

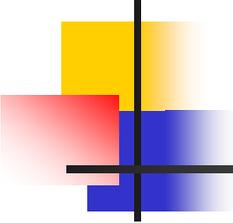


# IP地址地理定位算法及其有效性分析

杨云 龚俭

(东南大学计算机科学与工程学院, 南京, 211189)

2015-11-25



# 提纲

---

- 研究概述
- IP地址地理定位算法原理介绍
- 通过IP地址地理定位算法验证IPCIS现有的IP地址使用位置库的准确性
- 总结

# 研究概述

与利用GPS进行精确定位不同

## ■ 研究背景

### 1. IP地址地理定位概念

IP 地址地理定位是指在只知道用户或设备的 IP 地址的条件下，定位网络中用户和设备的实际使用位置信息，其中实际使用位置包括 IP 地址所在的国家、省（区域）、城市信息。

### 2. IPCIS

IPCIS是为CERNET主干网运行管理和安全保障系统提供 IP 地址综合信息的分布式系统

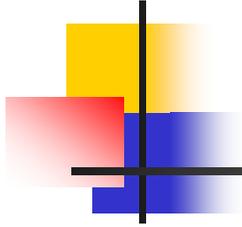
### 3. IPCIS的IP地址使用位置库

数据来源

数据规模

Maxmind IP2Location QQ纯真 CERNET whois

3,479,773个IP地址段  
IP地址个数 4,148,313,893

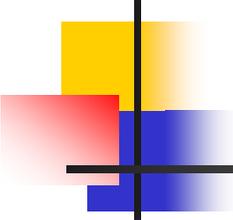


## ■ 研究背景

- IPCIS目前的IP地址使用位置库存在IP地址段定位信息不准确的情况，需要设计IP地址地理定位算法对其准确性进行验证

## ■ 研究内容

- 提出有效的IP地址地理定位算法
- 通过IP地址地理定位算法验证IPCIS现有的IP地址使用位置库的准确性



## IP地址地理定位算法原理介绍

### ■ 实验数据介绍

#### ➤ Traceroute拓扑路径信息

CAIDA 是最重要的互联网测量数据提供者之一，traceroute拓扑路径信息是通过利用路由追踪技术主动获取网络路由信息得到的。

$[IP_1, IP_2, IP_3, IP_4, IP_5, \dots, IP_{n-1}, IP_n]$

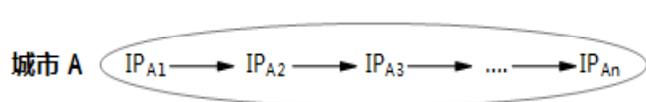
$IP_1$ 与 $IP_n$ 之间的路由线路

IP地址之间的通信时延

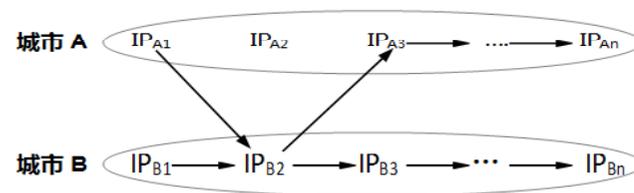
# 多跳IP地址关联分析算法

## 算法原理

定义序列 $A = [IP_1, IP_2, \dots, IP_{n-1}, IP_n]$  ( $n \geq 3$ ) 为一条traceroute路径的连续多跳IP地址序列(一条完整trace数据或者Trace数据的一部分), 若该序列的首尾IP地址 ( $IP_1, IP_n$ ) 的使用位置相同, 则trace序列首尾IP1和IPn之间的IP地址的使用位置均相同, 我们称序列A之为相关IP地址序列。



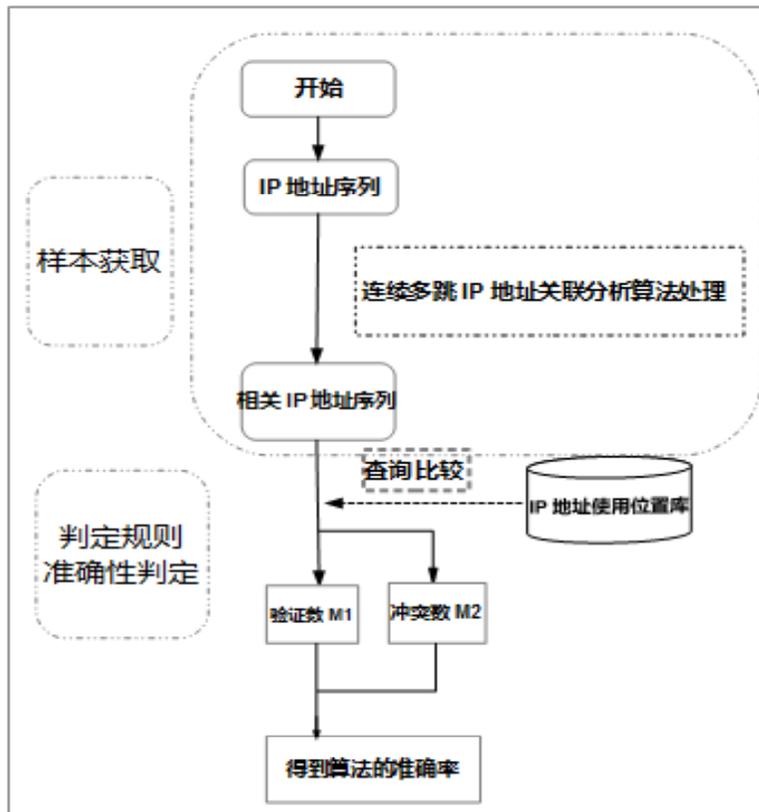
图一：相关IP地址序列



图二：通信时选择非最佳路由

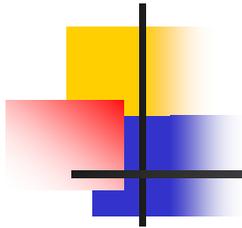
# 利用多跳IP地址关联分析算法验证IPCIS现有的IP地址使用位置库的准确性

## ■ 验证流程



(1).验证数：数据库中有M1个IP地址的使用位置验证为正确

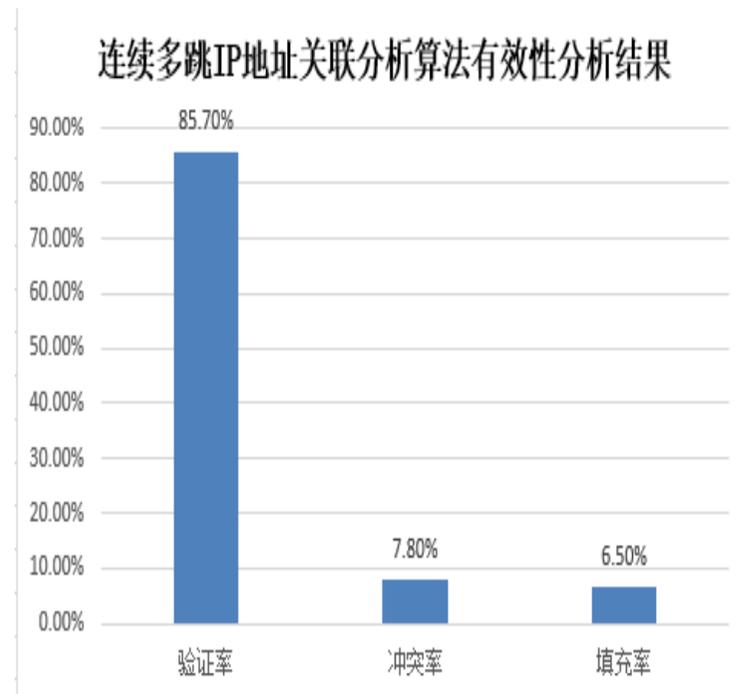
(2).冲突数：不能判断数据库中的使用位置信息准确与否



## ■ 实验结果

实验数据：  
10个cycle的traceroute路径信息

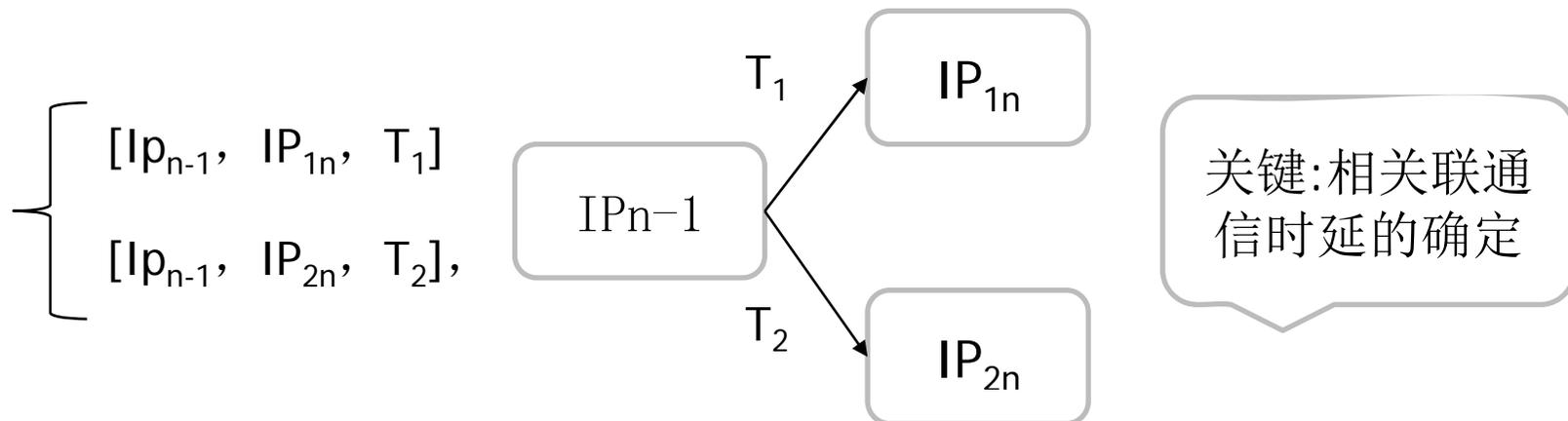
处理10个cycle的traceroute数据  
验证的IP地址数量为36035490  
个约占地址库数据总量的  
0.869%



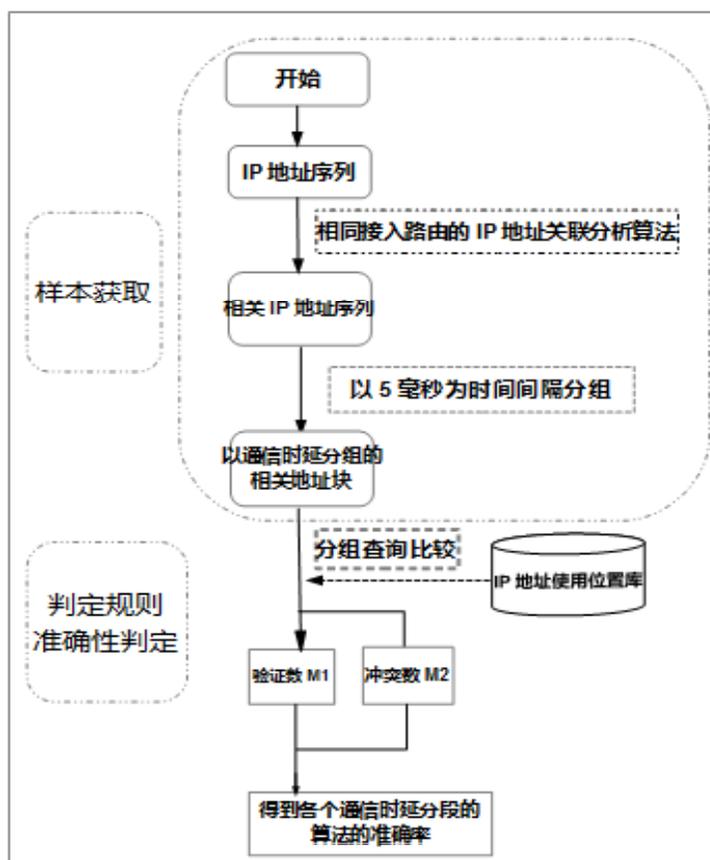
# 相同路由接入的IP地址关联分析算法

## 算法原理

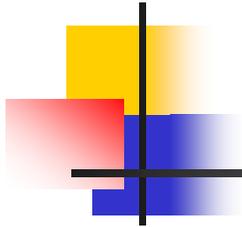
一般情况下，具有相同接入路由的多个主机 IP 地址的使用位置相同，此结论要求观察对象必须为主机 IP 地址，而 trace 中的 IP 地址不一定为主机 IP 地址，绝大多数是路由器 IP，因此，考虑增加通信时延作为约束，扩展该结论至全体 IP 地址，即具有相同上一跳 IP 地址且与上跳 IP 之间的通信时延小于相关联通信时延阈值  $T_m$  的多个 IP 地址的使用位置相同。



# 利用相同接入路由的IP地址地理定位算法验证IPCIS现有的IP地址使用位置库的准确性

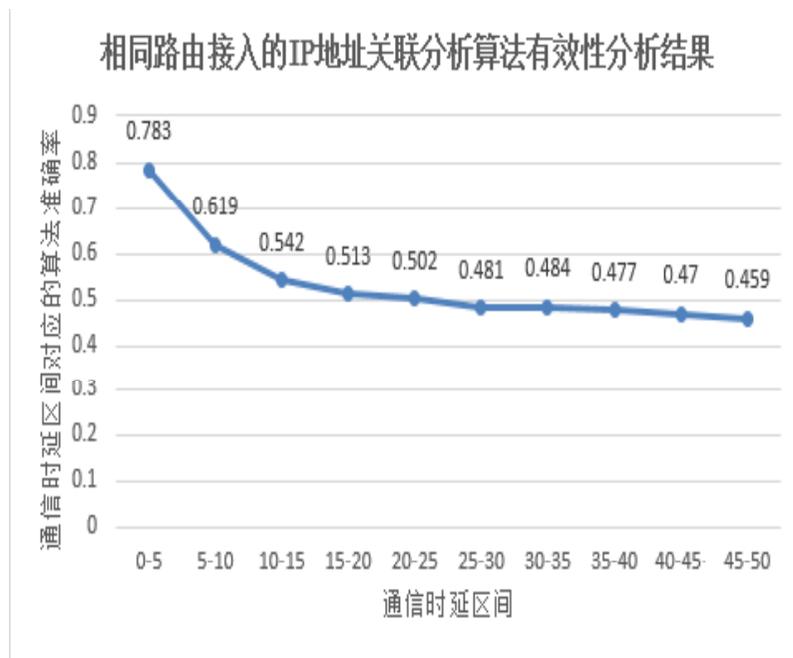


1. 确定通信时延阈值 $T_m$
2. 使用算法对数据库进行准确率验证

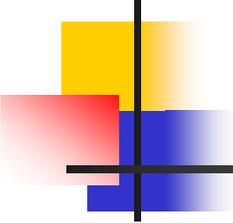


## ■ 实验结果

实验数据：十个cycle的traceroute数据



以5毫秒为通信时延阈值：  
处理十个cycle的trcaceroute数据后验证IPCIS现有的IP地址使用位置库中的IP地址数量为1475614222个，约占数据库IP总数的35.571%



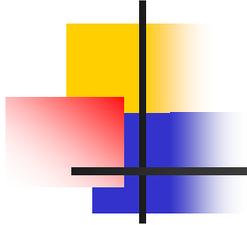
# 总结

---

- 总结:

本文基于traceroute数据设计了两种IP地址地理定位算法，用于对IPCIS现有的IP地址地理定位库的准确性进行验证，这个实验工作表明这个数据库的有效性还是相当高的，随机选10个Cycle的traceroute数据进行验证，验证率达80%左右不足：

对于算法得出的定位结果与IPCIS的数据库中的定位结果相冲突的情况，本文没有方法判断数据库中定位信息正确与否



The End  
&Thanks