一种基于聚类的路径伪造检测方法

杨 斌 陆余良 杨国正 张 亮
(电子工程学院 合肥 230037)

摘 要 提出一种基于聚类的路径伪造检测方法。该方法将相邻时刻路由路径的变化集作为检测对象,以前缀地址所属国家为依据,对路径变化集进行聚类,引入各变化自治域的AS链接概率偏离度、中间国家出现概率和中间国家地理偏离度的定义,在此基础上引入路径级异常检测指标,综合利用这些指标检测路由中的路径伪造异常行为。选用真实的路径伪造事件数据进行实验,结果表明该检测方法较以往的检测方法更为有效、可行。

关键词 异常检测,聚类,路由劫持,路径伪造,边界网关协议,AS路径
中图法分类号 TP393 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.08.035

Path Forging Detection Approach Based on Aggregation

YANG Bin LU Yu-liang YANG Guo-zheng ZHANG Liang
(Electronic Engineering Institute, Hefei 230037, China)

Abstract This paper presented a novel algorithm for detecting routing path forging based on aggregation. By selecting change of AS path as the detection object, using the country which the prefix belongs to as the standard, the change of AS path was converged. The definition of AS link probability deviance, intermediate country appearance probability, and intermediate country distance deviance were introduced. Based on these metrics, we introduced path-level detecting metrics and integrated these metrics to check routing path forging. The data of actual routing path forging event was tested by the proposed method. Experimental results demonstrate that the method is more valid and practical than previous methods.

Keywords Anomaly detection, Aggregation, Routing hijacking, Path forging, Border gateway protocol, AS path

1 引言

边界网关协议(Border Gateway Protocol,BGP)[1]是自治域之间用于交换路由选路信息的协议,然而,由于协议缺乏一个安全可信的路由认证机制,无法对传播路由信息的真实性和完整性进行验证[2],因此运行BGP协议的域间路由面临严峻的安全形势,其安全问题也日益突出。例如2008年的巴基斯坦YouTube事件[3]、2010年的China Telecom事件[4]以及2011年的Link Telecom劫持事件[5],这些事件对域间路由系统均造成了破坏性影响。

路由劫持是域间路由面临的最主要的威胁,通常由恶意的攻击或路由器配置错误而引发。依据路由劫持行为的不同,可以将路由劫持分为劫持前缀和路径伪造[6]。劫持前缀异常中,攻击者对外宣告自己是某个被攻击前缀的拥有者,路径伪造异常则通过对路由项中的AS-PATH属性进行攻击。两类攻击方法中,路径伪造异常更加隐蔽,较难检测,因而,针对路径伪造异常的检测一直是路由系统安全研究的重点。

近年来,针对路径伪造的检测已有许多研究。Christopher等[7]提出了一种基于拓扑结构的检测方法,边缘AS被划分到不同的聚类簇,认为一条有效的AS路径必须符合以下两条规则:①最多只能包含有一个子系列的核心AS;②AS路径中的连续边缘AS应隶属一个簇或所在簇之间存在一条非常近的连接。依照如上规则进行检测,实验结果验证了该方法的有效性,但是算法需要实时更新拓扑结构信息,同时边缘AS的分类结果也会对算法的结果产生较大的影响。文献[8]利用指纹来检测路径伪造异常,但该方法需要处理大量的返回数据信息,同时认为在网络中出现过的边均是有效合法的,并未考虑边的变动,使检测结果缺乏准确性。Li等[9]提出一种基于伙伴的路径伪造检测方法,认为路径被劫持后,待检测点与伙伴的相似性将被破坏,该方法具有较好的实时性,鲁棒性较高,然而并非所有待检测点都能找到足够数目的有效伙伴,且文中伙伴的判断依据仍有待丰富。上述研究均是对路由的AS-PATH属性独立地进行分析研究,处理的数据量较大,同时检测需要一定的先验知识作为支撑。针对上述问题,本文提出了一种基于聚类的路径伪造检测方法,该方法以目的前缀所属国家为依据聚类AS路径变化信息,有效地降低运算需要处理的数据量。实验结果表明,该方法能够有效地检测路径伪造异常行为。

到稿日期:2013-10-20 返修日期:2014-01-28 本文受安徽省自然科学基金(1208085QF107)资助。

杨 斌(1989-),男,博士生,CCF会员,主要研究领域为计算机网络安全,E-mail:yang810941186@163.com;陆余良(1964-),男,教授,博士生导师,主要研究领域为计算机网络安全;杨国正(1982-),男,讲师,主要研究领域为计算机网络安全;张 亮(1982-),男,讲师,主要研究领域为计算机网络安全。

2 路径伪造

路径伪造主要通过伪造虚假 AS-PATH 属性实现劫持目标前缀流量的目的，依据被劫持前缀的不同可以将路径伪造异常分为前缀路径伪造和子前缀路径伪造。为便于描述，本节设定攻击者 AS 为 A，被劫持的 AS 为 V，被劫持前缀为 p。

2.1 前缀路径伪造

前缀路径伪造的攻击方式表现为：A 向外宣告一条经过自己到达目标前缀 p 的虚假 AS-PATH 路由。该攻击方式不会增加互联网中前缀 p 的起始 AS，因而网络中不会检测到 MOAS(Multiple Origin AS conflict)冲突[10]，但实际网络中，该路由的 AS-PATH 属性是虚假无效的，该路由的存在将促使部分以目标前缀 p 为目的的流量信息经过 A，实现对流量的劫持。图 1 为该异常情况的一个实例。

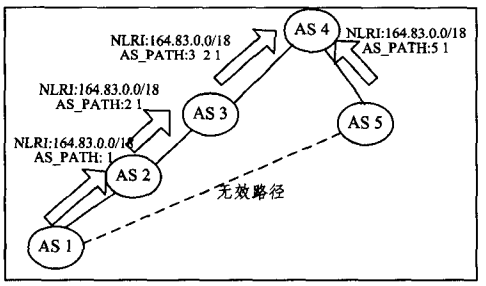

图 1 前缀路径伪造异常

图中 AS5 向 AS4 宣告了一条前缀为 164.83.0.0/18，AS-PATH 为 AS5、AS1 的虚假路由，同时 AS4 收到来自 AS3 的前缀为 164.83.0.0/18，AS-PATH 为 AS3、AS2、AS1 的真实路由，其中 164.83.0.0/18 是 AS1 所申请的 IP 地址块。由于这两条路由的前缀相同，因此 AS4 将根据路由选路策略筛选路由。由于 AS5 宣告的 AS-PATH 较短且 AS4 无法对 AS-PATH 中的所有链接进行验证，因此 AS4 将接受来自 AS5 的虚假路由，但事实上图中的 AS5 和 AS1 之间并不存在任何的连接。这样，到达 AS4 的以 164.83.0.0/18 为目的前缀的流量将会被转发到攻击者 AS5 中，进而实现对前缀 164.83.0.0/18 流量的劫持。

2.2 子前缀路径伪造

子前缀路径伪造的攻击方式与前缀路径伪造相似，不同的是攻击者仅劫持前缀 p 的部分前缀，即子前缀 p1，其中前缀 p 是自治域 V 所申请的 IP 地址块，如图 2 所示。

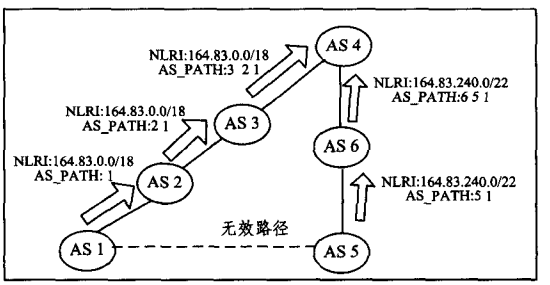

图 2 子前缀路径伪造异常

图中 AS4 将分别接收到来自 AS6 和 AS3 的前缀为 164.83.0.0/18，AS-PATH 为 AS3、AS2、AS1 的真实路由以及前缀为 164.83.240/22，AS-PATH 为 AS1、AS5、AS6 的虚假路由(其中 164.83.240/22 是 164.83.0.0/18 的子前缀)。与前缀路径伪造不同的是，由于这两条路由之间不存在冲突，即前缀不同，因此 AS4 将同时接受这两条路由，并对转发表进行更新操作。根据路由器转发的最长前缀匹配原则，到达 AS4 的以目的前缀 164.83.240.0/22 的流量均将被转发到攻击者 AS5。

显然，子前缀路径伪造较前缀路径伪造具有更好的攻击效果。其原因在于攻击者劫持前缀 p 的同时，网络中也存在着一条由自治域 V 宣告的关于前缀 p 的正常路由，若路由器同时接收到这两条路由，可能会由于 AS-PATH 较长的原因而过滤该虚假路由，而子前缀路径伪造所宣告的前缀 p1 与前缀 p 之间并不相同，路由器将同时接受这两条路由，因此关于子前缀 p1 的虚假路由将有可能对整个网络造成影响。

3 基于聚类的路径伪造检测算法

3.1 基本思想

本算法建立在以下分析的基础上：①BGP 在选路时，除某些已经设定的路由策略外，将优先选择低延迟或邻近的路由器作为下一跳地址，因此，路由存在路径伪造时，路径中伪造的自治域可能极度地偏离 BGP 路由器与目的前缀地址位置所在的直线，即伪造自治域距离 BGP 路由器与目的前缀所在直线较远。同理，该自治域也可能极其偏离其前后自治域地理位置所在的直线。同时，由于该自治域是伪造的，因此该自治域可能极少地出现于目的前缀所在位置与 BGP 路由器的路径当中。②路径中伪造的自治域有可能与路径中前后的自治域之间不存在连接，因此该自治域与前后自治域之间的链接将较少出现在其他路由项的路径中。

算法考虑存在冲突的路由路径，即相邻时刻的转发表中，以共同前缀为目的地址的存在路径冲突(即变化)的路由，主要原因在于，当路由存在路径伪造时，其与伪造发生前一时刻的正常路径存在不同的 AS 路径信息，显然，变化的 AS 路径信息极有可能就是路径中伪造的自治域。基于以上分析，只需要对存在冲突的 AS 路径中变化的 AS 进行分析，对其造成的距离偏离、链路的出现频度以及自治域出现的频度进行衡量分析，即可判断该变化 AS 所在的路由是否存在路径伪造。

单独分析各个变化 AS 无法定量准确地判断该 AS 是否造成了较大的距离差值，也无法确定该 AS 是否存在较低的出现频度，同时，如对所有的变化 AS 链路进行统计分析，各变化 AS 与其前后自治域的链接出现概率也无法直接用于检测分析，这是因为从全局的角度来看，该链接均具有极低的出现概率，不存在比较的标准。为此，提出了一种基于聚类的路径伪造检测方法，该方法首先将 AS 路径变化集依据其目的前缀所在国家聚类，将同一目的前缀所在国家的 AS 变化路径聚合到一个类别下，将各变化 AS 转化为其所在国家，取得变化 AS 集所对应的国家集合；接着依据 Z-Score 统计引入 AS 链接概率偏离度 LPD、中间国家出现概率 ICAP 以及中间国家地理偏离度 ICDP 分别对变化 AS 的链接出现概率、变化 AS 国家出现的频度以及变化 AS 所引入的距离偏离程度进行衡量；最后将各变化 AS 的指标转化为其所在路径的指标，并综合利用这些路径级指标筛选出存在路径伪造的路由。

3.2 指标定义

3.2.1 路径变化集聚类方法

设 BGP 路由器为 M，对于相邻时刻 t_1 和 t_2($t_1<t_2$)路由转发表的表项，以 $AP^{M,p}$ 代表 t_1 时刻提取的 M 到目标前缀 p

的 AS-PATH 信息(设前缀 p 的归属 AS 为 A_p),以 $AP^{M,p'}$ 表示 t_2 时刻提取的 M 到目标前缀 p' 的 AS-PATH 信息(设前缀 p' 的归属 AS 为 $A_{p'}$),其中($A_{p'}=A_p$)且 p 和 p' 之间存在相同的 IP 网段,即 p 和 p' 满足 $p\subset p'$ 或者 $p'\subset p$。$AP^{M,p}$ 与 $AP^{M,p'}$ 表示如下:

$$AP^{M,p}=\{M,AS_1^{M,p},\cdots,AS_k^{M,p},A_p\} \tag{1}$$

$$AP^{M,p'}=\{M,AS_1^{M,p'},\cdots,AS_k^{M,p'},A_{p'}\} \tag{2}$$

接着,由式(3)可得 M 在 t_2-t_1 时间内,关于前缀 p 和 p' 的 AS-PATH 的 AS 变化集 $NC^M(p,p')$:

$$NC^M(p,p')=AP^{M,p}\bigcup AP^{M,p'}-AP^{M,p}\bigcap AP^{M,p'} \tag{3}$$

设 $C=\{C_1C_2\cdots C_i\}$ 为国家集,对于 M 中路由转发表的各路径变化集,依据前缀地址所属国家 C^p 对路由变化集 $NC^M(p,p')$ 进行聚类,得到隶属于不同国家的变化路径集 $CP(C_d)$:

$$CP(C_d)=\bigcup\{NC^M(p,p')\},\forall p\in P^d,C^p=C_d \tag{4}$$

其中,P^d 表示位于国家 C_d 的 IP 前缀集。

3.2.2 链接概率偏离度——LPD

对于 $CP(C_d)$ 的 AS 变化集 $NC^M(p,p')$ 中的 AS 元素,获取该 AS 在其所属的 AS-PATH 中相互邻接的 AS 链接,并统计各个变化 AS 链接的出现数目,例如,$AS_i\in NC^M(p,p')$,若 $AS_i\in AP^{M,p}$,则其相邻的 AS 链接分别为$\langle AS_{i-1}^{M,p},AS_i^{M,p}\rangle$与$\langle AS_i^{M,p},AS_{i+1}^{M,p}\rangle$,其中 $AS_i=AS_i^{M,p}$,若 $AS_i\in AP^{M,p'}$,则其相邻的 AS 链接分别为$\langle AS_{i-1}^{M,p'},AS_i^{M,p'}\rangle$与$\langle AS_i^{M,p'},AS_{i+1}^{M,p'}\rangle$,其中 $AS_i=AS_i^{M,p'}$。取向量 IL^d 为 $CP(C_d)$ 中所有变化 AS 的 AS 链接出现次数,如式(5)所示:

$$IL^d=\begin{Bmatrix}N(AS_i,AS_j)\\ \vdots\\ N(AS_l,AS_m)\end{Bmatrix} \tag{5}$$

其中,$N(AS_i,AS_j)$为(AS_i,AS_j)链接出现的数目。

由于变化 AS 的相邻链接出现次数的分布较离散,因此无法在众多的链接中定位奇异链接(即可能并不存在于实际网络中的链接)。考虑到 $CP(C_d)$ 中正常变化 AS 的相邻链接相对虚假链接具备较高的出现概率,可以使用统计方法来定位这类虚假链接。由于 Z-Score 统计方法能够处理分布离散的数据,同时具有在众多数据中查找奇异点的能力,因此借助 Z-score 统计的定义,根据已知向量 IL^d 定义国家 C_d 的 $CP(C_d)$ 中各变化 AS 的相邻链接$\langle AS_i,AS_j\rangle$出现的链接概率偏离度,如式(6)所示:

$$LPD^d(AS_i,AS_j)=\frac{N(AS_i,AS_j)-E[IL^d]}{\sigma[IL^d]} \tag{6}$$

其中,$E[IL^d]$、$\sigma[IL^d]$分别为 IL^d 向量中 AS 链接出现次数的均值及其标准差。

依据链接概率偏离度的定义,结合 Z-Score 的性质,存在如下结论:①当 $LPD^d(AS_i,AS_j)<(>)0$ 时,则与到达国家 C_d 的其他变化 AS 的链接相比,$\langle AS_i,AS_j\rangle$出现的概率较低(较高);②当 $LPD^d(AS_i,AS_j)<0$ 且 $LPD^d(AS_i,AS_j)$ 极大时,说明$\langle AS_i,AS_j\rangle$链接出现的次数远远少于链接出现的平均次数,因此该链接极少地出现于路由器到国家 C_d 的路径中。

3.2.3 中间国家出现概率——ICAP

考虑 $CP(C_d)$,由式(7)可得从 M 到 C_d 变化路径集对应的自治域出现频度集 IA^d:

$$IA^d=\bigcup_{k=1}^{Z}\{N(AS_k)\} \tag{7}$$

对 IA^d 作进一步分析,统计获得从 M 到 C_d 的变化路径集中的各国家出现的频度(统计时应注意,若 $C(AS_i)=C(AS_j)$,将两者次数合并相加,即 $N(C(AS_i))=N(AS_i)+N(AS_j)$),构建向量 V_d,向量中的元素为国家 C_i 的出现次数,如式(8)所示:

$$V_d=\begin{Bmatrix}N(U_1^d)\\ \vdots\\ N(U_k^d)\\ \vdots\\ N(U_z^d)\end{Bmatrix} \tag{8}$$

其中,$U_k^d=C(AS_k)$并且$\forall x,y\in\{1,\cdots,z\},x\neq y,U_x^d\neq U_y^d$。

依据向量 V_d 定义 $CP(C_d)$ 中各中间国家 U_z^d 的出现概率 $ICAP(U_z^d)$,如式(9)所示:

$$ICAP(U_z^d)=\frac{N(U_z^d)}{\sum_{z=1}^{Z}\{N(U_z^d)\}} \tag{9}$$

显然,当 $ICAP(U_z^d)$ 极小时,表明中间国家 U_z^d 极少出现在 M 到国家 C_d 的变化集 $CP(C_d)$ 中。

3.2.4 中间国家地理偏离度——ICDP

对 $CP(C_d)$ 中的 AS 变化集 $NC^M(p,p')$ 而言,其集合中的各 AS 将不同程度地偏离 M 到国家 C_d 间的直线,同时也将对其相邻自治域地理位置间的直线造成偏离。由于各自治域隶属于不同的国家,因此参照自治域所处国家的地理位置,引入国家地理偏离度来衡量这种偏离的程度。

设 AS_i 为变化 AS,其所在的路由路径为 $AP^{M,T}$,路径中 AS_i 的相邻自治域分别为 $AS_{i-1}^{M,T}$,$AS_{i+1}^{M,T}$,引入式(10)度量 AS_i 所引入的中间国家偏离长度 $D(U_i)$:

$$D(U_i)=\frac{L(C(M),U_i)+L(U_i,U_{i-1})}{L(C(M),C_d)}+\frac{L(U_{i-1},U_i)+L(U_i,U_{i+1})}{L(U_{i-1},U_{i+1})} \tag{10}$$

其中,$L(C_x,C_y)$为国家 C_x 和国家 C_y 的地理距离,$C(X)$表示自治域(或路由器)X 所在的国家,U_{i-1}、U_i、U_{i+1}分别为自治域 $AS_{i-1}^{M,T}$、AS_i、$AS_{i+1}^{M,T}$所在的国家。$D(U_i)$表征了 AS_i 所处国家 U_i 引入的地理偏离长度,由两个部分组成:①U_i 对起始国家与目标国家造成的地理位置偏离长度,即 $C(M)\rightarrow U_i\rightarrow C_d$ 路径所引入的地理位置偏移;②U_i 对中间路径 $U_{i-1}\rightarrow U_i\rightarrow U_{i+1}$ 的地理位置偏移。

中间国家地理偏离长度的数值之间没有可比性,即无法衡量其对路径造成的偏离程度。考虑到路由器优先选择邻近路由器作为下一跳的性质,AS-PATH 中正常变化的 AS 不会对路径造成较大的偏离,相反虚假 AS-PATH 中变化的异常 AS 由于不遵循这一性质将会对路径产生较大的偏离。与 LPD 相似,为定位这类异常 AS,根据 Z-score 统计定义中间国家 U_i^d 的地理偏离度 $ICDP^d(U_i^d)$,如式(11)所示:

$$ICDP^d(U_i^d)=\frac{D(U_i^d)-E[D(U_i^d)]}{\sigma[D(U_i^d)]},U_i^d=C(AS_i),\ \forall AS_i\in CP(C_d) \tag{11}$$

其中,$E[D(U_i^d)]$、$\sigma[D(U_i^d)]$分别为以 C_d 为目的国家的所有中间国家地理偏离长度的均值和标准差。

显然，根据地理偏离度的定义，由 Z-Score 的性质可得如下结论：①当 $ICDP^d(U_i^d)<(>)0$ 时，相对于到达 C_d 的其他中间国家，国家 U_i^d 引入的地理偏离度小于(大于)中间国家引入的地理偏离度的均值；②当 $ICDP^d(U_i^d)>0$ 且 $|ICDP^d(U_i^d)|$ 极大时，中间国家 U_i^d 所引入的地理偏离度远远大于所有中间国家引入的平均地理偏离程度。

3.2.5 *路径级异常检测指标*

上文已经定义了基于目的国家 C_d 的各变化自治域的 AS 链接概率 LPD、中间国家出现频率 ICAP 以及中间国家地理偏离度 ICDP，同时，针对它们不同取值范围所代表的含义进行了分析描述，为检测各 AS-PATH 中存在的路径伪造异常行为，需要将这些指标转换到 AS 路径级。

本文认为路径伪造将导致 AS 路径的变化，其 AS 变化集 NC_j 的 AS 元素将引发极低的 ICAP、LPD 值以及极高的 ICDP 值。对应地，本文定义路径级异常检测指标，即最低国家出现概率 CAP、最低链接出现概率 CLP 以及最大地理偏离度 CGL，如式(12)、(13)、(14)所示：

$$CAP=\min\{ICAP(AS_f)\},\forall AS_f\in NC_j \tag{12}$$

$$CLP=\min\{LPD(AS_f,AS_i)\},\forall AS_f\in NC_j \tag{13}$$

$$CGL=\max\{ICDP^d(C(AS_f))\},\forall AS_f\in NC_j \tag{14}$$

其中，AS_i 代表在路径中与 AS_f 相邻的自治域。

3.3 路径伪造检测算法

3.3.1 *算法执行步骤*

基于上述指标，给出本文的路径伪造检测算法的主要步骤如下：

a)采集 BGP 路由器 M 相邻时刻的转发表 $T1\_Table$、$T2\_Table$，对 $T1\_Table$ 和 $T2\_Table$ 进行比较分析，得到所有 $NC^M(p,p')$。

b)分析 $NC^M(p,p')$的前缀地址，以其所属国家作为聚类的标准，将各个 $NC^M(p,p')$聚合到不同的国家类别 $CP(C_d)$ 中。

c)对属于国家类别 $CP(C_d)$下的各自治域 AS_i，依据定义分别计算其链接概率偏离度、其所在国家 $C(AS_i)$的中间国家出现概率以及中间国家地理偏离度。

d)针对每个隶属于国家类别 $CP(C_d)$的 $NC^M(p,p')$，分别计算路径 $AP^{M,p}$ 与 $AP^{M,p'}$ 的最低国家出现概率、最低链路出现概率以及最大地理偏离度。接着对路径进行分析，若路径级异常检测指标满足式(15)，则认为该路由存在路径伪造，标记该路由为路径伪造路由。

$$\begin{cases}CLP<\gamma\\ CAP\cdot\dfrac{1}{CGL}<\tau\end{cases} \tag{15}$$

其中 γ,τ 为事先设定的阈值，具体阈值的确定将在实验中阐述。

e)若已经完成对所有国家类别 $CP(C_i)$的检测，则检测过程结束，否则转 c)。

3.3.2 *算法修正*

上节的检测算法将 M 相邻时刻 AS 变化集中的所有 AS 元素作为分析对象，但是分析发现，处于顶层的 ISP 将会对检测结果造成严重的偏差。产生该结果的原因主要是顶层 ISP 的两个特点：①顶层 ISP 是互联网的核心，具有极强的连通性；②大部分 AS 都与其相连来实现网络流量的远距离传输。因此，AS-PATH 变化集中顶层 ISP 可能存在如下影响：①顶层 AS 可能具有较高的地理位置偏离值，可能会错误地将变化的顶层 AS 所在的路由记为恶意路由；②顶层 ISP 具有的强连通性将影响其他 AS 链接偏离度的计算结果。因此，为进一步提高算法的有效性，同时考虑到顶层 ISP 被攻击者利用来实施路由劫持，将路径伪造检测方法中 AS 变化集的顶层 ISP 去除，如式(16)所示：

$$NC^M(p,p')=NC^M(p,p')-Tier1\_ISP \tag{16}$$

其中，$Tier1\_ISP$ 代表位于层次 1 的 ISP。

4 实验和结果分析

为验证本文所提方法的有效性，选取 2011 年的 Russian 事件[11]作为实验对象，事件中，AS12182(美国)被报道通过路径伪造的攻击方式实现了对 AS31733(俄罗斯)的劫持，事件中共计有 5 个前缀被劫持，分别为 46.96.0.0/16、83.223.224.0/19、94.250.128.0/19、94.250.160.0/19、188.164.0.0/16。本实验以 RouteViews[12]的 linx 监测点为数据来源，分别以劫持前、劫持后的路由数据进行实验。

4.1 指标有效性分析

选取事件中的 83.223.224.0/19 以及 46.96.0.0/16 相关数据集来说明指标的有效性。由于事件中的路径伪造异常针对的目的国家为俄罗斯，计算 CP(RU)(其中 RU 是俄罗斯的缩写)的各变化路径中的指标值，获得 CP(RU)的 CLP-CGL 以及 CAP-CGL 的分布图，如图 3、图 4 所示(图中三角形标记的点代表存在路径伪造的路由，其余为存在 AS 路径变化但并未发生路径伪造的路由)。

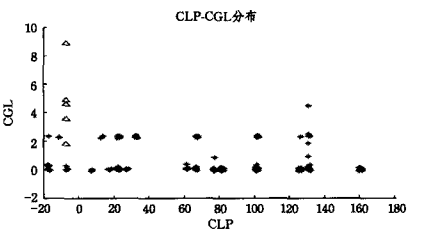

图 3 CLP-CGL 分布

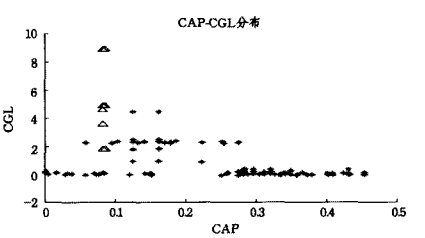

图 4 CAP-CGL 分布

从图中可以明显看出，存在路径伪造异常的路径大部分存在下述 3 个特征：①较大的 CGL 值，如图 3 所示，大部分存在异常的路径 CGL 值大于 3。②存在路径伪造的路由路径普遍具有较低的 CLP 值，在图 3 中存在异常的路径 CLP 值均为负数。究其原因，是由于路径伪造异常较难实现大规模爆发，因而在所有的 AS 变化中，该虚假 AS 链接出现概率较少，因此无法在极短的时间内传播到其他的路由中。③较低的中间国家出现概率 CAP，如图 4 所示，所有存在异常的路径 CAP 值均小于 0.1。综上所述，本文所提出的各项路径级

检测指标均能够标示存在路径伪造的路由。

然而进一步分析发现，部分正常路径也满足这 3 个特征中的 1 至 2 个，因而使用单一指标作为路径伪造检测的依据易将正常路径也归类为异常路由，为此需要综合分析这些指标。根据式(15)对 $CP(RU)$ 中各变化路径的路径伪造检测指标进行处理，得到图 5 所示的 CLP-CAP/CGL 分布(图中三角形标记的点代表存在路径伪造的路由)。

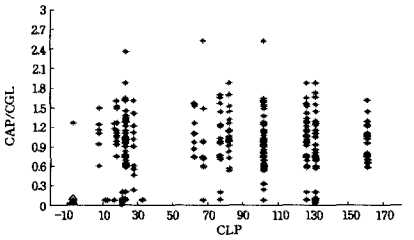

图 5　CLP-CAP/CGL 分布

从图 5 可见，对路径级异常检测指标按照式(15)操作后，存在路径伪造的路由的点均集中在图中的左下角，即存在路径伪造现象的路径信息同时具有极低的 CLP 值和 CAP/CGL 值，相反，正常的路由路径信息则不具备这两个特征。因此，本文所提出的检测方法能有效地从路由中提取出存在路径伪造的路由。

4.2　参数确定

选用 46.96.0.0/16 相关的路径伪造异常数据确定参数 γ,τ，图 6、图 7 分别显示了同一 τ、不同 γ 下检测结果的正确率曲线以及同一 γ、不同 τ 下检测结果的正确率曲线。

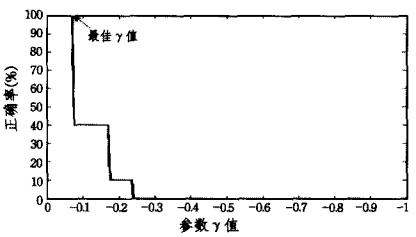

图 6　$\tau=0.55$ 时，正确率随 γ 的变化曲线

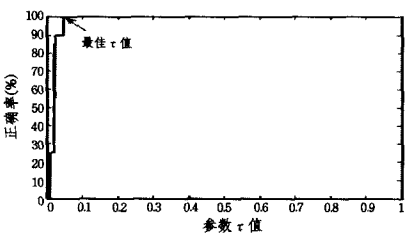

图 7　$\gamma=-0.07$ 时，正确率随 τ 的变化曲线

如图 6 所示，$\gamma<-0.07$ 时，算法的正确率随着 γ 的降低呈现递减的趋势，$\gamma\geq-0.07$ 时，路径异常检测的结果正确率均为 100%。虽然在这个范围内算法的检测正确率不变，但较高的 γ 可能会将转发表正常的路径项错误地判断为路径伪造异常路由，影响算法的检测结果；当 $\gamma<-0.07$ 时，将无法检测到存在的异常行为。因此，将算法中阈值 γ 的最优值定为 -0.07。同理，阈值 τ 取 0.047 为最优。

4.3　实验结果

依据上节所确定的 γ,τ 值，对 Russian 事件相关数据集进行检测，同时，将文献[9]提出的方法与本文提出的方法进行对比，将两者的检测结果按照被劫持前缀 83.223.224.0/19、94.250.128.0/19、94.250.160.0/19 以及 188.164.0.0/16 划分，实验结果如图 8 所示。

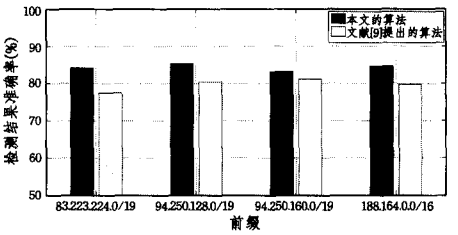

图 8　算法运行结果

从图中可以看出，本文提出的路径伪造检测方法较文献[9]所提的基于伙伴的检测方法具有更优的检测效果。其主要原因在于，基于伙伴的检测方法中将伙伴结点作为重要的检测依据，但是待检测前缀存在路径伪造时，部分伙伴已经被该异常路由所影响，因而与待检测前缀行为不一致的伙伴比例无法达到设定的报警阈值，导致无法检测出存在路径伪造的异常路由；而本文提出的方法是基于相邻时刻的路由变化集进行的操作分析，并不存在上述的问题，因此本文提出的算法具有一定的优越性。

5　相关研究比较

将本文提出的方法与前面提及的几种方法进行对比，分别在处理数据量、是否需要先验知识以及是否需要其他 BGP 转发表实现检测等方面进行比较。将需要处理的数据量分为适量、大量两个级别，两个级别的划分标准如表 1 所列。

表 1　数据量级别划分

级别	处理的数据
适量	单个路由器转发表数据或单个数据平面的探测数据
大量	多个路由器转发表数据或多个数据平面的探测数据

表 2 是本文方法与前面提及的 3 种检测方法的比较结果。

表 2　方法对比

检测方法	处理数据量	是否需要先验知识	是否需要其他 BGP 路由器
文献[7]	适量	是，实时拓扑信息	否
文献[8]	大量	是，指纹信息	否
文献[9]	大量	否	是
本文方法	适量	否	否

由表 2 可见，相比文献[7-9]中提出的检测方法，本文所提出的方法更为实用、可行，主要体现在以下几个方面：①由于处理的数据量较少，因而不会占用过多的内存，同时对 CPU 的要求不高；②不需要已有背景知识；③仅需要对单个 BGP 路由器中的转发表进行分析即可实现对路径伪造的检测。

结束语　本文针对路径伪造提出了一种基于聚类的路径伪造检测方法，该方法将 BGP 路由器相邻时刻转发表进行分析，得出 AS 路径变化集，接着依据其前缀地址所属国家不同聚类，对各类别中变化 AS 所在的路径得出最低国家出现概率、最低链路出现概率及最大地理偏离度，综合使用这 3 个指标对路径伪造进行检测。实验结果验证了所提出的方法的有

效性，与其他方法的比较也显示该方法具有处理数据量小、不需要先验知识以及不需要其他路由器信息的优点。

参 考 文 献

[1] Rekhter Y, Li T, Hares S. A Border Gateway Protocol 4(BGP-4)[EB/OL]. RFC4271. 2006

[2] 黎松，诸葛建伟，李星. BGP 安全研究[J]. 软件学报，2013，24(1)：121-138

[3] Brown M A. Pakistan hijacks YouTube. Renesys Blog [EB/OL]. http://www. renesys. c-om/blog/2008/02/pakistan-hijacks-youtube-1. shtml，2008

[4] Prefix hijacking by Michael Lindsay via Internap[EB/OL]. http://mailman. nanog. org/pipermail/nanog/2011-August/039379. html. 2011. 08

[5] Hiran R, Carlsson N, Gill P. Characterizin-g Large-scale Routing Anomalies A Case Study of the China Telecom Incident[C]// Passive and Active Measurement. 2012：229-238

[6] Hu X, Mao Z M. Accurate real-time identification of IP hijacking [C]// Proc. 07th Security and Privacy. Berkeley, CA, 2007：3-17

[7] Kruegel C, Mutz D, Robertson W, et al. Topology-based detection of anomalous BGP messages[C]// Proc. 6th Symp. Recent Advances in Intrusion Detection(RAID). 2007：17-35

[8] Hong S C, Hong J W K, Ju H. IP prefix hijacking detection using the collection of AS Characteristics[C]// Proc. 17th Network Operations and management symposium. Taipei, China, 2011：1-7

[9] Li J, Ehrenkranz T, Elliott P. Buddyguard：a buddy system for fast and reliable detection of IP prefix anomalies[C]// Proc. 20th IEEE International Conference(ICNP). 2012：1-10

[10] Zhao X, Pei D, Wang L, et al. An analysis of BGP multiple origin AS(MOAS) conflicts[C]// Proc. of the SIGCOMM Internet Measurement Workshop, 2001. San Francisco：ACM, 2001：31-35

[11] http://www. gossamer-threads. com/lists/nanog/users/144024 [OL]. 2011

[12] Route Views Project Page[OL]. http://www. routeviews. org. 2005

(上接第 157 页)

值的不足，通过对读写器增加两个不同的计数器，统计碰撞和空闲的时隙个数，差值运算后比较门限阈值，从而动态调整 Q 值的大小。该算法避免了读写器计算浮点的运算，同时将碰撞空闲时隙单独考虑，提高了系统效率和识别速度。仿真结果显示，与 EPC-C1G2 的 QA 算法相比，QA-DTCI 算法在不损耗系统吞吐率的情况下，提高了标签的读取速度，缩短了识别时延。在标签量为 1000、初始 Q 为 4 时，QA-DTCI 算法比 QA 算法在标签的读取速度上提升了 10%，在识别时延上降低了 4%，使系统的吞吐率维持在 34.8%左右。后续工作是优化 K_c 和 K_i 门限阈值，在保证当前碰撞时隙消耗的时间不变的情况下，继续提高系统的吞吐率，同时降低空闲时隙消耗的时间。

参 考 文 献

[1] Finkenzeller K. RFID Handbook：Radio-frequency Identification Fundamentals and Applications (Second Edition)[M]. England：John Wiley and Sons，2003

[2] Jihoon M, Wonjun L, Srivastava J. Adaptive binary splitting for efficient RFID tag anti-collision[J]. IEEE Communications Letters，2006，10(3)：144-146

[3] Lai Yuan-cheng, Lin C-C. A pair-resolution blocking algorithm on adaptive binary splitting for RFID tag identification[J]. IEEE Communications Letters，2008，12(6)：432-434

[4] Jihoon M, Wonjun L. Adaptive binary splitting：a RFID tag collision arbitration protocol for tag identification[C]// 2nd International Conference on Broadband Networks. Boston, United States，2005，1：347-355

[5] Law C, Lee K, Kai-Yeung S. Effcient memoryless protocol for tag identication [C] // Proceedings of the 4th In-ternational Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications. Boston, USA：ACM，2000：75-84

[6] Hwang T-W, Lee B-G, Kim Y-S. Improved anti-collision scheme for high speed identification in RFID system[C]// First International Conference on Innovative Computing, Information and Control. Beijing, China，2006，2：449-452

[7] Schoute F C. Dynamic frame length aloha [J]. IEEE Transactions on Communications，1983，31(4)：565-568

[8] Cha J R, Kim J H. Novel anti-collision algorithms for fast object identification in RFID system[C]// Proceedings of the 11th International Conference on Parallel and Distributed Systems. Washington D. C. , USA：IEEE，2005：63-67

[9] Lee S R, Joo S D, Lee C W. An enhanced dynamic framed ALOHA algorithm for RFID tag identification[C]// Proceedings of the 2nd Annual International Conference on Mobileand Ubiquitous Systems：Networking and Services. Washington D. C. , USA：IEEE，2005：166-174

[10] Chen W T. An accurate tag estimate method for improvingthe performance of an RFID anticollision algorithm based on dynamic frame length ALOHA[J]. IEEE Transactions onAutomation Science and Engineering，2009，6(1)：9-15

[11] EPCglobal. EPC radio-frequency identity protocols class-1 generation-2 UHF RFID protocol for communications at 860MHz-960MHz(version 1. 2. 0)[S]. 2008

[12] Lee D, Kim K, Lee Won-jun. Q+-Algorithm：An Enhanced RFID Tag Collision Arbitration Algorithm[J]. Ubiquitous Intelligence and Computing：Lecture Notes in Computer Science，2007，4611/2007：23-32

[13] Maguire Y, Pappu R. An optimal Q-algorithm for the ISO 18000-6C RFID protocol[J]. IEEE Trans. Automation Science and Engineering，2009，6(1)：16-24

[14] Fan X, Song I, Chang K. Gen2-based hybrid tag anti-collision Q algorithm using Chebyshev's inequa lity for passive RFID systems[C]// IEEE 19th International Symposium on Personal, Indoor, and Mobile Radio Communications. 2008：1-5