

并行计算调研

美国行星学会在 1999 年发表一项公告，呼吁因特网上的天文爱好者参与寻找地球外文明的科学实验。这个项目是美国加州大学伯克利分校(UC Berkeley, UCB)有关“搜寻地球以外智能”(Search Extraterrestrial Intelligent 简称 SETI)四个研究项目中的一项目，其全称是“在家中搜寻地球以外的智能”，缩写为 SETI@home。SETI@home 简单地说是一项旨在利用连入因特网的成千上万台计算机的闲置能力“搜寻外星文明(SETI)”的巨大试验。每一个参加者可以用下载并运行 SETI@home 屏幕保护程序的方式以自己的计算机参与检测外星文明信号的活动。

SETI@home 的工作原理

SETI@home 的工作由数据收集->数据传送->数据分析及回收->数据后处理->信息发布组成的。

1. 数据收集是通过波多黎哥国家天文和电离层中心建立在群山森林环抱中的、直径为 305 米(其面积相当于 26 个足球场大小)的巨型 Arecibo 射电望远镜进行的。Arecibo 将每天观测到的大约 35 GB 的数据记录在海量数字磁带上，并通过卫星传回 UCB。整个 SETI@home 项目的太空观测约需要 1100 盒数字磁带以记录 39 TB(terabytes, 1TB=1000GB)的数据。
2. SETI@home 把从 Arecibo 收集到的数据，经过计算分析之后根据客户的需要和电脑的情况，划分为小的工作单元即数据块。工作单元通过因特网传送到全球成千上万个客户端以进行数据处理。
3. SETI@home 传送数据结束后将自动切断连接，客户电脑便在 SETI@home 屏幕保护运行时开始对数据进行处理；SETI@home 应用程序对工作单元中的数据完成快速傅立叶变换的计算，其中大约要进行 1750 亿次运算，当一个工作单元分析完毕，闪烁的小图标便会提示客户回送并下载新的数据。
4. 所有客户端所获得的有价值的信号都将送回到 SETI@home。绝大多数客户端软件所找到的信号都是来自于地球的无线电频率干扰(RFI)，SETI@home 使用一大批算法和已知电信频率干扰资源的大数据库(SERENDIP IV 数据库)的数据来对比，从而排除所有可能的 RFI。对于极少数(可能只有<0.0001%)未被排除的信号，则将通过下一次观测太空中同一部位进行检测，如果该信号被再次确认，SETI@home 将要求给定望远镜使用时间，并再次观测这一最令人感兴趣的信号！假如一个上述信号被观测到多次，并确认它不是 RFI 和测试信号，SETI@home 将要求其他的天文研究组织使用不同的射电望远镜、接收器、电脑等再进行探测和辨识、确认。
5. 一旦信号被确认，SETI@home 将按照国际天文学联合会(International Astronomical Union, IAU)的电报发表公告，这是天文学界取得重大发现时公之于众的一种标准方式。而用其屏幕保护程序找到该信号的人(人们)，并将和 SETI@home 队伍中的其他成员一起被赋予“合作发现者”的称号。

实际上，SETI@home 是一次借助于因特网开展的大范围并行计算技术应用，那么什么是因特网上的并行或分布式计算呢？并行计算或分布式计算技术，一般是指用由成百上千个微处理器组成的大规模并行计算机系统或者用分布式计算机网络系统进行大任务数据处理的技术。并行和分布计算技术自 60 年代中期及 70 年代后期分别出现以来，一方面其并行处理方式已从阵列机(SIMD)、向量机及向量并行机、共享存储的对称多处理器系统(SMP)、以及近年来较热门的分布存储的大规模并行处理系统(MPP)逐步转向可伸缩并行机(Scalable

Parallel Computers)和各种类型的计算机机群系统(Clusters)。另一方面，在用通信线路连接的多计算机组成的分布式计算机网络系统中，并行和分布计算的应用也在日益增加。因特网属于分布式计算机网络系统之一，它是集计算机、计算机网络、数据库、多媒体以及分布计算模式于一体的一个网络综合体。因特网打破了时域和地域的局限，可以较低的费用充分调用散布于全球任何一个角落的可提供的 CPU 和内存资源。

在因特网上进行并行或分布式计算，一般是将任务的数据由安装了大型数据库的服务器使用根据特定算法编制的软件进行分割，然后分发到参与任务的多个客户机，客户机用户应用数据处理和分析软件去进行局部的数据处理；每一个客户机完成整个任务的一部分或多个任务中的相关部分，(这里客户机用户可以不用知道数据处理或分析的原理和过程)，任务完成后回送到服务器，再由服务器进行归并和进行更进一步的计算和综合分析以探求所需要的结果。

购置：

需满足如下要求。

大于 39TB 的硬盘存储数据。

至少 1750 亿次计算。

在 zol.com.cn 上查询 raid 相关资料，可以使用如下配置：

LSI Mega RAID SAS 9271-8i。售价约 ¥ 2000。

东芝(TOSHIBA)P300 系列 8TB 7200 转 64M SATA3 台式机硬盘(HDWD130)，购置 8 块，售价 $500 \times 8 = \text{¥} 4000$ 。

Intel XEON E3-1231V3 处理器，主频 3.4Ghz。售价约 ¥ 2000。需购置 $1750/34 = 51$ 个。总价约 ¥ 102000

综上，所有配置共计花费在 11 万元以内。