Huazhong University of
Science and Technology

華中科技大學

对等计算技术的应用及其安全问题

李瑞轩

网络与分布计算(IDC)实验室

华中科技大学计算机科学与技术学院

rxli@hust.edu.cn

http://idc.hust.edu.cn

主要内容

- 分布式计算经济学
- 对等计算概念与问题
- 对等计算的安全问题
- 对等计算技术的远景

1.从分布式计算经济学的观点看对等计算

等价计算概念

1$可以购买：

- 1 天CPU时间
- 4 GB 内存使用1天
- 1 GB 网络带宽
- 1 GB 磁盘存储3年
- 10 M 数据库访问
- 10 TB 磁盘访问
- 10 TB 局域网带宽
- 10 KWhrs == 1台计算机4天消耗的电能

计算假定：

- ✓ 2GH CPU, 2 GB RAM的计算机: $2,000
- ✓ 200 GB硬盘, 每秒访问100次或传输50MB数据: $200
- ✓ 1 Mbps宽带网: $100/月

按3年折旧，共计1000天计算.

分布式计算经济学

(Views of Jim Gray, 2003)

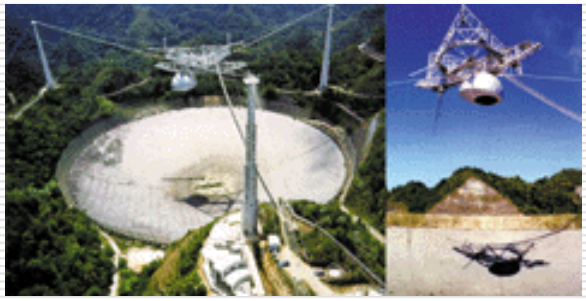
- 等价的计算代价：
  - 1次数据库访问
  - 10字节网络传输
  - 100,000条CPU指令
  - 10字节磁盘存储
  - 1M字节磁盘带宽
- 平衡点：
  - 10,000条CPU指令＝1字节网络传输

计算资源几乎是免费的

- Google一年免费提供在2PB数据库中的万亿次搜索.
- Hotmail一年免费收发万亿封E-mail.
- Amazon.com提供免费书籍搜索.
- 新浪等大量网络媒体提供免费新闻.
- ……

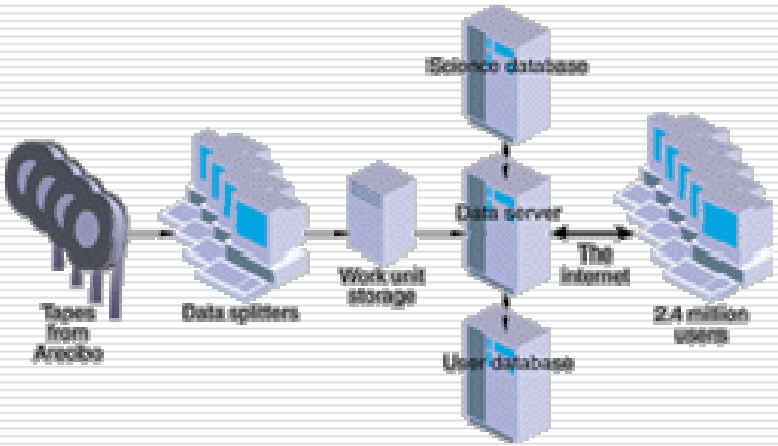
对等计算(Peer-to-Peer Computing)技术是如何发挥效能的？

SETI@Home外星生命搜索项目

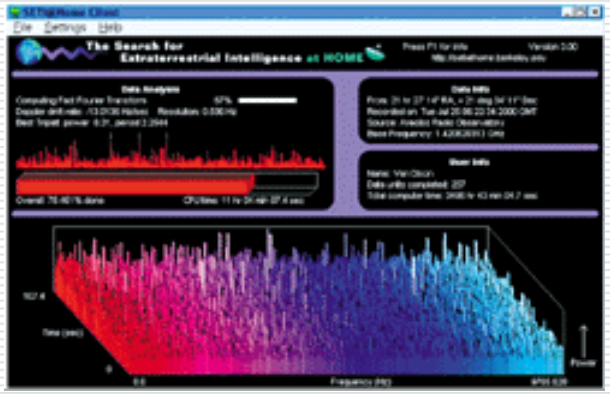


美国国家天文中心望远镜：305米

SETI@Home具有67万亿次浮点计算能力(67 Teraflops)



SETI@home客户程序



从经济学的观点看分布式计算：

对等(P2P)计算是一种经济的分布式计算

有关P2P的预测

- 福布斯预测 (2005.2)
  - 基于P2P拓扑的应用将成为Internet主流应用
- 美国高级智囊团 (2005.1)
  - 自组织的P2P系统将以分散控制方式提供高可用、可扩展、健壮的高性能计算网络

2.对等计算的概念与问题

什么是P2P?

- 定义1:
  - P2P是一类有效利用互联网边缘资源的分布式应用，这些资源包括存储、计算、内容、服务等 (Clay Shirky, 2000)
    - 边缘节点经常离线，没有固定IP，但却拥有资源
- 定义2:
  - P2P是一类分散的、自组织的分布式系统，这类系统中的通信或信息交互通常是对称的或对等的 (IPTPS'02)
- 定义3: ...

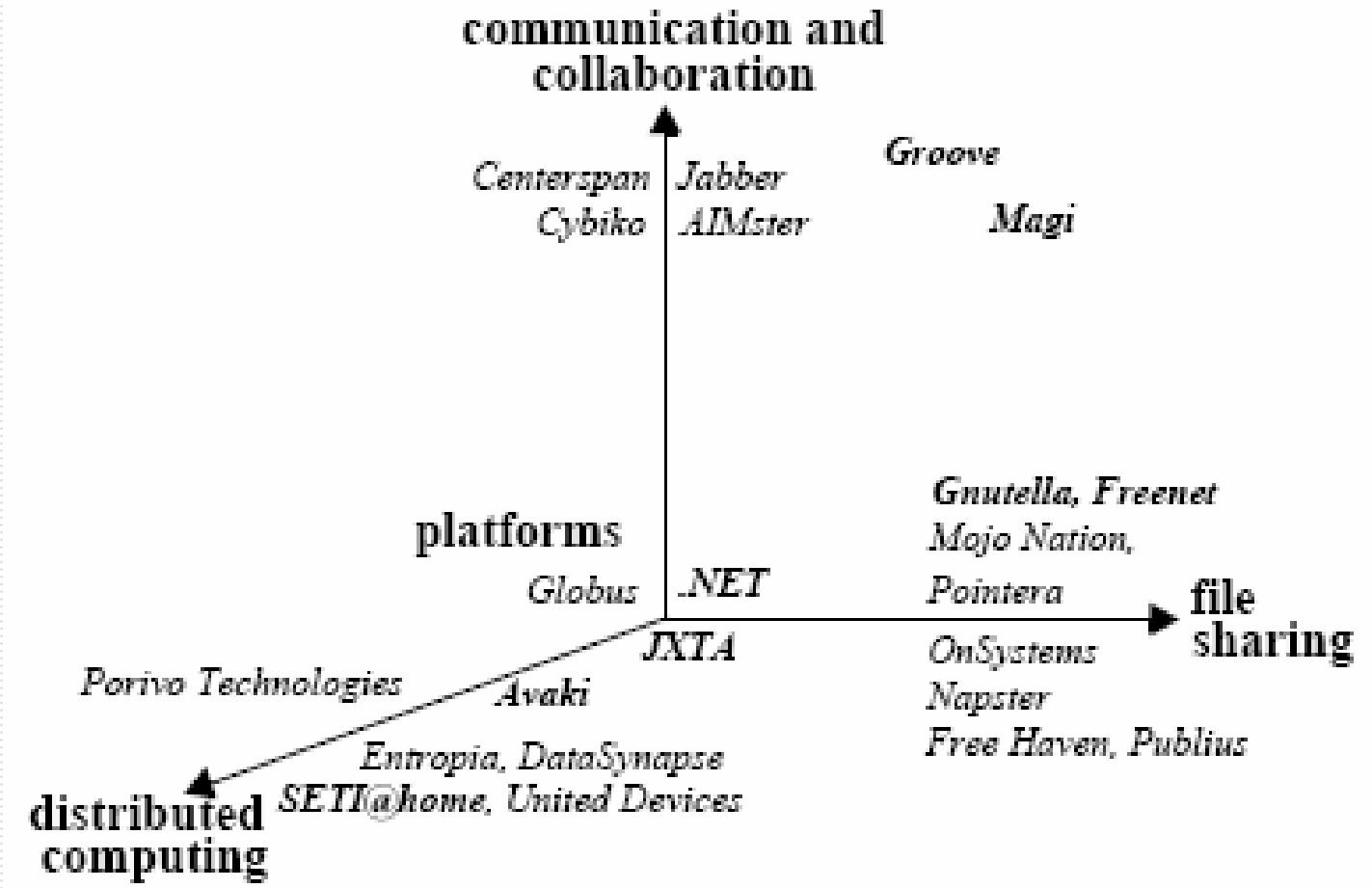
为什么选择P2P?

- 异常丰富的资源
- 天生的可扩展性
- 高度的可用性和健壮性
- 高性能的计算能力

P2P应用

- 文件共享 (Napster, FastTrack, KaZaA, iMesh, Gnutella, eDonkey, BitTorrent, eMule, Maze)
- 分布式计算 (SETI@home, UnitedDevices, DistributedScience)
- 分布式存储 (OceanStore, Farsite, HiveNet, Granary)
- 协同工作 (Groove, Engenia, Interbind)
- 即时消息 (MSN, Yahoo, AOL, ICQ, QQ)
- 流媒体服务 (Anysee, PPLive)
- Web服务社区 (uServ)
- 网络游戏
- 匿名Email
- 垃圾过滤

P2P分类 (Views of HP)

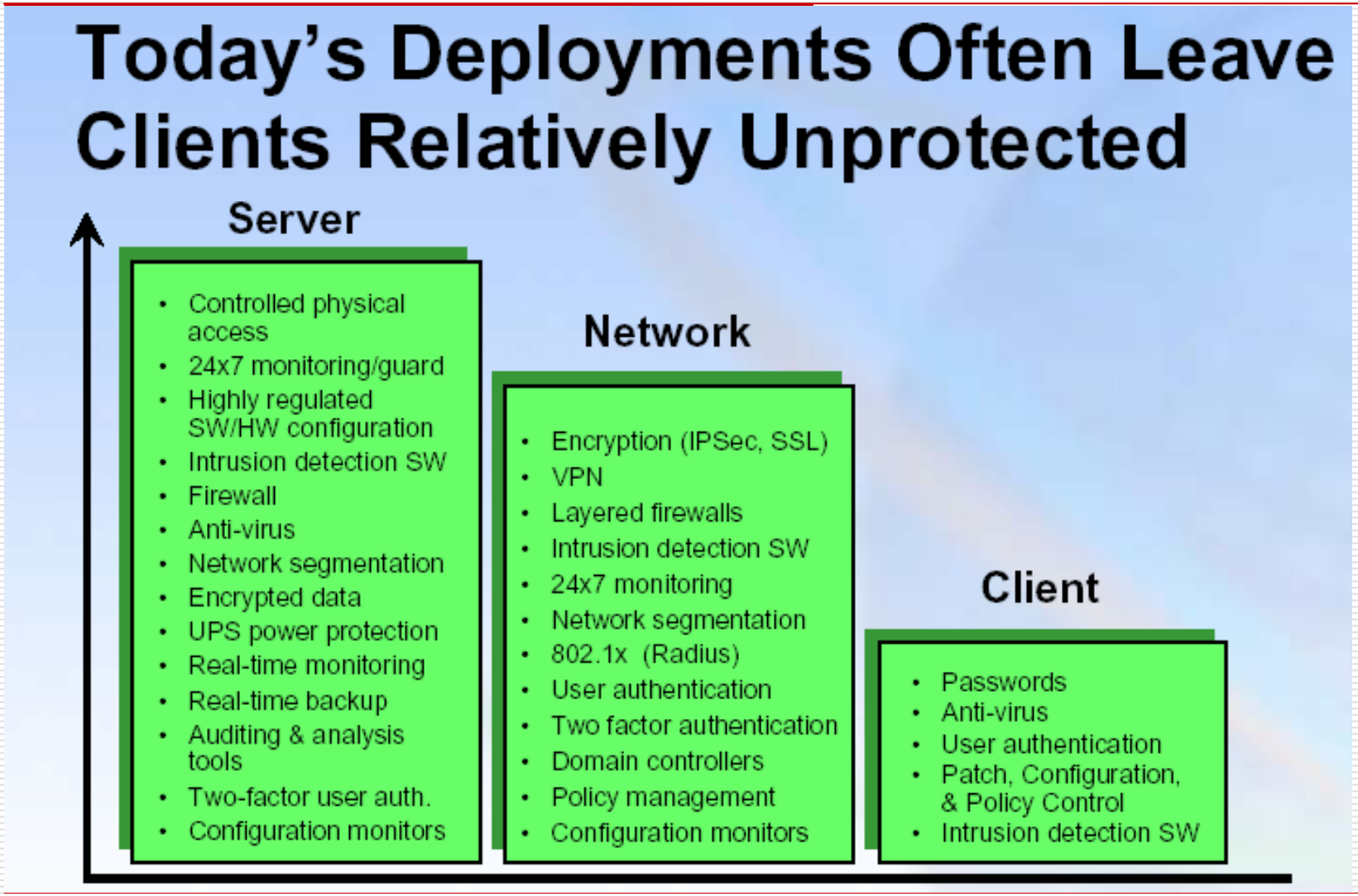


P2P研究的关键问题

- 基本问题
  - 资源放置
  - 资源发现
  - 资源获取
- 更多问题
  - 可扩展性
  - 可用性
  - 匿名性
  - 安全性
  - 容错能力
  - 自组织
  - 互操作性
  - 复杂查询
  - 搜索机制
  - 负载平衡
  - 版权保护
  - 服务质量

3.对等计算的安全问题

3.1 分布式计算的安全现状



实际应用中的客户安全需求

- 客户端安全在现代信息系统及应用中越来越重要，例如：
  - 分布式分发控制 (Distributed Dissemination CONtrol)
    - *例如，一个病人的健康记录可能需要在某一特定时间从他的主治医师传给他的健康咨询师，但不允许其他人访问*
  - 基于P2P的VoIP应用
    - 声音数据的实时保护
      - 通话内容的不能被窃听或非法录制
    - 声音对象的传递控制 (e.g., 声音邮件)
  - 基于P2P的电子商务
    - P2P平台中电子货币的安全管理
    - 基于P2P的电子支付系统
  - ……

P2P系统中的安全需求 (续)

- 需要新的安全模型和架构
  - 客户和服务器之间信任关系的改变
    - ✓ 没有强安全保护的中央服务器
    - ✓ 数据位于Peer或客户方
  - 策略执行位置改变
    - ✓ 客户端的策略执行需要安全控制
  - P2P平台和应用中的安全
    - ✓ 环境的分布性与动态性
    - ✓ 各种基于软件的攻击
  - 客户端可信的用户身份认证与授权
  - Peer与Peer之间的信任路径
    - ✓ 传输过程中的窃听、欺骗、篡改攻击
    - ✓ Peer之间的安全信息交互

3.2 P2P安全问题

- P2P病毒
  - 2005年1季度比2004年1季度P2P威胁增加了271%，其中，即时通讯病毒和蠕虫占82%，利用即时通讯工具传输文件功能的攻击方式占14%
  - 绕过防火墙，风险难以控制
  - 用户群巨大，传播速度惊人：可以在30秒内感染50万台电脑！

P2P安全问题 (续)

- 泄密
  - 通讯过程泄密：包括IM、FS等大多数P2P软件的通讯都是明文方式，通讯内容极易被第三方窃取
  - 间谍软件泄密：通过文件共享等方式，将间谍软件分发到个人电脑，收集相关资料
  - 人为泄密：绕过防火墙的监控，随意收发各种信息

P2P安全问题 (续)

- 欺诈与攻击
  - 非法访问
  - 假冒服务
  - 路由攻击
- 系统漏洞
  - 非法占用系统资源
  - 版权和法律问题

3.3 P2P系统安全控制方法

- 身份标识 (Identity)
- 身份认证 (Authentication)
- 授权 (Authorization)
- 资源保护 (Resource Protection)
- 通信加密 (Secure Communications)
- 风险评估 (Risk Assessment)

身份标识

Identity

由一些组合在一起的特征组成，这些特征的集合使得某个人或某个对象与其他人或对象区别开来。

在任何环境下，一个需要身份标识的系统都必须通过身份识别过程对系统的所有对象进行区分。

- 提供认证和授权等安全措施的基础
- 进行审计等可追溯工作的前提
- 计算参与者的行为声誉度的依据

身份认证

Authentication

如何才能确信某一对象向你宣称的身份就是该对象的真实身份

- 密码技术
- 令牌
- 数字化证书
- 使用第三方节点实现不可否认认证
- 基于生物学特征的认证方法

授权

Authorization

确定用户能执行何种操作或用户能访问哪些资源。

- 使用资源访问能力列表，将一个用户或一个密钥与一个访问权限集合关联起来。
- 使用访问控制列表，将资源和一个被授权访问该资源的用户集合关联起来。
- 基于密钥或信任状的P2P系统更有价值。

资源保护

Resource Protection

- 资源的标识
- 机密性保护
- 一致性保护
- 可用性保护

通信加密

Secure Communications

- 通讯的保密性
- 通讯双方的不可欺诈性
- 通讯内容的不可篡改性

风险评估

Risk Assessment

- 分析系统的潜在威胁
- 评估用户交互的风险
- 提出风险规避建议

3.4 P2P信任管理

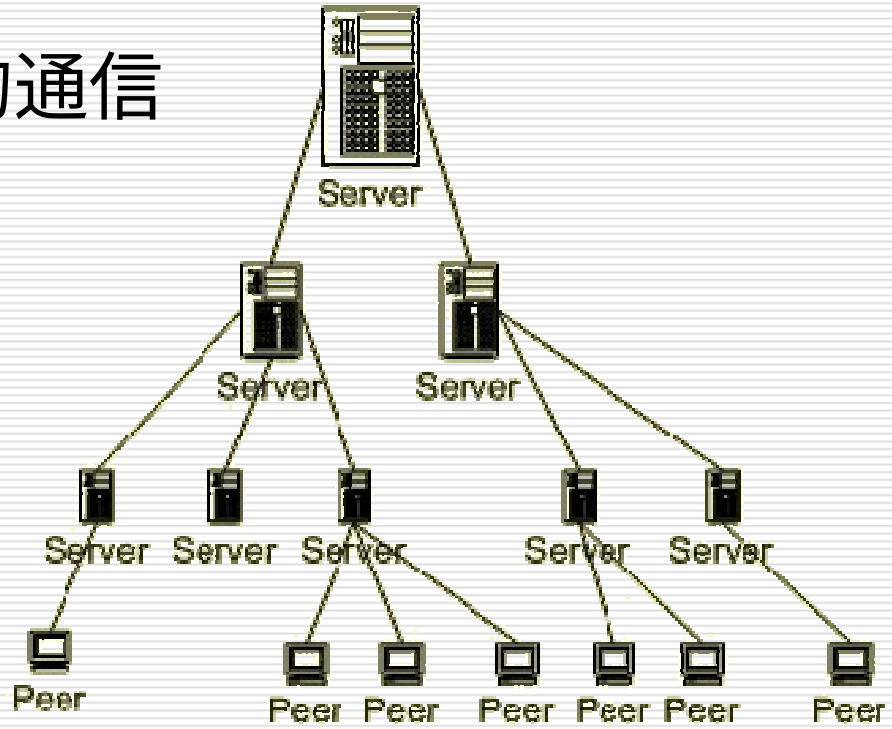
- 信任模型 (Trust Model)
- 声誉管理 (Reputation Management)

信任模型

分层信任模型 (Hierarchical)

- 树状结构 (tree)
- 适合层次对象之间的通信

  e.g. PKI

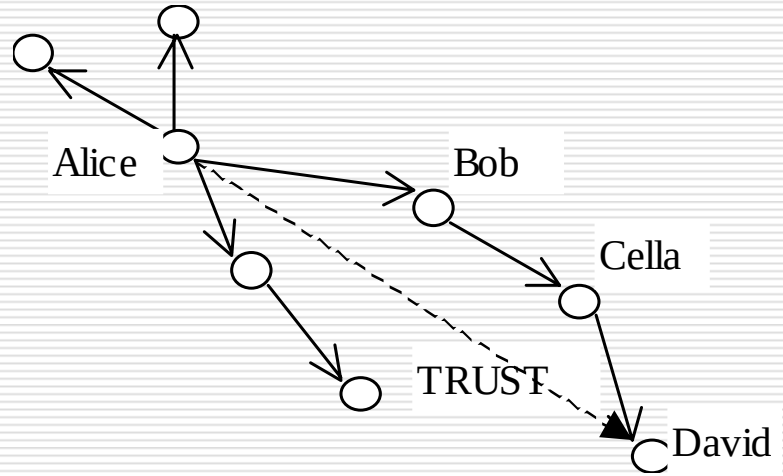


信任模型 (续)

信任网络 (Web of Trust)

- 网状结构 (net)
- 信任在Peer之间传播

  e.g. PGP, SPKI



信任模型 (续)

基于声誉的信任模型 (Reputation-based Trust)

- 依据Peer过去的行为，赋予相应的权力
- 需要保留历史记录
- 具有可追溯性

声誉管理

集中式声誉管理 (Centralized Reputation Management)

- Peer的声誉记录统一管理
- 每次发生信息交互行为，即可记录在案

  $R=(\Sigma Sat-\Sigma UnSat)/Count()$

- 中心服务器维护工作惊人
- 身份标识的更换
- 联合欺诈

  应用实例：eBay

声誉管理 (续)

分布式声誉管理 (Distributed Reputation Management)

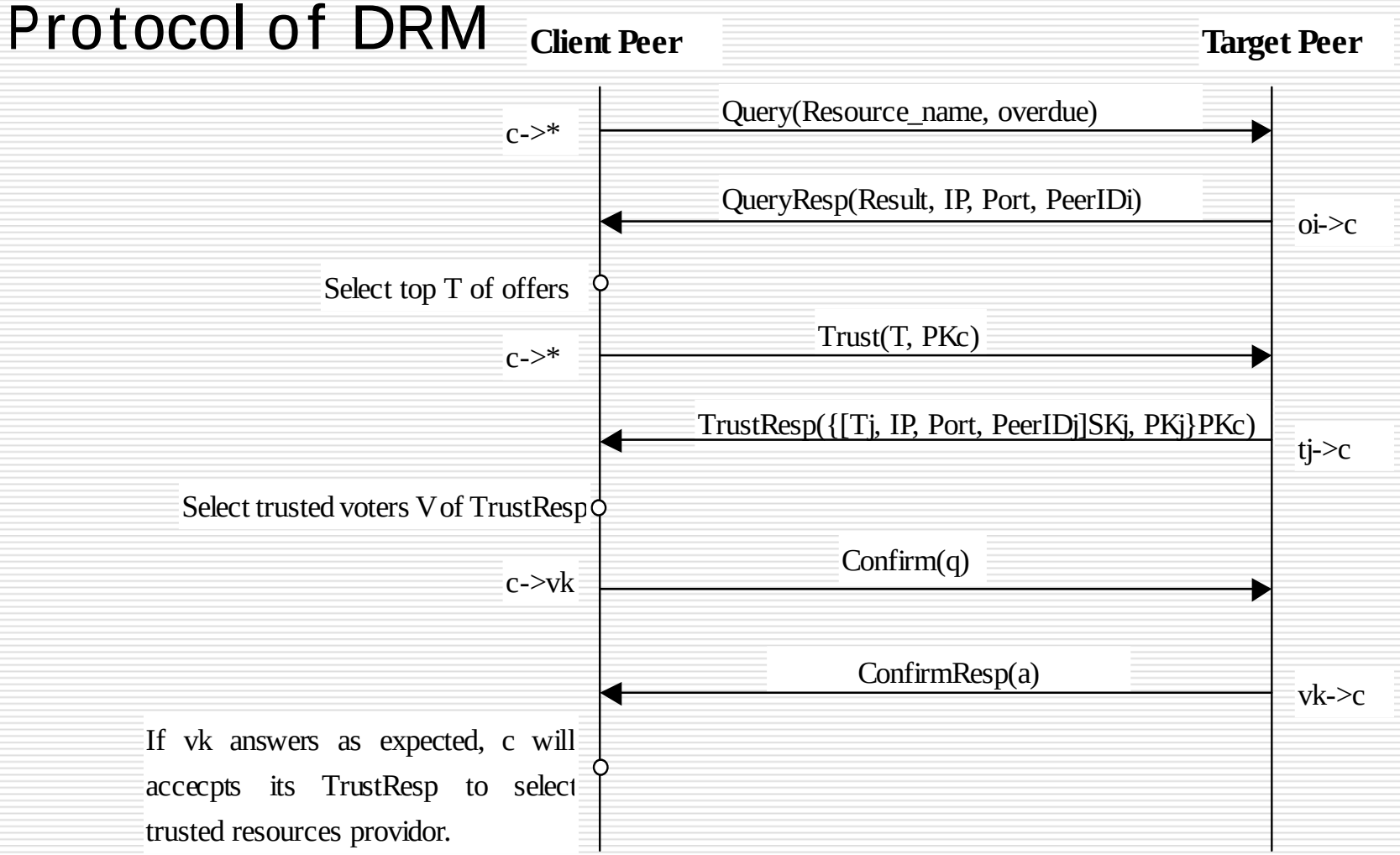
- Peer保留自己的访问历史和声誉评估库
- 在需要对某个Peer作出评估时，向P2P网络的其他Peer发出查询该Peer声誉评价的请求
- 综合多方面的声誉评价，得到该Peer的声誉

- 结果的收敛性难以保证
- 查询范围的控制

Stanford：eigenRep

声誉管理 (续)

Protocol of DRM



4.对等计算技术的远景

对等计算技术的远景

- 远景1: 对等计算技术与网格计算技术的结合
- 远景2: 对等计算技术将为服务的边缘化提供支撑
- 远景3: 对等计算技术将成为普适计算环境中的基本技术

Internet的异构现状

- 数十亿形式各异的客户终端通过Internet或无线网在使用服务
  - 个人机，便携机，PDA，移动电话，智能家电
- 数百万的终端成为超级节点
- 数百万的集群提供本地服务
- 数百万的可信网格节点在独立提供服务
- 数百万的可信网格节点在协作提供服务
- 数百台超级计算机在为科学计算提供服务

P2P与Grid的结合

- 目标:
  - 构造高可靠的、高安全的、可扩展的资源共享系统
- 关键:
  - 资源管理
- 方法:
  - 利用网格的高安全、高可靠服务
  - 利用P2P的高可扩展性，避免单点失效问题
  - 在网格与P2P之间进行平衡

边缘计算 (Edge Computing)

- 目标:
  - 自动及时或暂时地将计算资源或内容从中央服务器移动到离终端用户更接近的站点（即网络的边缘）
- 优点:
  - 大大提高服务质量 (如极小的服务响应时间)
  - 资源的高效利用
  - 服务的高可用性
  - 服务的高性价比

普适计算 (Pervasive/Ubiquitous Computing)

- 目标：
  - 使人们能够使用任意设备、通过任意网络、在任意时间都可以获得一定质量的网络服务
- 服务内容：
  - 计算、管理、控制、资源浏览等
- 关键技术：
  - 场景识别、资源组织、人机接口、设备无关性技术、设备自适应技术等

核心技术及挑战

- 动态资源配置：
  - Internet应用的按需部署
- 挑战性问题：
  - 资源配置的自动化
  - 资源配置的优化
  - 高效的服务发布
  - 灵活的安全控制

总结

- 对等计算是一种经济的分布式计算
- 客户端安全在现代信息系统及应用中越来越重要
- P2P系统需要综合应用多种安全控制技术
  - One size fits all – NOT!!!!
  - Tradeoff between security and availability
- P2P技术将是未来分布式计算的基础技术之一

Thanks!