一种新的告警关联聚类算法*

刘冬生，曾小荟，唐卫东，肖晓红，郭恺强
（井冈山大学 电子与信息工程学院，江西 吉安 343009）

摘　要：针对网络中的告警泛洪和故障处理复杂问题，提出一种结合元胞学习自动机（CLA）和决策树 ID3 的新告警关联聚类算法。在 CLA 算法中使用学习自动机对告警信号进行分簇，但是在一个簇内如果出现任何子群或交错，则决策树 ID3 学习算法通过分割数据样本训练在该簇上来优化决策边界，从而大大减少分簇告警数目以及完成对根源性告警的定位。仿真表明，该算法能有效地对大量告警信号进行分析，并且能比较准确地鉴定出根源性告警。

关键词：告警泛洪；告警关联；根源性告警；聚类算法
中图分类号：TP309；TP301.6　　**文献标志码**：A　　**文章编号**：1001-3695(2013)12-3786-04
doi：10.3969/j.issn.1001-3695.2013.12.069

Novel alarm correlation clustering algorithm

LIU Dong-sheng, ZENG Xiao-hui, TANG Wei-dong, XIAO Xiao-hong, GUO Kai-qiang
(*School of Electronic & Information Engineering, Jinggangshan University, Ji' an Jiangxi* 343009, *China*)

Abstract: Aiming at the alarm floods and the complexity of network fault, this paper proposed a novel alarm correlation clustering algorithm. The algorithm combinatorial method based on cellular learning automata(CLA) clustering and ID3 decision tree learning algorithm. The CLA used learning automata to cluster the alarms and events, however if there were any subgroups or overlaps within a cluster, the ID3 decision tree trained on that cluster refined the boundaries, it could reduce a number of alarms and locating the root cause alarm. The simulation result show that this algorithm can analyse alarms efficiently and indentify the root cause alarm accurately.

Key words: alarm floods; alarm correlation; root cause alarm; clustering algorithm

0　引言

告警是来自设备的一个不确定的信息，说明系统出现问题。故障是产生告警的原因，一个故障可能会产生大量的来自有效网络组件的告警。实际上，一个故障会以反应链的形式而导致另一个故障产生，这样会产生无数的告警，从而会掩盖真正的重要故障。而且大多数的告警并没有包含故障真正原因的信息，但一个故障产生时会出现大量的告警，其中一些告警是冗余的，这样使得故障的处理更复杂。由此可见，找出根源性的告警是进行故障分析的关键和基础。告警关联分析是指对告警进行合并和转换，能确定反映根本原因的告警。因此通过告警关联来进行网络故障管理，许多研究者认为这是一种最有效的解决办法。其主要目的是减少显示给网络管理者的信息量，提高信息的用处，以及及时地确定最可能引起告警的故障，甚至提出正确的措施[1]。总的来说，告警关联提供了一个概念上的理解，在一个有大量告警的网络中，用于有效的故障诊断、隔离和识别、积极维护以及趋势分析等[2]。

目前有很多的算法应用到网络告警关联中，例如将相似告警进行分簇的告警聚类算法[3,4]，通过应用改进的 K-means 分簇法[5]和对假阳性及低息告警进行过滤[6]等来提高入侵检测准确性的算法，以及考虑告警数据库信息量大的问题而提出的增量式关联规则挖掘算法[7]等。在考虑告警的严重性、可信度不足、网络动态适应性差、关联误报率高和对根源性告警难以定位等问题上，文献[8～11]也分别提出了一些比较好的解决办法。其目的在于减少告警数量和提供其中与规则学习有关的算法，贝叶斯置信网络和聚类算法很流行[12]。在相关学习算法中，其挑战在于对于大型数据集来说，是在挖掘模式的有效性与完整性之间产生的权衡问题而带来的不稳定性。作为一个鲁棒性概率模型，尽管贝叶斯学习算法越来越引起关注，但在复杂实时系统中，其带来的复杂性和低效率性是一个问题且代价昂贵。基于聚类算法的确定特性比如告警的序列和间隔值已经提出和应用，在趋势分析和告警风暴探测中，作为一个有效的方法去分类网络告警。然而，由于缺少一个有效的能准确计算在一个临时空间中告警之间的距离量化相似度标准，它们很少应用到序列模式发现中[13]。基于此，本文提出一个结合基于元胞学习自动机（cellular learning automata, CLA）和决策树 ID3 的告警关联聚类算法，通过对告警和事件进行聚类和优化决策边界来对告警信号进行分簇和优化，从而检测出产生故障的根源性告警。

收稿日期：2013-03-14；**修回日期**：2013-04-28　**基金项目**：江西省自然科学基金资助项目(20122BAB201038)；江西省科技支撑计划项目(2012BBG70161)；江西省教育厅 2012 年度科技计划资助项目(GJJ12465)

作者简介：刘冬生，男，江西吉安人，讲师，硕士，主要研究方向为无线传感网络、网络安全(liuds7615@163.com)；曾小荟，男，副教授，博士，主要研究方向为计算机网络、信息安全；唐卫东，男，副教授，博士，主要研究方向为计算机信息可视化等；肖晓红，女，副教授，硕士，主要研究方向为网络安全；郭恺强，男，讲师，博士，主要研究方向为计算机网络体系结构.

1　元胞学习自动机和决策树 ID3 学习算法

1.1　元胞学习自动机(CLA)

CLA 是由学习自动机(learning automata,LA)与元胞自动机(cellular automata,CA)相结合的一个功能强大的数学模型[14]。LA 是一种在未知的随机环境中完成自适应决策系统,它通过一个学习过程自适应提高它们的性能。其学习过程是,在每个阶段,LA 基于动作概率矢量从一组有限的动作集合中选择一个动作,并输入到随机环境中去,环境对每个由自动机所选择的动作进行性能评价,以决定该动作的有效性;然后环境为自动机提供一个动作响应,自动机在这个阶段根据该动作响应去更新内部动作概率矢量。通过重复该过程,LA 最终学习选择到来自环境的期望响应的最佳动作。假定 a_i 是在第 n 轮中根据概率矢量 $p(n)$ 选择的动作,环境中有两个输出 $\beta \in \{0,1\}$,当 $\beta=0$ 时表示环境响应对于该动作是有利的,此时概率矢量根据式(1)更新:

$$p_j(n+1)=\begin{cases} p_j(n)+a[1-p_j(n)] & \text{if} \quad i=j \\ p_j(n)(1-a) & \text{if} \quad i\neq j \end{cases} \tag{1}$$

当 $\beta=1$ 时表示环境响应对于该动作是不利的,此时概率矢量可根据式(2)来更新:

$$p_j(n+1)=\begin{cases} p_j(n)(1-b) & \text{if} \quad i=j \\ b/(r-1)+p_j(n)(1-b) & \text{if} \quad i\neq j \end{cases} \tag{2}$$

其中:a 为奖励参数;b 为惩罚参数;r 为动作数目大小。

CA 是定义在一个具有离散、有限状态的元胞组成的元胞空间上,并按照一定的局部规则,在离散的时间维上演化的动力学系统。元胞自动机能进行复杂的计算,具有高效性和鲁棒性。一个 CA 包含有限的元胞空间,其状态是一组有限的整数集 $\varphi=\{1,2,\cdots,k-1\}$。每个元胞根据一个局部规则离散地更新它们的状态,每个元胞的新状态取决于之前的该元胞和邻居元胞的状态。

CLA 的基本原理是利用 LA 去调整 CA 的随机状态传输概率。一个 CLA 是一个分配了一个 LA 或多个 LA 的元胞,居住在元胞中的 LA 基于它的动作概率矢量决定元胞的状态(动作)。每个元胞基于它过去的经历以及它的邻居元胞的行为来演化。CLA 开始于初始状态(每个元胞的内部状态),然后在每个阶段,自动机根据它的概率矢量选择一个动作,接下来根据 CA 的局部规则产生一个强化信号给居住在元胞的 LA。基于接收到的强化信号,每个 LA 更新它的内部概率矢量。这个过程不断重复,直到得到最佳结果。

1.2　决策树 ID3 学习算法

ID3 是一种基于信息熵的决策树学习算法,根据属性集的取值选择示例的类别[15]。其核心是在决策树中各级节点上选择属性,用信息增益(information gain)作为属性分类的标准,递归地拓展决策树的分支,完成决策树的构建。其基本原理是:设 S 是 s 个数据样本的集合,假设类别属性有 m 个不同的取值,定义 m 个不同的类 $C_i(i=1,2,\cdots,m)$。设 s_i 为 C_i 的样本数,对一个给定的样本分类所需的期望信息如式(3)所示:

$$I(s_1,s_2,\cdots,s_m)=-\sum_{i=1}^{m}P_i\log_2(P_i) \tag{3}$$

其中:p_i 是任意样本属于 C_i 的概率。设属性 A 具有 v 个不同值 $\{a_1,a_2,\cdots,a_v\}$,可以用属性 A 将 S 分成 v 个子集 $\{a_1,a_2,\cdots,a_v\}$,如果 A 选择测试属性(即最好分裂属性),则这些子集对应于由包含集合 S 的节点生长出来的分支。设 s_{ij} 是子集 s_j 中类 C_i 的样本数,由属性 A 划分的熵 $E(A)$ 由式(4)得到:

$$E(A)=\sum_{j=1}^{v}(s_{1j}+s_{2j}+\cdots+s_{mj})/(s\times I(s_{1j},s_{2j},\cdots,s_{mj})) \tag{4}$$

其中:$(s_{1j}+s_{2j}+\cdots+s_{mj})/s$ 是充当第 j 个子集的权值,并且等于子集(即属性 A 值为 a_i)中样本个数除以 S 中的样本总数。其中对于给定子集的期望信息为

$$I(s_{1j},s_{2j},\cdots,s_{mj})=-\sum_{i=1}^{m}p_{ij}\log_2(p_{ij}) \tag{5}$$

式中:$p_{ij}=s_{ij}/|S_j|$ 是样本属于 C_i 类的概率。由期望信息和熵值可得到对应的信息增益值,即在属性 A 上分支获得的增益可表示为

$$\text{gain}(A)=I(s_{1j},s_{2j},\cdots,s_{mj})-E(A) \tag{6}$$

把信息增益最大的属性选为树的根节点,接下来一个新的决策树通过使用训练样本子空间根据信息增益值来递归地构建。当所有样本在可利用的训练样本子空间来自同一类时就形成叶节点或决策点。算法用正常的数据构建 ID3 决策树,异常检测使用决策树通过用测试样本的特征以遍历该树来实施。如果能到达叶节点,那么测试样本检测是正常的,否则是异常的。

2　结合 CLA 和 ID3 的告警关联聚类算法

2.1　聚类算法

聚类是把数据群分类或分簇的处理过程,它有分块、分层、基于密度、基于网格和基于模型等技术。聚类有时偏向于得到圆形簇,并且其延展性也是一个问题。使用欧几里德(Euclidean)距离或曼哈坦(Manhattan)距离测量可趋于得到有相似大小和密度的球形簇,但也有可能会出现其他形状。有时聚类对输入数据的次序比较敏感以及不吸纳新插入的数据。聚类结果的可解释性和使用性、数据的高维数性、噪声和欠测值都是聚类的问题[16]。聚类算法是用来缓和告警泛洪特别是当需要处理大量告警时是一个比较好的算法。对于一个数据空间,聚类算法的目的是将数据空间分割为簇,这样在一个簇内的数据对象之间有很高的相似度,而与另一个簇内的数据对象相比,其相似度几乎完全不同[17]。这样使得在每个簇内的数据对象具有相互类似的特性,其相似度由某个距离标准来衡量。对于告警模式如关联、趋势和结合等是基于它们的行为相似度来提出告警聚类,相似度经常是由一组有意义分布的告警来定义。另外对于大多数告警来说,由于不包含带故障根本原因的信息,也即当一个故障出现时会出现很多的告警,而这里很多的告警是偶然的,这样给故障的处理带来更多的困难。

在基于 CLA 的聚类算法中,可把已分配了 n 维数据的 LA 称为一个 agent[18]。在数据分簇中,每个 agent 基于它的方向概率矢量来动作,如果当前 agent 和邻居 agent 的相似度高,则 agent 根据相似度向邻居 agent 的方向调整,即当前的 agent 基于本地信息通过学习获得邻居 agent 的方向,且这个方向是适宜的,从而找到相似 agent。在 CLA 中,每个学习自动机相当于一个 agent,且包含有一个 n 维数据,每个元胞又包含一个或多个 agent,一个元胞学习机假定为一个 agent 的外部环境,每个 agent 基于一个先前的标准和邻居及不同的方向概率来选择一个方向,如果当前 agent 有邻居 agent,则当前 agent 根据与其邻居 agent 的相似度来给予是奖励还是惩罚以更新它的概率矢量。另外,假定每个邻居 agent 分别影响各自当前 agent 的移

动方向以及相似度的计算是基于欧几里德距离或曼哈坦距离。如果相似度高,则当前 agent 给予奖励,且与邻居 agent 有相同的方向,否则给予惩罚。

2.2 告警关联聚类算法

在实际运转中,网络告警在前续告警与后续告警之间以暂时关系的复杂性连续产生。一个事件指的是某个特定的故障所产生的一组告警,但一个故障能导致包含不同告警的全局事件或几个局部事件。一个事件由几个确定的告警组成,一个告警有且仅有一个与其相关的事件。发现在这些告警之间的关联,能有助于找到在不同网络故障之间的内在因果关系。本文为了研究其关联性,首先对告警给出相关定义[19]:a)告警定义,$A=(R,T)$,$[(t,S,I)]$,其中 R 表示资源,T 表示告警类型,t 表示时间,S 表示刚度,I 表示信息;b)告警事件,$E_i=\langle e_i,t_n\rangle[\cdots](i,n=1,2,\cdots,)$,其中 e_i 表示告警类型,t_n 表示告警发生的时间;c)告警类型:$e_i=\langle$ object_class, object_instance, alarm_num, edsc$\rangle[\cdots](i,n=1,2,\cdots,)$,其中 object_class 表示对象分类的序列号,object_instance 表示目标示例的序列号,alarm_num 表示告警类型的序列号,而 edsc 表示告警信息包括告警的优先权和告警描述。当告警有相同的 object class 和 object_instance 时,可归纳到同一类别中去。在本文算法中,基于 CLA 算法的聚类过程包含以下三个步骤:

a)输入一组数据样本,样本产生的告警类型在元胞中接受奖励或惩罚,产生的每个告警类型给出奖励,其他告警类型在同一元胞和邻居元胞中给出惩罚。

b)通过捕获在前阶段的奖励参数和其他参数如时间和维度来找到初始事件。

c)通过学习自动机形成最后事件聚类,即将许多相似度高的告警归结为单一的告警,从而大大减少告警数目。

假设每个通信节点为一个元胞,每个元胞包括一个自动机,m 为在一个元胞中的告警类型的个数,G 表示产生的告警,H 表示一个元胞中除了产生的告警外其他告警的集合,则 CLA 对告警信号给出奖励或惩罚的概率矢量更新规则可用式(7)表示:

$$\begin{cases} p_{n+1}(G)=p_n(G)+a(1/r-p_n(G)) \\ p_{n+1}(H)=p_n(H)(1-b) \end{cases} \quad \forall h\in H\Rightarrow h\neq G,|H|\leqslant m \tag{7}$$

其中:a 是奖励参数;b 是惩罚参数;r 是通信节点数目。当告警类型大量产生时,告警概率立即增大;当产生告警时,在元胞和它的邻居中更新奖励和惩罚。在元胞中,每个节点通过接收告警信号而激活,而自动机选择一个合适的动作,然后根据式(7)对告警信号进行惩罚或奖励。这里要指出的是,在一个元胞中所有的概率之和应该等于 $1/r$,以便在下一个阶段中轮回使用学习自动机。现在自动机应找到元胞与它的邻居之间的告警关系,通过告警之间的关系则可以找到系统中的故障,并且能确定下来。这样基于 CLA 可把数据样本空间分类成 k 个不相交的簇,CLA 保证每个样本只相关于一个簇,但是在一个簇内如果出现任何子群或交错,则决策树 ID3 用蕴涵规则在特征空间上通过分割数据样本训练在该簇上来优化决策边界,通过关联性规则形成带有根源性告警的分簇。

在优化决策边界算法中,包括两个阶段:a)计算 CLA 中一个衡量异常的数值(ADM)以及获得决策树 ID3 的决策数值;b)结合 CLA 分簇算法和决策树 ID3 两种算法的结果,得到最终的 ADM 值。具体过程如下:

在第一阶段中,对于每个数据样本 $S_i(1\leqslant i\leqslant n)$,其包括如下三个步骤:

a)计算欧几里德距离 $D(S_i,r_j)(1\leqslant j<n)$,找到离 S_i 最近的 f 个分簇,其中 f 是预先定义的参数。

b)对这最近的 f 个分簇计算 CLA 中的 ADM 值以及获得决策树 ID3 的决策数值。

c)对 S_i 返回 ADM 矩阵。

算法得到数据样本 $S_i(1\leqslant i\leqslant n)$ 和 f 作为输入参数,且对每个数据样本给出矩阵 ADM$[f\times 2]$。设 $DT_1,DT_2,\cdots,DT_k$ 是通过运用 CLA 算法在数据样本形成分簇 $C_1,C_2,\cdots,C_k$ 上的 ID3 决策树,$r_1,r_2,\cdots,r_k$ 分别是分簇 $C_1,C_2,\cdots,C_k$ 的中心,然后考虑到一个样本,从 f 个候选分簇 $T_1,T_2,\cdots,T_f$ 选取 ADM 值,所选择的这 f 个候选分簇是根据在 S_i 与分簇中心之间的欧几里德距离最近的 $C_1,C_2,\cdots,C_k$ 中的 f 个分簇。

假定 $z_1,z_2,\cdots,z_f$ 是候选分簇 $T_1,T_2,\cdots,T_f$ 的中心点,在最佳矢量 S_i 和 f 个候选分簇的欧几里德距离可表示为

$$D(S_i,z_j)=d_j \quad 1\leqslant j\leqslant f \tag{8}$$

则对每个 f 个候选分簇通过 CLA 分簇算法可计算出 ADM 值,其值为

$$\mathrm{ADM}_j=\frac{N_{C_j}}{\sum_{i=1}^{k}N_{C_i}}F_i \quad 1\leqslant i\leqslant n,1\leqslant j\leqslant f \tag{9}$$

其中:N_{C_j} 表示分簇 C_j 中的数据样本的数目;$N_{C_j}/\sum_{i=1}^{k}N_{C_i}$ 表示分簇 C_j 的概率;$F_i=1-d_j/\sum_{i=1}^{k}D(S_i,r_j)$,这是一个换算因子,它表示分簇 j 和样本 S_i 之间的距离与样本 S_i 和分簇 $C_1,C_2,\cdots,C_k$ 之间距离之和的比值权衡。F_i 通过来自最佳矢量 S_i 用距离来惩罚分簇 C_j 的概率,这样小的距离可得到高的 ADM 值,反之亦然。用决策树 ID3 来对 f 个候选分簇进行“0”或“1”的判决,“0”或“1”分别代表样本的正常或异常。该结果的输出是依据决策树能否遍历到一个叶节点来判定。

在第二阶段中,对每个数据样本 $S_i(1\leqslant i\leqslant n)$,通过 CLA 分簇算法和决策树 ID3 的判决计算出 ADM 值来构成矩阵 ADM$[f\times 2]$。ADM 矩阵包含结合 CLA 和 ID3 算法在 f 个分簇上产生的值 $\mathrm{ADM}_1,\mathrm{ADM}_2,\cdots,\mathrm{ADM}_f$。然后,算法在 ADM 矩阵中对 f 个分簇 $T_1,T_2,\cdots,T_f$ 进行排序,使得在 S 和候选分簇 $T_1,T_2,\cdots,T_f$ 之间的距离 $d_1,d_2,\cdots,d_f$ 满足关系 $d_1<d_2<\cdots<d_f$。在有序 ADM 矩阵中,如果满足 $\mathrm{DT}_j=0$ 且 $\mathrm{ADM}_j\leqslant 0.5$ 和 $\mathrm{DT}_j=1$ 且 $\mathrm{ADM}_j>0.5$,则 CLA 算法中的初始值 ADM_j 作为结合 CLA 和 ID3 算法的最终取值。其中 $\mathrm{DT}_j=0$ 或 1 表示 j 分簇中的 ID3 决策树将数据样本归类为正常或异常。最后,对于每个测试样本 S_i,得到来自两种算法的结合所产生的在封闭区间[0,1]的一个 ADM 值。对最佳测试样本分类应用阈值规则来判断是否正常或异常,即如果 ADM 值大于一个预先定义的一个阈值,则该测试样本是异常的,否则是正常的。

3 性能评估

为了对告警关联算法进行性能评估,并且分析算法的有效性以及对产生告警的根本原因进行准确识别,本文给出如图 1 所示的网络拓扑仿真图。

图中有 26 个节点,其中人为给出 6 个根源性告警,包括 3 个 IP down、2 个 link down 和 1 个 port down 等。假定这 6 个根源性告警分别出现在图中黑色箭头指向的节点上,并且所有的节点都与服务器相连。比如图 1 中的节点 N_2 通过路由器连接到所有的服务器,这样 N_2 产生与服务器相关的告警如 HTTP

unavailable 或 FTP unavailable 等告警类型,同样路由器也会同时产生相邻路由器的告警信号,如 neighbor loss 这样的告警类型。根据 TCP/IP,网络的体系结构由上而下分为应用层(application layer)、运输层(transportation layer)、网络层(network layer)和数据链路层(data link layer)四层,在仿真中使用 120 个人为的告警数据集作为告警关联算法的输入数据,随机分布在网络节点的这四层中,如图 2 所示。这里要指出的是,在图 1 中,由于 R_8 路由器链路发生故障,则所有的告警信号都无法到达 N_{10}、N_{11}和 N_{12}这三个节点,因此这三个节点不会出现告警信号。

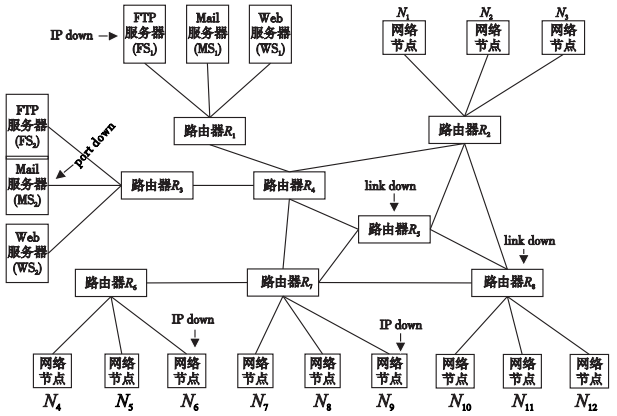

图1　仿真模型

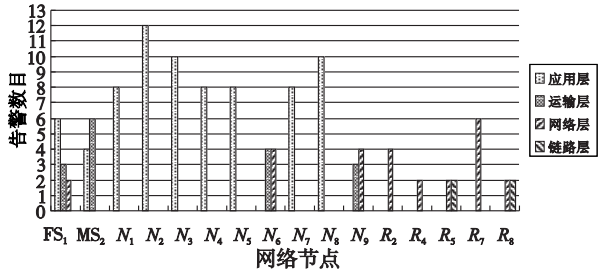

图2　仿真前告警在各节点的数量分布

图 3 是告警信号在各个节点的数量分布在经过 CLA 算法后的结果。由图 3 可知,告警数量明显减少,已经由最初给出的 120 个减少到 32 个,因为对于同类型的告警信号归纳为 1 个单一的告警信号。比如对于 MS_2 节点,原来在应用层有 4 个同类型的告警信号,则根据算法的判定规则,归类为 1 个告警信号。而对于 R_2 或 R_7 来说,由于它们都与 R_5、R_8 相连,则会同时产生分别由 R_5 和 R_8 带来的不同类型的 neighbor loss 告警信号。同样,由于 N_1、N_2 等站点都通过路由器连接 FS_1 和 MS_2 服务器,也产生两种类型的相似告警信号如 FTP unavailable 和 HTTP unavailable 的告警信号。这样 R_2、R_7、N_1 和 N_2 等就可归纳为分别有两种不同类型的告警信号。

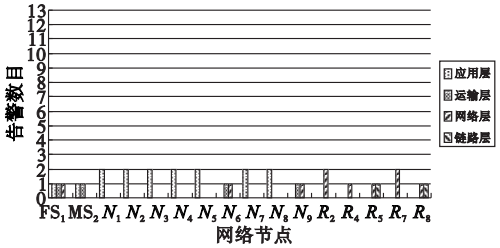

图3　CLA算法后告警数目的分布

在经过本文的聚类算法后,告警数目又将大大减少,因为通过聚类算法,可将具有相似性且有关联性的告警信号归结为一个告警分簇。比如对于 FS_1 节点,由于 IP down 是引起其在运输层产生告警信号 port down 的根本原因,且引起在应用层出现 FTP unavailable 和 FTP down 2 个告警信号。由于这些告警信号的产生都是基于同样的原因且有关联性,从而可把这些告警信号归结为一个分簇,这样在 FS_1 中,就有 4 个告警信号,它们构成一个分簇。同样在 R_5 中,由于 link down 是产生告警信号的初始事件,则在 R_5 处检测到的 IP down 和由 R_4 和 R_7 产生的 neighbor loss 告警信号均是由初始事件产生,它们属于相似性告警,具有关联性,故把这 3 个告警信号也归结为一个告警分簇。据此可将告警信号分成六个分簇,数量也由原来的 32 个再减少到 17 个,其分簇情况如图 4 所示。由于根源性告警由各节点产生,其他告警信号的产生均由其初始事件引起,因此它们构成的分簇就是节点上的根源性告警以及由其引起的其他告警信号所形成,故在图 4 中用节点来表示这六个分簇。

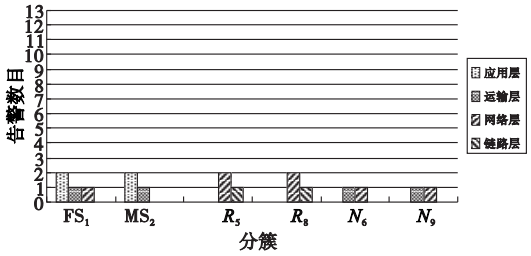

图4　聚类算法后各簇的告警数目分布

根据分簇后的情况以及告警信号在各层的分布,可确定根源性告警。例如在 FS_1 分簇中,有 4 个告警信号,分别为应用层中的 2 个 unavailable、FTP down 告警信号和运输层中的 port down 告警信号以及网络层中的 IP down 告警信号等。显然在该簇中,根据协议规则可知,出现在最底层中的告警信号 IP down 应为根源性告警。对于其他分簇,同样可检测出根源性告警,如 MS_2 中的根源性告警为 port down、R_5 中的根源性告警为 link down 等。因此在通过本文算法后,网络中出现的告警信号由最初给出的 120 个最终减少到 6 个,同时可确定这 6 个告警信号为根源性告警。

4　结束语

故障管理系统越来越要求确保有鲁棒性和质量保证,告警信息是出现故障的直接表现,尽管告警信息在系统中对于用来确定故障是有一定的作用,但产生的告警泛洪却会掩盖造成故障的根源性告警。本文针对网络中一旦出现故障则会出现的大量告警信号难以处理的问题,提出了一种新的告警关联综合算法,该算法对告警泛洪的分簇和对根源性告警的识别具有比较好的效果。当然,该算法没有考虑到时间窗的问题,如何及时有效地发现根源性告警和使算法更具鲁棒性是今后研究的方向。

参考文献:

[1] THANGAVEL M, Dr THANGARAJ P. Cluster based statistical anomaly intrusion detection for varied attack intensities[J]. International Journal of Computer Applications, 2011, 24(9): 27-33.

[2] SHARADA K A, HEMANT, PRASHANTH, *et al*. A model proposed for reducing the false positive alarm rate using the feature of event correlation[J]. International Journal of Advanced Research in Computer Science and Software Engineering, 2012, 2(8): 103-108.

(下转第 3793 页)

者即使获取了一次通信的会话密钥也无法利用该信息去获取另一次通信的会话密钥,以此实现了已知密钥安全。

4　结束语

本文结合自验证公钥密码学和口令信息,提出了一个用户直接通信的跨域口令认证密钥协商协议。域服务器间通过自验证公钥密码学实现相互认证和密钥协商。域服务器通过口令认证实现对域内用户的认证,并协助用户完成认证和生成密钥,但不能获取最终的会话密钥。分析结果表明,与同类协议相比,新协议计算代价和通信代价均较低,且能抵抗字典攻击和未知会话密钥共享攻击。协议同时保证了会话密钥的公平性、前向安全和已知会话密钥安全等属性的实现。

参考文献:

[1] BYUN J W, JEONG I R, LEE D H, *et al*. Password-authenticated key exchange between clients with different passwords[C]//LNCS, vol 2513. 2002:134-146.

[2] CHEN L. ISO/IEC JTC 1/SC27 N3716, A weakness of the password-authenticated key agreement between clients with different passwords scheme[S]. 2003.

[3] KIM J, KIM S, KWAK J, *et al*. Cryptanalysis and improvement of password-authenticated key exchange scheme between clients with different passwords[C]//LNCS, vol 3043. 2004:895-902.

[4] PHAN R C W, GOI B M. Cryptanalysis of an improved client-to-client password-authenticated key exchange (C2C-PAKE) scheme [C]//LNCS, vol 3531. 2005:33-39.

[5] WANG Shu-hong, WANG Jie, XU Mao-zhi. Weaknesses of a password-authenticated key exchange protocol between clients with different password[C]//LNCS, vol 3089. 2004:414-425.

[6] BYUN J W, LEE D H, LIM J I. EC2C-PAKA: an efficient client-to-client password-authenticated key agreement [J]. Information Science, 2007, 177(19):3995-4013.

[7] WANG Feng-jiao, ZHANG Yu-qing. Cryptanalysis of a client-to-client password-authenticated key agreement protocol, Report 2008/248[R/OL]. (2008). http://eprint. iacr. org/.

[8] PHAN R C W, GOI B M. Cryptanalysis of two provably secure cross-realm C2C-PAKE protocols[C]//Proc of INDOCRYPT. 2006:104-117.

[9] FENG Deng-guo, XU Jing. A new client-to-client password-authenticated key agreement protocol[C]//Proc of IWCC. 2009:63-76.

[10] 陈安林,潘进,郭超,等. 移动自组网中跨域两方认证密钥协商协议研究[J]. 计算机应用研究,2011,28(7):2734-2737.

[11] 刘小琼,潘进,李国朋. 基于无证书的两方跨域认证密钥协商协议[J]. 计算机应用研究,2012,29(2):646-649.

[12] 王涛. 跨域客户间口令认证的密钥交换协议[J]. 计算机工程与设计,2012,33(7):2546-2549.

[13] CHEN Li-qun, HOON W L, YANG Guo-min. Cross-domain password-based authenticated key exchange revisited[C]//Proc of the 32nd IEEE International Conference on Computer Communications. 2013:1052-1062.

[14] GIRAULT M. Self-certified public keys[C]//Proc of EUROCRYPT. 1991:490-497.

[15] PETERSEN H, HORSTER P. Self-certified keys—concepts and applications[C]//Proc of Communications and Multimedia Security. Boca Raton: Chapman and Hall. 1997:102-116.

[16] SAEEDNIA S. Identity-based and self-certified key-exchange protocols[C]//Proc of ACISP. 1997:303-313.

[17] KOBLITZ N, MENEZES A, VANSTONE S. The state of elliptic curve cryptography[J]. Journal of Design, Codes and Cryptography, 2000, 19(2-3):173-193.

(上接第 3789 页)

[3] AI-MAMORY S O, ZHANG Hong-li. Intrusion detection alarm reduction using root cause analysis and clustering[J]. Journal of Computer Communication, 2009, 32(2):419-430.

[4] GADDAM S R, PHOHA V V, BALAGANI K S. K-means + ID3: a novel method for supervised anomaly detection by cascading K-means clustering and ID3 decision tree learning methods[J]. IEEE Trans on Knowledge and Data Engineering, 2007, 19(3):345-354.

[5] 韩占柱. 改进 K-means 算法的网络数据库入侵检测[J]. 微电子学与计算机,2012,29(3):144-150.

[6] PIETRASZEK T. Using adaptive alert classification to reduce false positives in intrusion detection [C]//Lecture Hotes in Computer Science, vol 3224. 2004:102-124.

[7] 李金凤,王怀彬. 基于关联规则的网络故障告警相关性分析[J]. 计算机工程,2012,38(5):44-46.

[8] VIINIKKA J, DEBAR H. Processing instrusion detection alert aggregate with time series moding[J]. Information Fusion, 2009, 10(2):312-324.

[9] 梅海彬,龚俭. 多 IDS 环境中基于可信度的警报关联方法研究[J]. 通信学报,2011,32(4):138-145.

[10] 张亚普,孟相如,张立,等. 基于 SVM 和模糊逻辑的告警相关性分析[J]. 计算机应用研究,2011,28(2):685-688.

[11] 闫生超,唐云善,张春平,等. 基于网络和时间关联关系的告警相关性分析[J]. 电力系统自动化,2011,35(9):78-87.

[12] KIM D S, SHINBO H, YOKOTA H. An alarm correlation algorithm for network management based on root cause analysis[J]. International Journal of Computer Science and Network Security, 2011, 9(12):1233-1238.

[13] YASAMI Y, MOZAFFARI S P. A novel unsupervised classification approach for network anomaly detection by K-means clustering and ID3 decision tree learning methods[J]. Journal of Supercomputing, 2010, 53(1):231-245.

[14] HARIRI A, RASTEGAR R, NAVI K, *et al*. Cellular learning automata based evolutionary computing for intrinsic hardware evolution [C]//Proc of NASA/DoD Conference on Evolvable Hardware. 2005:294-297.

[15] MITCHELL T. Machine learning[M]. New York: McGraw-Hill, 2004.

[16] CHEN Yan, LEE J. Autonomous mining for alarm correlation patterns based on time-shift similarity clustering in manufacturing system [C]//Proc of IEEE Internation Conference on Prognostics and Health Management. 2011: 365-371.

[17] HAN Jia-wei, KAMBER M. Data mining: concepts and techniques [M]. San Mateo: Morgan Kaufmann, 2006.

[18] IMANAZARI M, NOORHOSSEINI S M, AKBARI F. Dynamic alarm correlation based on cellular learning automata in telecommunication networks[J]. Internation Journal of Computer Science and Network Security, 2009, 9(12):168-173.

[19] WALLIN S. Chasing a definition of "alarm"[J]. Journal of Network and Systems Management, 2009, 17(4):457-481.