互联网拓扑知识
在流量优化和网络抗毁中的应用

中国科学院计算技术研究所

报告人：张国清

gqzhang@ict.ac.cn

中科院计算所

INSTITUTE OF COMPUTING TECHNOLOGY

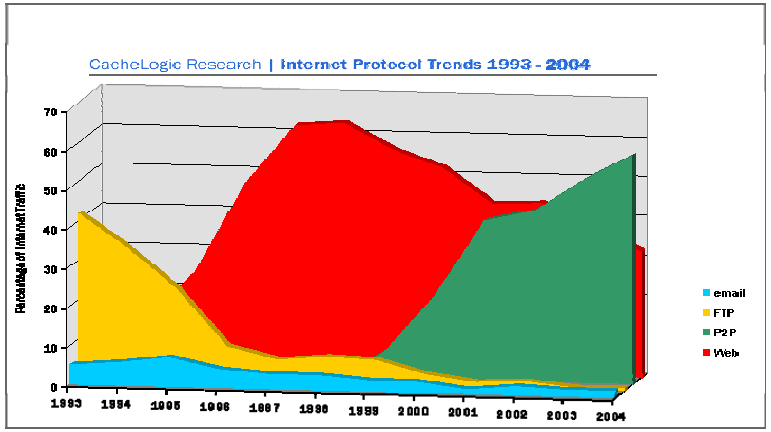
第一部分：流量优化

- 背景-**P2P流量优化的动机**
- 我们的工作-**P2P流量优化**
  - P2P网络拓扑的优化
  - P2P内容分发的优化
  - 实验结果与分析
  - 接入网的P2P传输优化

} **提高P2P流量的局部性**

背景

- P2P消耗了互联网的大部分带宽
  - 互联网60%以上的流量来自P2P程序



http://www.cachelogic.com

- P2P引起的问题
  - 加大了网络的压力
  - 增长了网络的经营开支
  - 降低了互联网上其它应用的性能

- P2P造成网络过载的主要原因
  - P2P常常对底层网络状态不太关心
  - 在Peer关系建立上比较随意，导致P2P覆盖网拓扑与底层网络拓扑的不匹配
  - 内容分发没有考虑网络的拓扑信息和ISP的流量优化策略
  - 在域间和域内产生了很多不必要的往返流量，降低了网络效率

  ☆统计发现**BT用户中50%~90%的数据块从网外下载。**
- ISP与P2P运营商之间的“战争”
  - 限制与反限制，容易导致“双输”的局面

- ISP、P2P内容提供商、终端用户之间的和谐需要共赢
  - ISP：节约网络资源和提高网络效率
  - P2P：提高吞吐量和降低时延；降低被ISP限速和封杀的危险
  - 终端用户：性能和费用

  ☆要达到三方共赢必须优化P2P流量。
- P2P流量优化
  - 利用网络信息影响P2P系统，优化P2P流量。获取网络信息的方式有两种：

    1)用测量或推算方式获得网络信息;

    2)ISP主动提供网络信息。比较准确。

我们的工作

提高P2P流量的局部性(locality)

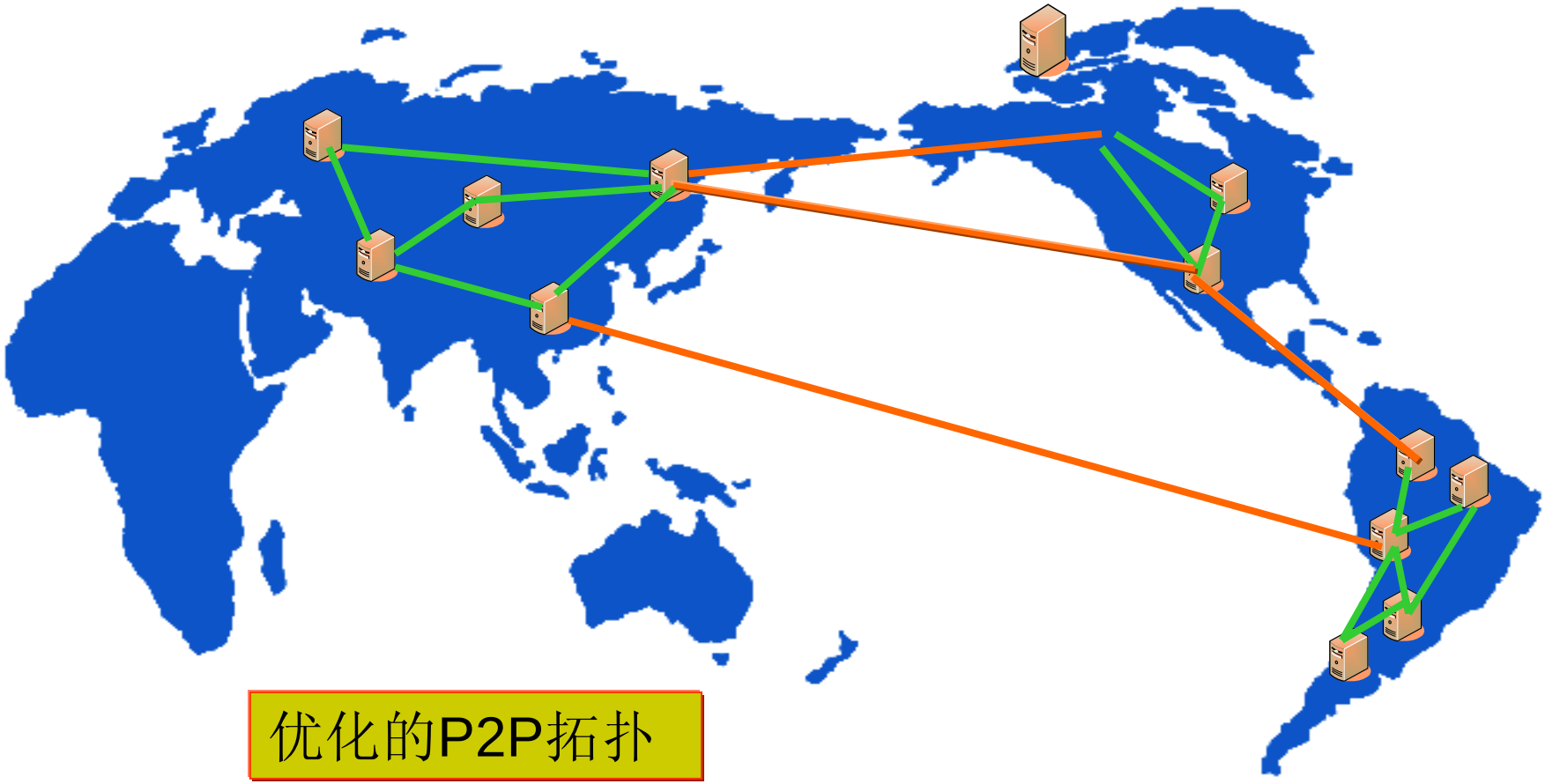
- 利用网络拓扑信息优化P2P
  - 网络拓扑信息获取
  - 优化P2P网络拓扑
  - 优化P2P内容分发
    - 数据调度算法
    - 缓存置换算法
  - 实验结果与分析
- 接入网的Peer-to-Peer传输优化

网络拓扑信息获取

- 网络拓扑信息
  - IP到IP前缀的映射
  - IP前缀到AS的映射
  - 互联网AS (自治系统)级拓扑
- 获取方式
  - **WHOIS方式**：如使用pWhoIs.org提供的查询服务可获得IP的前缀和所属AS等详细信息
  - **BGP方式**：处理BGP路由表或BGP更新消息，提取IP前缀和AS拓扑，如RouteViews项目
  - **Traceroute方式**：通过Traceroute对网络主动测量，获得路由器级或AS级拓扑，如CAIDA的Skitter项目。我们用这种方式多次测量了中国大陆的**AS级拓扑**

P2P覆盖网拓扑的优化

提高P2P网络拓扑的局部性是优化P2P流量的第一步



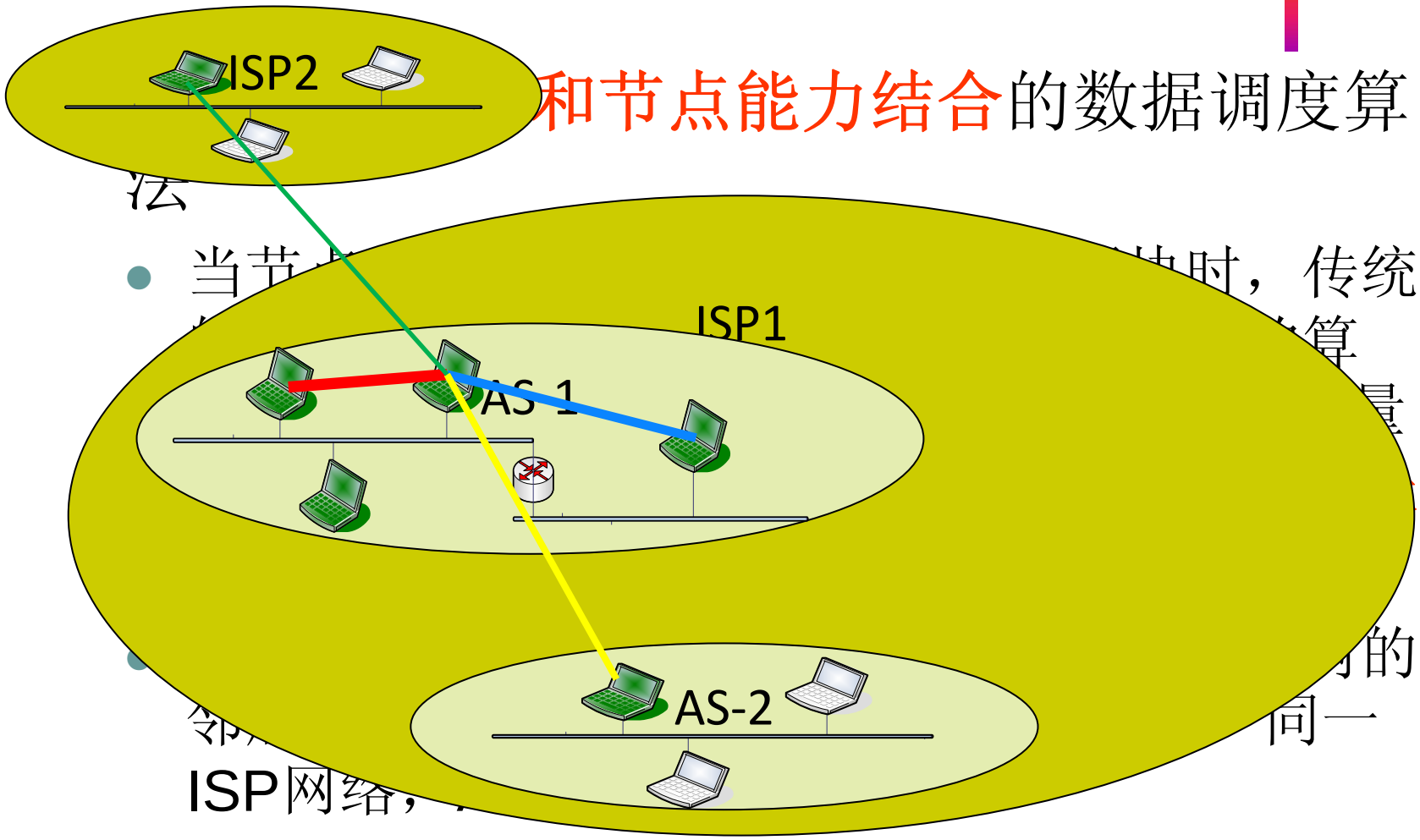
优化的P2P拓扑

内容分发策略的优化

提高内容访问的局部性

- 数据调度算法
- P2P缓存置换算法

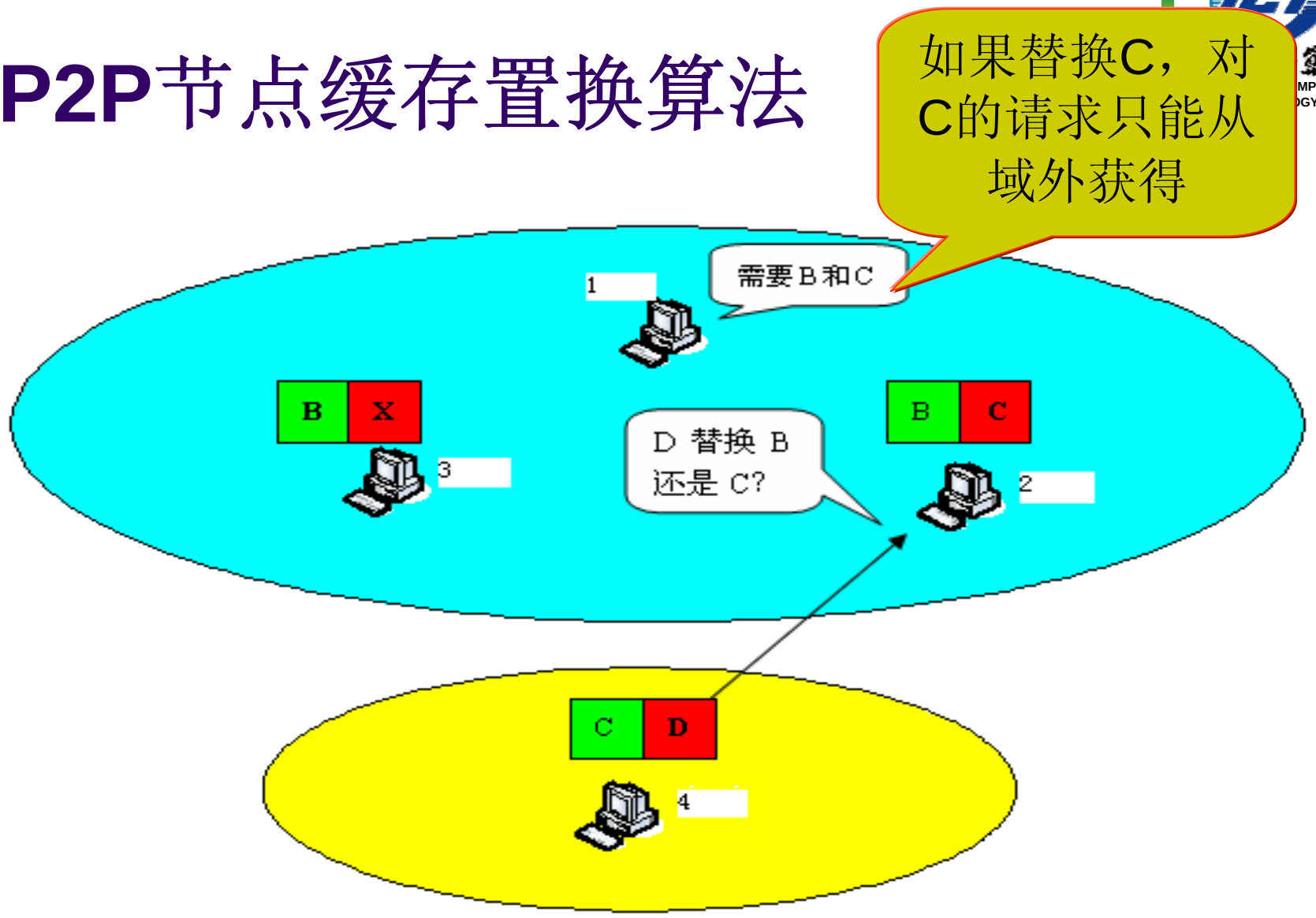
数据调度算法



和节点能力结合的数据调度算法

- 当节点[illegible]时，传统[illegible]算[illegible]量[illegible]
- [illegible]的[illegible]同一ISP网络，[illegible]

P2P节点缓存置换算法

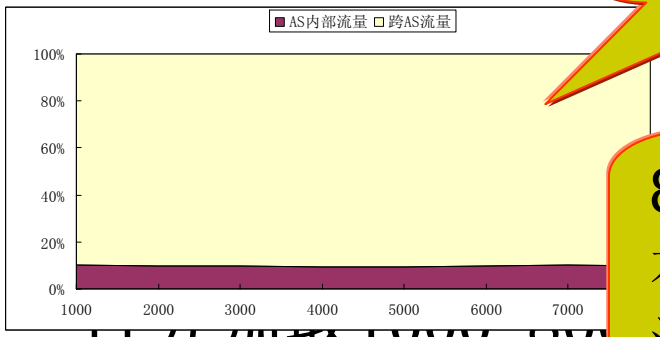


仿真实验

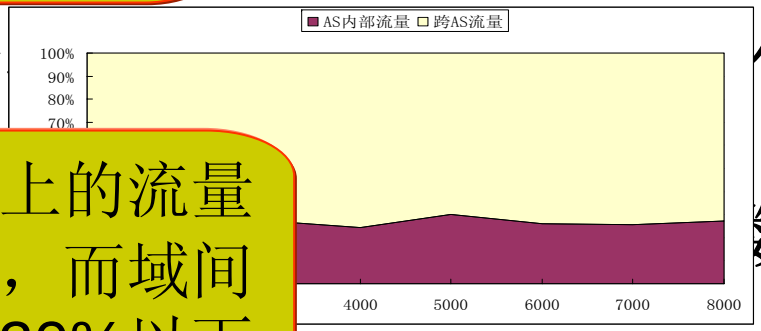
- **实验1：**以P2P视频直播系统作为测试对象，比较不同邻居分配策略和数据调度算法下的效果

	节点选择	数据调度
方法一	Random 策略	带宽优先策略
方法二	Random 策略	邻居能力结合节点间拓扑关系策略
方法三	AS 优先策略	带宽优先策略
方法四	AS 优先策略	邻居能力结合节点间拓扑关系策略

10%左右的流量在域内，90%左右的流量在域间

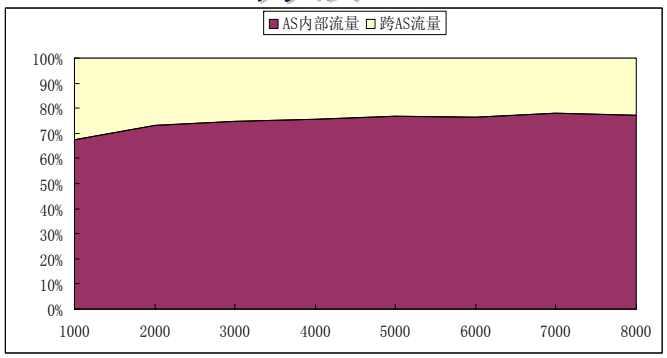


方法一

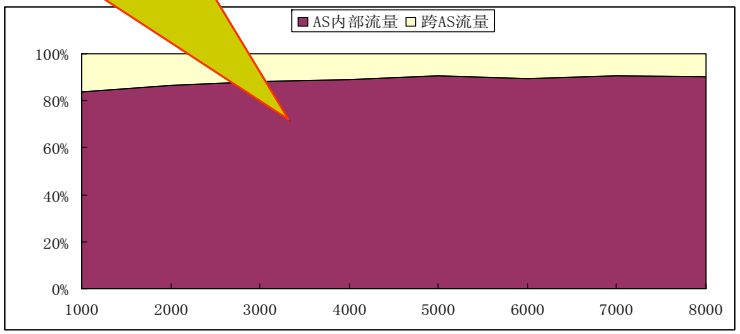


方法二

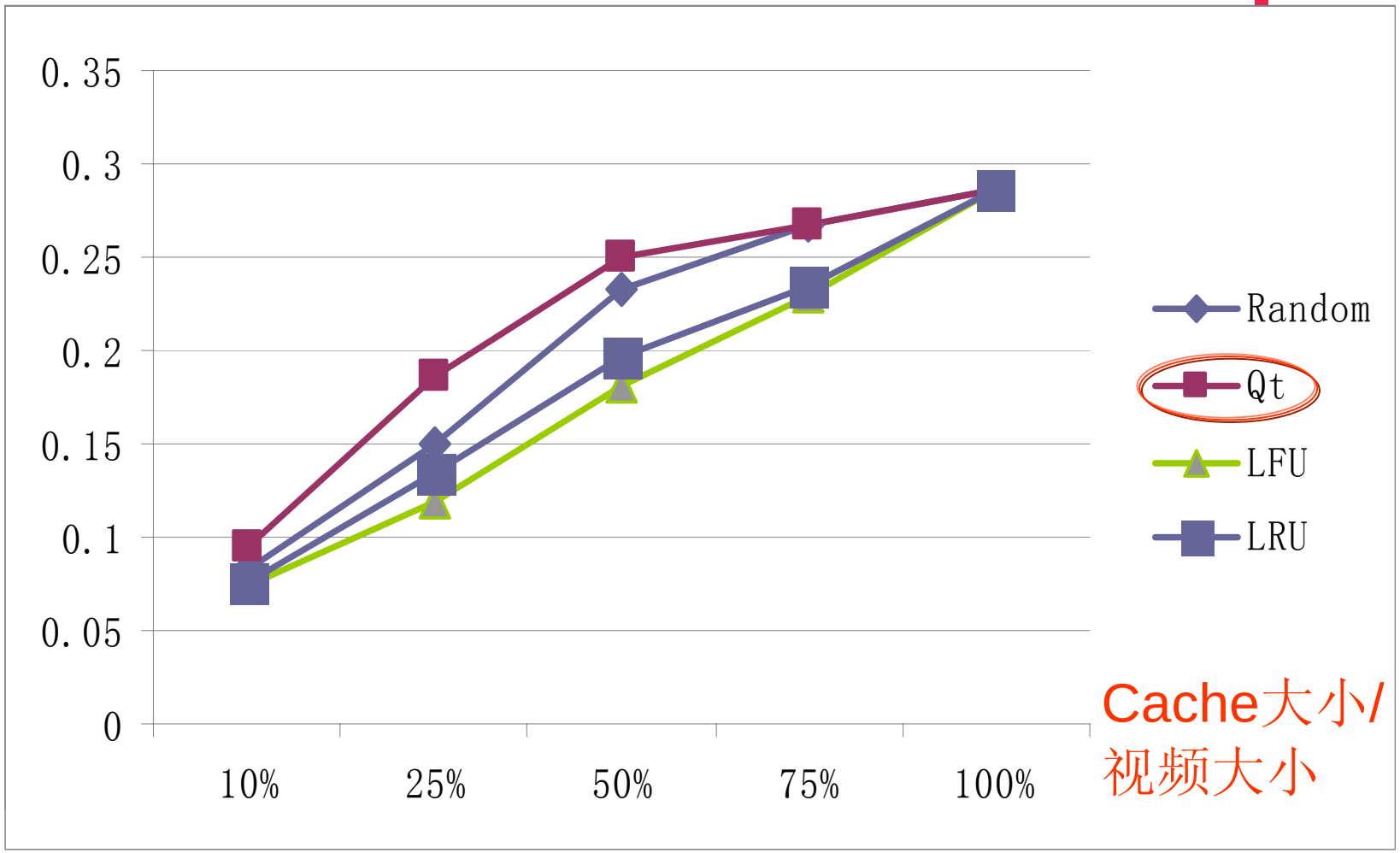
80%以上的流量在域内，而域间流量在20%以下



方法三



方法四



域从内源节点下载的数据所占流量的比例

实验结果评价

- 利用网络拓扑信息优化P2P的拓扑和内容分发，可以大幅减少P2P流量，尤其是网间流量
- 进一步改进
  - 仿真实验跟真实网络上的实验有一定差异
  - 获取的网络拓扑信息不一定完整和准确
  - 没有考虑利用ISP网络的其它信息，如网络使用策略等

我们的工作

- 利用网络拓扑信息优化P2P
  - 网络拓扑信息获取
  - 优化P2P网络拓扑
  - 优化P2P内容分发
    - 数据调度算法
    - 缓存置换算法
  - 实验结果与分析
- 接入网的Peer-to-Peer传输优化

接入网络对P2P传输的影响

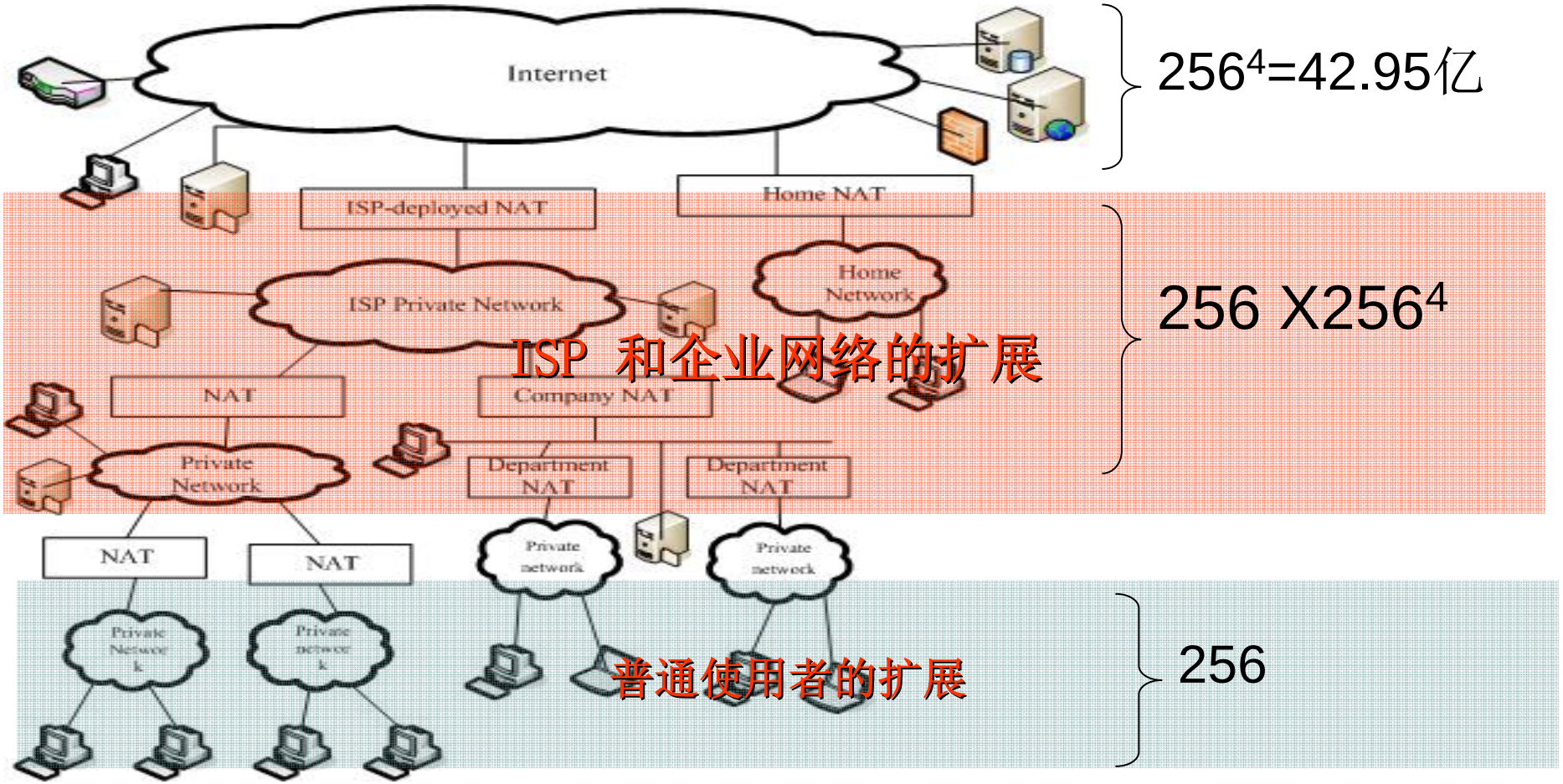
- 由于Firewall、NAT等设备的存在造成了peer与peer间数据传输的不对称性
- 由于Firewall、NAT等设备的级联造成了peer与peer间寻址的不确定性
- 接入对象的安全需求导致接入网络具有封闭性
  - 例如：企业、政府、小区、商业中心等网络

NAT网络中P2P传输优化方案

- 单层NAT网络的P2P传输优化方案
  - 通过ping的方式测试连通性
- 多层NAT网络的P2P传输优化方案
  - NAT的级联是一种不可忽视的网络扩展方式
  - IETF协议的空白
  - 不能进行连通性测试
  - 需要将单层NAT测试和多层NAT测试统一

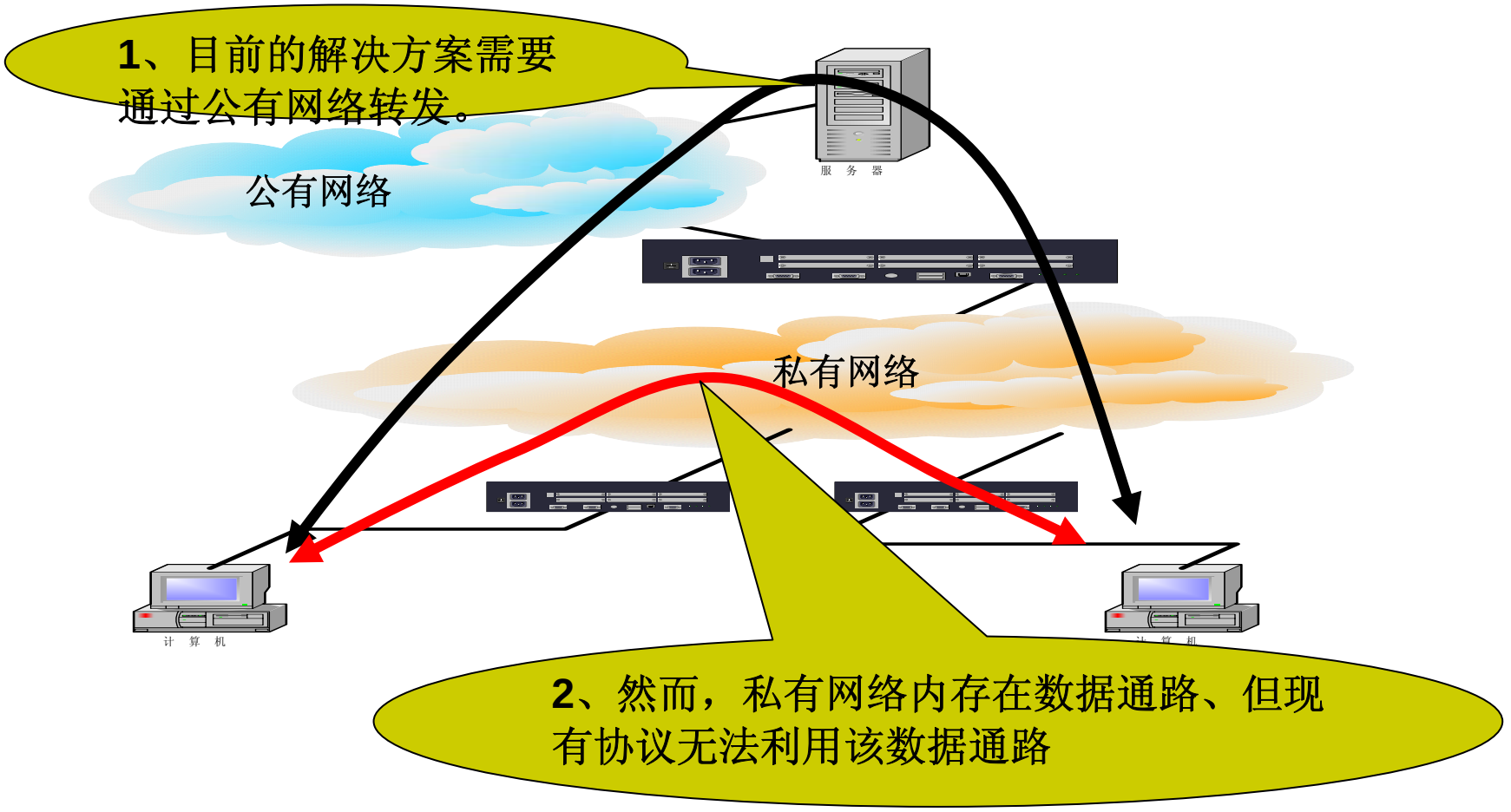
多层NAT网络环境下P2P传输优化的必要性

存在大量隐藏的接入者，提供了相对集中的存储资源



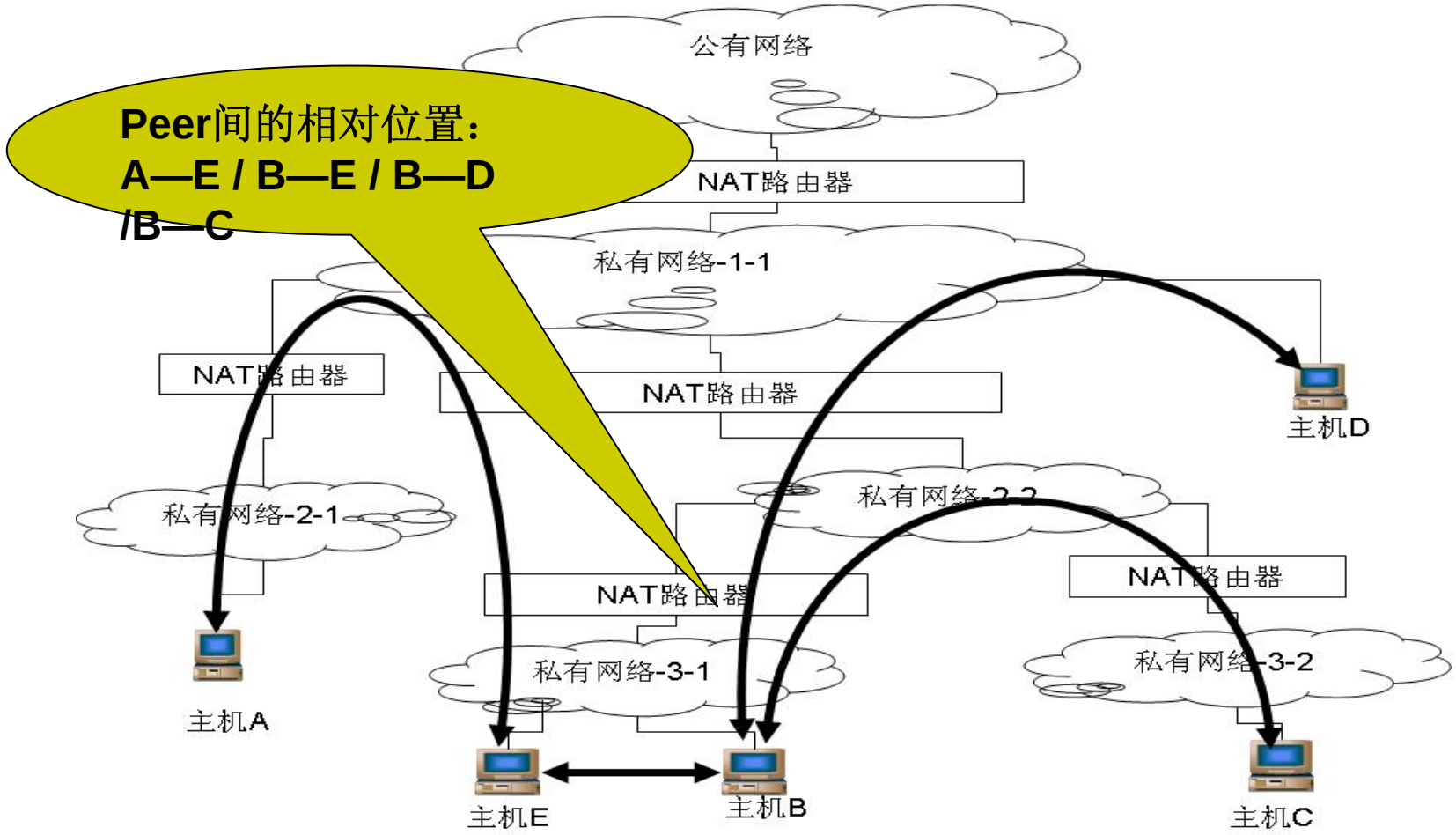
多层NAT网络环境下P2P传输优化的必要性

存在数据传输通道，提高数据传输质量

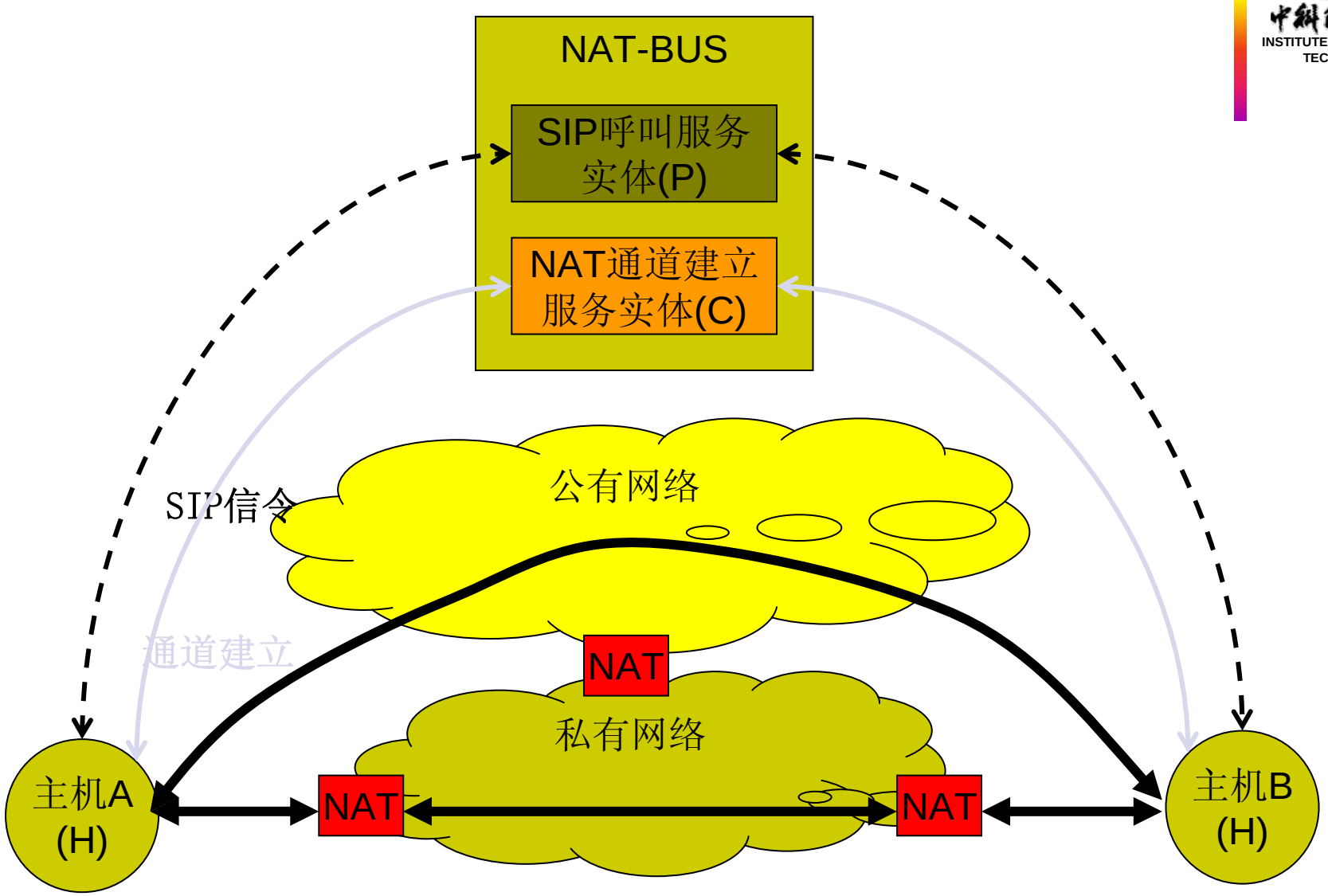


多层NAT网络环境下P2P传输优化的必要性

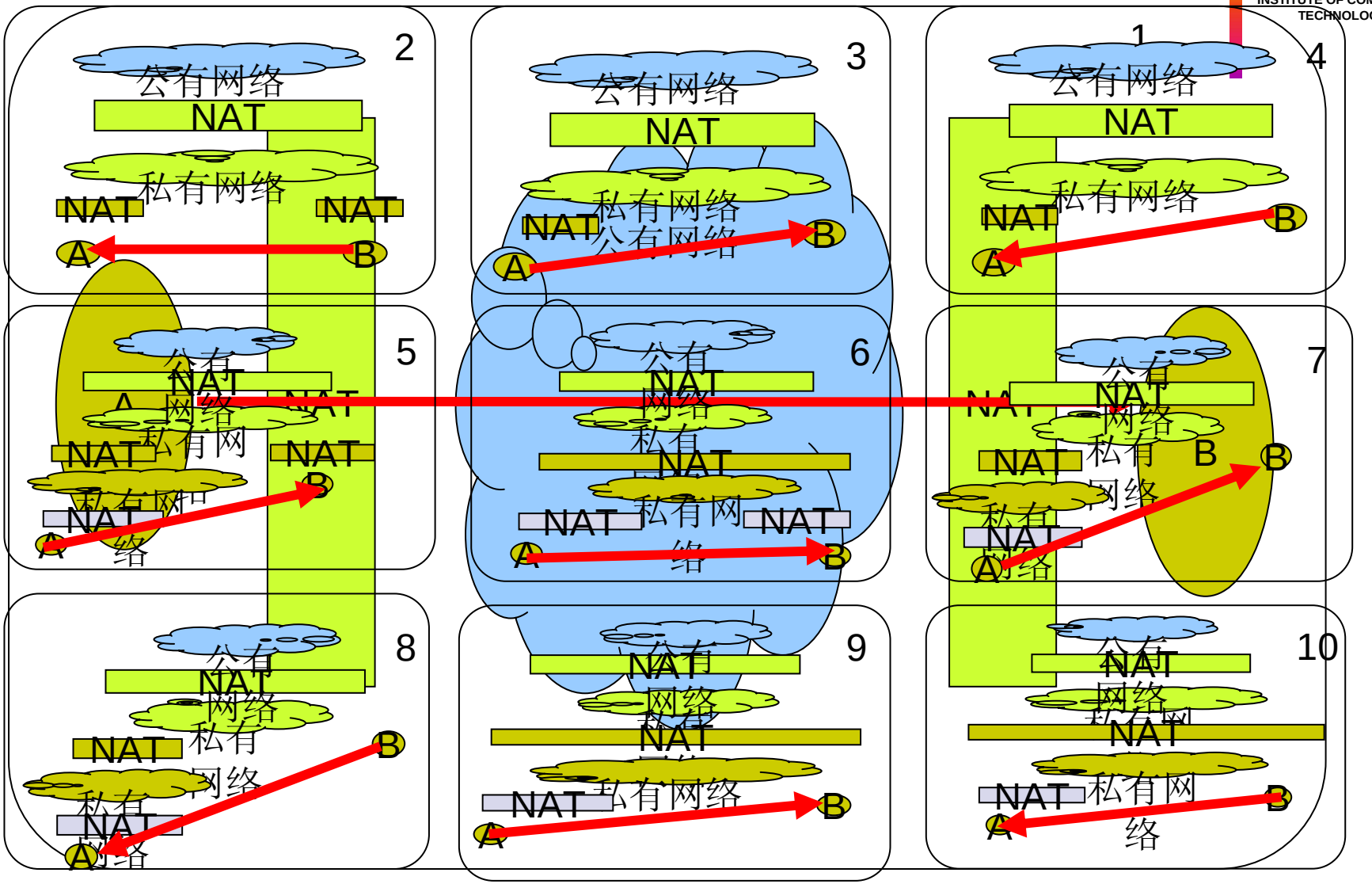
Peer间的相对位置复杂，需要多方协作



NAT-BUS系统概念模型

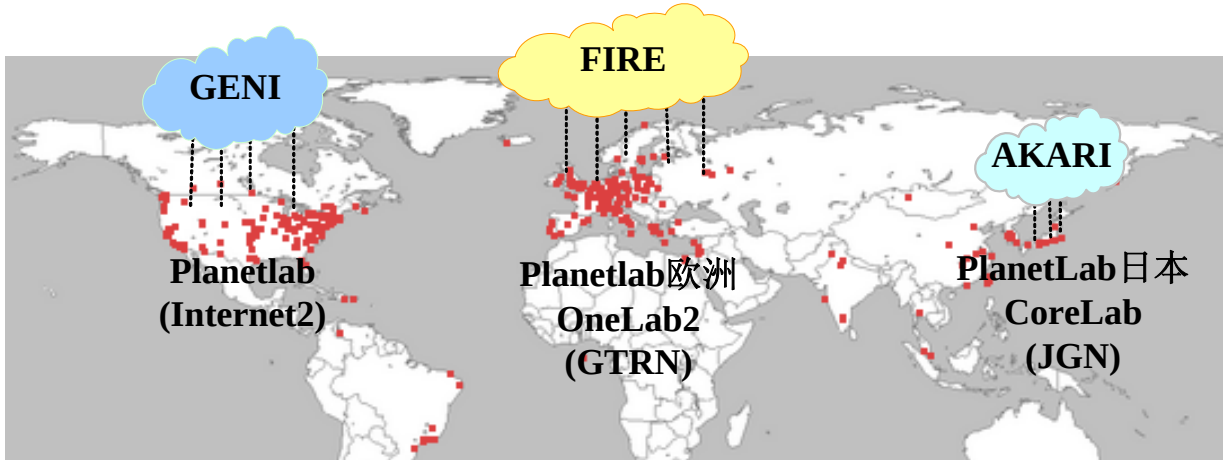


私有网络穿越的典型场景



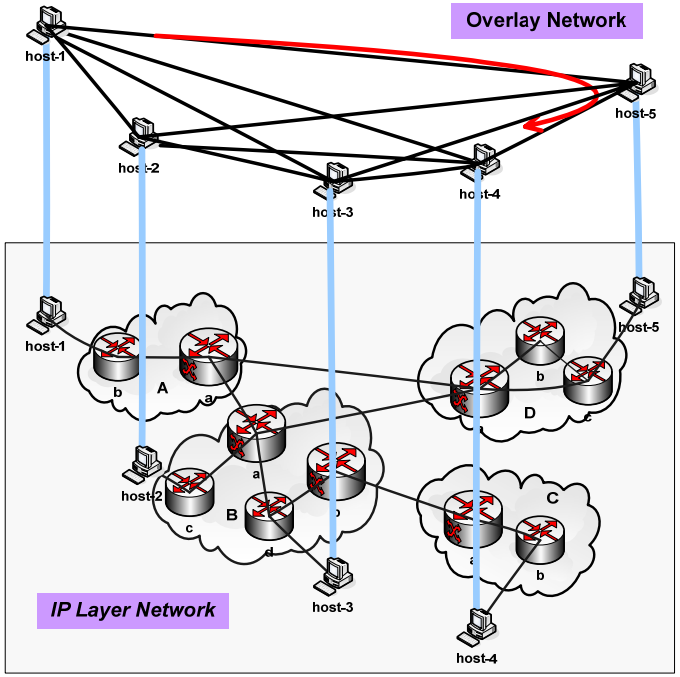
第二部分：网络抗毁

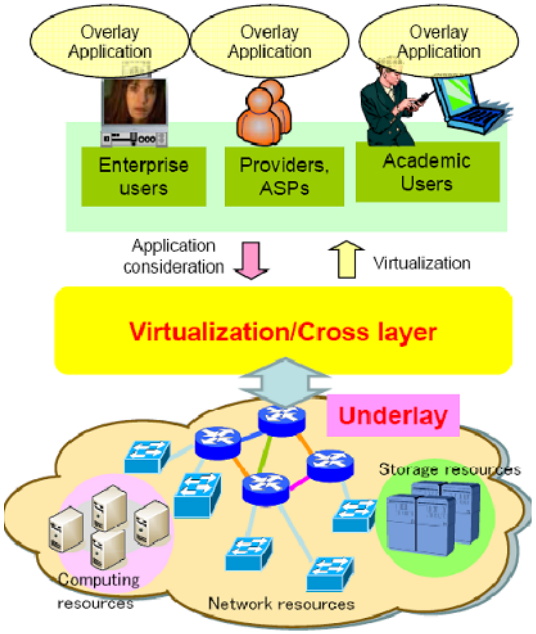
网络路由优化解决方案

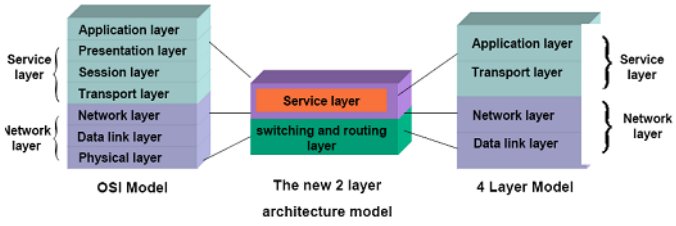


Clean-Slate?

GENI、FIRE、AKARI



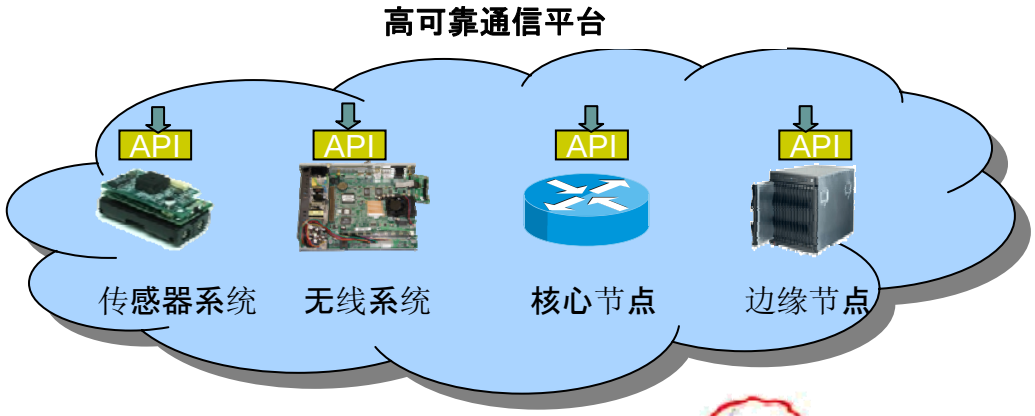


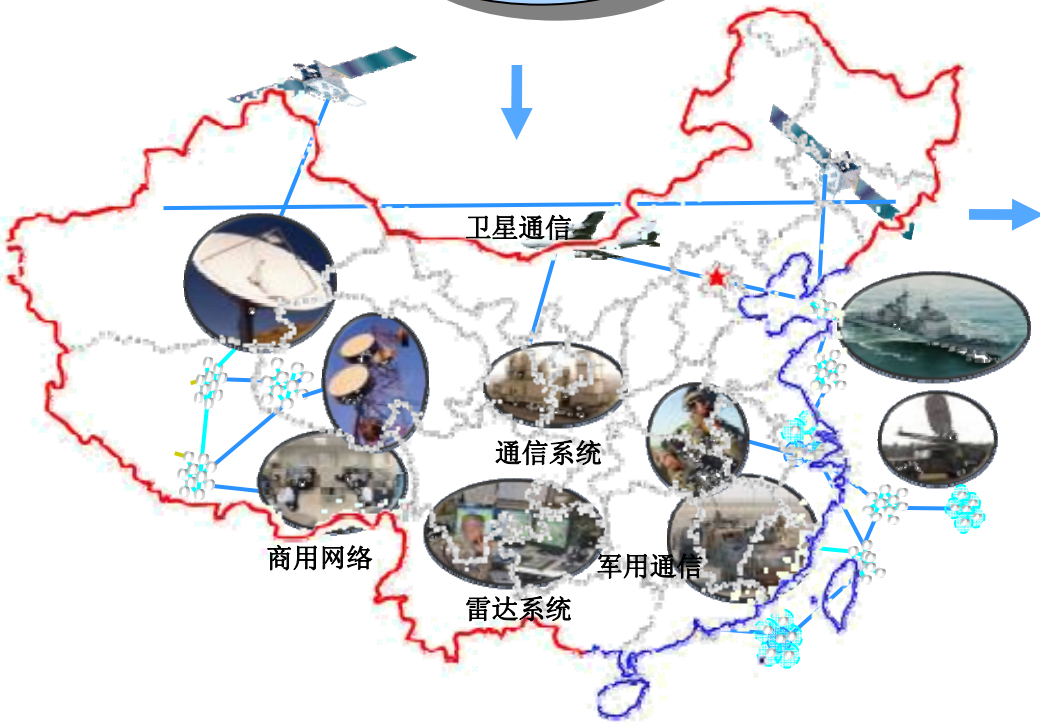


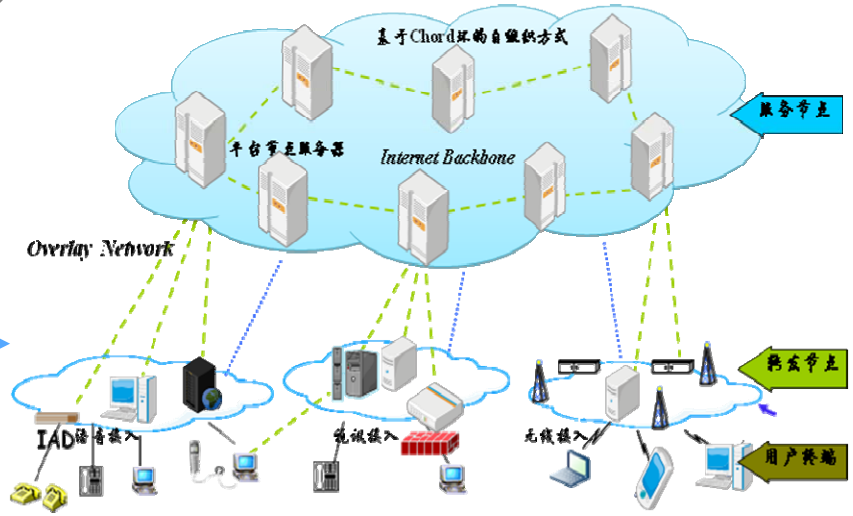
Clean-Slate+Overlay!

ITU—T的未来网络架构

基于底层信息感知的抗毁覆盖网实现



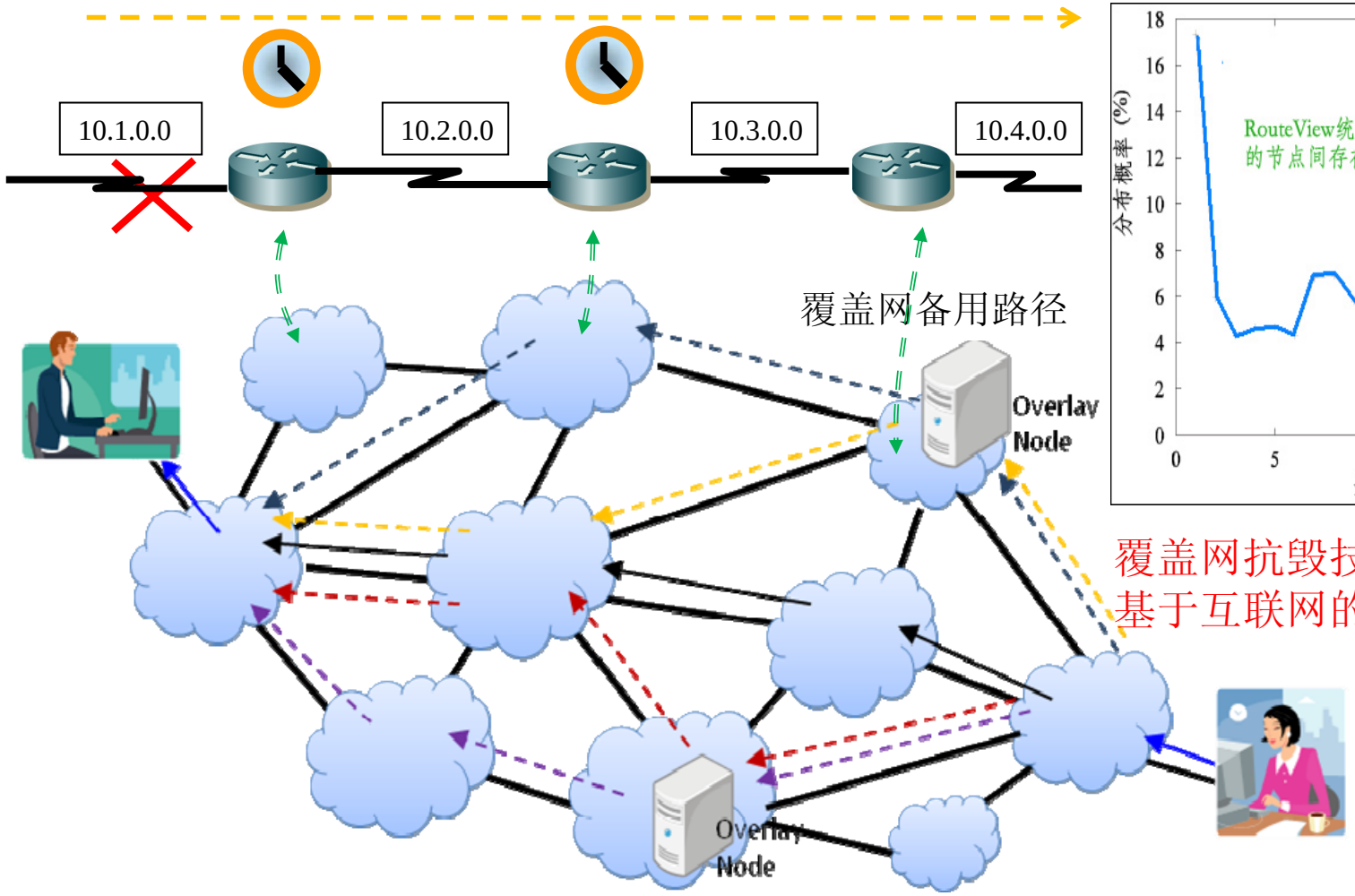


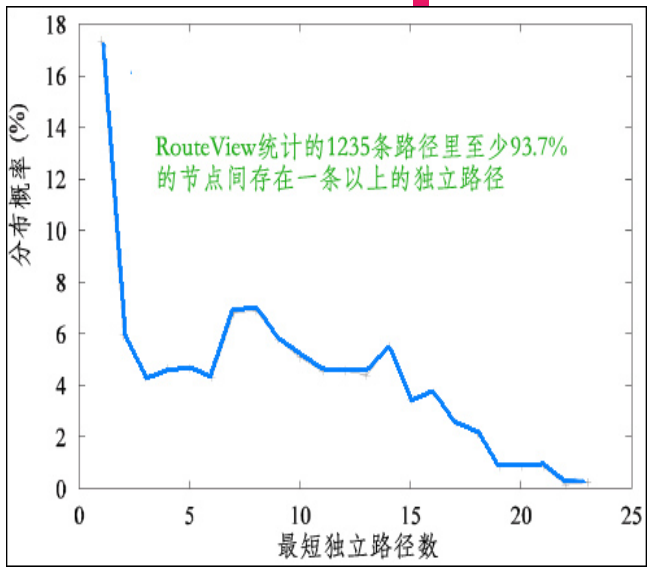


分层组网机制可降低消息开销：

$max\{O(m^2), O(k^2), O(m \times k)\}$

路由抗毁协议原理

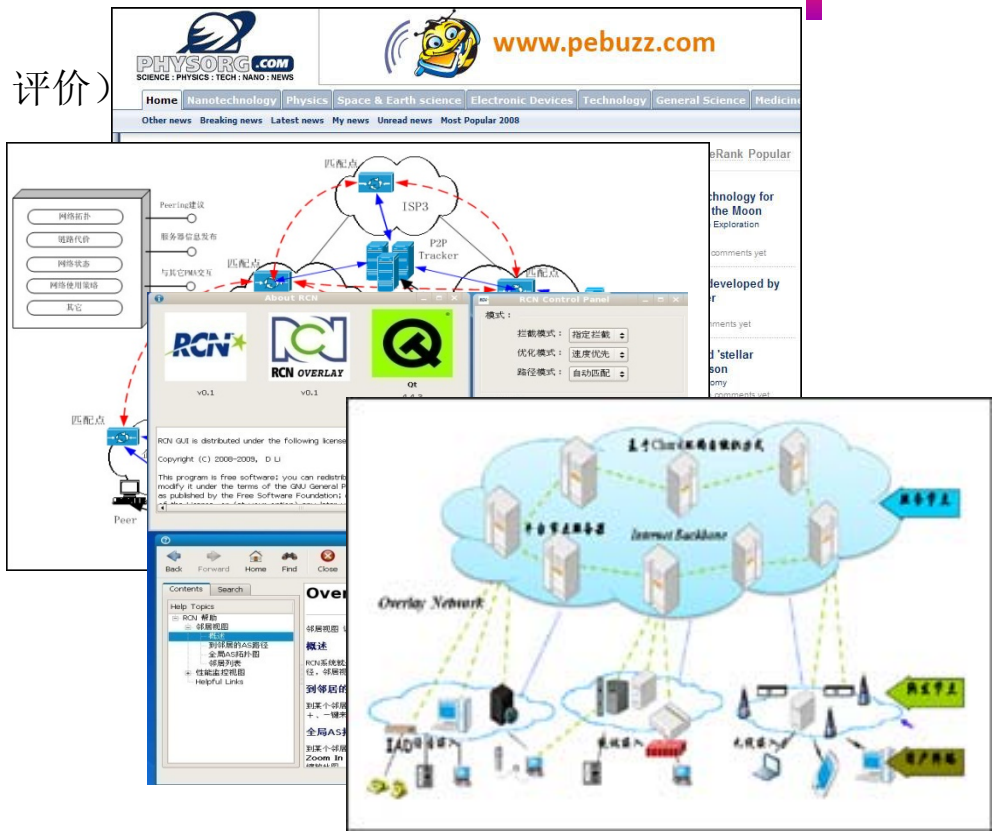




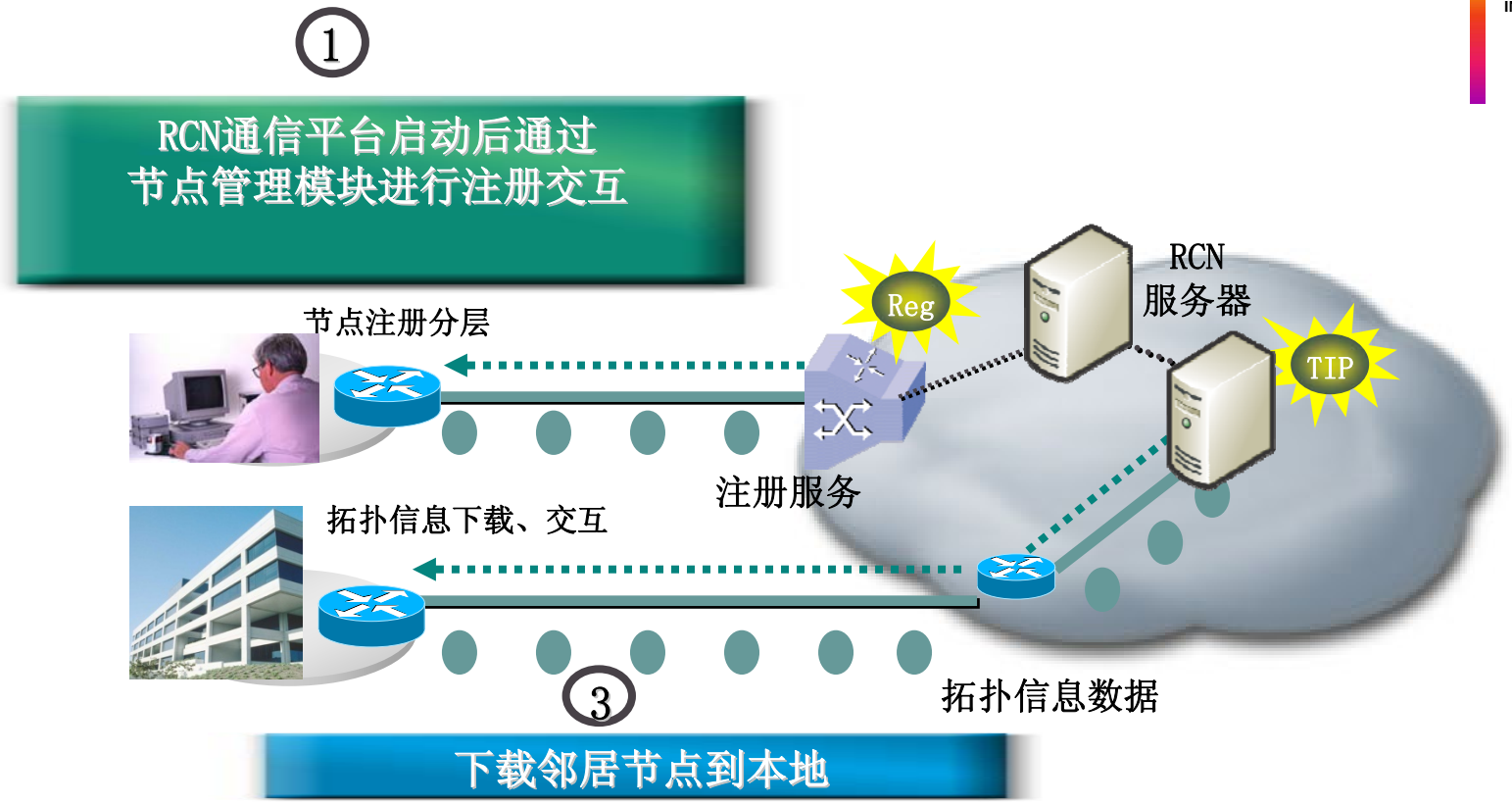
覆盖网抗毁技术的可行性是
基于互联网的多路径结构特点！

与网络结构理论相关的覆盖网技术成果

- 基于拓扑感知的快速组网
  - 系统动态扩展（节点扩展、管理、评价）
  - 结合底层信息（合理分域、分层）
- 基于拓扑结构的转发优选
  - 终端节点维护量小
  - 转发节点质量高
  - 转发成功率高（路径相关性小）
- 多径路由
  - 拓扑特征利用
  - 流量均衡

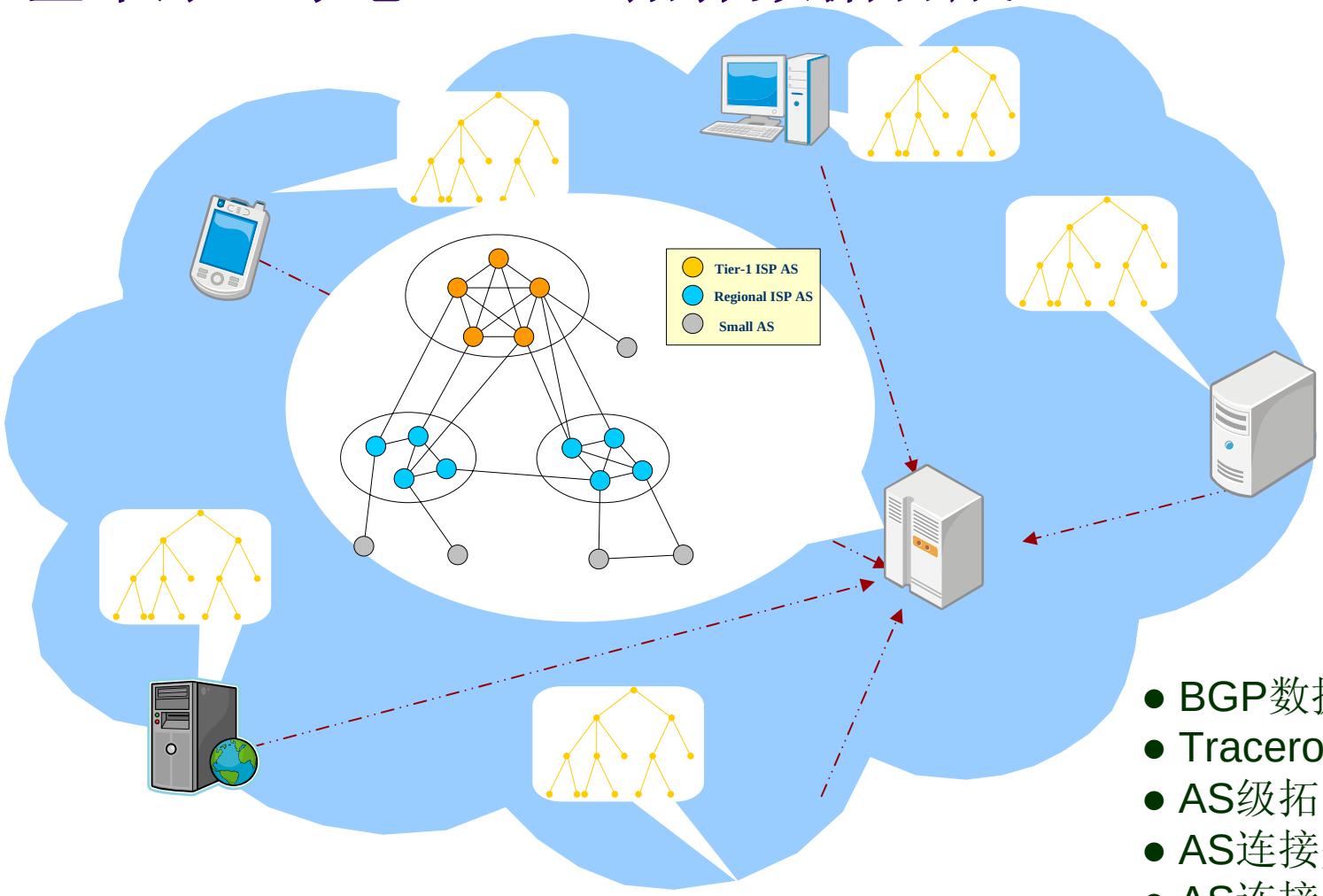


基本原理示意------底层拓扑信息感知过程



通过分布式探测、集中式解析的自动拓扑感知技术，系统可为转发优化提供精确的信息“地图”

基本原理示意-------拓扑数据合成



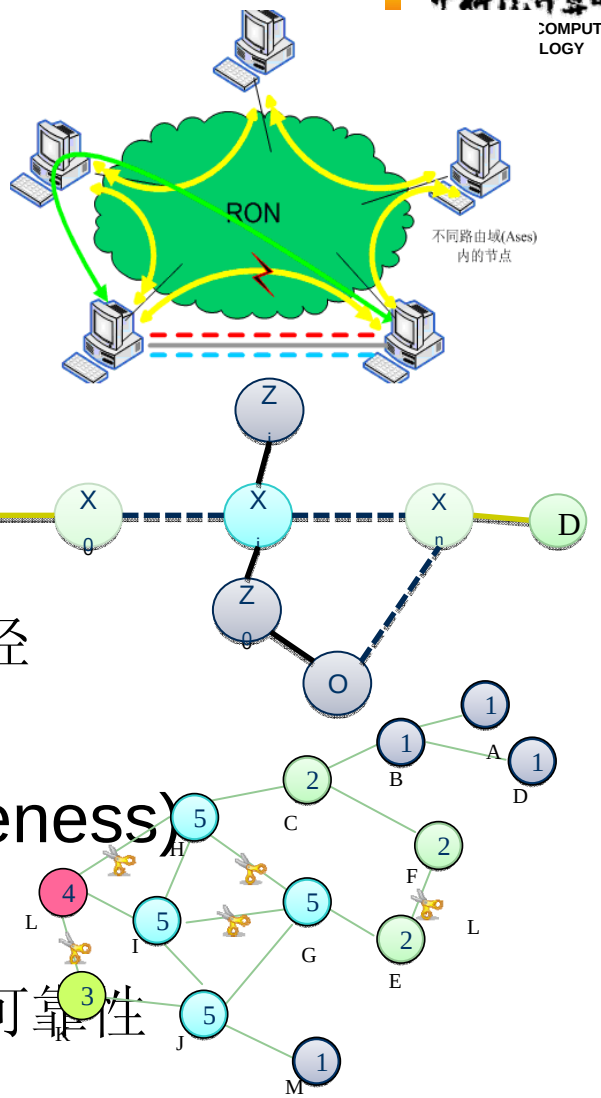
- BGP数据采集
- Traceroute数据采集
- AS级拓扑视图构建
- AS连接关系分析算法
- AS连接路径性能探测

基于拓扑结构信息的转发节点选取

- 随机选取或Fullmesh探测方式
  - 优点：简单快速
  - 缺点：开销大、效率低、可扩展性很差

- 最早分叉选取方式
  - 优点：开销低、扩展性好
  - 缺点：可能会选中转发可靠性低的长路径

- 基于拓扑结构参数(coreness\Betweeness)的FSRI选取方式
  - 优点：开销低、扩展性好、提高了转发可靠性
  - 缺点：计算量稍大于最早分叉



FSRI选取算法的转发可靠性比较

- 判断备用路径性能的两个重要指标
  - 路径相关性和路径长度
- 路径转发成功率RS(Relay success rate)

$$RS = (1 - K / Lp_0) * P_{loss}^{Lp_b}$$

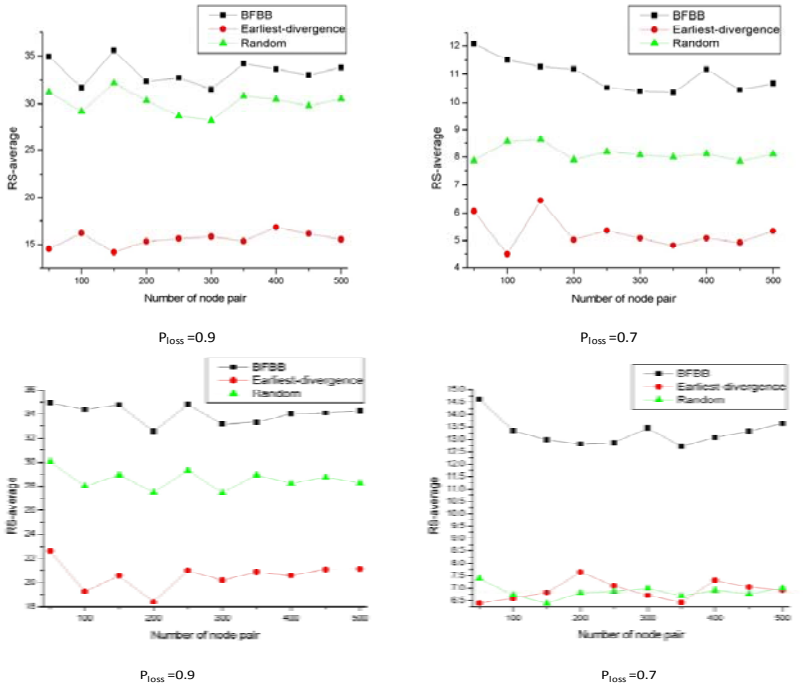
Lp_0：原始路径的长度

Lp_b：备用转发路径长度，

K：重叠的链路数目

P_{loss}：链路性能下降率（小于1）

实验数据	ChinaAS	PFP-1000	PFP-2000	PFP-3000
节点数	135	1000	2000	3000
边数	338	3000	6000	8996



基于拓扑结构信息的分域组网

- 综合转发性能相似度算法
  - 转发节点对n个Landmark服务器进行探测，主要探测延迟、丢包率和瓶颈带宽
  - 计算两个节点之间的延迟相似度 D_{AV}，丢包率相似度L_{AV}和瓶颈带宽相似度 Bw_{AV}

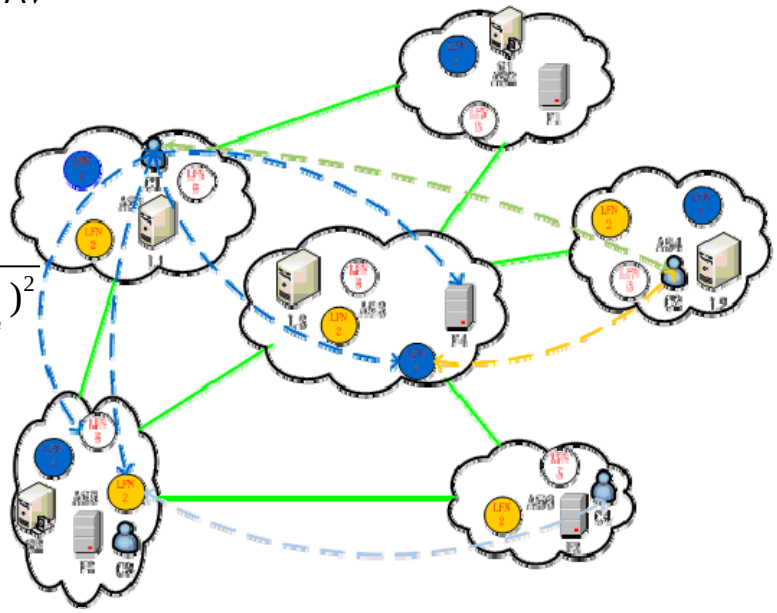
$$D_{Av} = \sqrt{(D_{A_{L_1}} - D_{B_{L_1}})^2 + (D_{A_{L_2}} - D_{B_{L_2}})^2 + \ldots + (D_{A_{L_n}} - D_{B_{L_n}})^2}$$

$$L_{Av} = \sqrt{(L_{A_{L_1}} - L_{B_{L_1}})^2 + (L_{A_{L_2}} - L_{B_{L_2}})^2 + \ldots + (L_{A_{L_n}} - L_{B_{L_n}})^2}$$

$$Bw_{Av} = \sqrt{(Bw_{A_{L_1}} - Bw_{B_{L_1}})^2 + (Bw_{A_{L_2}} - Bw_{B_{L_2}})^2 + \ldots + (Bw_{A_{L_n}} - Bw_{B_{L_n}})^2}$$

  - 计算对应的综合性能相似度

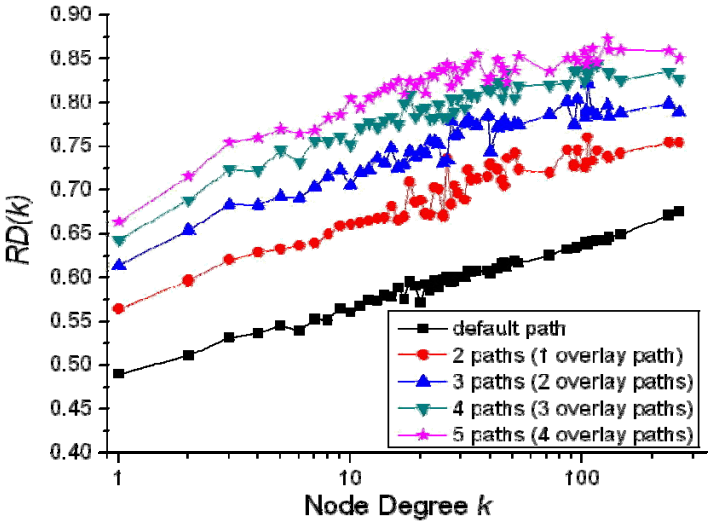
$$I_{AB} = \sqrt{(\alpha D_{Av})^2 + (\beta L_{Av})^2 + (\gamma B_{Av})^2}$$



- 依据拓扑信息计算路径相关性

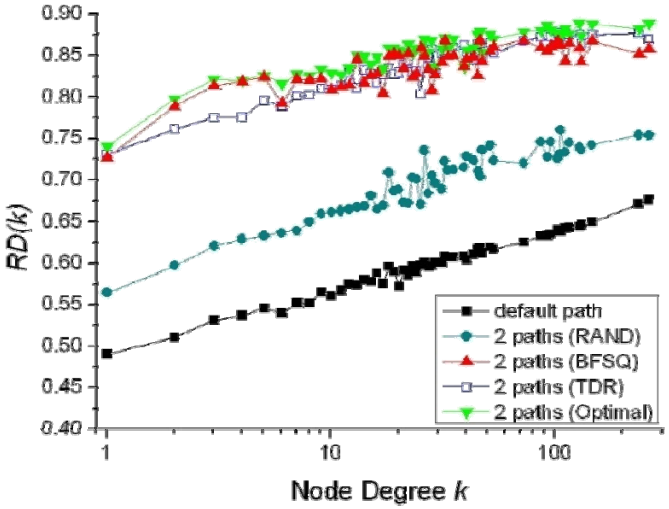
$$Cov(X_{dOLPN}, Y_{dOLPN}) = \frac{E(XY) - E(X)E(Y)}{STD(X)STD(Y)}$$

基于拓扑信息的路径性能评估与选取

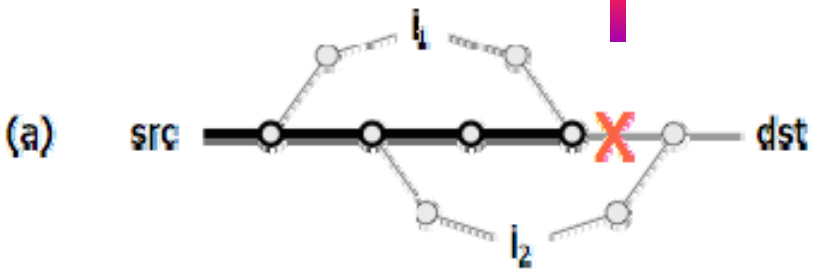


路径选择算法

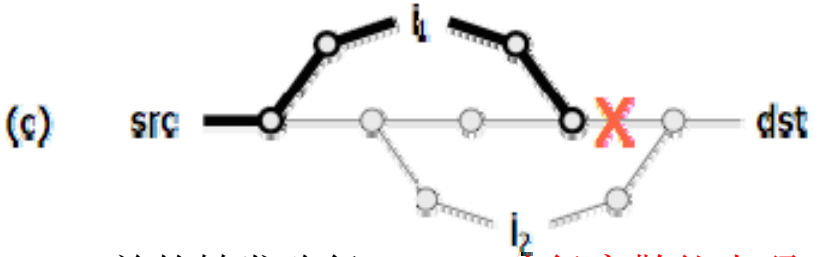
- 随机、BFSQ、TDR
- 算法仿真比较



通过运用适度的拓扑信息利用BFSQ算法选择高效的overlay路径，可有效提高路由可靠性



I_2 “好的转发路径” ------> 路径离散能力强

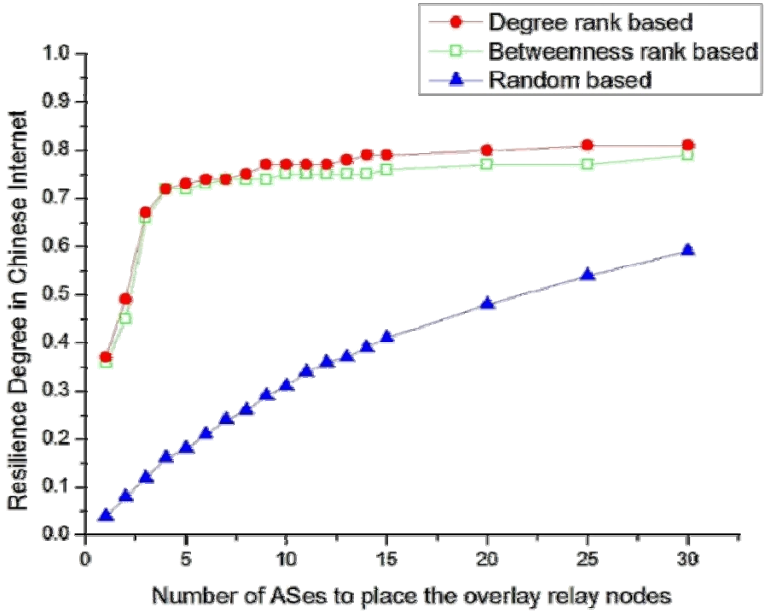


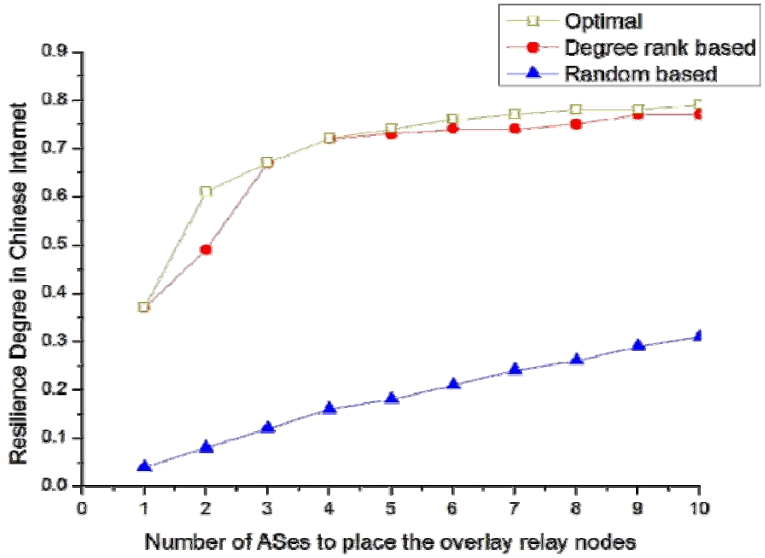
I_1 “差的转发路径” ------>路径离散能力弱

基于拓扑参数的覆盖网节点部署算法

- 基于度节点的覆盖网节点部署方案
  - 将节点按降序排列
  - 按度等级加入候选节点
  - 寻求满足RD指标的折中

$$RD = \frac{\sum_{s \neq d} I_{sd}\{P' \geq 2\}}{\sum_{s \neq d} I_{sd}\{P \geq 1\}}$$





相关工作的部分成果

- 应用互联网结构知识的相关专利20多项
- 相关标准
  1. 计算所牵头，“基于承载网感知的P2P流量优化技术框架”， 2009H103
  2. 计算所牵头，“基于承载网感知的P2P流量优化技术：网络匹配服务器发现协议”，已立项
  3. 声学所牵头，计算所参与，“基于承载网感知的P2P流量优化技术：网络匹配服务协议”， 已立项
  4. 中国电信牵头，计算所参与，“基于DNS协议的IP位置解析服务的技术要求”，已报批
  5. 中兴通讯牵头， P2P流量优化(安全)

中科院计算所
INSTITUTE OF COMPUTING TECHNOLOGY

谢谢！