# 3.1 System model

总体来说，Our system 主要涉及以下三个主体：

1. Recruitment Center **(RC)**: RC由三台半可信服务器组成，其中两台为普通服务器，一台为主服务器,p代表着platform。三台服务器间采用安全多方计算实现真值隐私保护。RC会为DR招募工人，并向工人发出感知任务，进行真值发现后，将数据报告回DR。RC会从DR处收取佣金，并选取合适的Worker派发酬劳，从中赚取差价，实现平台收益最大化。
2. **Data Requester（DR）**: DR将一组感知任务task打包为一个Recruitment request发送给RC，通过付出一定的佣金，获取感知数据，用于后续分析。
3. **Sensing Users (SU)**：***SU***是一大群注册在RC平台上的数据采集工人。他们通过一系列设备在执行特定的spatiotemporal sensing task 来换取酬劳。

In this paper, 我们试图构造一个隐私保护的RC，面对DR的一次Recruitment request，为其招募到特定数量和比例的SU，在确保真值可信的情况下，实现平台收益最大化。Recruitment request and Type of SU are defined as follows:

**Definition 1. （*Recruitment request，Task*）** RC允许DR发起k Recruitment request，denoted by . and each Recruitment request 包含distributed spatiotemporal Task ,which denoted by .

**Definition 2. （*Trusted SU，Untrusted SU, Unknown SU*）** RC will divide n registered ***SU*** into three set： trusted ***SU*** set , untrusted SU set and unknown SU set 。 包含着那些总是准确报告正确sensing data的可信任SU，则包含那些由于设备误差较大、骗取佣金等原因而不可被信任的SU， 则代表着未被系统识别分类的SU。

Our system will use trust-based score ——denoted by Degree Of Trust (DOT)——to区分这三类worker. After 每次Recruitment request，we 更新DOT，当DOT大于或小于某些threshold时，SU从 unknown ***SU*** set 被分到或中。DOT and threshold are defined as follows:

**Definition 3. (*DOT*)** DOT是RC用来评估SU是否可靠的指标。DOT越高，则代表该SU越可信。RC会根据每个***SU***在多个Task上的表现，为每一个SU在每次Recruitment request更新一次DOT。并根据DOT对SU进行分类，并由于后续的招募。DOT is denoted by , 其中代表着***SU***，代表着recruitment request 中的sensing task. DOT的具体计算方式会在接下来的文章中介绍。

**Definition 4. (*DOT Threshold*)** DOT Threshold是区别Trusted SU，Untrusted SU, Unknown SU的分界线。DOT Threshold包含两个值，and . 当时，则认为该SU为Trusted SU；当时，则认为SU为 untrusted SU；否则，该SU为unknown SU。

在这个系统中，DR会向RC支付定额的Brokerage。SU会将执行sensing task 后得到感知数据sensing data报给RC来换取compensation。RC对这些sensing data进行统计分析后，将最后的Estimated Truth Data报告给DR，完成此次招募。sensing data， estimated truth data，brokerage and compensation are defined as follows:

**Definition X. (*Sensing Data, Estimated Truth Data*)** sensing data is the value reported by ***SU*** who is recruited to execute a sensing task, which is denoted by , 其中代表着***SU***，代表着recruitment request 中的sensing task. ***Estimated Truth Data*** 是经过RC统计处理过的Sensing data，作为真值发现的结果发还给DR，which is denoted by ,代表着 recruitment request 中的 sensing task.

**Definition 5. (Brokerage)** Brokerage 是某一轮recruitment request 中DR支付给RC的commission。在本文中，我们假设brokerage是一笔仅根据sensing task数量变化的金钱，which is denoted by , 其中代表着recruitment request。

**Definition 6. (Compensation)** compensation是RC为SU 上报 ***Sensing Data***所支付的报酬，which is denoted by 。这里为了简化模型，本文假设每一次支付的compensation都是一样的，即当一名SU完成一次sensing task并上报sensing data时，RC会支付 compensation。

为了方便读者的阅读，we将以上符号列举如下：

# HERE IS A TABLE

|  |  |
| --- | --- |
|  | number for sensing tasks in request |
|  |  |
|  | number of recruitment requests |
|  |  |
|  |  |

# 3.2 Problem statement

在一般的MCS系统中，RC会招募大量的SU来完成DR提出的sensing task。而这些SU并不总是诚实的，设备误差、测量误差都会导致sensing data失真。Worsely，SU 中可能会混入一批恶意的用户，他们通过提交完全虚假的sensing data来骗取compensation。如果RC雇佣了这些恶意用户，将会浪费大量金钱的同时得到一个完全错误的 ***Estimated Truth Data***. 而这种情况可以通过识别并雇佣少数trusted SU来解决。

我们考虑一个存在一群untrusted worker的例子. As shown in Fig x(a), we assumption GTD of sensing task is 50. 假设我们雇佣了五位工人，并支付给了这五位工人每人3$ compensation. Four of them are untrusted worker, report sensing data respectively as 58, 65, 60, 75. While only one worker can be trusted, who report results 51. 设采用mean method as staistical analysis , we incorrectly get 61.6 as ***Estimated Truth Data*** while 花费了大量的金钱在不可信任的人身上。

但如果我们只雇佣少数但可信工人，say 三个用户，who report 49, 52, 51 as sensing data. 假设我们仍然以每名工人C = 3$的报酬雇佣他们，那我们只花费了15$.

# HERE IS A FIG

但我们需要注意的是，DR发布的Task往往要求收集SU的大量隐私数据。例如在消费数据收集场景中，用户上传的数据又可能泄露用户的消费习惯、消费额度和流水明细等。此外，不加以保护的感知数据也有可能因为被攻击而产生隐私泄漏。因此，如何避免真值发现过程中的隐私泄漏问题也成为本文的研究重点。

根据上述分析，本文试图建立一个联结DR和SU的 privacy-preserving Computations Recruitment Center。这个平台应该达到以下目标：

1. 隐私保护：平台应该是隐私安全的。这意味着，真值发现应该在一个分布式的服务器集群上实现。所有SU提交的sensing data 会被加密存储在不同的服务器上，当单一服务器被攻击时，攻击者无法得到有效的数据。同时，系统应该确保DR不能得到某一特定SU提交的原始数据，而只能得到经过platform处理过的统计数据。
2. 可靠真值发现：平台应该可以识别出untrused SU，通过排除这一部分用户的fake数据，来保证汇报给DR的***Estimated Truth Data***的准确性。
3. 平台最大收益：平台将以profit最大化为目标来运营。这意味着，当DR付出的brokerage一定时，RC需要在确保真值可靠的情况下，雇佣尽可能少的trusted SU 来实现成本的最小化，从而赚取尽可能多的差价。

我们设计了以下指标来衡量truth discovery的可靠性和平台的收益：

（1）truth discovery metrics

Before we define truth discovery, we need a benchmark to vertify the truth and reliable data. Here we have ground truth data(GTD)，witch is defined as bellow:

**Definition X.(GTD)** GTD是指在群智感知中通过验证和校准获取的准确的参考数据。在实验中，这可以被认为是一种已知的数据。we denote it as ，其中代表着recruitment request 中的sensing task.

to measure the the accuracy of truth discovery, we should calculate the gap between GTD and our ***Estimated Truth Data***. Root Mean Square Error (RMSE) is a 经典的指标，对于较大误差，RMSE越高。RMSE曲线可以很好的反映truth discovery 的质量。For particular recruitment request ， we define as bellow:

For all requests, we have:

(2) platform revenue metric

In this paper, we 将RC简单的看做一个收支系统。RC向DR收取Brokerage作为委托费用，向SU支付compensation作为激励。另外，我们设定作为维持RC运转的必要费用。Now we can have profit as defined:

**Definition X. (Profit)** The difference between brokerage and the sum of compensation and is called profit. specifically，we define it as:

# 4. The design and analysis of XXX

The high level design of the XXXX is depicted in Fig. X. RC response a recruitment request as the following nine steps:

(1) DR向platform发起任务需求；

(2) platform release sensing tasks；

(3) SU看到任务后，向platform申请任务；

(4) platform从响应的SU的工人中择优招募工人；

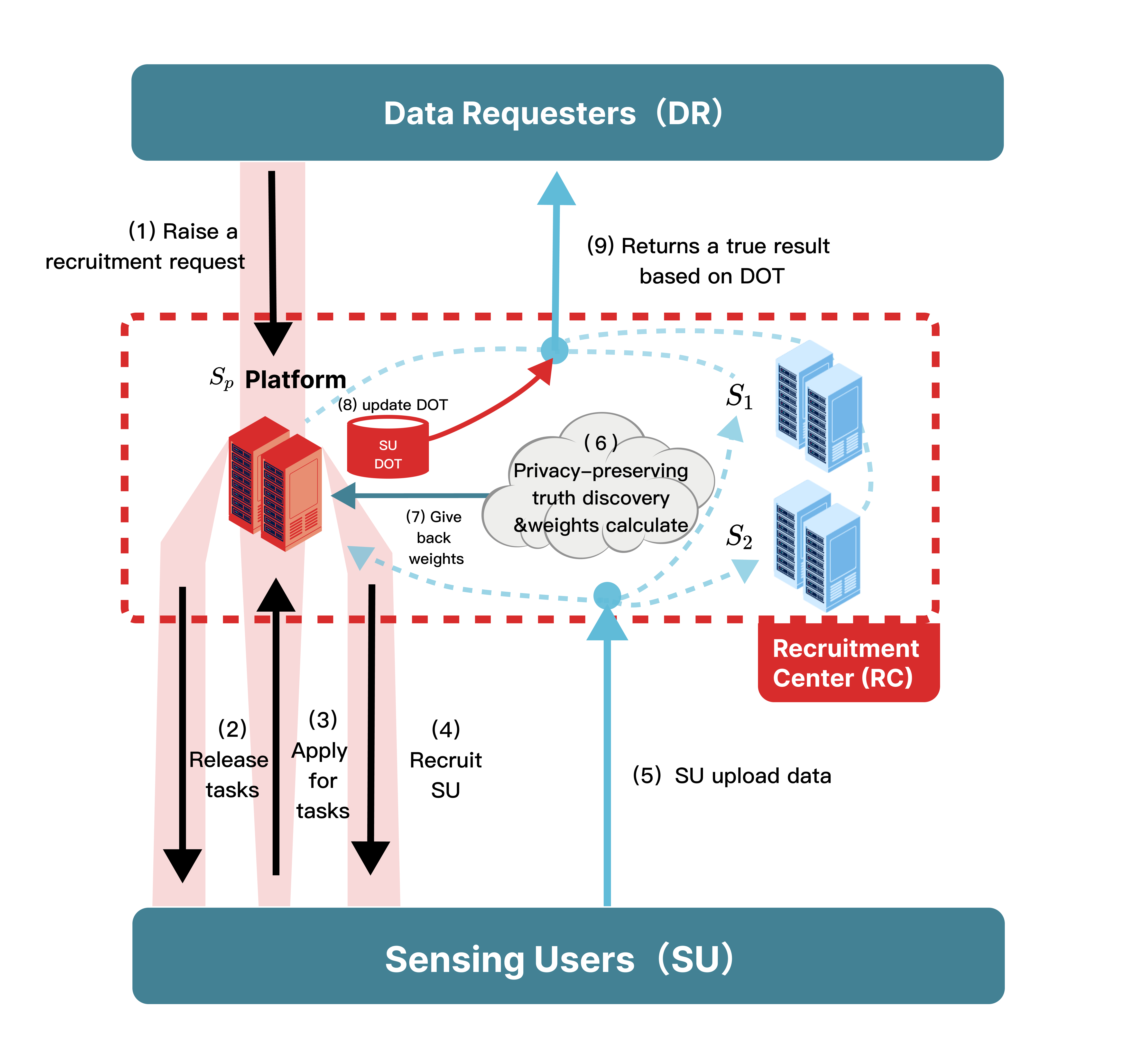
(5) SU完成sensing tasks 并 upload sensing data to RC. These data will be divide into three shares storing in three servers.

(6) 基于Privacy-Preserving Share Computing Protocols，三个服务器协作完成权重计算；

(7）权重被汇总到特别服务器platform上；

(8) platform得到权重后，计算每位SU的DOT，然后在维护的数据库中更新SU的DOT，并重新区分SU到中，为下一次工人招募提供更准确的帮助。

(9) 根据得到的DOT，修正得到的权重，并依据分布处理的truth share，向DR报告统计后的***Estimated Truth Data***。



### 4.1 一般的privacy-preserving computations truth discovery

Zhong. and Sharemind [XX}，提出了一种面向隐私保护的多方安全计算真值发现协议。

#### 4.1.1 data upload

After users 完成了sensing tasks, their sensing data will be divide into three addictive secret shares, and then upload each part to 对应的 servers. Say RC recruit SUto fulfill sensing tasksin recruitment request, and the SU gets data as . For particular sensing data , XX generate two random numbers ,, and calculate three secret sharesas:

Then we 分别send secret shares to servers , , . Each of server have a local database for persistent storage.

#### 4.2.2

### 4.2 基于DOT改进的truth discovery

前面那个方法不好使，后面做改进

### 4.3 平台收益最大化的雇佣机制

尽可能的招trusted SU，并计算的得到的方差，如果方差足够小，就只招这些人。 但是有一个上限不能高于多少人。

并且，在每次招聘中都选用一定数量的unknown SU。这个比例要调参可能。

特别情况下，

如果该格子没有Trusted Worker的话，系统会雇佣旁边N个格子的可信用户来进行插值

# 5. Experiment analysis

为了进行实验，we design a background data generation simulator。

对比算法：

数据可信度-对比一般方法

收益：对比随机找人