现代支付系统信息安全的反欺诈监测模型研究

吴敬花[1,2] 周启海[1] 刘家芬[1]

(西南财经大学经济信息工程学院 成都 610074)[1] (四川农业大学信息与工程技术学院 雅安 625014)[2]

摘 要 计算机技术、通讯技术的迅猛发展与金融支付方式的信息化创新,使中国现代支付系统既越来越高效便捷,也面临日益加剧且监测颇难的金融信息安全威胁。这种威胁会影响我国现代支付系统信息化进程,还将影响国家金融命脉的信息安全与稳健发展。为此,提出了一种现代支付系统信息安全的反欺诈监测模型,该模型基于计算机链路挖掘新技术对现代支付系统海量信息进行动态反欺诈监测。对现代支付系统主要支付工具之一的信用卡进行反欺诈监测模拟的结果表明,该模型对提高信用卡欺诈判别的动态性、准确性和有效性,降低现代支付系统金融风险具有积极的意义。

关键词 现代支付系统,信息安全,链路挖掘,反欺诈监测模型

Research of Anti-fraud Detection Model for Advanced Payment System Information Security

WU Jing-hua[1,2] ZHOU Qi-hai[1] LIU Jia-fen[1]

(School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu 610074, China)[1]

(School of Information and Engineering Technology, Sichuan Agriculture University, Ya'an 625014, China)[2]

Abstract The rapid development of computer science and communication technology, as well as informational innovation on financial payment methods, make national Advanced Payment System(APS) more and more efficient and convenient, and they also bring many threats to financial information security which is very difficult to be detected. These threats will affect the informationalization course of national APS, and also influence the information security and steady development of national finance system. Therefore, this paper proposed an anti-fraud detection model for APS information security. Based on the link-mining as a new technology in computer science, this model detects the mass information of APS and finds fraud information dynamically. Through simulation on anti-fraud detection to credit-card which is one of the main tools of APS, its simulating results show that the model is significant to improve the dynamic detection, accuracy and validity of credit card fraud, and also reduces the financial risk in APS.

Keywords Advanced payment system, Information security, Link mining, Anti-fraud detection model

1 引言

目前,现代支付系统已成为金融领域的重要支柱平台,媒体技术渗透和支付方式的创新在给现代支付系统带来高效的同时,也带来更多的信息安全的威胁[1],特别是现代支付系统资金流上的欺诈支付问题,它将通过现代支付系统直接影响着金融秩序的稳定。因此,研究现代支付信息安全的反欺诈监测模型来监测并控制现代支付系统资金流上的恶意信息与金融欺诈势在必行。现代支付系统的海量信息处理与计算机新技术——链路挖掘技术[2]的发展,使准确监测和控制现代支付系统的资金流上的欺诈问题成为可能。但目前针对现代支付系统的信息安全问题大都集中在协议或制度等方面的研究[3],即便是现代支付系统主要支付工具之一的信用卡,其欺诈研究也主要集中在神经网络对账户数据的挖掘[4,5]。迄今,现代支付系统的信息安全反欺诈监测模型研究,仍显不足。为了对现代支付系统资金流进行动态高效的监控,本文融合互补了神经网络和链路挖掘技术优势,提出了一种现代支付系统信息安全的反欺诈监测模型(link-mining based anti-fraud model, LAFM);该模型以信用卡交易欺诈进行监测的模拟,分别用神经网络、链接挖掘、LAFM(以链路预测作为主要算法特征)3个模型,对信用卡交易记录从单案例数据、链接以及网络图进行了欺诈监测。其结果显示LAFM模型对提高信用卡欺诈监测的准确性与有效性具有重要意义。

2 LAFM设计

2.1 问题定义

为建立LAFM,需要定义两方面的内容。

(1)资金流过程

现代支付系统在进行支付时,通常需要多个参与者参与,这些参与者分布在现代支付系统不同环境中,任何一个环节

到稿日期:2010-01-21 返修日期:2010-04-02 本文受国家自然科学基金项目(60903201),四川农业大学双支计划项目(00570914),四川农业大学教学质量工程项目(00509018)资助。

吴敬花(1976-),女,博士生,讲师,主要研究方向为电子商务、电子支付,E-mail:hhhjjjww@gmail.com;周启海(1947-),男,教授,博(硕)士生导师,主要研究方向为计算几何、算法研究与应用、财经计算、同构化信息处理等;刘家芬(1980-),女,博士,讲师,主要研究方向为网络信息安全、隐私保护、形式化理论。

进行欺诈都将在资金流上表现异常，故监测欺诈行为时可由资金流监测情况进行判定。以信用卡为例，使用信用卡的资金流的过程如图 1 所示，它表示了信用卡在现代支付系统中完成一次支付的资金流过程。

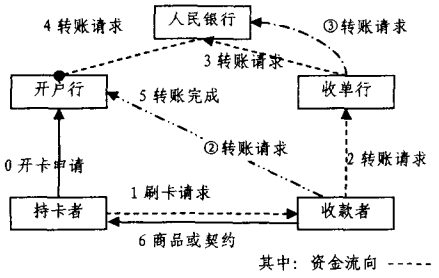

图 1　使用信用卡完成支付的资金流过程

图 1 中，②表示持卡者和收款者在同行支付的一次资金流过程，若持卡者和收款者不是同行，则需经过步骤 2－步骤 4；③表示持卡者和收款者跨国支付的一次资金流过程，该过程中涉及跨国的清算组织，若持卡者和收款者都是在我国现代支付系统进行，则经过步骤 3；其余表示正常状态下持卡者和收款者利用现代支付系统的资金流环节。信用卡的欺诈表现在用信用卡支付过程中资金流上的支付信息异常，包括违规套现、欺诈使用、账户冒用以及恶意透支等等，收集信用卡交易数据和建立信用卡使用的社会网络时，可参考图 1 的资金流过程。

(2)模型的判定准则

模型在判定资金信息流的过程中，根据文献[6]中 ROC 技术二值判别法的原理，使用 FRP(错误肯定率)、TRP(正确肯定率)度量模型的性能，其具体内容如表 1 所列。

表 1　信用卡交易判别矩阵

		预测值	
		正确	错误
实际值	正确	TP(正确肯定)	FN(错误否定)
	错误	FP(错误肯定)	TN(正确否定)

定义 FRP，TRP 分别为：

$$FRP=FP/(FP+TN) \tag{1}$$

$$TRP=TP/(TP+FN) \tag{2}$$

LAFM 模型建好后，根据 FRP，TRP 判定模型的性能。

2.2　LAFM 系统结构

现代支付系统资金流上运行着各类信息，现以信用卡支付信息为例设计 LAFM 系统结构。LAFM 模型建立在改进传统对单案例数据基于神经网络的数据挖掘和基于链接的链路预测模型基础之上。因此，LAFM 的设计与实现上分 3 个步骤：(1)对单案例数据使用神经网络进行信用卡的欺诈监测；(2)对链接数据用链接挖掘进行信用卡的欺诈监测；(3)对上述两次监测中不一致的数据用 Markov 链进行链路预测。LAFM 系统结构如图 2 所示。

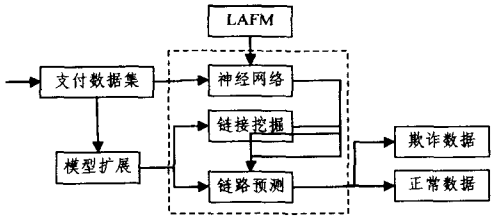

图 2　LAFM 系统结构图

图 2 中，支付的数据集来源于现代支付系统对信用卡支付的数据收集；模型的扩展是为链接挖掘和链路预测服务；神经网络主要对单个支付的案例数据进行挖掘，作初步判断；链接挖掘主要对案例数据的链接进行挖掘，从另一角度对案例数据进行判断。由于两种方法针对同一案例数据的利用具有不同角度与层面，故其对欺诈的判定结果具有对象同一性与结果互证性(但它未必总一致；若两次判定结果不一致者，则需采用链路预测技术进行再判别处理)，从而提高 LAFM 的预测精度。

2.2.1　神经网络下的单案例数据挖掘

神经网络的案例数据分析是监测信用卡欺诈的最常用的方法。对于交易数据集：$S=\{S_0,S_1,S_2,\cdots\}$，$S_i=(p_1,p_2,\cdots,Flag)(i=1,\cdots,n$，其中 $Flag$ 是欺诈标志)。对于收集数据，属性 p_i 是否对信用卡的欺诈 $Flag$ 有影响，由式(3)进行判定。

$$p(Flag|p_i)=p(Flag) \tag{3}$$

通过式(3)剔除与信用卡欺诈使用无关的节点数据的属性信息，并将影响信用卡欺诈的属性作为神经网络的输入元。属性 $p_i,p_j(i,j=1..r)$之间是否具有相关性，由式(4)进行判定。

$$\begin{aligned} &cov(x,y)=E((x-E(x))(y-E(y))) \\ &D(x)=E(x-E(x))^2 \\ &\rho=cov(x,y)/\sqrt{D(x)D(y)} \end{aligned} \tag{4}$$

式(4)表示如果两个属性之间具有相关性，则 ρ 值较大($\rho>0.5$)，认为属性之间相关，计算公共的并用公因子个数作为神经网络隐藏层的神经元个数，输出为记录欺诈的得分值，由此可得其三层神经网络；神经网络误差传播采用后向传播神经网络(BP)，神经网络的初始权重选[－1，1]的随机数，学习速率为样本数量的倒数，由反向传播网络对权值 δ 进行调整，当 $|\Delta\delta|<0.001$ 或执行次数大于样本数时，计算结束。

2.2.2　链接挖掘

链接挖掘是针对节点之间存在关系的社会网络数据进行挖掘的，在研究复杂的社会网络中具有重要的价值，其特例之一的 Web 链接挖掘目前已有广泛应用[7]。链接挖掘，一方面重视关注节点数据的链接信息，使数据挖掘的结果更具有合理性；另一方面，它将众多的链接信息简化为链接是否存在的概率信息，其规则定义为：

Rule 1：信用卡社会网络中，u_i,u_j 之间链接的存在与链接概率的关系如式(5)所示：

$$Link(u_i,u_j)=\begin{cases}0, & u_i,u_j \text{ 无链接存在，不属于一类}\\ 1, & u_i,u_j \text{ 有链接存在，同属于一类}\end{cases} \tag{5}$$

$$\begin{aligned} &\text{Prob}(Link(u_i,u_j)=1|(w_k,u_{ik},u_{jk}))=F(w_k,u_{ik},u_{jk});\\ &\text{Prob}(Link(u_i,u_j)=0|(w_k,u_{ik},u_{jk}))\\ &\qquad=1-F(w_k,u_{ik},u_{jk})\\ &\qquad=1-\text{Prob}(Link(u_i,u_j)\\ &\qquad=1|(w_k,u_{ik},u_{jk})) \end{aligned}$$

式中，$k\in[1,r]$，r 代表节点数据的属性个数。

而 $F(x,\beta)$ 值为：$F(w_k,u_{ik},u_{jk})=\sum_{k=1}^{r}w_k*\min\{\frac{u_{ik}}{u_{jk}},\frac{u_{jk}}{u_{ik}}\}/r$，显然：$0\leqslant F(w_k,u_{ik},u_{jk})\leqslant 1$。

信用卡链接挖掘基于该规则进行。为便于链接挖掘和神经网络数据挖掘结果的比较，根据信用卡基本模型，在上述交易记录节点上对模型进行扩展，建立信用卡支付的链接模型。

(1)信用卡交易记录的链接数据模型

由于链路挖掘关注链接信息，将信用卡原有的数据模型进行扩展[8]，信用卡由单个案例数据按照时间和空间等顺序将模型扩展为有向带权图 $G=(V,E)$，$(v_i \in V, e_{ij} \in E)$。其中：$v_i$ 是网络结构图中的顶点，该顶点表示信用卡交易的一个交易记录、扩展节点 v 的属性、信用卡资金的支付帐号等，并通过交易记录的共有信息形成案例数据之间的链接 e_{ijk}（由第 k 个属性引起的第 i 个节点到第 j 个节点的边），由链路预测的规则，计算 $P(e_{ij})$ 此时链接表示抽象的信息，即表示了节点 i 和节点 j 是否属于同一类别，于是信用卡的社会网络模型转为概率图 $G=(V,\{P_{ij}\})$，链挖掘则是对概率图进行数据挖掘。

(2)信用卡交易记录的链接参数计算

将信用卡支付记录数据形成的链接作为训练集，计算链接形成过程中的参数，即链接权重 w_k。针对每个链接的预测，误差为式(6)所示：

$$e_{ij}=1-f(w_k,u_{ik},u_{jk})=F(w_k) \tag{6}$$

构造平方和函数：

$$L(w_k)=\sum_{k=1}^{n} e_{ij}^2$$

当 $\frac{dL}{dW_k}=0$ 时，$L(W_k)$ 取极小值，即此时误差最小，得到 W_k，并由此得到链接挖掘的基本模型。

2.2.3 信用卡数据网络的链路预测

基于神经网络的单案例数据挖掘模型和基于最小残差平方和的链接挖掘模型，从不同的角度对信用卡支付的记录数据进行挖掘，两者在判定信用卡支付上具有对象同一性与结果互证性。但从支付数据构成网络的视角，链路预测模型因 Markov 链的广泛应用[8]而更为准确，进而可充分利用节点的结构信息，将单案例的数据挖掘和链接数据挖掘进行整合。当 LAFM 的两种模型处理结果出现结果歧义时，链路预测须采用 Markov 链对源数据进行再判定处理，而其关键是确定参数建立 Markov 链。

Markov 模型记为 $\langle S,A,\lambda\rangle$，$S$ 代表了状态空间，由一系列的支付记录构成的节点，A 是转移支付矩阵，它使得节点由一种状态转为另一种状态；λ 是 S 的分布参数。如果 $S(t)$ 表示 t 时刻的状态，$S(t)$ 由式(7)计算所得：

$$S(t)=S(t-1)A \tag{7}$$

此处，A 的维度由 S 的状态数决定，Maximum Likelihood 由于是参数估计中最常用的方法，将用来估计 A 值和 λ 值，其每个元素估计分别由式(8)、式(9)计算：

$$A(s,s')=C(s,s')/\sum_{s} C(s,s'') \tag{8}$$

$$\lambda(s)=C(s)/\sum_{s} C(s,s') \tag{9}$$

此处，$C(s,s')$ 是状态 s 到状态 s' 的状态数目，$A(s,s')$ 是 s 到 s' 的一步转移的概率矩阵，$A*A$ 是两步转移的概率矩阵，并以此类推。

基于 Markov 链的信用卡欺诈的链路挖掘为 $P(s(t)\mid s(t-1))=P(S(t-1))A$，因信用卡记录反应了两种状态：支付欺诈和支付非欺诈，所以，A 是 $2*2$ 的概率矩阵，将 $S(t-1)$ 状态下 $S(t)$ 概率值最大的作为 $S(t-1)$ 的下一目标状态，即对于一个记录，如果下一记录的信用卡支付欺诈概率大于信用卡支付正常使用的概率，则信用卡的下一记录作信用卡欺诈处理。鉴于 Markov 链计算复杂，采用上述训练集数据中所有欺诈记录以及时间上最临近的与欺诈记录等额的正常记录共同组成用于 Markov 链的节点构成的图作为训练集，计算 Markov 链的参数。

3 LAFM 测试数据生成与处理

为便于模型的评价，需要收集信用卡的交易数据，但信用卡数据涉及到隐私和安全等因素，因此使用基于信用卡模型（参照支付网关属性）的数据产生器产生信用卡交易记录，根据支付网关提供的信用卡交易信息属性以及信用卡风险 3% 的调查结果，产生 10000 个交易记录，其中 7000 个记录作训练集，3000 记录作测试集，3% 的异常记录平均分布在 10000 个数据中。样本随机抽取的部分案例数据，如图 3 所示。

编号	支付时间	支付地点	支付商户号	商户类型	支付金额	是否欺诈
946	2008-3-29 12:44:56	10101111	11111	011	205582	0
947	2008-3-17 19:08:03	00110100	11001	001	7028	0
948	2010-4-9 19:28:49	10010011	11011	010	127489	0
949	2009-11-21 23:20:15	10000000	11111	001	74646	0
950	2008-10-24 1:39:09	00011110	11111	000	36151	0
951	2009-9-1 9:30:06	01111100	00110	100	1829	0
952	2008-12-9 20:28:15	00110001	00000	011	43505	1
953	2009-5-22 17:21:58	00110011	10110	101	15383	0
954	2010-4-17 8:09:29	01111111	01100	010	100913	0
955	2009-3-1 10:14:23	10110100	01111	000	8993	0
956	2010-6-16 8:43:31	01101100	10000	101	11606	1
957	2008-4-29 6:38:16	10000011	01010	010	43686	0
958	2010-4-2 11:27:53	01100000	10011	101	73524	0
959	2008-8-14 5:27:15	00101010	11111	001	178139	0
960	2007-9-11 20:17:25	00001100	00110	111	36977	0

图 3 基于信用卡模型产生的部分仿真数据

图 3 中，支付地点、支付商号、支付类型等都是字符串类型，代码代表一个具体的地点或商号或类型。为便于计算，对数据进行标准化（无量纲化）处理，并映射到区间[0,1]，其处理方式如下：

(1)对离散型的数据，如信用卡刷卡地点（交易帐号），标准化过程为：设 $X=\{x_1,x_2,\cdots,x_n\}$，$x_i(i=1,2,\cdots,n)$ 是第 i 次消费的刷卡地点，以 m_{xi} 表示在地点 x_i 刷卡的次数，标准化计算如式(10)所示：

$$C(x_i)=m_{xi}/(\max\{m_{xi}\})(m_{xi}\geqslant 0) \tag{10}$$

显而易见：$C(x_i)\in[0,1]$。

(2)对连续型数据，如信用卡交易的金额，$X=\{x_1,x_2,\cdots,x_n\}$，$x_i(i=1,2,\cdots,n)$ 是第 i 次消费的交易金额，$x_i\sim N(\mu,\delta)$，δ 为该消费者的刷卡金额标准差，μ 为平均值，标准化过程为：

$$C(x_i)=e^{-\frac{1}{2}(x_i-u)^2} \tag{11}$$

当 $x_i=u$ 时，$f'(C(x_i))=0$，$f''C(x_i))<0$，$x_i=u$ 时，$C(x_i)$ 取极大值 1，所以，$C(x_i)\in[0,1]$，δ,μ 为参数。

4 LAFM 的实现与评价

4.1 神经网络下的单案例数据挖掘

由仿真数据经过信用卡属性数据进行的因果关系判定，得到神经网络输入为：卡支付时间、卡支付地点、卡支付商户类型、卡支付的商户号、卡支付金额；通过因子分析，得到 4 个公共因子，从而选择神经网络的隐藏层包含 4 个节点，神经网络的输出为异常的得分，以此判断信用卡是否有欺诈行为(1-欺诈，0-正常使用)，同时对不同模型的误差进行比较，当神经网络结构模型为 5-4-1 时，输出节点平均误差小。神经网络的初始权重选[−1,1]的随机数，学习速率 l=1/7000，选择反向传播网络对权值 δ 进行调整，当 $|\Delta\delta|<0.001$ 或执行次数 $n>7000$ 时，计算结束。实验中不断调整阈值，当阈值取 0.43 时，系统获得最佳性能，此时系统对信用卡消费行为检测的正确肯定率 $TRP=91.34\%$，错误肯定率 $FRP=12.22\%$，该结果较文献[5]更具有合理性。在链路分析中，将以此结果来进行模型和算法比较。

4.2 链接挖掘

由7000个信用卡交易记录作为节点，根据信用卡模型生成7000个交易记录的链接（如同一信用卡帐号、同一刷卡商户、同一刷卡时间等）作为计算对象，计算链接形成过程中的参数，即链接权重 w_k，针对每个链接的预测，误差为：

$$e_{ij}=1-f(w_k,u_{ik},u_{jk})=F(w_k)$$

针对7000个信用卡记录形成的图的测试链接求得 w_k，使其对链接误差绝对值最小，模型参数为：

$$w_k=\{0.54,0.43,0.32,0.47,0.65\}$$

将模型用于对测试集的节点形成的链接进行预测，试验数据分为两类：正常的信用卡交易记录和具有欺诈行为的信用卡交易记录。对于新的信用卡交易记录，其判别方式如图4所示。

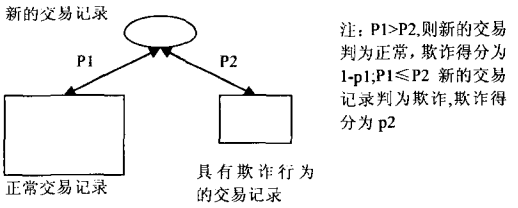

图4 信用卡交易的链接预测

图4表示了在交易记录监测的过程中，判定交易记录是否存在欺诈行为，通过计算，此时系统对信用卡消费行为检测的正确肯定率为 $TRP=93.16\%$，错误肯定率为 $FRP=14.4\%$。

4.3 链路预测

在训练集7000个交易节点数据的训练集中随机选择210个正常交易记录和所有210个欺诈交易记录形成的网络共同作为Markov链的训练集。对于初始支付节点，正常时其初始状态概率值为(1,0)，异常时初始状态概率值为(0,1)，经过计算后求得同一账号按时间的一步转移支付矩阵为：

$$A=\begin{pmatrix}0.819 & 0.181\\ 0.363 & 0.637\end{pmatrix}$$

在预测分析时，综合前两次的信用卡欺诈监测的结论，将两次监测结果不一致的作为进一步链路挖掘的测试数据，其结果如表2所列。

表2 信用卡两次监测的结果分析

		预测值			
		两次正确	一次正确	两次错误	总和
实际值	正确	2537	271	102	2910
	错误	3	15	72	90
	总和	2540	286	174	3000

该实验中，用于进一步做链路挖掘预测的交易记录节点数据为286个。通过对286个记录监测，得到 $TRP=88.19\%$，$FRP=13.3\%$，并进一步计算得到LAFM对信用卡欺诈行为检测的正确肯定率为 $TRP=95.4\%$，错误肯定率为 $FRP=5.6\%$。

4.4 LAFM评价

LAFM综合利用了信用卡交易记录的信息，通过社会网络的方法对数据进行分析，提高了信用卡欺诈监测的准确度。其对测试集中3000个数据的预测结果（节点按照对欺诈的概率值已排序）如图5所示。

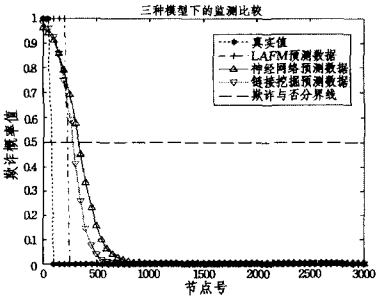

图5 测试集在各种模型下的预测结果

图5表示，LAFM在信用卡欺诈监测中，其预测结果与实际值最近，由此得到LAFM的预测精度相对其他模型更高。

LAFM和其它监测模型的综合性能如表3所列。

表3 信用卡欺诈各方法比较

	LAFM	神经网络	链接挖掘
数据挖掘方法	混合神经网络和链接挖掘以及Markov链	神经网络	OLS（最小平方二乘）
数据	图（节点、链接形成的网络图）	节点（交易数据）	链接、部分节点信息
复杂性	复杂	简单（适用大量数据）	复杂
精度（TRP,FRP）	高（95.4%,5.6%）	低（91.34%,12.2%）	一般（93.2%,14.4%）
判别次数	3	1	1
合理性	高	一般	一般

结束语 现代支付系统的广泛使用和媒体技术的发展以及支付方式的创新，使现代支付系统的信息安全面临更大的挑战。本文对现代支付系统信息安全的反欺诈监测模型进行研究，目的是提高现代支付系统的安全性，但要真正做到现代支付系统的安全，还需要对监测的欺诈行为进行控制，这将在后期研究中进行。

现代支付系统信息安全包含诸多方面，本文以信用卡的欺诈监测为例设计的现代支付系统信息安全的反欺诈监测模型（LAFM），其监测对象是现代支付系统的信息流的信息，具有一定代表性。LAFM采用混合策略的方法特别是引入链路挖掘的思想，不仅考虑了支付数据本身，也考虑了支付数据之间的联系，这使得现代支付系统的欺诈监测更合理，结果更容易被解释，也提高了现代支付系统信息安全欺诈监测的准确率，这为从新的角度研究现代支付系统的欺诈提供了新的视角，也对有效进行反欺诈控制，促进现代支付系统的发展，从而进一步促进金融秩序的稳定具有重要的意义。

参 考 文 献

[1] Asokan N, Janson P A, Steiner M, et al. The state of the art in electronic payment systems [J]. Computer, 1997, 30(9): 28-35

[2] Clauset A, Moore C, Newman M E J. Hierarchical structure and the prediction of missing links in networks [J]. Nature, 2008, 453: 98-101

[3] Vincent O R, Folorunso O, Akinde A D. Improving e-payment security using Elliptic Curve Cryptosystem[J]. Electronic Commerce Research, 2010, 10(1): 27-41

[4] Zhu Yan-li, Guo Xiao-Juan, Wang Shun-Ping, et al. A Study of

Credit Card Risk Detection Based on Chaos Theory and Neural Network[C]//CESCE. Wuhan, China, 2010: 29-31

[5] 郭涛,李贵洋,袁丁. 基于置信度和神经网络的信用卡异常检测[J]. 计算机工程,2008,34(15):205-207

[6] Fawcett T. An introduction to ROC analysis[J]. Pattern Recogn Lett., 2006, 27: 861-874

[7] 葛唯益,程龚,瞿裕忠. 语义 Web 链接结构分析之综述[J]. 计算机科学,2010,37(3):17-21

[8] Getoor L, Diehl C P. Link mining: a survey[J]. ACM SIGKDD Explorations Newsletter, 2005, 7(1): 3-12

(上接第 69 页)

实际上,能把数据流切割成报文段的主要原因是 TCP 在 RTT 或稍小于 RTT 的时间段内的突发特性。我们知道,TCP 发送端通过一次或几次的突发,发送拥塞窗口大小的报文,而在每 RTT 或稍小于 RTT 的剩余时间内都在空闲等待。这主要是 TCP 机制中的 ACK 压缩、慢启动和其它因素造成的。TCP 的这种突发性使得一个长的 TCP 流可以看成是被空闲时间分隔的一些小的报文段的串联。

5.2 FSLB 算法的性能仿真

我们对 FSLB 算法采用基于 trace 驱动的仿真方式,利用表 1 中列出的 trace 作为分派数据,通过与包粒度分派的轮循算法(RR)、流粒度的静态哈希表算法(S-Hash)的比较,说明 FSLB 算法具有良好的均衡性能。

5.2.1 性能参数

通过以下两个参数衡量 FSLB 算法的性能:

· 负载分派误差 LI(Load inaccurate)。均衡系统中 F_i 为路径 i 的期望负载分配比,且有 $\sum_{i=1}^{K} F_i = 1$,K 为路径数。一个最佳的负载均衡系统,在任何时间,其负载分派算法都应该保证路径 i 的负载分配比为 F_i。但是实际中是不可能达到这种理想效果的。设路径 i 的实际负载分配比为 F_i',我们定义负载分派误差为:

$$LI=\frac{1}{K}\sum_{i=1}^{K}|F_i-F_i'|$$

· 报文乱序程度。TCP 机制中当报文超时、重传计时器超时或收到 3 个相同的 ACK 时,发送端重传报文,并认为此时网络发生拥塞,启用慢开始算法或快恢复算法,降低自己的拥塞窗口 cwnd。我们用重传的报文个数来衡量数据报文的乱序程度。

5.2.2 仿真结果及分析

均衡系统从发散节点到聚合节点有 3 条路径,它们期望的负载分配比为 $\vec{F}=(0.3, 0.3, 0.4)$。基于 trace 驱动的仿真得到的各负载分配方法的平均负载分派误差 LI 如图 6 所示。

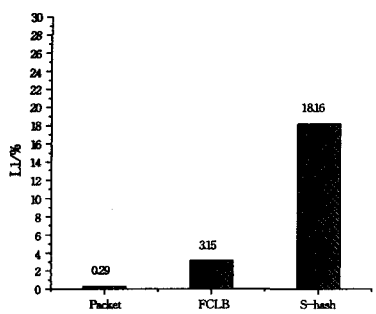

图 6 3 种负载分派算法的负载分派误差 LI

基于 trace 驱动的仿真可以体现出 Internet 网络中数据的特征,但是这种方法不能得到 FSLB 算法对 TCP 拥塞控制机制的响应。利用 ns2 对图 7 所示的网络拓扑进行 FSLB 算法仿真,通过改变链路时延 i,j,获得不同的最大路径时延差值,结果如图 8 所示。重传报文的个数与 β 和路径最大时延差值有关。当时延差值大于取定的 β 值时,重传的报文个数增加;反之,重传的报文非常少。这说明,只要 β 大于各路径时延差值的最大值,FSLB 算法就可以保证报文不出现乱序。

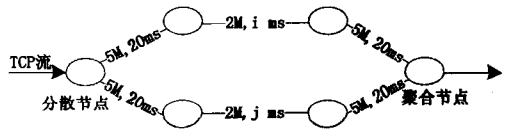

图 7 仿真拓扑

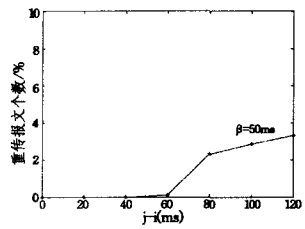

图 8 重传报文个数的百分比

结束语 本文提出采用基于相邻数据报文的时间间隔对同一业务流进行分割。分割后,一个业务流可以看成是由多个报文段串接而成。在有多条路径的均衡系统中,提出了 FSLB 负载分派算法,其使均衡系统在保证 TCP 流保序的同时,各路径能得到较理想的负载分配比。通过基于 trace 驱动的分析,FSLB 算法的均衡效果远远优于基于流水平的负载分派机制,稍逊于基于包水平的负载分派机制。

参 考 文 献

[1] Berkeley V J. TCP Evolution from 4. 3-Tahoe to 4. 3-Reno [C]// Proceedings of the 18th Internet Engineering Task Force. 1990: 365-374

[2] Brakmo L, O'malley S, Pererson L. TCP Vegas: New Techniques for Congestion Detection and Avoidance[C]// ACM SIGCOMM 1994. 1994: 24-35

[3] Thaler D, Hopps C. 2991 RFC[S]. November 2000

[4] Hopp C. 2992 RFC[S]. November 2000

[5] Cao Zhiruo, Wang Zheng, Zegura E. Performance of Hashing-Based Schemes for Internet Load Balancing[C]//Proceedings of IEEE INFOCOM. 2000: 332-341

[6] Rost S, Balakrishnan H. Rate-aware splitting of aggregate traffic [R]. MIT, 2003

[7] Sinha S, Kandul S, Katabi D. Harnessing TCP's Burstiness with Flowlet Switching[C]// 3rd ACM SIGCOMM Workshop on Hot Topics in Networks. San Diego, November 2004

[8] Kandul S, Katabi D, Davie B, et al. Walking the Tightrope: responsive yet stable traffic engineering[C]// SIGCOMM. 2005

[9] Wang H, Xie H, Qiu L, et al. Cope: Traffic engineering in dynamic networks[C]// ACM SIGCOMM. 2006

[10] Elwalid A, Jin C, Low S H, et al. MATE: MPLS Adaptive Traffic Engineering[C]// INFOCOM. 2001

[11] Lee Youngseok, Park Ilkyu, Choi Yanghee. Improving TCP Performance in Multipath Packet Forwarding Networks[J]. Communication and Networks, 2002, 4(2): 1-10