◎网络、通信、安全◎

基于路由更新的BGP路由稳定性监测工具设计

李 玮[1,2],张大方[1,2],张元媛[1],曾 彬[2]
LI Wei[1,2],ZHANG Da-fang[1,2],ZHANG Yuan-yuan[1],ZENG Bin[2]

1.湖南大学 软件学院,长沙 410082
2.湖南大学 计算机与通信学院,长沙 410082
1.Software School,Hunan University,Changsha 410082,China
2.School of Computer and Communication,Hunan University,Changsha 410082,China
E-mail:rj_wli@hnu.cn

LI Wei,ZHANG Da-fang,ZHANG Yuan-yuan,et al.Design of BGP routing stability monitoring tool based on routing updates.Computer Engineering and Applications,2010,46(8):75-77.

Abstract: Currently most of the methods and tools of monitoring inter-domain routing stability available are from the macro point of view and the parameter is simple,some of the metrics only count the number of updates which can not offer particular information for eliminating routing faults.The paper designs and implements an effective monitoring tool for inter-domain routing stability,which analyzes the basic point based on binary group of (peer,prefix) or prefix according to routing updates and can reflect the status of inter-domain routing stability.The tool is good to meet requirements by applying the data collected from practice routing environment.

Key words: active measurement;route monitoring;Border Gateway Protocol(BGP);routing stability

摘 要: 目前域间路由稳定性监测分析方法和工具大多从宏观角度出发,反映路由稳定性状况的参数指标单一,有的甚至仅对路由更新信息进行了简单的分类统计,不能为故障排查提供更为详尽的依据。设计并实现了一种有效的域间路由稳定性监测工具,在更新消息层面上针对(对等体、前缀)二元组或前缀为基点展开分析,多方面反映了域间路由稳定状况。通过应用实际的相关路由信息进行分析,验证了工具的有效性。

关键词: 主动测量;路由监测;边界网关协议(BGP);路由稳定性
DOI: 10.3778/j.issn.1002-8331.2010.08.021 **文章编号:** 1002-8331(2010)08-0075-03 **文献标识码:** A **中图分类号:** TP393

路由系统是网络互联的核心部分，控制着网络流量的传输,路由不稳定将导致网络性能下降,如丢包率上升、包乱序增多、网络连接不稳定、网络不可达、网络延时,严重时会导致一定程度的网络连接中断。目前BGP[1]是实际网络中大多应用的域间路由协议。BGP路由不稳定[2]主要指BGP路由表中路由的不断更新,即已经存在于路由表中的路由消失或被新的路由取代,或有新的路由加入路由表。一旦某些BGP路由属性发生改变、网络不可达或有新的网络被通告,路由更新消息就会在网上传播,这就意味着BGP路由不稳定事件发生。研究[2]表明BGP不稳定的原因有很多,主要包括:链路或路由器失败与撤销、路由软件实现上的问题、路由协议配置错误、路由器存储能力和CPU处理能力缺乏。

为了提升网络路由系统的稳定性能,有必要对网络路由性能指标进行有效的监测,给网络管理者提供及时、有效的路由状况监测指标数据。网络路由监测技术的研究将使得网络维护者能够及时发现网络中存在的不稳定状况并做出合适的处理,从而全面提升网络的稳定性[3-5]。

目前域间路由稳定性监测分析方法和工具大多从宏观角度出发,反映网路稳定性状况的参数指标单一,有的甚至仅对路由更新信息进行了简单的分类统计,不能为故障排查提供更为详尽的依据,因此在相关研究的基础上设计并实现了一种有效的域间路由稳定性监测工具,在更新消息层面上针对(对等体、前缀)二元组或前缀为基点展开分析,从多方面反映了域间路由的局部和整体路由状况。

基金项目: 国家自然科学基金重大研究项目(the Grand National Natural Science Foundation of China under Grant No.90718008);国家自然科学基金(the National Natural Science Foundation of China under Grant No.60673155,No.60703097)。

作者简介: 李玮(1972-),男,博士生,讲师,主要研究方向为计算机网络,生物计算;张大方(1959-),男,博士,教授,博士生导师,主要研究领域为可信系统与网络、容错计算、网络测试、软件容错、软件测试、网络安全、软件工程,电子政务;张元媛(1982-),女,硕士,主要研究领域为网络测试;曾彬(1979-),男,博士,主要研究方向为网络性能监测。

收稿日期: 2009-11-04 **修回日期:** 2009-12-22

1 相关工作

早期的相关研究大多从微观角度出发,例如在研究路由收敛[6]对稳定性影响时引入了最短路由宣告间隔时间 MARI、链路延迟 LD、撤销使用限制率 WRATE 等,并由此衍生出 M_{up}、M_{down}、M_t、$P_{\min}$、$P_{\max}$ 等参数指标。文献[7]对上述参数进行了拓展,产生了 $MRAI_{long}$、$MARI_{short}$、MIN_PROC_TIME、MAX_PROC_TIME、t_p 长度等;文献[8]在对路由不稳定性进行研究时,仅对 BGP 的路由宣告和路由撤销信息进行了简单的统计,计算了路由宣告信息在路由表中所占有的比例,如以 prefix 和 AS 为单位简单测算了路由宣告信息累计比例;并将路由事件进行了简单的定义和统计,分析了事件在时间轴上数量分布特性。第一,分析可以更好地了解这些特性怎样影响端到端通信;第二,分析可以影响到路由协议的设计和路由分布机制以及路由器的设置和应用;第三,更好地了解路由系统可以提高网络拓扑、端到端延迟以及丢包特征的建模。目前的分析或测量工具功能单一缺乏全局性,为了能多角度地反映出域间路由的稳定状况[9]。其中指标监测指标单一、路由时间间隔过长、监测层面较少、所需信息量大或存储需要的资源过多等都是这些工具的缺陷。

2 监测指标

2.1 指标制定原则

通过收集和观察当前的路由更新信息来评估 BGP 的运行状态,该层面的分析可以展示特定前缀 prefix 和特定对等体 peer 的具体路由状况。对等体针对特定前缀发布的 BGP 更新信息中,大部分都能快速显示出路由习性。因此基于 BGP 更新消息分析路由的不稳定性,可以考察某个前缀到对等体的路由更新情况。把 BGP 更新消息按(对等体,前缀)二元组分组,每组的信息按时间先后顺序进行排列,记为(*peer*,*prefix*)的 *BGP* 时间序列。定义 $G(peer,prefix)$代表以 *peer* 为基点的某项域间路由监测指标集,其中 $peer \in$ 对等体集合,$prefix \in$ 到达 *peer* 的前缀集合。从二元组的角度出发,可以考察路由事件、路由平均持续时间、路由有效时间、前缀可达时间、前缀持续时间等一系列的路由指标。

二元组的分析是以对等体为出发点,考察该对等体到达指定前缀的路由状况,该角度限定只能从某一个对等体来监测,如果对等体的数量较多时进行逐一考察是不切实际的。除了关注对等体外,还需要把握更多的指定前缀的不稳定性状况。同理也可以前缀为出发点,考察前缀到达目的地的路由状况。定义 $G(prefix)$代表以 *prefix* 为基点的某项域间路由监测指标集,其中 $prefix \in$ 指定前缀集合。

为了能够较好地对域间路由状况进行监测和分析,分析周期的选取至关重要,合理地选取周期 T 将有利于捕捉路由不稳定状况。选择的周期 T 太长,大大超过了路由波动持续的时间;选择的周期 T 过短,过度密集的采集数据将给数据处理带来较大的不便,导致实时效果不理想。考虑到监测指标的特性,对局部路由指标选取 30 秒为周期,对整体路由指标选取 30 分为周期。

2.2 主要测量指标

(1)路由平均有效时间

BGP 的路由更新信息中,AS_PATH 属性的 AS 序列表示某个对等体 *peer* 到达某个前缀 *prefix* 的一条 AS 级的路径 R,记为(*peer*,*prefix*)的路由 R。如图 1 所示,被宣告的路由 R 的生存周期为从 R 被宣告时刻起到其被显示或隐式地撤销止,即路由 R 在时间轴上连续出现时间记为 K_m,其中 m 代表该路由 R 第 m 次出现。若 K_m 在各个时间周期 T 中出现一次,即认为路由 K_m 在周期 T 内有效。若路由 K 在时间 T 内出现 n 次,称 R 在 T 内的路由平均持续时间,记为 $Savg(R)=\sum_{m=1}^{n} K_m/n$。

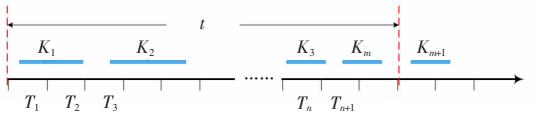

图 1 路由时间刻度表

一条路由的平均有效时间的长短可以反映出路由的不稳定性,如果平均路由时间很短且该路由多次被选为主路由,则需要增加对其的关注。

(2)前缀热度

若 AS 之间的路由不稳定,将导致大量的路由更新消息产生。这些不稳定的 *prefix* 在网络中无疑会给路由造成不良影响,及时发现这些无用且过于活跃的 *prefix* 将是前缀热度指标要反映的问题。

在图 2 中,$N(d,p)$为在时间 T 内 *prefix* P 出现次数记录,将 *prefix* 在某段时间内的活跃程度称为前缀热度,记为 $A(d,p)$,取值如下所示:

$$A(d,p)=\begin{cases}0, N(d,p)<T_u\\ 1, N(d,p)\geqslant T_u\end{cases}, T_u \text{ 为阈值}$$

活跃前缀生存期为:

$$C(p)=\sum_{d=0}^{n-1} A(d,p), n \text{ 为观测周期数}$$

文献[9]通过对连续 3 年的数据分析研究表明 T_u 取值为 46 时,可以捕获前 1%的活跃前缀,T_u 取值可以根据实际需要适当地放大或缩小。病态路由事件、蠕虫攻击等均有可能引发前缀过热,前缀热度指标可以在一定程度上反映路由不稳定性。

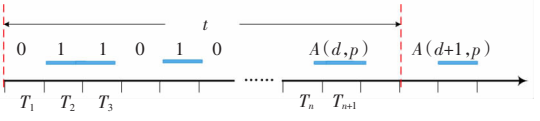

图 2 前缀热度时间刻度图

(3)前缀可达时间

在一个二元组中从 *peer* 到 *prefix* 存在着一条或多条不同的路由,可以以前缀为基点,考察 *peer* 到前缀的可达状况。如图 3 所示,路由 R_m 为在时间 t 内出现的从 *peer* 到 *prefix* 路由,每条路由在周期 T 内出现一次即被视为在 T 内可达。设 $Route=\{r|r\in R_m, R_m$ 为时间 T 内 *peer* 到 *prefix* 的有效路由$\}$,其在 T 内前缀的可达时间记为 $SP(R)=\sum_{i=1}^{m} S(R_i)$;若在 T 内共有 n 条路由出现,则该前缀的平均可达时间记为 $SP_{avg}=\sum_{i=1}^{m} S(R_i)/n$。

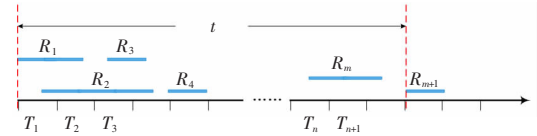

图 3 前缀可达时刻图

一般而言只要有路由存在则该前缀是可达的。与路由有效时间不同的是该指标不只是考虑某条路由 R 的状况,而是从

整体出发监测路由的可达时间。若可达时间过短则反映出该前缀的可达途径有异常；可达时间越长则说明该前缀的路由较为稳定，可达持久度高。因此，前缀可达指标可以从路由的可达时间情况来反映路由的不稳定状况。

(4)平均路由长度

如图 4 所示，设 $h(R)=\{h_1,h_2,\cdots,h_n\}$表示路由 R 经过的路由对等体个数。由指标 1 可知 $S(R)$为路由 R 在时间 T 内的路由有效时间，该指标越长路由被采用的概率越大，则可设 $S(R)/T$ 为权值。在$(peer,prefix)$时间 T 内所有到达指定对等体路由的平均路由长度指标为 $H_{avg}=\sum_{i=0}^{n}\frac{S(R_i)}{T}h(R_i)$。通常情况下路径长度越长不稳定的概率越大。

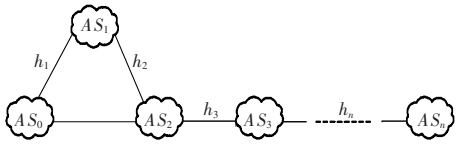

图 4　路径长度示意图

3　测量工具的设计

模型结构如图 5 所示，数据采集采用对等体方式，即在实际的网络中插入一个接收路由更新信息的对等体，可在该对等体内形成路由表且能接收大量的路由更新信息。模型读取 BGP 路由原始数据后，以二元组$(peer,prefix)$和 $prefix$ 为基点，分别从局部和整体两个层面来分析路由信息。局部指标模块包括路由有效时间等；整体指标模块包括前缀热度、前缀可达时间、平均路由距离等。其分析处理的数据可以输出、保存、查看。

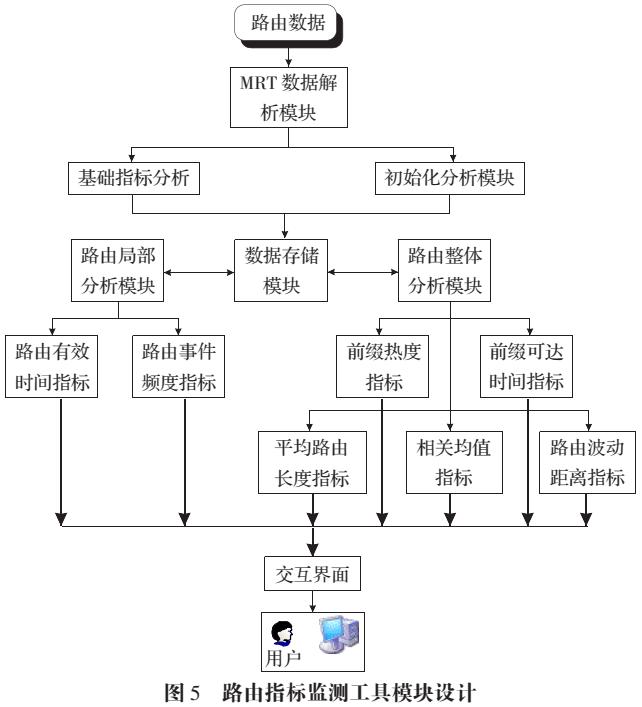

图 5　路由指标监测工具模块设计

路由指标监测工具包括如下主要模块：

(1)MRT 数据解析模块：为了便于路由软件的研究有统一的数据格式标准，IETF 定义了 MRT 格式的数据，该模块可将 MRT 格式的“0”、“1”数据转化为容易读懂的文本数据。

(2)基础指标分析模块：该模块可提供基本指标的分析、查询等功能，基础指标有 updates 信息统计、peer 信息统计、prefix 信息统计等简单数据。

(3)初始化分析模块：实现对用于查询数据的哈希表进行初始化，为进一步的分析奠定基础。

(4)数据存储模块：该模块可以将各个指标分析后的结果存储，可供用户查询、导出等。

(5)路由局部分析模块：该模块实现了局部路由指标的分析，包括路由有效时间指标分析、路由事件频度分析等。

(6)路由整体分析模块：该模块实现了整体路由指标的分析，包括前缀热度指标分析、前缀可达时间指标分析、平均路由距离指标分析、路由波动距离指标分析、相关均值指标分析等。

界面是用户与工具后台的中介，通常包括配置选择、结果显示、用户管理、日志记录等，部分界面如图 6 所示。

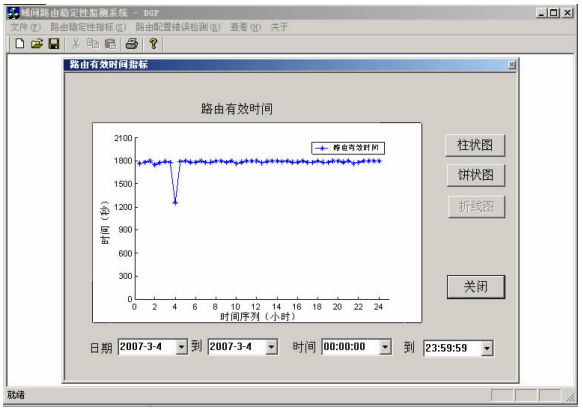

图 6　工具模型主界面

4　数据分析应用

该文实际选取了路由信息采集器 rrc00.ripe.net，AS12654(IP：193.0.4.28)对等体与 14 个核心网络骨干路由器建立了连接。选取了从 2007 年 1 月 1 日 00：00 至 2007 年 1 月 2 日00：00 期间的路由数据。应用工具所提供的指标分析实际路由状况。

首先选取了 AS4608 与前缀 59.37.2.0/23 的路由状况为对象，考察了二元组(202.12.29.64，59.37.2.0/23)其中一条路由 R(4608　1221　4637　4134　4809)在 24 小时内的通达状况，如图 7 所示虽在 4 时处有较大的波动但不足以中断该路由，其他时间基本保持在满值附近，说明该路由保持较好的通畅。

选取 194.42.208.0/20、59.37.20.0/23、58.49.108.0/24 三个较为活跃的前缀作为考察对象。图 8 中蓝、绿两组数据较为平缓，可反映出这两个前缀的活动比较均匀，而 194.42.208.0/20 的线条起伏比较的大。大部分的路由事件都是由一小部分 $prefix$ 引起的，例如受到蠕虫攻击时 BGP 经受着大量的路由更新被注入因特网，因此为了减少该前缀所受影响，可以结合其他指标进行综合分析。

在域间路由更新信息中包含的前缀比较多，可以根据需要选取特定前缀进行考察。文中任意选取了 194.42.208.0/20 和 203.135.27.0/24 进行分析，如图 9 所示这两个前缀的可达程度很高，基本在 900 s 以上。这两个前缀具有较多的姊妹前缀，因此可达时间已经可以满足路由需要。

5　总结

和域内路由有所不同，域间路由涉及的域较多，转发的路

(下转 136 页)