.3%	93.1%	82.0%	89.7%
10	89.4%	92.6%	82.0%	89.6%

通过上述实验结果可以分析出，本文的语义 Web 服务发现策略在查全率/查准率方面高于文献[6]，因此，认为本文提出的多阶段匹配的语义 Web 服务发现框架结构是可行的、有效的。

参 考 文 献

[1] Massimo P, Takahiro K, Payne Terry R, et al. Importing the semantic Web in UDDI[C]// Proceedings of Web Services, E-business and Semantic Web Workshop. Toronto, Canada, 2002: 225-236

[2] Martin D, et al. OWL-S: Semantic Markup for Web Services [R]. Damlconsortium, 2009-3-9. http://www.daml.org/services/owl-s/1.0/owl-s.pdf

[3] 杜小勇，李曼，王珊. 本体学习研究综述[J]. 软件学报，2006，17(9)：1837-1847

[4] 吴健，吴朝辉，李莹，等. 基于本体论和词汇语义相似度的 Web 服务发现[J]. 计算机科学，2005，28(4)：595-602

[5] 张明宝，马静. 一种基于知网的中文词义消歧算法[J]. 计算机技术与发展，2009，2(19)：9-15

[6] 胡建强，邹鹏，王怀民，等. Web 服务描述语言 QWSDL 和服务匹配模型研究[J]. 计算机学报，2005，28(4)：505-513

[7] 邓汉成，王瑛，王敏芳. 从检索实例看查全率与查准率之间的关系[J]. 情报学报，2000，19(3)：237-241

[8] 田晓珍，尚冬娟. Web 的个性化服务[J]. 重庆工学院学报：自然科学版，2008，22(7)：76-80

(上接第 108 页)

[4] Yi P, Jiang X H, Wu Y, et al. Distributed Intrusion Detection for Mobile Ad Hoc Networks [J]. Journal of Systems Engineering and Electronics, 2008, 19(4): 851-859

[5] Sen S, John A C. A grammatical evolution approach to intrusion detection on mobile ad hoc networks[C]// Proceedings of the Second ACM Conference on Wireless Network Security. 2009: 95-102

[6] Liu Y, Li Y, Man H, et al. A hybrid data mining anomaly detection technique in ad hoc networks[J]. International Journal of Wireless and Mobile Computing, 2007, 2(10): 37-46

[7] Wang C H, Chin S. Reputation based intrusion detection system with threshold cryptography for wireless mobile ad hoc networks[C]// The 2008 International Computer Symposium (ICS 2008). 2008: 503-509

[8] Ramachandran C, Misra S, Obaidat M S. FORK: a novel two-pronged strategy for an agent-based intrusion detection scheme in ad-hoc networks [J]. Computer Communications, 2008: 31(16): 3855-3869

[9] Zhou B, Shi Q, Merabti M. Balancing intrusion detection resources in ubiquitous computing networks[J]. Computer Communications, 2008, 31(15): 3643-3653

[10] Vigna G, Gwalani S, Srinivasan K, et al. An intrusion detection tool for AODV-based ad hoc wireless networks[C]// Proceedings of the 20th Annual Computer Security Applications Conference. 2004: 16-27