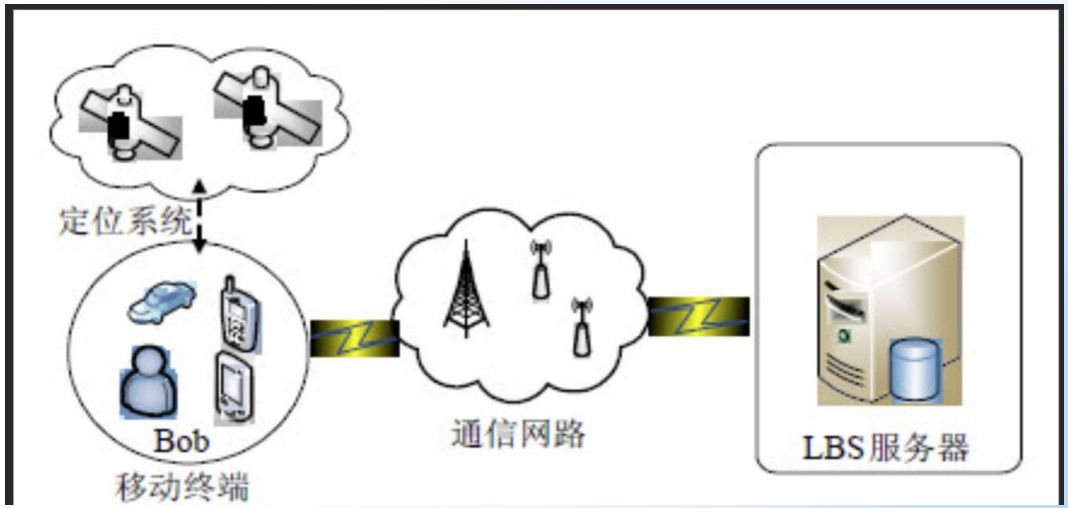
chap4 移动通信中的位置隐私问题
Privacy for Location-based Services

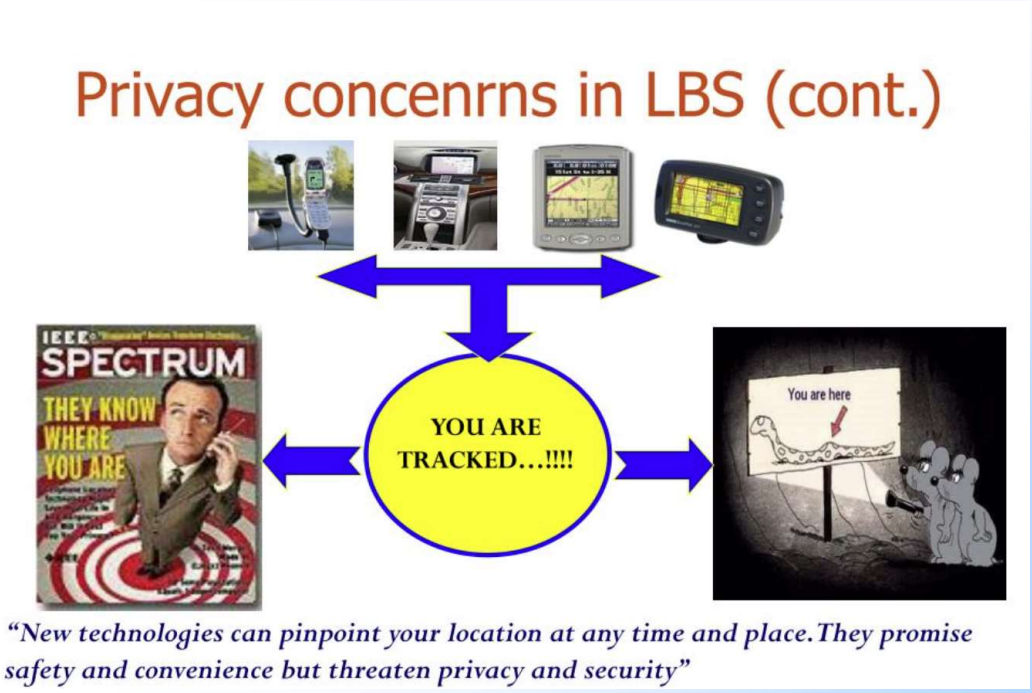
4.1 LBS概念

⑩ 基于位置服务 (Location Based Services，LBS)是指围绕地理位置数据而展开的服务，其由移动终端使用无线通信网络 (或卫星定位系统)，基于空间数据库，获取用户的地理位置坐标信息并与其他信息集成以向用户提供所需的与位置相关的增值服务。



带来隐私问题?

- 人们在享受具有诱惑力的各种服务的同时，也意识到自己可能随时随地被人跟踪，被人获知曾经去过哪里、做过什么或者即将去哪里、正在做什么，换句话说，人们的隐私和安全受到了威胁。



4.2 位置隐私研究面临的挑战

（1）保护位置隐私与享受服务是一对矛盾。

位置信息越精确，服务质量越高，隐私度却越低，位置隐私和服务质量之间的平衡是一个难处理却又必须考虑的问题。这里考虑的服务质量包含响应时间、通讯代价等等，与具体的环境有关。

（2）位置匿名的即时性特点。

在位置隐私中，通常处理器面临着大量移动对象连续的服务请求以及连续改变的位置信息，使得匿名处理的数据量巨大而且频繁的变化。在这种在线（Online）的环境下，响应时间也是用户的满意度的一个重要衡量标准

（3）基于位置匿名的查询处理。

经过匿名处理的位置信息，通常是对精确的位置点进行模糊化处理后的位置区域。得到的查询结果跟精确的位置点的查询结果是不一样的。如何找到合适的查询结果集，使得真实的查询结果被包含在里面，同时也没有浪费通讯代价和计算代价，是匿名成功之后需要处理的主要问题。

（4）位置隐私需求个性化。

不同的用户具有不同的隐私需求，即使相同的用户在不同的时间和地点隐私需求也不同。让用户自定义个性化的隐私需求，从而动态地协调隐私保护与服务质量的平衡点。

（5）位置隐私还要考虑对用户的连续位置保护的问题。

因为攻击者有可能积累用户的历史信息来分析用户的隐私。

4.3 位置隐私保护技术

第一，发布假位置，但query的内容是真的

方法一：采用附近的一个标志物来替换用户的位置。

优点：

- 采用标志物的方式实现location privacy

缺点：

- 标志物确定不当则有可能会收到不合理的answer
- 没有标志物该怎么办

4.3 位置隐私保护技术

方法二：False Dummies:用户向LBS发送m个不同的位置，其中只有1个是正确的，但LBS会对每一个位置进行应答，正确的位置的设备能够辨别出哪一个reply可用。

优点：

- 利用不存在的用户来提高隐私的等级
- 即使LBS不可靠也无法识别出哪一个query是来自用户的
- 可以防止被偷听

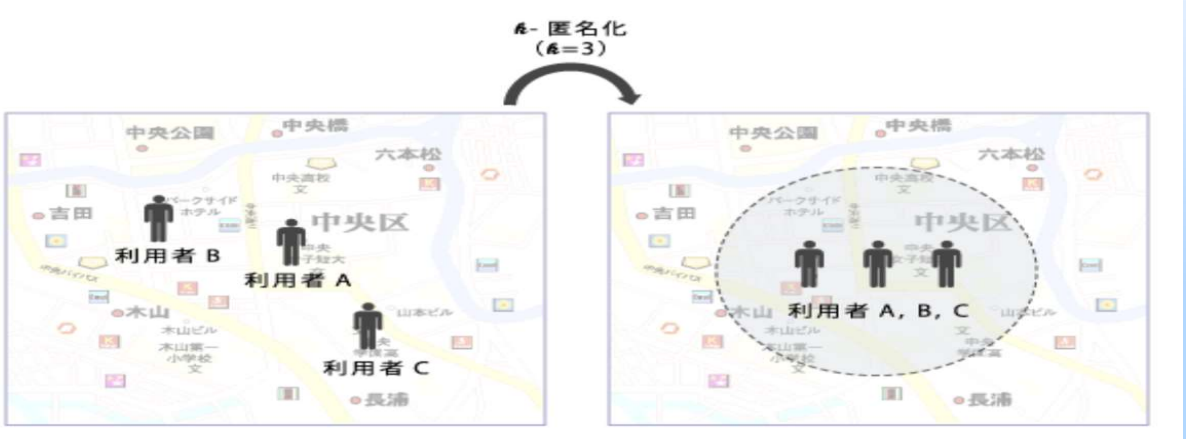
缺点：

- 对于资源的消耗太大。LBS要答复太多的query，而最终只有一个query有用。

4.3 位置隐私保护技术

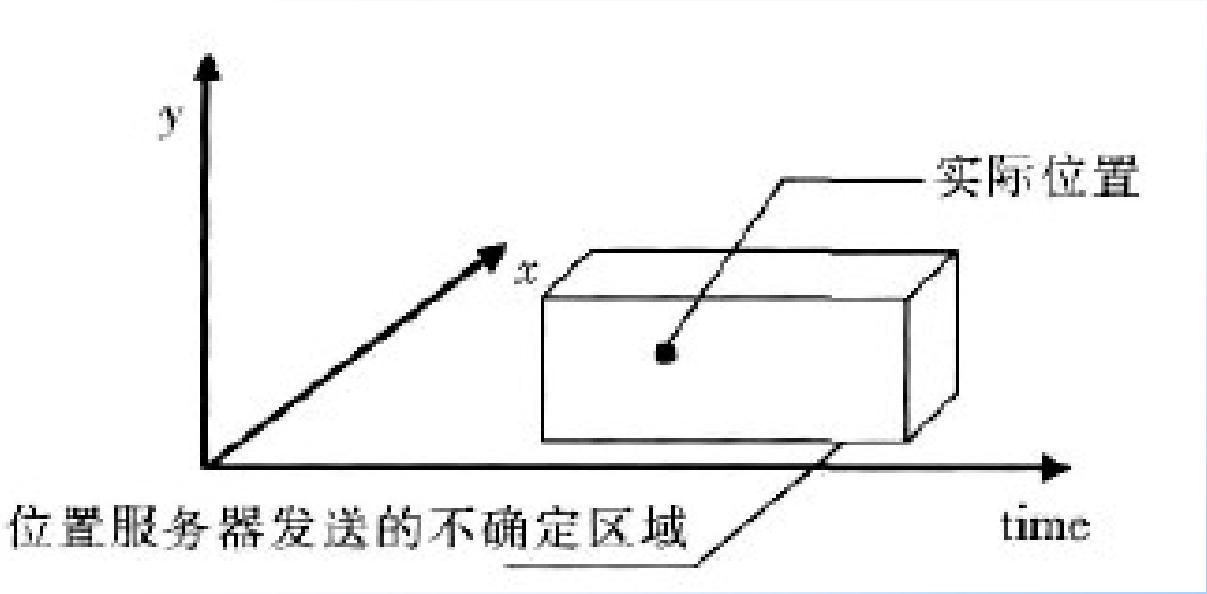
第二，空间匿名

- 本质上是降低对象的空间粒度，即用一个空间区域来表示用户的真实的精确位置。区域的形状不限，可以是任意形状的凸多边形，现在普遍使用的是圆和矩形，称这个匿名的区域为匿名框。
- 所以在位置 k- 匿名模型中，k 值越大，匿名度越高。
- 以匿名集的大小表示匿名度。



第三，时空匿名（Spatio- Temporal Cloaking）

- 在空间匿名的基础上，增加一个时间轴。在扩大位置区域的同时，延迟响应时间。



4.4 系统结构

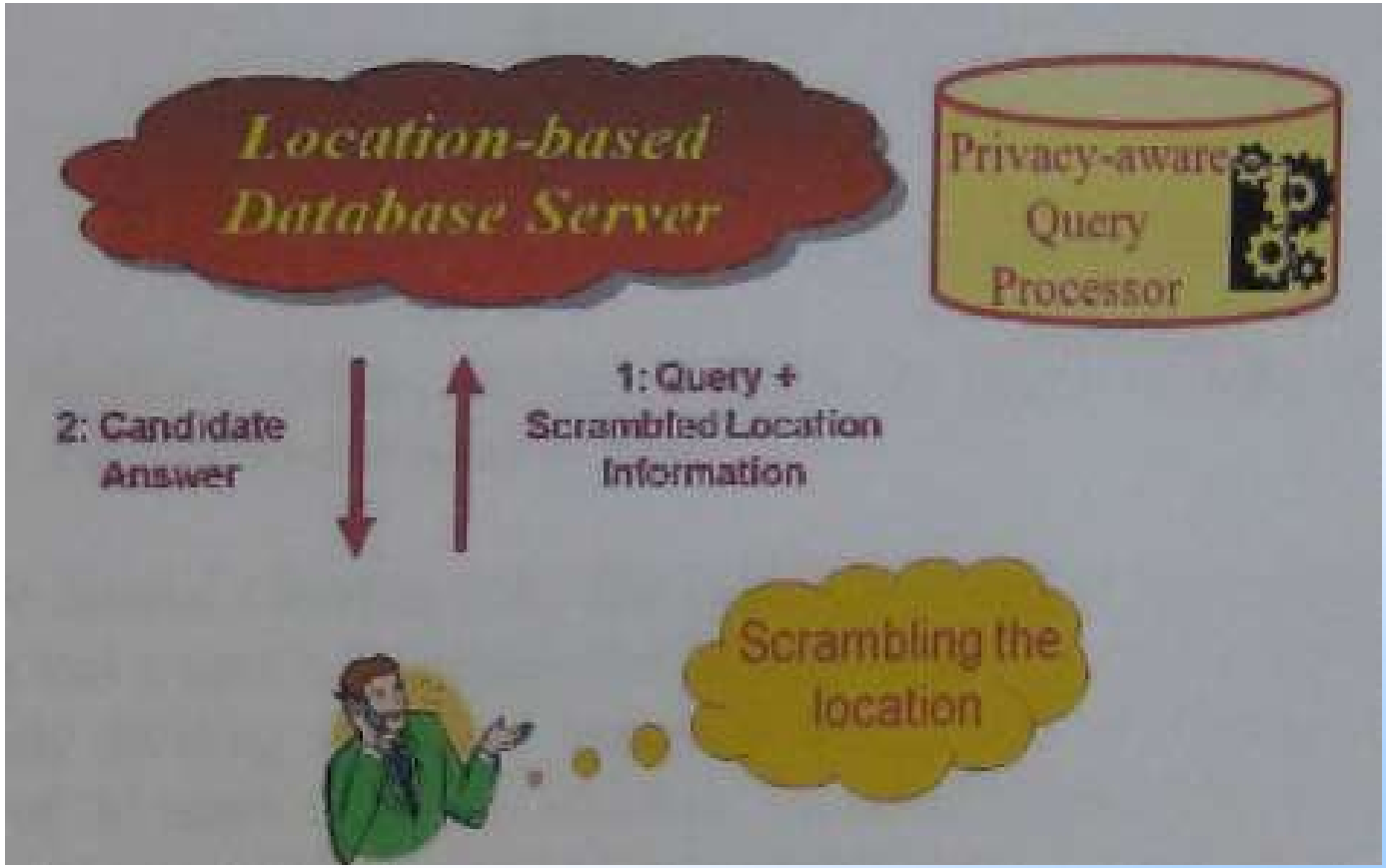
1 独立式结构

2 中心服务器结构

3 分布式点对点结构

- Non- corporative Architecture
- Centralized Trusted Party Architecture
- Cooperative (peer-to-peer)Architecture

Non- corporative Architecture



Centralized Trusted Party Architecture

- 一般采用空间k-匿名技术；
- 数据表的k-匿名化(k-anonymization) 是数据发布时保护私有信息的一种重要方法,它要求发布的数据中存在一定数量(至少为k) 的在准标识符上不可区分的记录,使攻击者不能判别出隐私信息所属的具体个体,从而保护了个人隐私.
- k-匿名通过参数k指定用户可承受的最大信息泄露风险。
- k-匿名化在一定程度上保护了个人的隐私,但同时会降低数据的可用性。

Centralized Trusted Party Architecture

- 存在一个可信赖的第三方准确的收到用户信息，将用户的位置blur（模糊化）之后，再将信息传输给LBS。
- TTP用来将用户的ID进行匿名化，在这种框架下，系统十分依赖于TTP的可靠性。
- 分类：
  - Quad tree 四叉树Spatial cloaking
  - Nearest – Neighbor k- Anonymizing
  - Mix Zone

Quad tree Spatial cloaking

思想：实现k-Anonymity，即在一个空间范围内，一个用户无法于另外k-1个用户相区分。

方法：将空间递归的划分成若干个区域，直到user所在的区域能够实现k-Anonymity，该区域被看做cloaked region

优点：

- 利用TTP将用户的所在位置以一个范围的方式来标示，可以隐藏用户的准确位置
- 在区域内能够实现k-Anonymity

缺点：

- 认为TTP是完全值得信任的。
- 需要平衡提供服务的质量与隐私等级之间的关系

Nearest – Neighbor k- Anonymizing

思想：从周围用户中随机选择k-1个用户实现k- Anonymity。

方法：

- 以用户为中心划定区域S(S中包括k个人)
- 在S中随机选择一个用户，进行同样的操作，划分区域S’
- 将S’作为cloak region

目的：随机选择一个用户替代当前用户成为区域的中心。

优点：

可以一次性的对大量用户实现位置隐私。

Mix Zone

用户可以在Mix Zone中更换假名，当用户数量少于k时，Mix Zone中的用户可以拒绝发送自己的更新信息。当用户走出Mix Zone的时候，偷听着无法将现在的用户与刚才在Mix Zone中的假名联系起来。

优点：

- 在Mix zone中的用户使用假名，并且经常跟换假名，使得偷听着无法将假名联系起来。
- 使用假名的形式来隐藏用户的真正身份
- 建立Mix Zone来对于用户的身份进行隐藏
- 在中间件处采用匿名的方式来使用application
- 利用anonymity set和熵来评估匿名性

缺点：

- 在某些情况下可以将假名进行联系
- Mix Zone的边界设定方法没有讨论

Cooperative (peer-to-peer) Architecture

- 在这种结构之下，用户之间通过相互协作来实现隐私保护，隐私的等级取决于在Architecture中用户的数量。
- K个用户交换他们的位置信息，然后计算出坐标的平均值来替换用户的位置以实现k-anonymity

Cooperative (peer-to-peer) Architecture

方案1：中心用户通过计算中心点的坐标，将信息以明文的形式发送出去。

缺点：除了中心用户之外，其余用户的查询结果都无法得到保障。

方案2：Adding noise：向原来的数据中加入噪声来隐藏位置信息，但是在加噪声的过程中保证整个group中噪声之和为0，即中心不会发生改变。

- 缺点：如果用户的位置不改变，则中心用户就会在去掉噪声干扰之后找出该用户的真实位置。

方案3： 用户将数据加密后传输给中心用户，中心用户计算经过加密的用户位置平均值。随后将平均值和k加密传输给LBS

- 缺点：中心用户工作量太大，资源消耗太大。

Summary：保护隐私构建一个匿名框

- 匿名框的大小从一个侧面表示匿名程度

- 但是匿名框的大小也与用户提出服务所在位置的周围环境有关。

个性化位置K-匿名

k-匿名

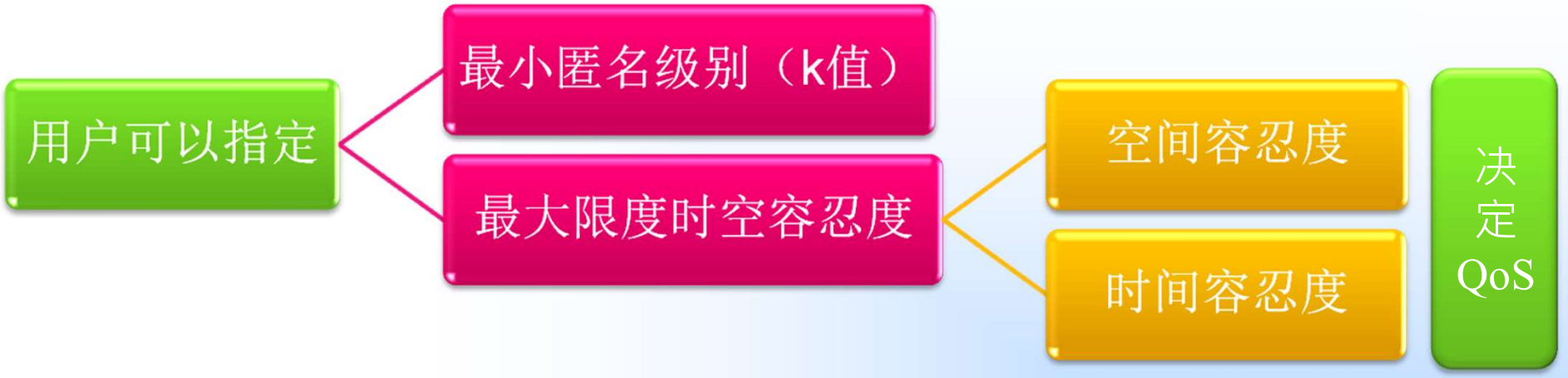
- 一条数据表示的个人信息和至少其他k一l条数据不能区分。

位置k-匿名

- 当一个移动用户的位置无法与其他k-1个用户的位置相区别时，称此位置满足位置k-匿名。

个性化位置K-匿名

- 以消息或者个人为单位，让用户自定义个性化的隐私需求，从而动态地协调隐私保护与服务质量的平衡点


用户可以指定
最小匿名级别（k值）
最大限度时空容忍度
空间容忍度
时间容忍度
决定QoS

4.5 匿名算法评价标准

（1）匿名成功率（Success Rate）

- 匿名成功率越高，匿名算法就越好。
- 匿名成功率可以定义为：成功匿名的消息数在所有移动用户提出的匿名请求的消息数中所占的比例，

（2）相对匿名度（Relative Anonymity Level）

- ✓ 相对匿名度是指用户实际匿名度与他所要求的匿名度 k 的比值。
- ✓ 相对匿名度的值一定大于等于 1。
- ✓ 一般情况下，认为较高的相对匿名度意味着更好的匿名效果，由于匿名与服务质量之间的平衡问题，较高的相对匿名度可能意味着较高的查询代价。

（3）相对空间粒度（Relative Spatial Resolution）

- 相对空间粒度表示匿名算法获得的匿名空间粒度的一个参数，其定义为匿名请求所定义的可容忍的最大匿名空间与匿名算法所获得的匿名空间的比。
- 相对空间粒度越大越好。相对空间粒度越大，说明在满足隐私需求的前提下，匿名空间越小，更接近最优解

（4）相对时间粒度（Relative Temporal Resolution）

- 它表示的是匿名算法时间粒度的一个参数。
- 其定义为匿名请求所定义的可容忍的最大匿名延迟时间与匿名框中时间维长度的比。
- 相对时间粒度越大，说明在满足隐私需求的前提下，较短的时间范围内完成了匿名，在时间轴上更接近最优解。
- 所以，相对时间粒度越大越好。相对时间粒度和相对空间粒度均大于等于 1。
- 相对匿名度、相对空间粒度和相对时间粒度反映的是服务质量。

（5）消息处理时间（Message Processing Time）

- 消息处理时间反映的是匿名算法的运行效率，它指的是一定规模移动用户的所有查询请求在多长时间内可以得到匿名处理。
- 当然，处理时间越短越好，说明了匿名算法的高效性。

一种基于聚类的位置k匿名解决方案

相关基本概念

k-匿名

- 一条数据表示的信息和至少其他k-1条数据不能区分[1]

LBS

- Location-Based Services 基于位置服务器

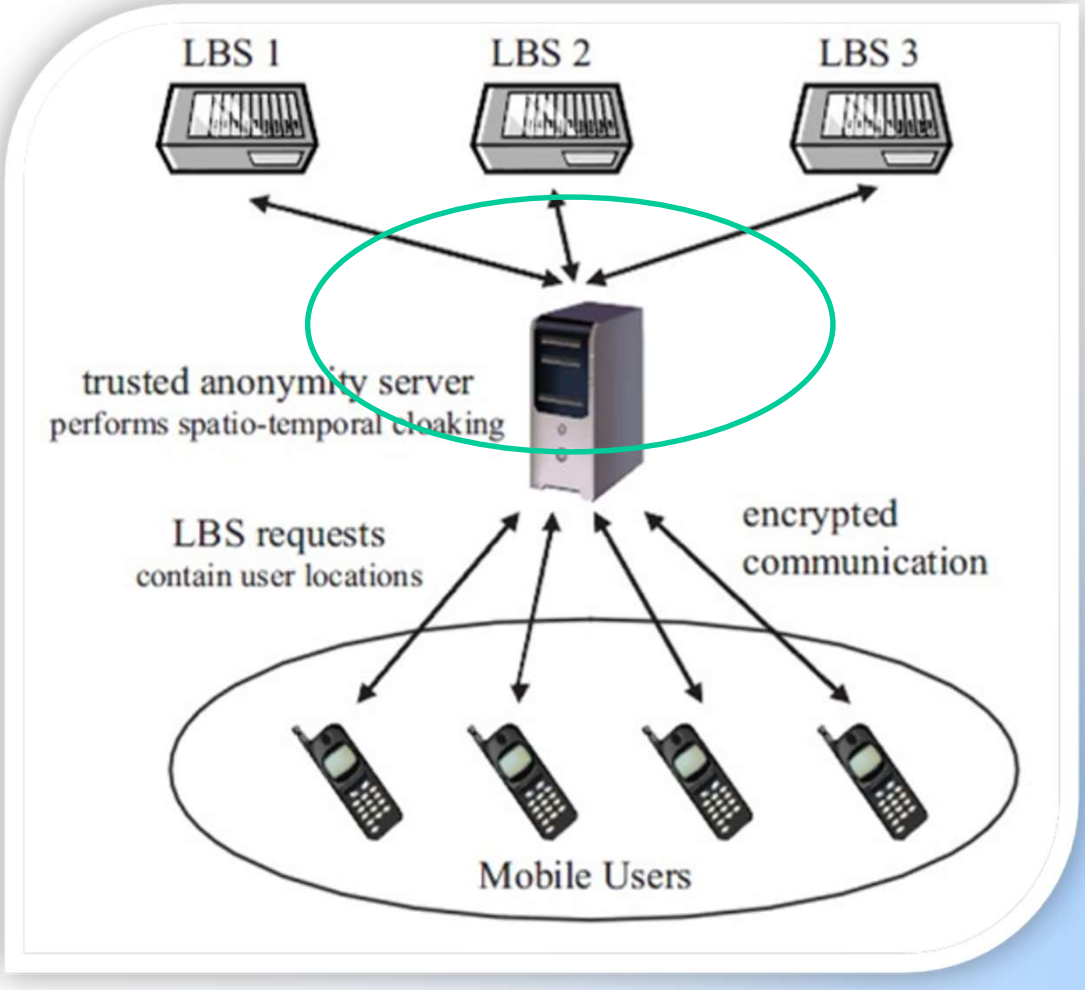
MBR

- Minimum Boundary Rectangle 最小矩形框

主要思想

- 聚类
  - 分成几个簇
  - 递归
  - MBR替代用户真实位置
- 大量移动用户
  - 区域内所有用户
- K-匿名
  - 用户指定
- OoS
  - 比用户期待的隐私级别更高

系统结构


LBS 1
LBS 2
LBS 3
trusted anonymity server
performs spatio-temporal cloaking
LBS requests
contain user locations
encrypted
communication
Mobile Users

可信任匿名服务器（trusted anonymity server）

- 作用：将从移动客户端收集到的信息，转化为一种安全地形式发送给LBS，即进行匿名处理。

位置信息混淆器（location perturbation engine）

- 用来实现上述位置隐私架构
- 运行在匿名服务器上的可信任第三方平台（TTP）上，对移动用户的请求信息在转发到LBS之前进行匿名化。

聚类算法（Clustering Algorithms）

$$m_s \in \mathrm{S} : \{\mathrm{u_{id}}, \mathrm{n_{id}}, (\mathrm{x}, \mathrm{y}), K, C\}$$

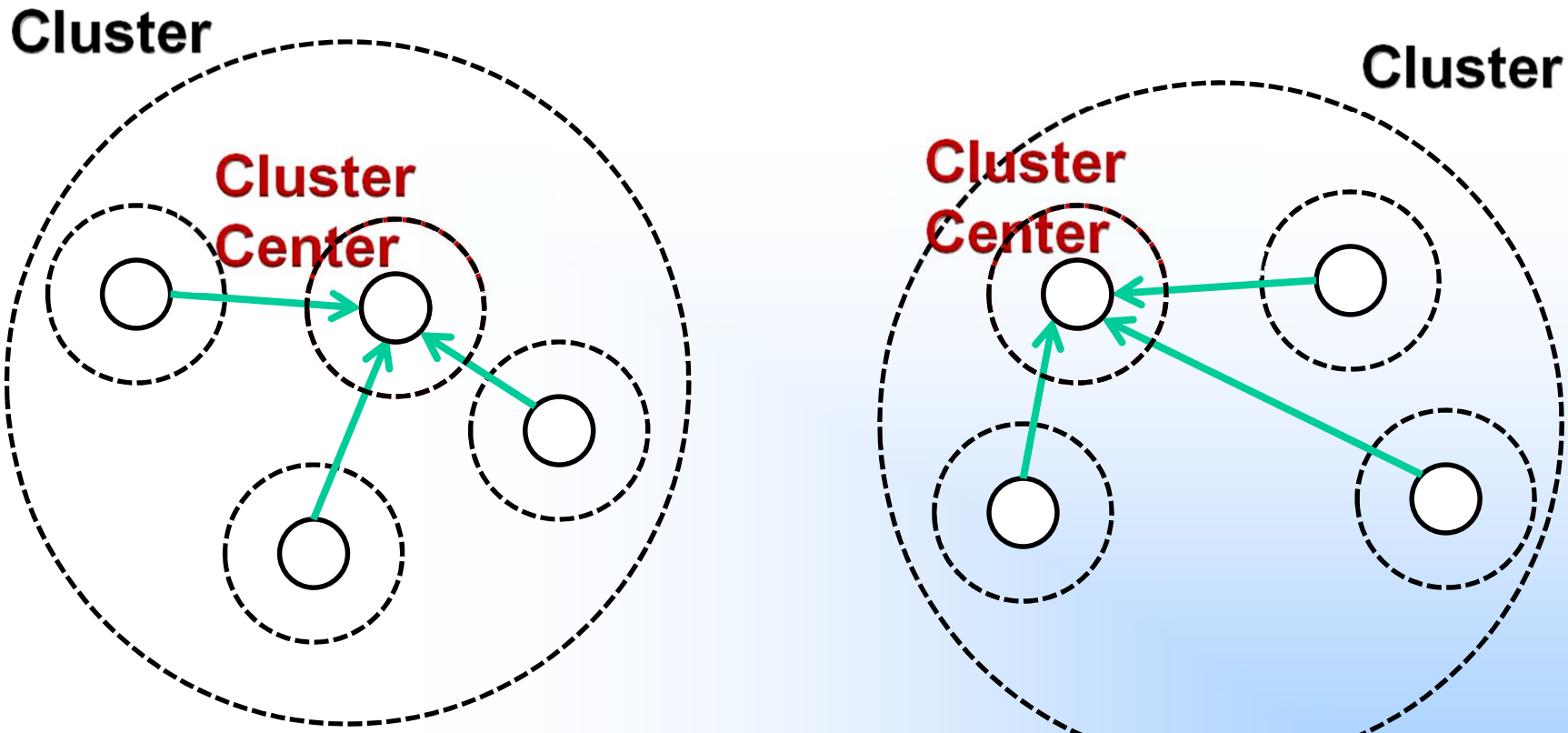
$$\mathrm{m_t} \in \mathrm{T} : \{\mathrm{u_{id}}, \mathrm{n_{id}}, \mathrm{X} : \phi(\mathrm{cx}, \frac{1}{2}\mathrm{W_{MBR}}), \mathrm{Y} : \phi(\mathrm{cy}, \frac{1}{2}\mathrm{H_{MBR}}), \mathrm{C}\}$$

建立簇

- 初始化簇中心
  - MN：MBR中选择垂直或水平方向最近的点
  - NR：一个点随机选，另一个距它最近
  - RP：两个点都随机选
  - RS：所有点水平分两部分，分别选其中一随机点

建立簇

初始化

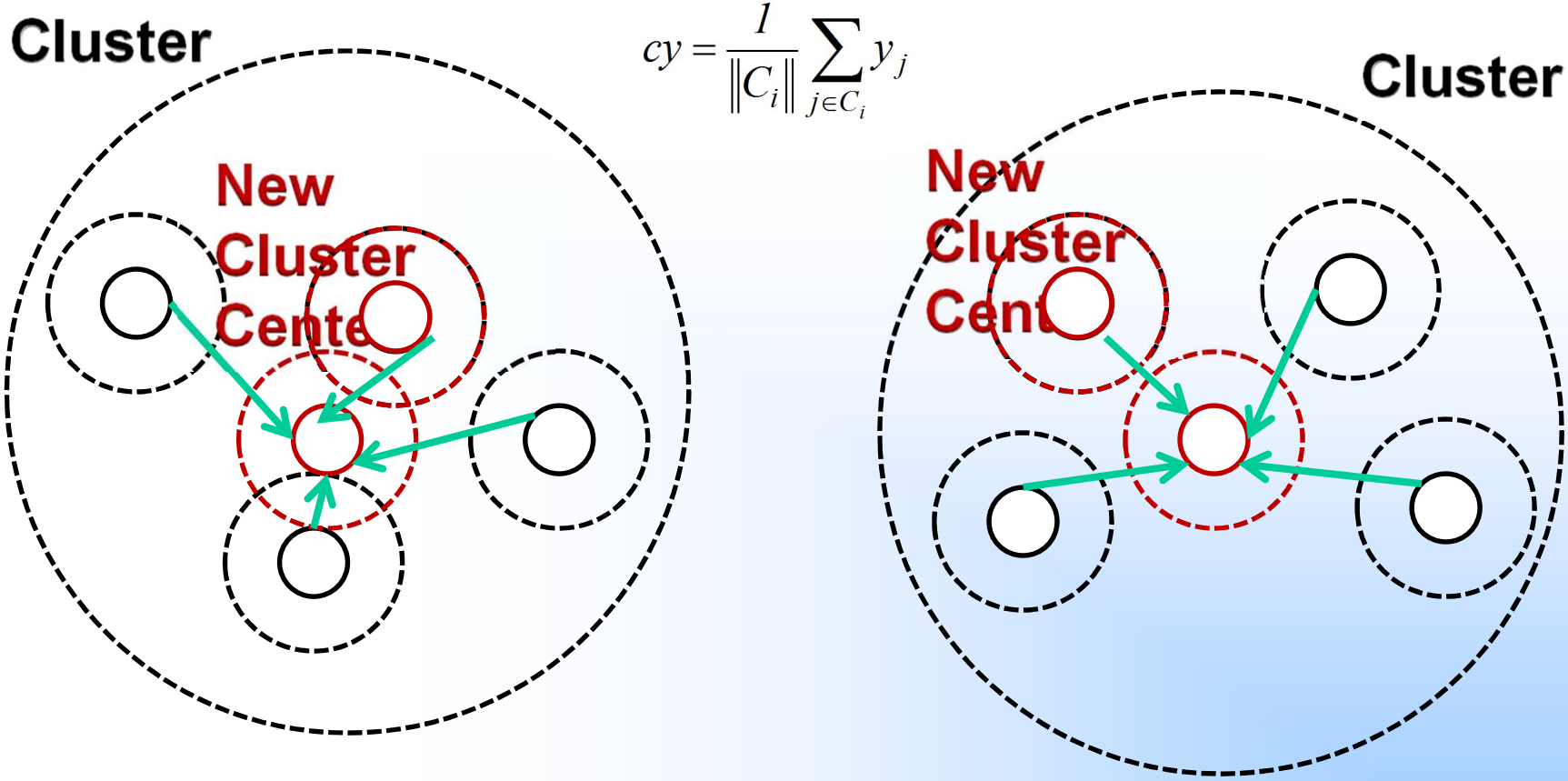


建立簇

递归调整

$$cx = \frac{1}{\|C_i\|}\sum_{j \in C_i} x_j$$

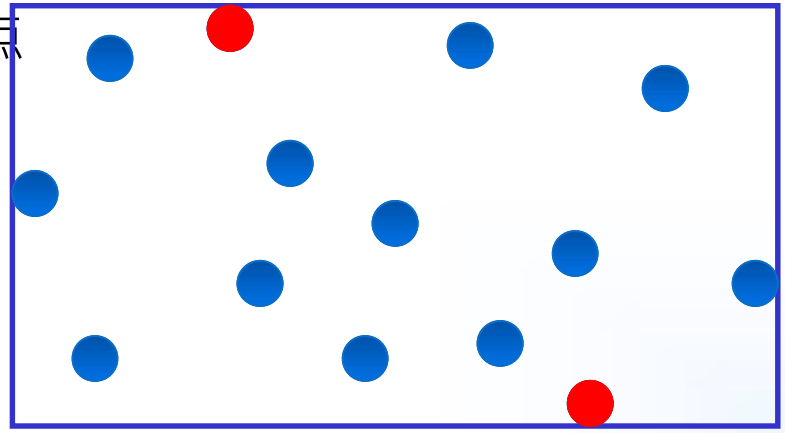
$$cy = \frac{1}{\|C_i\|}\sum_{j \in C_i} y_j$$



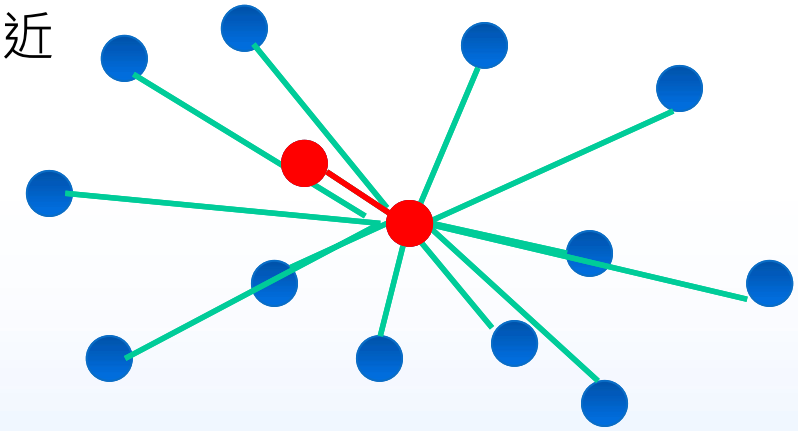
ClusterCloak算法框架

初始中心的选择

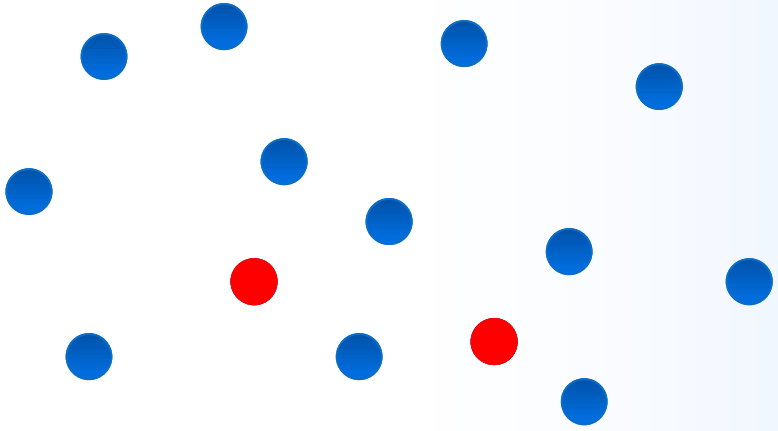
MN--MBR中水平或垂直方向最近的两点



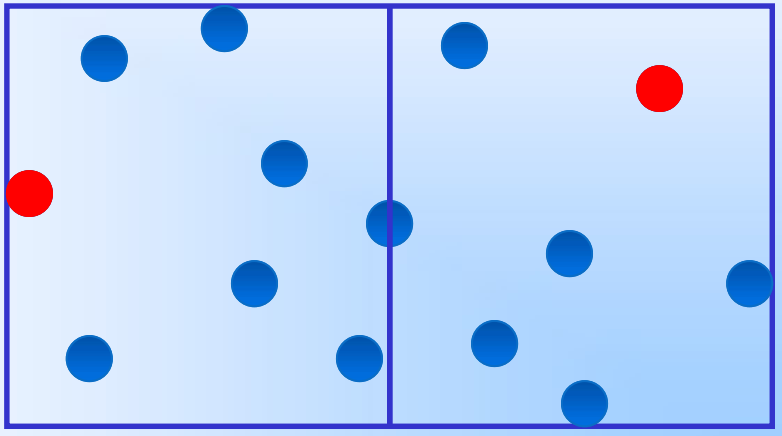
NR--一点随机选择，另一点距其最近



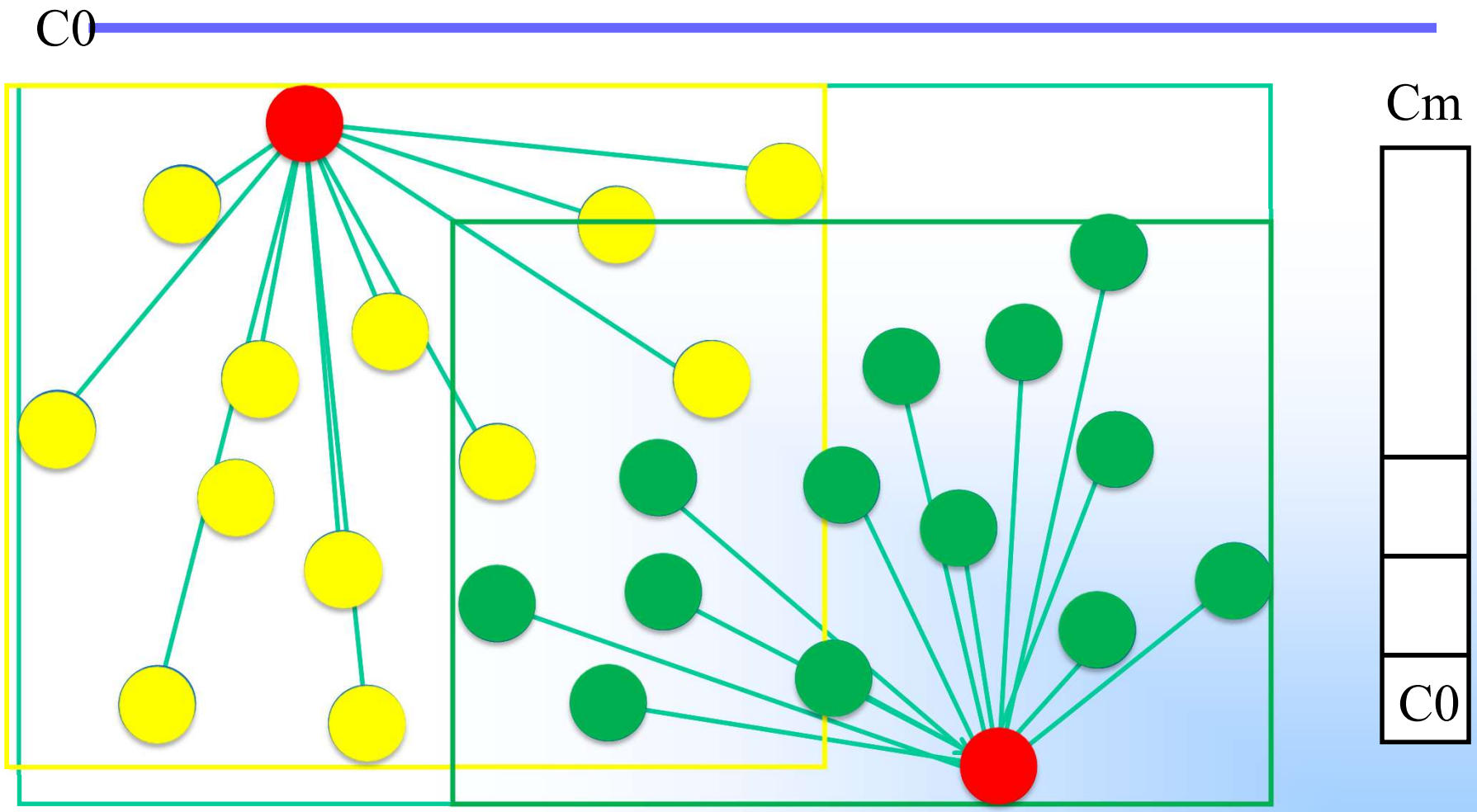
RP--两点均随机选择



RS—MBR水平分割，分别随机选择

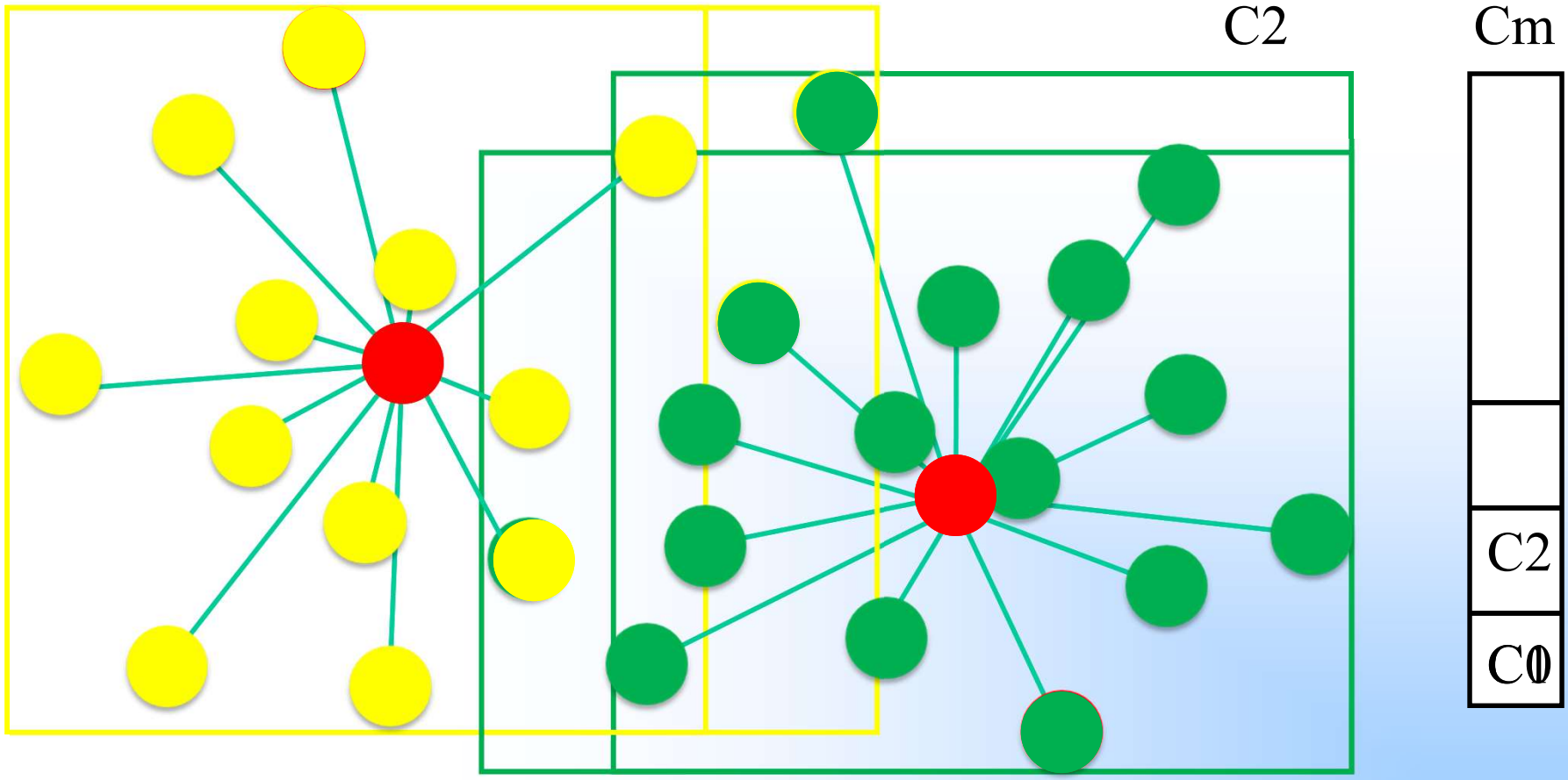


建立簇结构

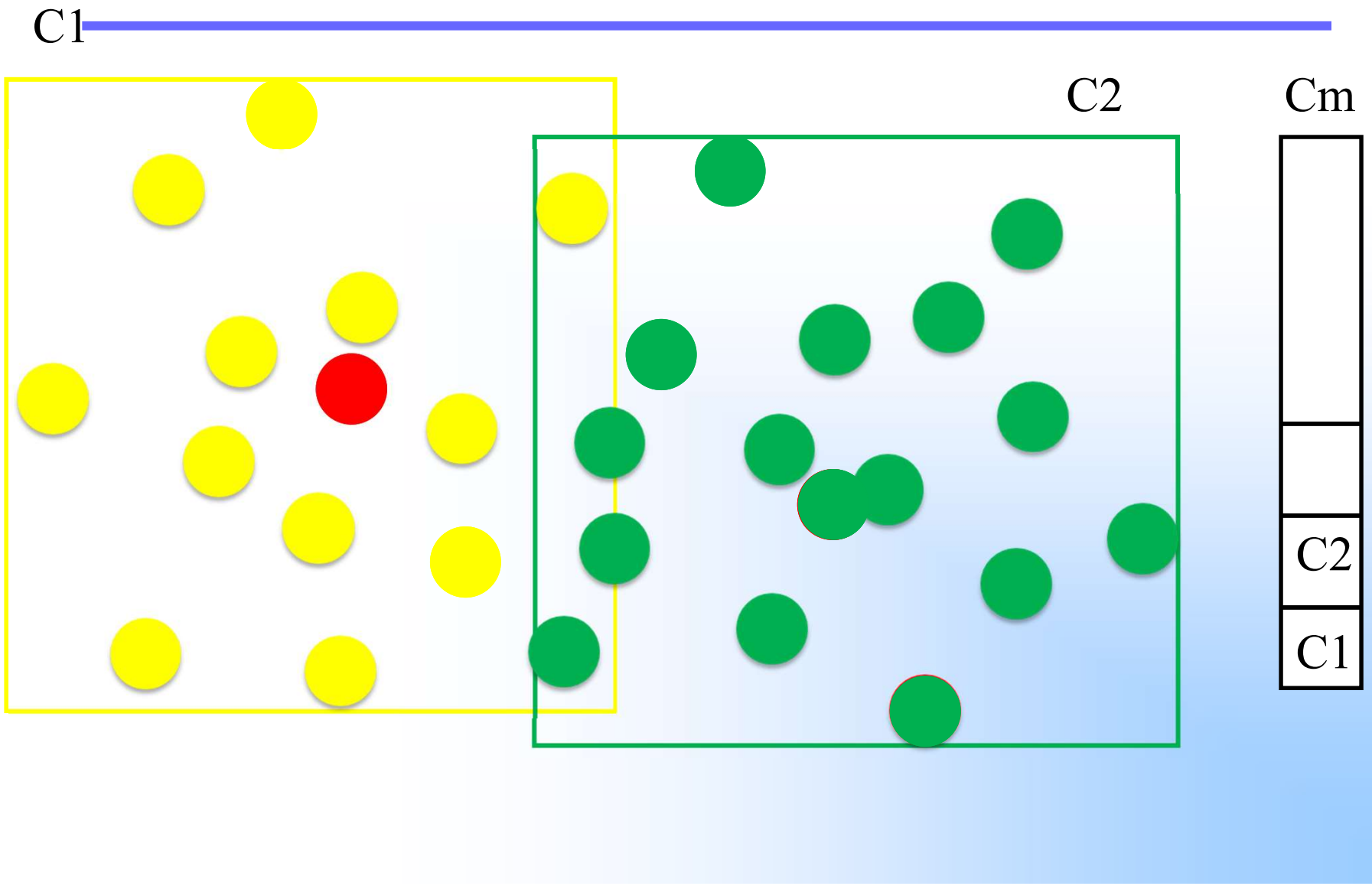


Advanced Network Technology

建立簇结构

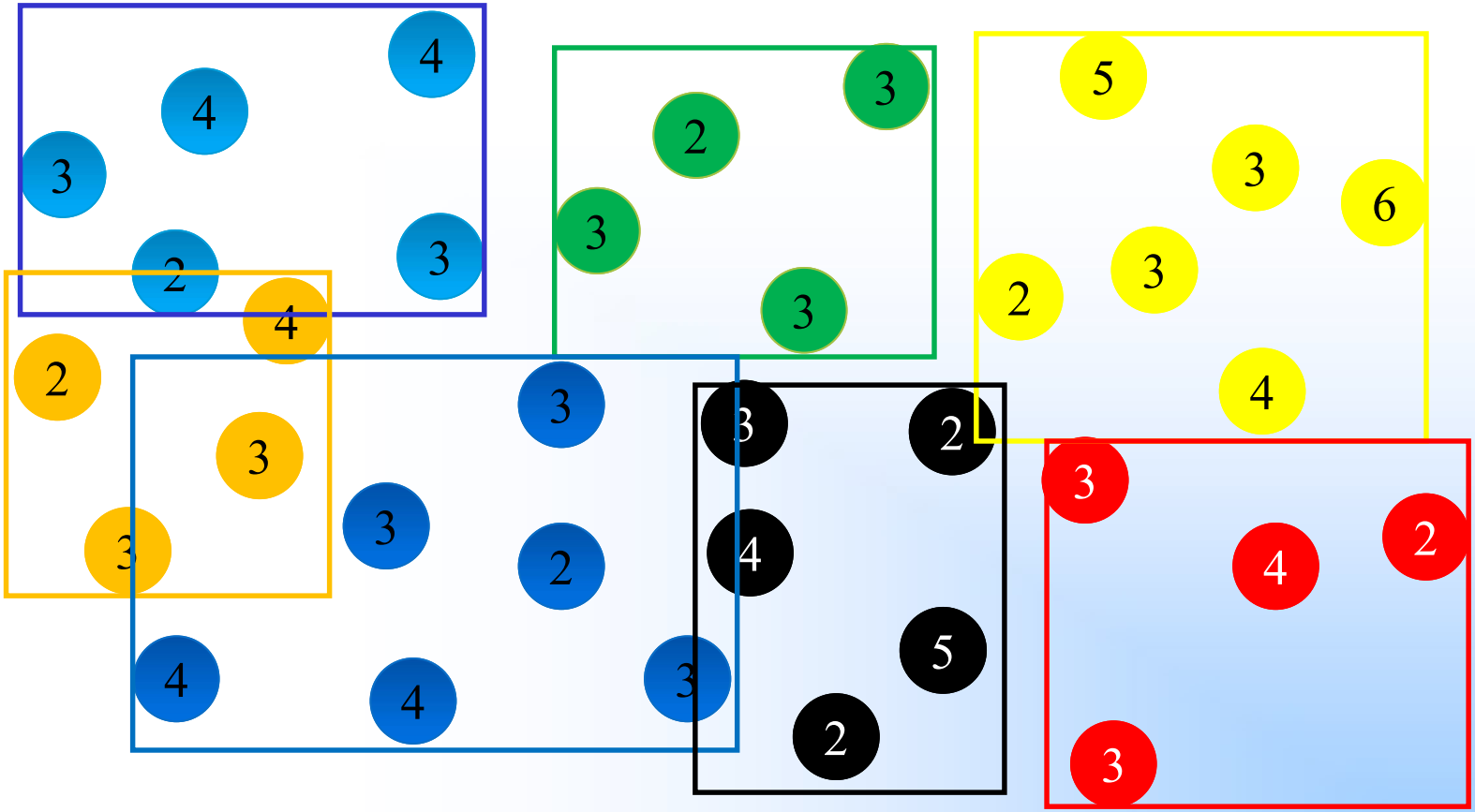


建立和调整簇结构

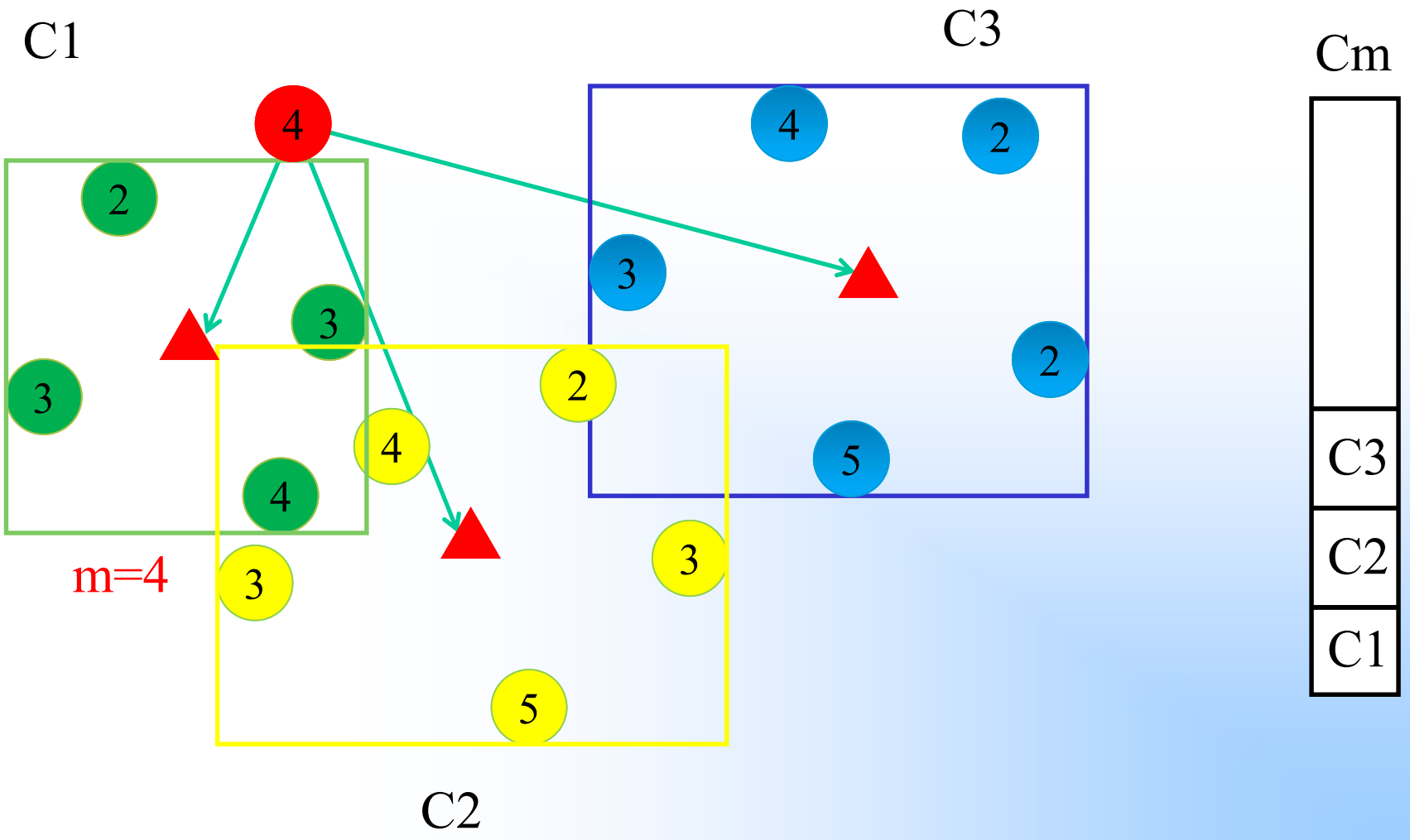


建簇的结果

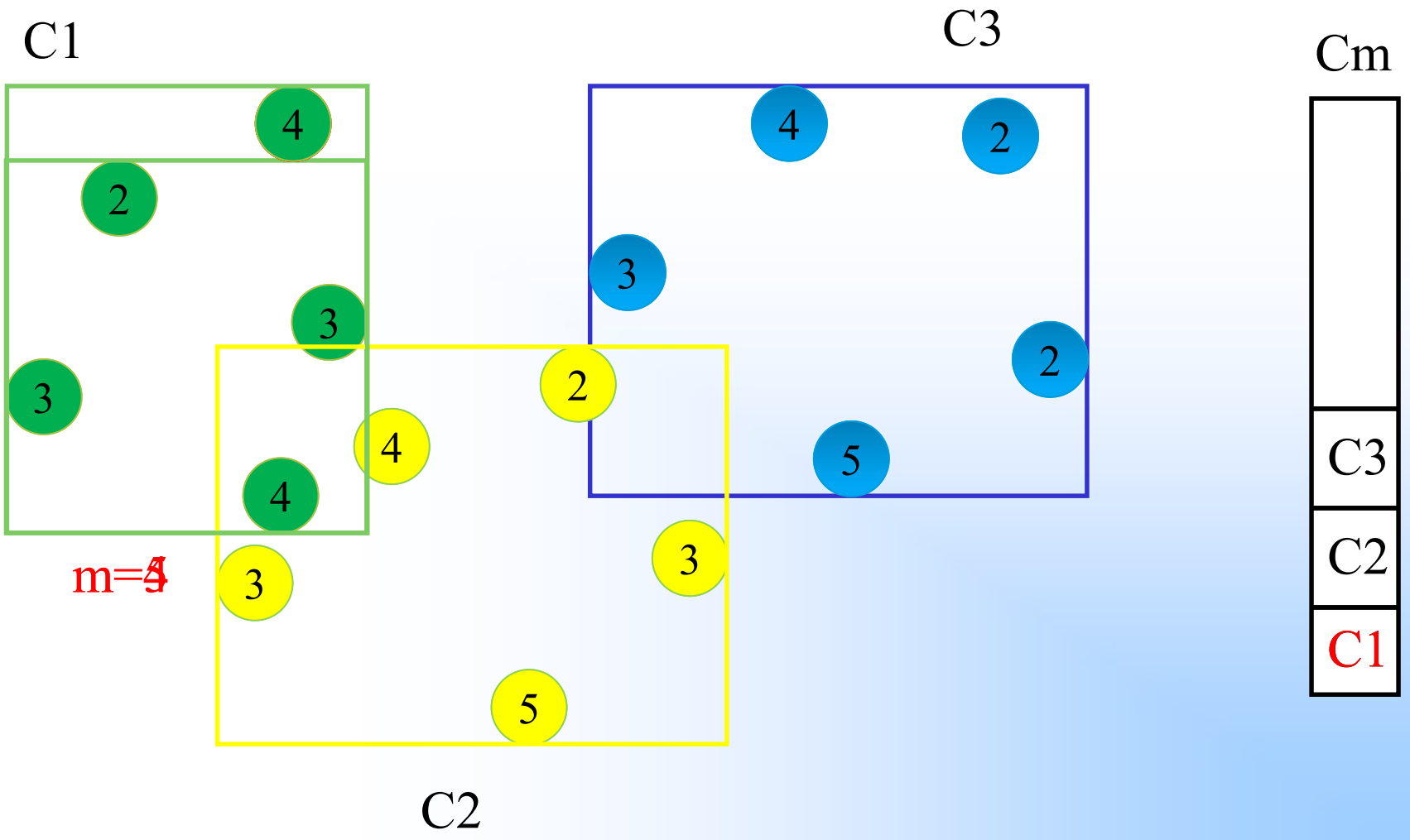
每个簇都保证：用户的数量大于等于该簇中最大的k值



簇结构调整—单用户加入

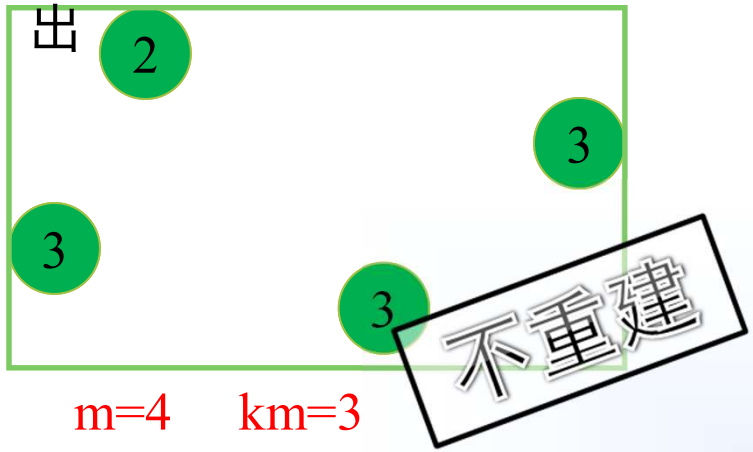


簇结构调整—单用户加入



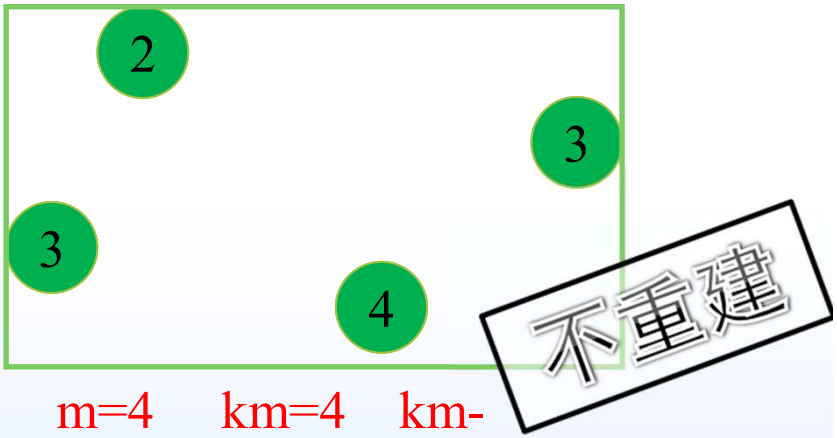
簇结构调整—单用户退出

m>km 任一点退出



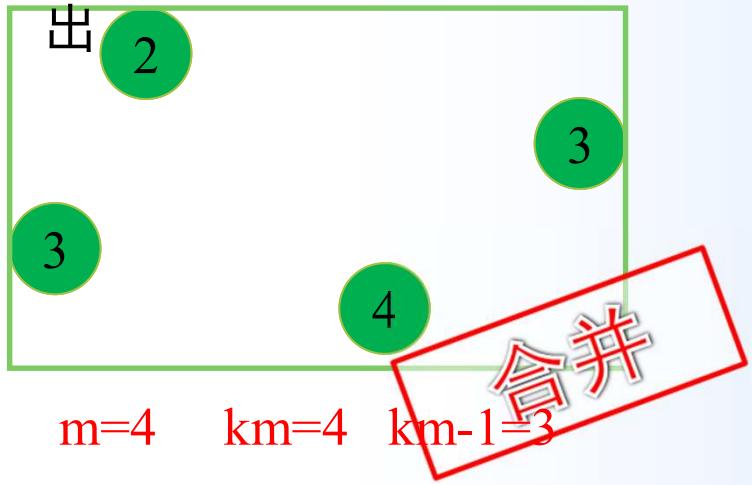
m=4 km=3

m=km>km-1 km退出



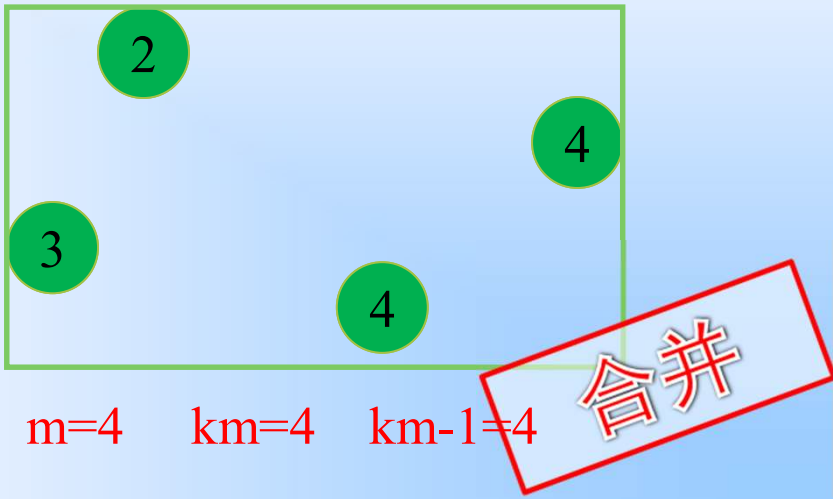
m=4 km=4 km-1=3

m=km>km-1 除km点退出



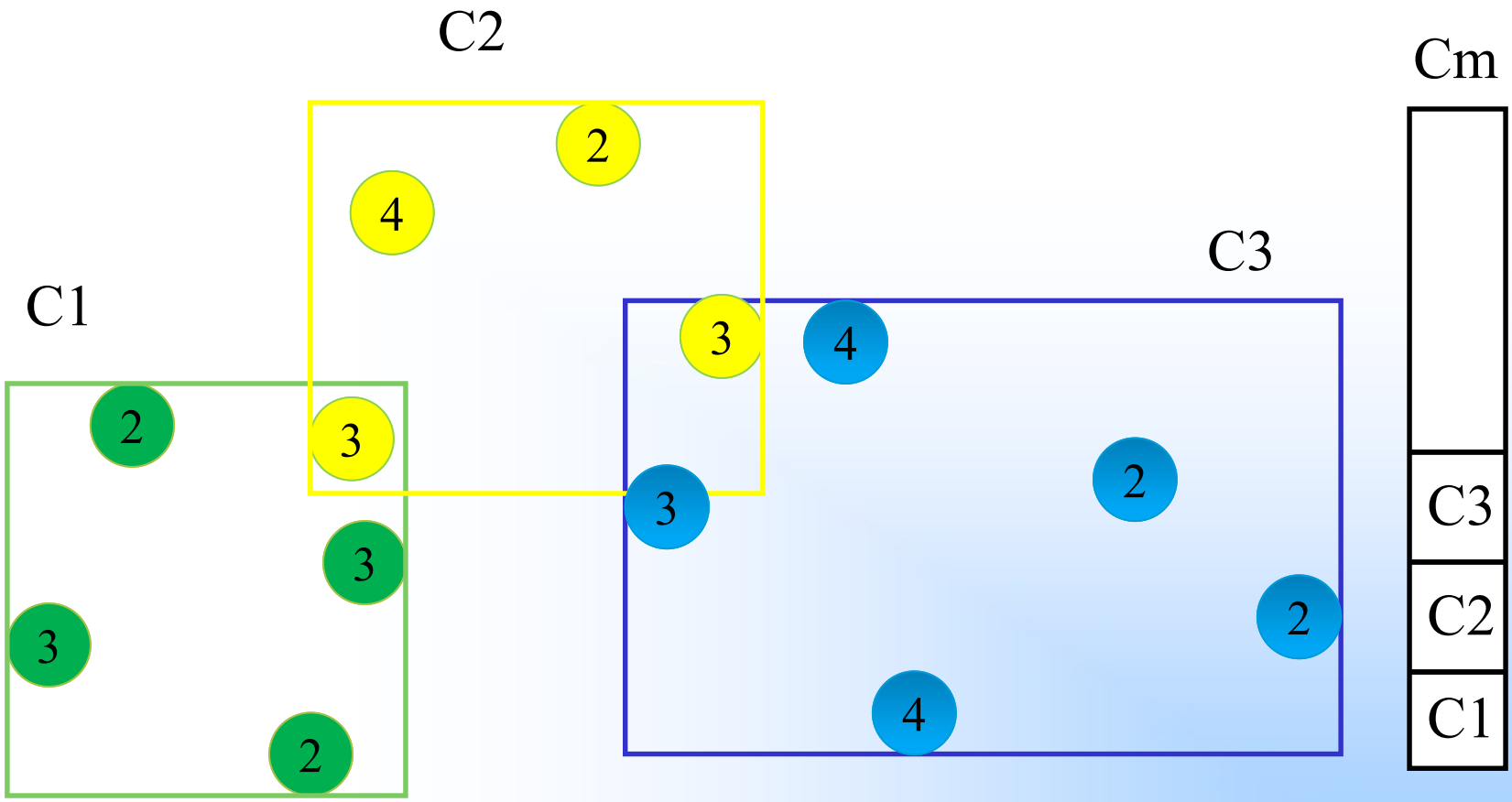
m=4 km=4 km-1=3

m=km=km-1 任一点退出



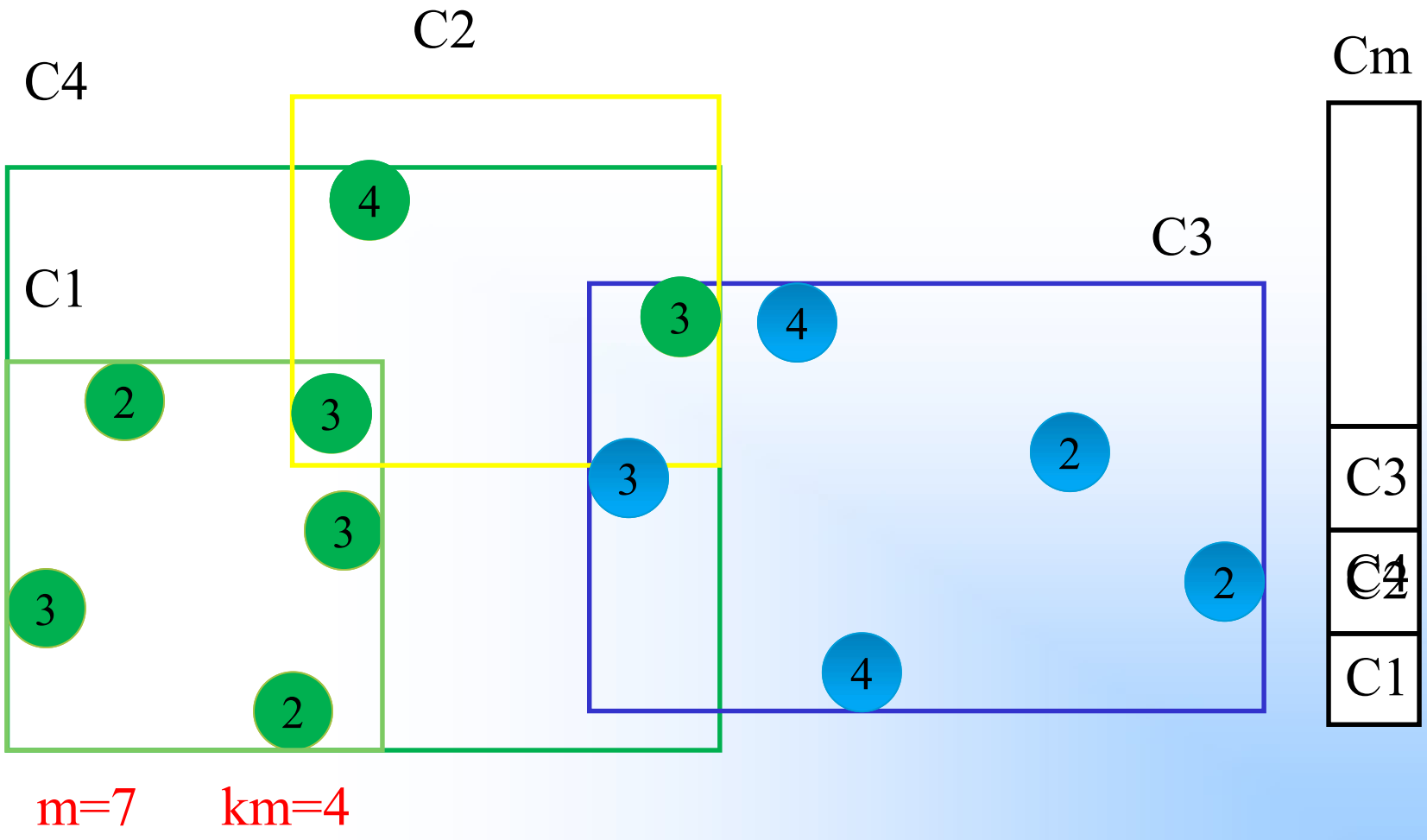
m=4 km=4 km-1=4

簇结构调整—簇合并

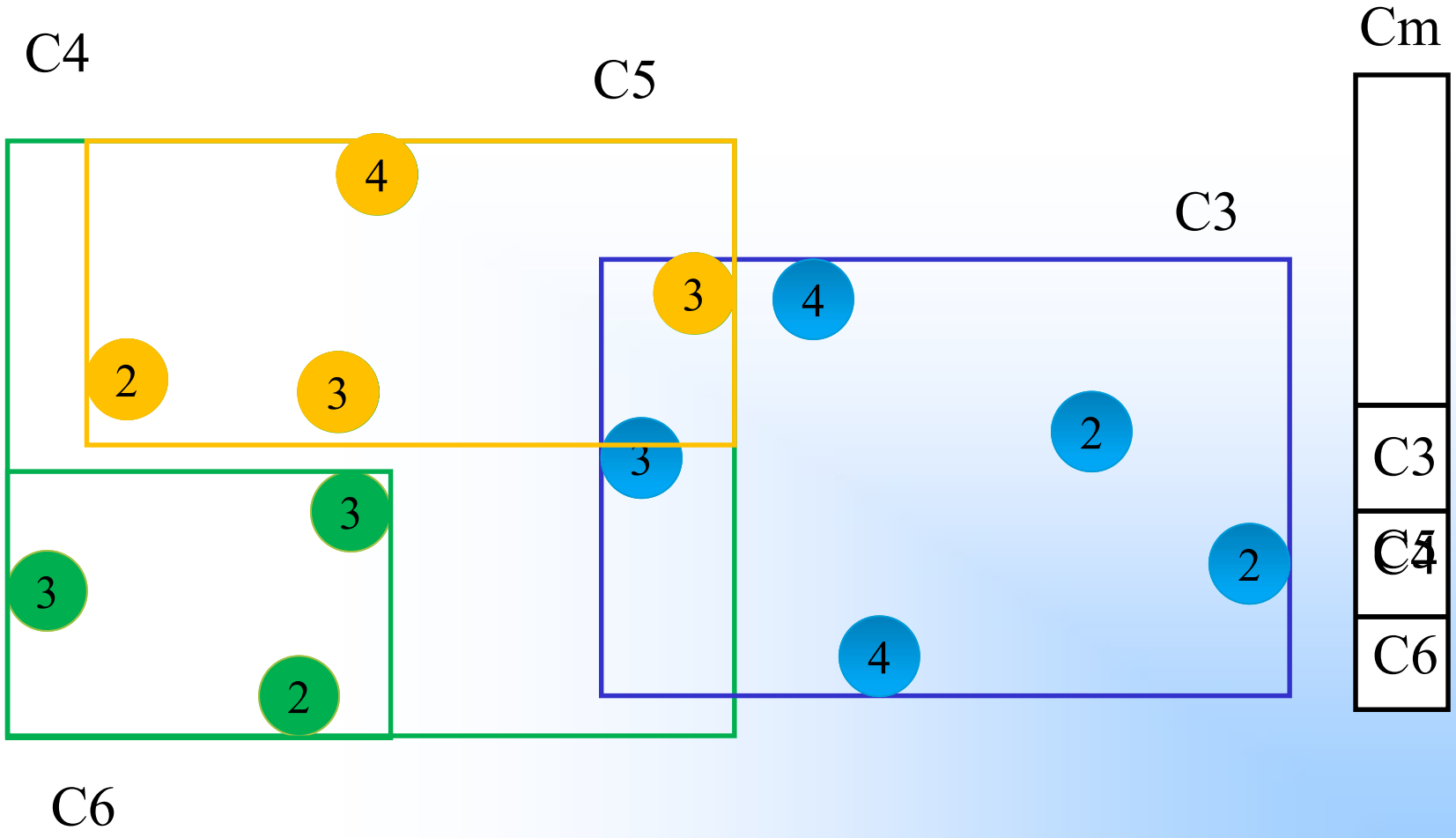


Advanced Network Technology

簇结构调整—簇合并

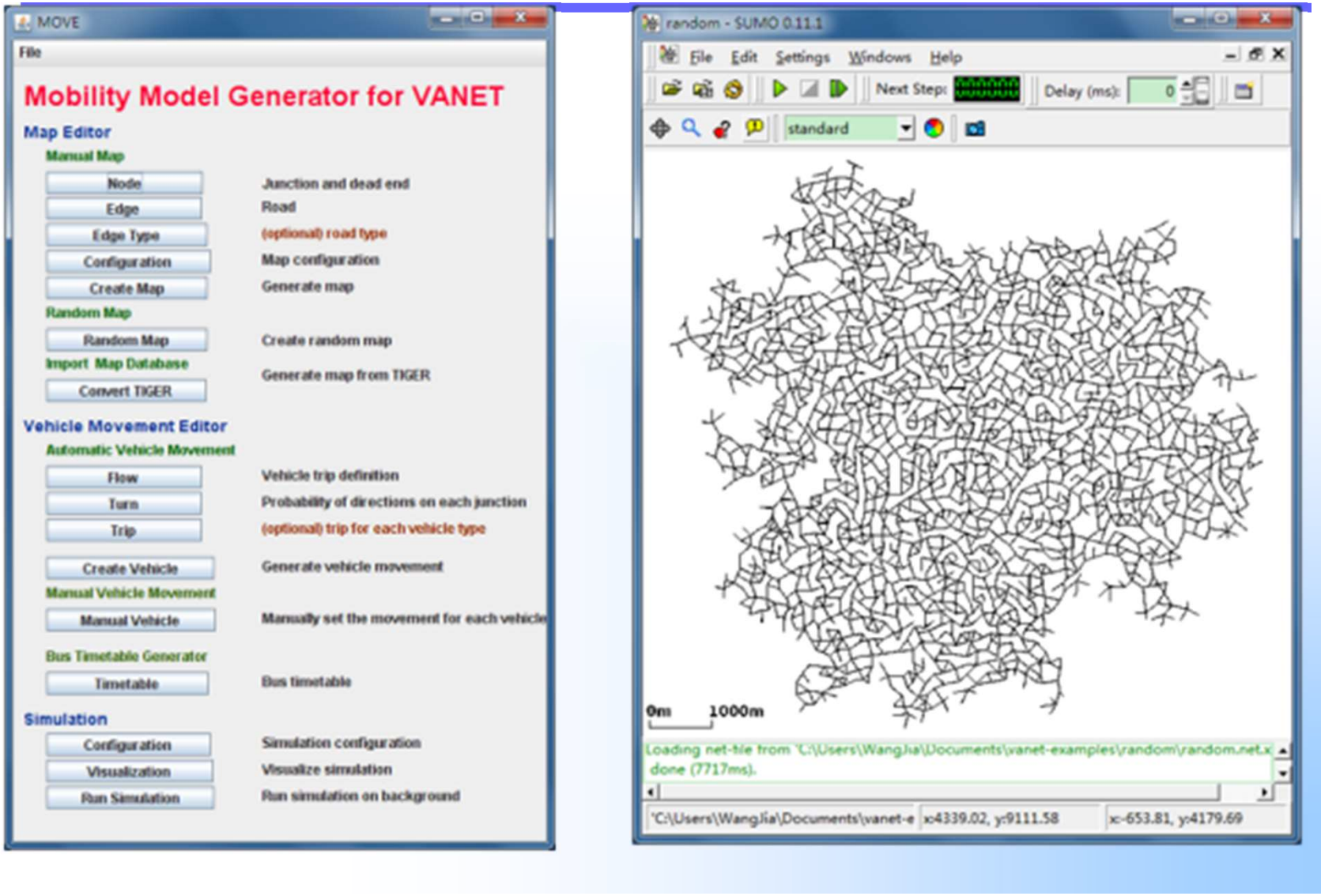


簇结构调整—簇合并



Advanced Network Technology

仿真环境



主要结论—算法分析

建立簇的复杂度

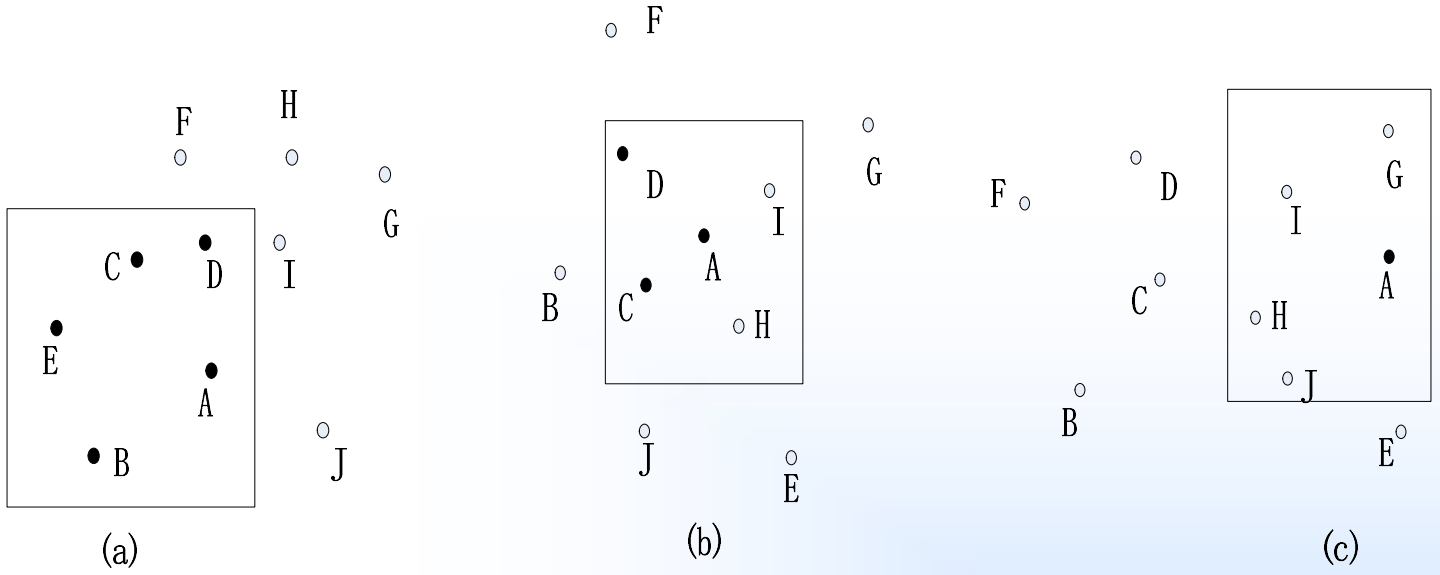
算法	复杂度
Nbr-k	$O(n^2)$
Local-k	$O(n^2)$
ARNN	$O(n^2)$
ClusterCloak	$O(n\lg n)$
Casper	$O(n\lg n)$
HilbertCloak	$O(n\lg n)$

簇结构调整的复杂度

算法	复杂度
Nbr-k	$O(n^2)$
Local-k	$O(n^2)$
ARNN	$O(n^2)$
ClusterCloak	$O(1)$
Casper	$O(n\lg n)$
HilbertCloak	$O(n\lg n)$

（1）匿名成功率
（2）相对匿名度
（3）相对空间粒度
（4）相对时间粒度
（5）匿名完成时间（算法复杂度）

4.6 问题讨论：Continuous Queries Attack



满足k-匿名？

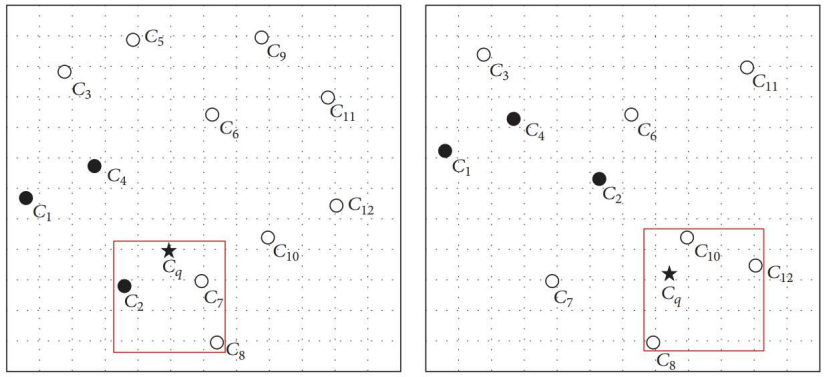
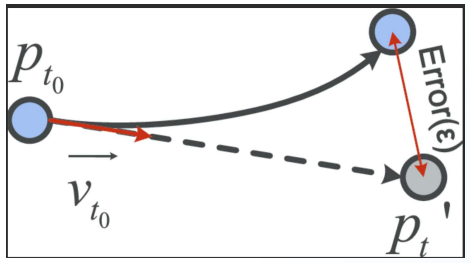


FIGURE 2: A K-anonymization problem related to continuous spatial queries.

解决方案

利用历史轨迹信息构建匿名区域. 在连续位置服务查询中，利用马尔可夫预测模型对每个用户下一个时刻可能到达位置进行判断，距离较近的用户为一个匿名区域；



灰色代表预测位置
蓝色代表实际位置

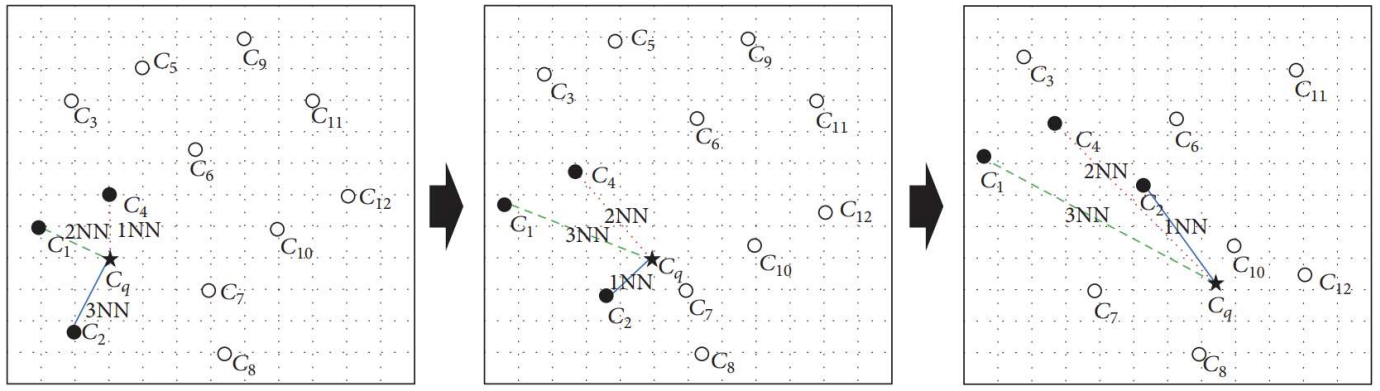


FIGURE 3: An example of adaptive-fixed 2-anonymization (C_N = 4).

谢谢聆听