用于动态服务均衡的抗扰动集合

Jin Sima，**Chao Pan** 和**Olgica Milenkovic**电气与计算机工程系，伊利诺伊大学厄巴纳-香槟分校，美国 {jsima,chaopan2,milenkov}@illinois.edu

arXiv:2303.12996v1 [cs.IT] 23 Mar 2023

摘要——**平衡且交换鲁棒的极小交易于 [1]，中提出，对于研究数据流行度变化下服务器访问请求协议的均衡性和稳定性至关重要。迄今为止，这类交易的构造依赖于通过迭代组合较小集合而得到的成对集合，这些较小集合具有可证明的稳定性保证，同时还需要进行穷举计算机搜索。目前，所得总动态均衡差异与已知理论下界之间仍存在不可忽视的差距。我们给出了流行度变化受限时总服务请求差异的新上界和下界。我们提出的构造性近最优方法采用了一类新的成对图，其顶点是两个平衡集合，边（弧）则刻画了受限幅度的流行度变化（交换）所引起的均衡状态及其潜在变化。**

I. 引言

在现实世界的分布式存储系统中，人们采用服务（访问）请求控制方法来平衡对服务器的请求，并防止出现服务时间瓶颈 [2]–[6]。平衡服务的一种方法是将不同的文件组合分配给不同的服务器，以确保在发生磁盘故障时能够高效地重构/再生内容，同时平衡存储在各服务器上的文件“平均流行度”。为执行这种分配，文献 [2], [7], [8] 提出了一类新的 Steiner系统及相关组合设计，称为 *MinMax* Steiner系统。除了设计中常见的区块交集约束外，还必须根据流行度值为区块中的元素制定标记规则，以确保服务器访问均衡。

在实践中，数据流行度会随时间变化，而每次流行度变化后都在服务器间重新分配文件的成本很高，尤其是在变化幅度较小时。因此，设计组合设计，使得在数据流行度发生小幅扰动时，各服务器平均流行度得分之间的差异仍受到严格限制，这一点十分重要。一种可能的解决方案基于组合交易 [9], [10], ——这类数学对象在编码理论文献中鲜为人知。顾名思义，交易允许交换一组违反流行度约束的问题区块，并以能够减小服务请求差异的区块取而代之。交易以精巧的组合构造为基础，其中一种构造见文献 [11] ，之后在文献 [1] 中被改编用于当前问题。

文献 [1] 介绍了如何构造均衡且具备流行度交换鲁棒性的交易，以确保服务器访问频率在数据流行度变化幅度有限时保持稳定。具体而言，作者证明了可以通过谨慎选择所谓的“定义集”，构造阶数随整数参数 t 线性增长的均衡且交换鲁棒的交易；此外，还可以构造此类集合，并为总流行度差异给出可证明的上界。该方法依赖于对较小的最优定义集进行精心的递归分组，以及穷举计算机搜索。然而，由于问题具有复杂的组合性质，穷举搜索所能解决的最大实例规模为 24（即 24 个流行度评分各不相同的文件），这导致流行度差异的理论下界与上界之间存在不可忽视的差距。

本文的贡献有两方面。首先，我们提出一种递归构造（第三节），在幅度为一的流行度变化下，可实现显著更小的总差异，$_{8\,5}(t-\frac{1}{4})$，优于目前报告的最佳差异 $2t+O(1)$。其次，我们建立了一个显著更紧的下界$_{3\ 2}(t-\frac{2}{3})$（第四节），用于刻画集合总差异，相较于 [1]中证明的下界$_{2\,3}(t-\frac{1}{3})$更为紧密。我们的证明技巧采用了新的一类成对图，它们能够同时描述服务器的均衡状态和差异，以及流行度置换可能导致的失衡增幅。

II. 预备知识

受篇幅限制，我们不提供访问均衡协议构造的细节，而是请有兴趣的读者参阅 [1]。我们转而聚焦于优化交换鲁棒交易的构建模块，称为伴随集合。为此，假设给定整数集合 $T=[1,4t]=\{1,\ldots,4t\}$的一个划分，该集合对应于 $4t$ 文件的流行度排名，并将其划分为 $2t$个（互不相交的）集合（S_1，.....，S_{2t}），且满足 $\forall i\in[t]$

$$|S_{2i-1}|=|S_{2i}|=2,\ \Sigma(S_{2i-1})=\Sigma(S_{2i}).$$

此处我们用 $\Sigma(S_{2i-1})=\sum_{j\in S_{2i-1}}j$ 表示集合 S_{2i-1} 中元素之和。配对 (S_{2i-1},S_{2i}) 称为均衡伴随集合，而所有均衡伴随集合的集合 $(S_1,\ldots,S_{2t})$ 称为均衡定义集 [1]。若不满足求和约束，则称伴随集合（及相应的定义集）不均衡。总不均衡度由总差异衡量，其定义为 $\sum_{i=1}^{t}|\Sigma(S_{2i-1})-\Sigma(S_{2i})|$. 存在超指数数量的有效均衡定义集。

集合，而关键问题是在存在幅度有限的数据流行度变化的情况下，找到最坏情况总集合差异最小的均衡伴随集合（稍后将对此进行形式化描述）。流行度变化会改变伴随集合中元素的标签，进而改变其元素之和。因此，由流行度变化引起标签重新分配后，新形成的数对（S'_{2i-1}，S'_{2i}）之和可能不再相同，从而导致访问失衡。

为简单起见，我们只考虑变化幅度为 1 的数据流行度变化，也就是说，我们考虑标签或流行度值的相邻交换$I = \{(i_1, j_1)$、(i_2, j_2) .. 、$(i_m, j_m)\}$，其中 $|i_\ell - j_\ell| = 1$和 i_1、…、i_m、j_1、…、j_m 均互不相同。例如，对于 $t = 2$ 和 $T = [1, 8]$，允许的人气交换是 {(1, 2)、(2, 3)、(3, 4)、(4, 5)、(5, 6)、(6, 7)、(7, 8)}的子集，且同一个元素不能同时出现在两个不同的交换中，因为包含共同元素的交换对应于非相邻交换，因此会导致更大的流行度变化幅度。例如，{(1, 2)、(3, 4)、(5, 6)} 是允许的，而 {(1, 2)、(2, 3)、(5, 6)、(7, 8)} 则不允许。

在没有人气交换的情况下，均衡定义集的集合总差异为 0。给定定义集的最坏情况总平衡差异是任何有效交换集合 I可能造成的最大集合总差异。我们的目标是找到最坏情况集合总差异最小的（近似）最优定义集。例如，当 $t = 2$时，伴随集合

$$S_1 = \{1, 8\}, S_2 = \{3, 6\}, S_3 = \{2, 7\}, S_4 = \{4, 5\},$$

构成一个最优定义集的示例。交换{(1，2)，(5，6)} 会导致伴随集合发生变化

$$S'_1 = \{2, 8\}, S'_2 = \{3, 5\}, S'_3 = \{1, 7\}, S'_4 = \{4, 6\},$$

并产生 $|(2 + 8) - (3 + 5)| + |(1 + 7) - (4 + 6)| = 4$的集合总差异，这也是此情形下的最坏情况集合总差异。对于其他定义集选择，例如 $S_1 = \{1, 4\}$，$S_2 = \{2, 3\}$，$S_3 = \{5, 8\}$，$S_4 = \{6, 7\}$，最坏情况集合总差异为 ≥ 4（给定示例中为 6），因此这些定义集并非最优。

形式化地，若将针对 $4t$ 个元素的所有允许的相邻交换集合记为 $\mathcal{I}_t$，将所有有效的定义集集合记为 $\mathcal{S}_t$，那么目标是找到一个定义集 $(S_1^*, \ldots, S_{2t}^*)$，使其最坏情况下的总差异 $\max_{I_t \in \mathcal{I}_t} D(S_1^*, \ldots, S_{2t}^*; I_t; t)$ 尽可能小，即

$$(S_1^*, \ldots, S_{2t}^*) = \arg\min_{(S_1, \ldots, S_{2t}) \in \mathcal{S}_t} \max_{I_t \in \mathcal{I}_t} D(S_1, \ldots, S_{2t}; I_t; t), \quad (1)$$

$$D^*(t) = \min_{(S_1, \ldots, S_{2t}) \in \mathcal{S}_t} \max_{I_t \in \mathcal{I}_t} D(S_1, \ldots, S_{2t}; I_t; t), \quad (2)$$

$$D(S_1, \ldots, S_{2t}; I_t; t) = \sum_{i=1}^{t} \left| \Sigma(S'_{2i-1}) - \Sigma(S'_{2i}) \right|, \quad (3)$$

其中 $(S'_1, \ldots, S'_{2t})$ 表示经过人气交换后的集合。

请注意，当 t 的规模即使只是中等时，通过穷举搜索找到 $(S_1^*$, ……, $S_{2t}^*)$ 在计算上也很困难，因为搜索空间的大小会超指数增长。一种递归构造方法在 [1], 中被提出，其构造区块是小规模 t（即 $t \in \{1, 2, 3, 4, 5, 6\}$）的最优定义集，可通过计算机搜索找到。作者还推导出了最坏情况下的总差异下界，其形式为 $\frac{2}{3}(t - \frac{1}{3})$。

接下来，我们将这个下界提升至 $\frac{3t-2}{2}$，并提出一种完全解析的递归方法，该方法能够保证实现比先前报道的方法更小的最坏情况总集合差异。我们的结果总结在下一个定理中。

定理1。对于任意整数 $t > 0$，有

$$D^*(t) \geq \frac{3t - 2}{2}. \quad (4)$$

此外，对于任意正整数 $z \geq 2$ 和 $t = 5 \cdot 2^{z-2} - 1$，有

$$D^*(t) \leq 2^{z+1} - 2, \quad (5)$$

这意味着 $D^*(t) \leq \frac{8t-2}{5}$。上界和下界仅相差一个常数因子1。07。

III. 递归构造

接下来，我们为满足 $t = 5 \cdot 2^{z-2} - 1$,$z \geq 2$ 的 t 值构造定义集 $(S_1, \ldots, S_{2t}) \in \mathcal{S}_t$，使得总差异的上界为

$$\max_{I_t \in \mathcal{I}_t} D(S_1, \ldots, S_{2t}; I_t; t) \leq \frac{8t}{5} - \frac{2}{5}.$$

首先，考虑 $z = 2$ 且 $t = 5 \cdot 2^{z-2} - 1 =$4 的情形，并为集合 S 添加上标以表示 z 的值。通过穷举搜索，我们可以证明，对于 $t = 4$，唯一的最优定义集选择为

$$S_1^2 = \{1, 16\}, S_2^2 = \{8, 9\}, S_3^2 = \{2, 7\}, S_4^2 = \{4, 5\}, \quad (6)$$
$$S_5^2 = \{10, 15\}, S_6^2 = \{12, 13\}, S_7^2 = \{3, 14\}, S_8^2 = \{6, 11\}.$$

其差异（执行最坏情况交换后）等于$6 = 2^{z+1} - 2$。接下来，我们描述一种递归构造，为 $t_1 = 5 \cdot 2^{(z+1)-2} - 1$ 构造定义集 $(S_1^{z+1}$,. .., $S_{2t_1}^{z+1})$，该构造基于 $t_2 = 5 \cdot 2^{z-2} - 1$ 的 $(S_1^z, \ldots, S_{2t_2}^z)$:

$$S_i^{z+1} = S_i^z + 1, \forall i \in [5 \cdot 2^{z-1} - 1],$$
$$S_i^{z+1} = S_{i-5 \cdot 2^{z-1}+2}^z + 5 \cdot 2^z - 1, \forall i \in [5 \cdot 2^{z-1} - 1, 5 \cdot 2^z - 4],$$
$$S_{5 \cdot 2^z - 3}^{z+1} = \{1, 5 \cdot 2^{z+1} - 4\},$$
$$S_{5 \cdot 2^z - 2}^{z+1} = \{5 \cdot 2^z - 2, 5 \cdot 2^z - 1\},$$

其中，对于整数集合 S 和整数 a，我们定义 $S + a = \{x + a : x \in S\}$。图 1 展示了 $z = 4$ 的一种构造示例。构造背后的直观思路如下：一旦我们将最后的伴随集合固定为 $S_{5 \cdot 2^z - 3}^{z+1} = \{1, 5 \cdot 2^{z+1} - 4\}$ 和 $S_{5 \cdot 2^z - 2}^{z+1} = \{5 \cdot 2^z - 2, 5 \cdot 2^z - 1\}$，其余 $t_1 - 1$ 个伴随集合中使用的元素就必须来自两个互不相交的对称区间 $[2, 5 \cdot 2^z - 3]$ 和 $[5 \cdot 2^z, 5 \cdot 2^{z+1} - 5]$。因此，对于 $t_2 = 5 \cdot 2^{z-2} - 1$，我们可以重复使用伴随集合 $(S_1^z$，……, $S_{2t_2}^z)$ 的构造。



图1。 $z = 4$， $t = 19$ 的递归构造示例。我们使用式 (6) 中的伴随集合（记为 B_{16}）作为构造区块。同一伴随集合中的整数具有相同的颜色。

接下来，我们对递归构造的伴随集合族的最坏情况下的总差异进行如下上界估计。令

$$d_z = \max_{I_t \in \mathcal{I}_t} D(S_1^z, \ldots, S_{2t}^z; I_t; t),\ t = 5 \cdot 2^{z-2} - 1.$$

我们有如下引理。

引理 1。 对于 $z \geq 2$，我们有 $d_{z+1} \leq 2d_z + 2$。

证明： 证明基于对允许的交换集合 I_t 所导致的差异变化情形进行考察：

1) $|\Sigma(S'^{z+1}_{5\cdot 2^z-3}) - \Sigma(S'^{z+1}_{5\cdot 2^z-2})| = 0.$
2) $|\Sigma(S'^{z+1}_{5\cdot 2^z-3}) - \Sigma(S'^{z+1}_{5\cdot 2^z-2})| = 1.$
3) $|\Sigma(S'^{z+1}_{5\cdot 2^z-3}) - \Sigma(S'^{z+1}_{5\cdot 2^z-2})| = 2.$

情形1)： 在此情形下，我们有

$$\begin{aligned}&|I_t \cap \{(1,2), (5\cdot 2^z - 3, 5\cdot 2^z - 2)\}| \\ =&|I_t \cap \{(5\cdot 2^z - 1, 5\cdot 2^z), (5\cdot 2^{z+1} - 3, 5\cdot 2^{z+1} - 4)\}|.\end{aligned}$$

若 $|I_t \cap \{(1, 2), (5\cdot 2^z - 3, 5\cdot 2^z - 2)\}| = 0$，则与集合 $(S'^{z+1}_1, \ldots, S'^{z+1}_{5\cdot 2^{z-1}-2})$ 以及集合 $(S'^{z+1}_{5\cdot 2^{z-1}-1}, \ldots, S'^{z+1}_{5\cdot 2^z-4})$ 对应的差异均至多为 d_z。因此，总差异至多为 $2d_z$。

若 $|I_t \cap \{(1,2),(5\cdot 2^z - 3, 5\cdot 2^z - 2)\}| = 1$，则与集合 $(S'^{z+1}_1, \ldots, S'^{z+1}_{5\cdot 2^{z-1}-2})$ 以及集合 $(S'^{z+1}_{5\cdot 2^{z-1}-1}, \ldots, S'^{z+1}_{5\cdot 2^z-4})$ 对应的差异至多为 d_z+1。因此，总差异至多为 $2d_z + 2$。

同样，若 $|I_t \cap \{(1,2),(5\cdot 2^z - 3, 5\cdot 2^z - 2)\}| = 2$，则与 $(S'^{z+1}_1, \ldots, S'^{z+1}_{5\cdot 2^{z-1}-2})$ 和 $(S'^{z+1}_{5\cdot 2^{z-1}-1}, \ldots, S'^{z+1}_{5\cdot 2^z-4})$ 对应的集合差异至多为 d_z。那么，$(S'^{z+1}_1, \ldots, S'^{z+1}_{5\cdot 2^z-2})$ 的总差异至多为 $2d_z$。

情况2)： 对于这一情形，以下两个条件之一成立：

$$\begin{aligned}&|I_t \cap \{(1,2), (5\cdot 2^z - 3, 5\cdot 2^z - 2)\}| = 1, \\ &|I_t \cap \{(5\cdot 2^z - 1, 5\cdot 2^z), (5\cdot 2^{z+1} - 3, 5\cdot 2^{z+1} - 4)\}| = 1.\end{aligned}$$

根据对称性，我们不妨假设 $|I_t \cap \{(1, 2), (5\cdot 2^z - 3, 5\cdot 2^z - 2)\}| = 1$。那么，以下情况之一成立：

$$\begin{aligned}&|I_t \cap \{(5\cdot 2^z - 1, 5\cdot 2^z), (5\cdot 2^{z+1} - 3, 5\cdot 2^{z+1} - 4)\}| = 0, \\ &|I_t \cap \{(5\cdot 2^z - 1, 5\cdot 2^z), (5\cdot 2^{z+1} - 3, 5\cdot 2^{z+1} - 4)\}| = 2.\end{aligned}$$

因此，由集合 $(S'^{z+1}_1, \ldots, S'^{z+1}_{5\cdot 2^{z-1}-2})$ 引起的差异至多为 $d_z + 1$，由 $(S'^{z+1}_{5\cdot 2^{z-1}-1}, \ldots, S'^{z+1}_{5\cdot 2^z-4})$ 引起的差异至多为 d_z。因此，所有集合 $(S'^{z+1}_1, \ldots, S'^{z+1}_{5\cdot 2^z-2})$ 的总差异至多为 $d_z + 1 + d_z + 1 = 2d_z + 2$，其中额外的 1 来自 $(S'^{z+1}_{5\cdot 2^z-3}, S'^{z+1}_{5\cdot 2^z-2})$ 的差异。

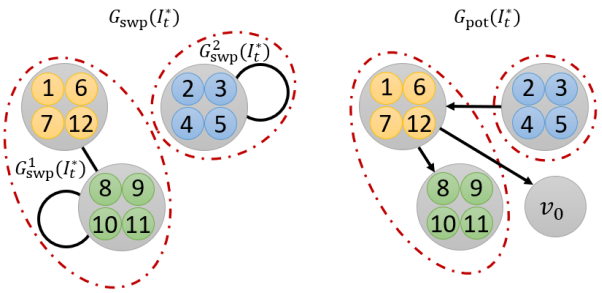


图 2. 为最坏情况流行度交换 $I_t^* = \{(2, 3), (8, 9), (11, 12)\}$ 和 $t = 3$定义的两个辅助图 $G_{\text{swp}}(I_t^*)$ 和 $G_{\text{pot}}(I_t^*)$。每个包含 4 个元素的灰色圆圈表示一对伴随集合，同时表示两个图中的一个节点。$G^1_{\text{swp}}(I_t^*)$ $G^2_{\text{swp}}(I_t^*)$这两个连通分量在 $G_{\text{swp}}(I_t^*)$ 中被圈出。为方便分析，无论底层图是 $G_{\text{swp}}(I_t^*)$ 还是 $G_{\text{pot}}(I_t^*)$，连通分量的组成始终为 $G^1_{\text{swp}}(I_t^*)$ $G^2_{\text{swp}}(I_t^*)$。

情况3)： 以下情况至少有一种成立：

$$\begin{aligned}&|I_t \cap \{(1,2), (5\cdot 2^z - 3, 5\cdot 2^z - 2)\}| = 2, \\ &|I_t \cap \{(5\cdot 2^z - 1, 5\cdot 2^z), (5\cdot 2^{z+1} - 3, 5\cdot 2^{z+1} - 4)\}| = 2.\end{aligned}$$

与情况2)类似，利用对称性，我们可以假设$|I_t \cap \{(1,2), (5\cdot 2^z - 3, 5\cdot 2^z - 2)\}| = 2$，因此 $|I_t \cap \{(5\cdot 2^z - 1, 5\cdot 2^z), (5\cdot 2^{z+1} - 3, 5\cdot 2^{z+1} - 4)\}| = 0$。那么，由 $(S'^{z+1}_1, \ldots, S'^{z+1}_{5\cdot 2^{z-1}-2})$ 引起的差异至多为 d_z，而由 $(S'^{z+1}_{5\cdot 2^{z-1}-1}, \ldots, S'^{z+1}_{5\cdot 2^z-4})$ 引起的差异至多为 d_z。因此，所有集合 $(S'^{z+1}_1, \ldots, S'^{z+1}_{5\cdot 2^z-2})$ 的总差异至多为 $d_z + d_z + 2 = 2d_z + 2$，其中额外的 2 来自 $(S'^{z+1}_{5\cdot 2^z-3}, S'^{z+1}_{5\cdot 2^z-2})$ 的差异。 ■

根据引理 1 以及 $d_2 = 6$ 这一事实，可得

$$d_z \leq 2^{z+1} - 2 = \frac{8t}{5} - \frac{2}{5}. \tag{7}$$

IV. 关于 D(t) 的下界

我们现在证明 (4)。在此之前，先给出证明所需的定义。

固定伴随集合 $(S_1, \ldots, S_{2t}) \in \mathcal{S}_t$。对于任意一组允许的交换 $I_t \in \mathcal{I}_t$，定义无权无向图 $G_{\text{swp}}(I_t) = (V_{\text{swp}}(I_t)\ E_{\text{swp}}(I_t))$ 来描述交换集合 I_t。具体而言，节点集为 $V_{\text{swp}}(I_t) = \{v_i\}_{i=1}^t$，其中节点 v_i 对应于伴随集合 (S_{2i-1}, S_{2i})。对于任意两个节点 $v_{i_1}, v_{i_2} \in V_{\text{swp}}(I_t)$，如果存在交换 $(i, i+1) \in I_t$，使得 i 和 $i+1$ 分别属于集合 $S_{2i_1-1} \cup S_{2i_1}$ 和 $S_{2i_2-1} \cup S_{2i_2}$，那么 v_{i_1} 与 v_{i_2} 之间就存在一条边。注意，i_1 和 i_2 可以相同，这意味着 $G_{\text{swp}}(I_t)$ 可以有自环。此外，同一对节点之间也允许存在重边。图 2 展示了 $G_{\text{swp}}(I_t)$ 的一个示例，其中 $t = 3$ 和 $I_t = \{(2, 3), (8, 9), (11, 12)\}$。注意，$G_{\text{swp}}(I_t)$ 的类似（但不同）定义也曾用于 [1]。

除 $G_{\text{swp}}(I_t)$外，我们还定义有向图$G_{\text{pot}}(I_t) = (V_{\text{pot}}(I_t)\ E_{\text{pot}}(I_t))$，用于描述未包含在 I_t中的交换 $(i, i+1)$ 所引起的潜在差异变化。节点集 $V_{\text{pot}}(I_t) = V_{\text{swp}}(I_t) \cup \{v_0\}$ 等于

由节点集 $V_{\text{swp}}(I_t)$扩充而成的集合，其中加入了一个辅助节点 v_0，该节点充当虚拟伴随集，稍后将对此进行讨论。为定义有向边集（即弧集） $E_{\text{pot}}(I_t)$，我们考虑 $i \in [4t-1]$的所有潜在交换 $(i, i+1)$。请注意，潜在交换 $(i, i+1)$ 与 I_t 中的选定交换可以共享一个元素。对于交换 $(i, i+1) \notin I_t$，若 $(i, i+1)$ 满足以下任一条件，则存在一条从 v_{i_1} 指向 v_{i_2} 的弧 (v_{i_1}, v_{i_2})：

1) $i \in S_{2i_1}, i+1 \in (S_{2i_2-1} \cup S_{2i_2}) \backslash S_{2i_1}$，以及在 I_t 中执行交换后 $\Sigma(S'_{2i_1-1}) < \Sigma(S'_{2i_1})$。 2) $i+1 \in S_{2i_1}, i \in (S_{2i_2-1} \cup S_{2i_2}) \backslash S_{2i_1}$，以及在 I_t 中执行交换后 $\Sigma(S'_{2i_1-1}) > \Sigma(S'_{2i_1})$。 3) $i \in S_{2i_1-1}, i+1 \in (S_{2i_2-1} \cup S_{2i_2}) \backslash S_{2i_1-1}$，以及在 I_t 中执行交换后$\Sigma(S'_{2i_1-1}) > \Sigma(S'_{2i_1})$。 4) $i+1 \in S_{2i_1-1}, i \in (S_{2i_2-1} \cup S_{2i_2}) \backslash S_{2i_1-1}$，以及在 I_t 中执行交换后$\Sigma(S'_{2i_1-1}) < \Sigma(S'_{2i_1})$。 5) $i \in S_{2i_1-1}, i+1 \in (S_{2i_2-1} \cup S_{2i_2}) \backslash S_{2i_1-1}$，以及在 I_t 中执行交换后$\Sigma(S'_{2i_1-1}) = \Sigma(S'_{2i_1})$。 6) $i+1 \in S_{2i_1}, i \in (S_{2i_2-1} \cup S_{2i_2}) \backslash S_{2i_1}$，以及在 I_t 中执行交换后 $\Sigma(S'_{2i_1-1}) = \Sigma(S'_{2i_1})$。

直观地说，如果 i 和 $i+1$ 分别属于 $S_{2i_1-1} \cup S_{2i_1}$ 和 $S_{2i_2-1} \cup S_{2i_2}$，且交换 $(i, i+1)$ 会进一步增大伴随集合 (S'_{2i_1-1}, S'_{2i_1}) 的差异 $|\Sigma(S'_{2i_1-1}) - \Sigma(S'_{2i_1})|$（当该差异非零时），那么 $(v_{i_1}, v_{i_2}) \in E_{\text{pot}}(I_t)$。此外，如果伴随集合 (S'_{2i_1-1}, S'_{2i_1}) 的差异为 0，那么只有当交换 $(i, i+1)$ 能产生正的集合差异 $\Sigma(S'_{2i_1-1}) - \Sigma(S'_{2i_1})$ 时，我们才有 $(v_{i_1}, v_{i_2}) \in E_{\text{pot}}(I_t)$。请注意，引入条件 5) 和 6) 是为了保证对应于多条弧的交换能够同时增大同一对伴随集合的集合差异。同样，$i_1 = i_2$，即自环，在 $G_{\text{pot}}(I_t)$ 中是允许的。此外，同一节点对 (v_{i_1}, v_{i_2}) 之间也允许存在多条弧 $(v_{i_1}, v_{i_2}) \in E_{\text{pot}}$。

除了节点 $\{v_i\}_{i=1}^t$ 在 $G_{\text{pot}}(I_t)$ 中的弧之外，如果满足以下任一条件，我们还添加弧 $(v_{i_1}, v_0) \in E_{\text{pot}}(I_t)$:

1) $1 \in S_{2i_1-1} \cup S_{2i_1}$ 且交换 $(0, 1)$ 可以增大伴随集合 (S'_{2i_1-1}, S'_{2i_1}) 的差异，当$\Sigma(S'_{2i_1-1}) \neq \Sigma(S'_{2i_1})$ 时；2) $1 \in S_{2i_1}$ 当 $\Sigma(S'_{2i_1-1}) = \Sigma(S'_{2i_1})$ 时。

请注意，交换 $(0, 1)$ 在 I_t 中不允许，仅为简化分析而纳入 $E_{\text{pot}}(I_t)$。类似地，我们为交换 $(4t, 4t+1)$ 定义一条弧 $(v_{i_2}, v_0) \in E_{\text{pot}}(I_t)$。当 $t=3$ 且 $I_t = \{(2, 3), (8, 9), (11, 12)\}$ 时，$G_{\text{pot}}(I_t)$ 的一个示例如图 2 所示。

给定定义集（S_1、.....、S_{2t}），令 I_t^* 为所有能够导致最坏情况下的总差异的交换集中规模最小的交换集，即，

$$I_t^* \in \arg\max_{I_t \in \mathcal{I}_t} D(S_1, \ldots, S_{2t}; I_t; t).$$

接下来我们证明

$$D(S_1, \ldots, S_{2t}; I_t^*; t) = 2|I_t^*|, \tag{8}$$

这意味着每一对 $(i, i+1) \in I_t^*$ 都会使总差异增加 2。这是因为交换 $(i, i+1)$ 会使两个伴随集合的差异增加或减少 (S_{2i_1-1}, S_{2i_1}) 和 (S_{2i_2-1}, S_{2i_2})，它们分别包含 i 和 $i+1$，差异变化量为 1；除非 i和 $i+1$位于同一个集合 S_{2i_1-1}或 S_{2i_1}中，这种情况可以忽略。因此，交换 $(i, i+1) \in I_t^*$对总差异的贡献为 2、0 或 -2。由于 I_t^* 的大小最小，I_t^* 中的每次交换都会使集合总差异增加 2。因此，(8) 成立。

由于后续分析不依赖于 t 的取值，以下我们省略记号中的下标 t。为简便起见，我们也记 $G_{\text{swp}} = G_{\text{swp}}(I^*) = (V_{\text{swp}}、E_{\text{swp}})$，以及 $G_{\text{pot}} = G_{\text{pot}}(I^*) = (V_{\text{pot}}、E_{\text{pot}})$。图 G_{swp} 可划分为一组连通分量 $G_{\text{swp}}^j = (V^{j\text{交换}}\ E_{\text{swp}}^j)$、$j \in [J]$，如图 2 所示。对于 $v_i \in V_{\text{swp}} \cup \{v_0\}$，令

$$d_{\text{in}}^{\text{pot}}(v_i) = |\{(v_j, v_i) : (v_j, v_i) \in E_{\text{pot}}, v_j \in V_{\text{swp}} \backslash \{v_i\}\}|$$

为节点 v_i 的入度，并且

$$d_{\text{out}}^{\text{pot}}(v_i) = |\{(v_i, v_j) : (v_i, v_j) \in E_{\text{pot}}, v_j \in V_{\text{swp}} \backslash \{v_i\}\}|$$

为图 G_{pot} 中节点 v_i 的出度。注意，除辅助节点 v_0 外，假设 V_{swp} 和 V_{pot} 具有相同的顶点集 $\{v_i\}_{i=1}^t$。下文中，我们交替使用 v_i 表示 V_{swp} 中的节点或 V_{pot} 中的节点。此外，对于任意顶点子集 $V_\subseteq \subseteq \{v_i\}_{i=0}^t$，我们使用

$$d_{\text{in}}^{\text{pot}}(V_\subseteq) = \sum_{v \in V_\subseteq} d_{\text{in}}^{\text{pot}}(v), d_{\text{out}}^{\text{pot}}(V_\subseteq) = \sum_{v \in V_\subseteq} d_{\text{out}}^{\text{pot}}(v).$$

引理 2. 对于所有 $j \in [J]$，其中 J 表示连通分量的数量，我们有

$$d_{in}^{pot}(V_{swp}^j) - d_{out}^{pot}(V_{swp}^j) \leq |V_{swp}^j| + 4(|E_{swp}^j| - |V_{swp}^j|). \tag{9}$$

注意，由于 $d^{\text{pot}}{}_{\text{out}}(v_0) = 0$且 $d^{\text{pot}}{}_{\text{in}}(V_{\text{swp}} \cup \{v_0\}) - d^{\text{pot}}{}_{\text{out}}(V_{\text{swp}} \cup \{v_0\}) = 0$，因此 $d^{\text{pot}}{}_{\text{in}}(V_{\text{swp}}) - d^{\text{pot}}{}_{\text{out}}(V_{\text{swp}}) \geq 0 - d^{\text{pot}}{}_{\text{in}}(v_0) \geq -2$。于是，我们有

$$-2 \leq |V_{\text{swp}}| + 4(|E_{\text{swp}}| - |V_{\text{swp}}|),$$

这意味着

$$2|E_{\text{swp}}| \geq \frac{3|V_{\text{swp}}|}{2} - 1. \tag{10}$$

根据 G_{swp} 的定义，由于 $E_{\text{swp}} = |I^*|$，(8) 和 (10) 可推出

$$\max_{I \in \mathcal{I}} D(S_1, \ldots, S_{2t}; I) \geq \frac{3t-2}{2}$$

对于任意 $(S_1, \ldots, S_{2t})$，因此我们有 (4)。

现在我们证明引理 2。注意，对于任意 $j \in J$，我们有

$$\begin{aligned} & d_{\text{in}}^{\text{pot}}(V_{\text{swp}}^j) - d_{\text{out}}^{\text{pot}}(V_{\text{swp}}^j) \\ =& |\{(v', v) : (v', v) \in E_{\text{pot}}, v' \notin V_{\text{swp}}^j, v \in V_{\text{swp}}^j\}| \\ & - |\{(v, v') : (v, v') \in E_{\text{pot}}, v' \notin V_{\text{swp}}^j, v \in V_{\text{swp}}^j\}|, \end{aligned}$$

即，$d^{\text{pot}}{}_{\text{入弧}}(V_{\text{swp}}^j) - d^{\text{pot}}{}_{\text{出弧}}(V^{j\text{交换}})$ 等于从 V_{swp}^j 外部的节点指向 V_{swp}^j内部节点的弧数，减去从 V_{swp}^j 内部节点指向 V_{swp}^j外部节点的弧数。此后，对于任意子集 $V_\subseteq \subseteq V_{\text{swp}} \cup \{v_0\}$，令

$$\text{in}(V_\subseteq) = |\{(v', v) : (v', v) \in E_{\text{pot}}, v' \notin V_\subseteq, v \in V_\subseteq\}|$$

表示 E_{pot} 中的入弧数量，这些弧从 $V_{\subseteq}$ 外部的节点发出并指向 $V_{\subseteq}$内部的节点，并令

$$\text{out}(V_{\subseteq}) = |\{(v,v') : (v,v') \in E_{\text{pot}}, v' \notin V_{\subseteq}, v \in V_{\subseteq}\}|$$

分别表示从 $V_{\subseteq}$ 中的节点指向 $V_{\subseteq}$之外节点的出弧数量。那么，(9)等价于

$$\text{in}(V^j_{\text{swp}}) - \text{out}(V^j_{\text{swp}}) \le |V^j_{\text{swp}}| + 4(|E^j_{\text{swp}}| - |V^j_{\text{swp}}|). \quad (11)$$

现在证明 (11)。对于任意顶点集 $V_{\subseteq} \subseteq V_{\text{swp}}$，令

$$d(V_{\subseteq}) = |\{\{v,v'\} : \{v,v'\} \in E_{\text{swp}}, \{v,v'\} \cap V_{\subseteq} \neq \varnothing\}|$$

表示 E_{swp} 中至少与 $V_{\subseteq}$中的一个节点关联的边数。

命题 1. 对于任意子集 $V_{\subseteq} \subseteq V_{swp}$，有$(V_{\subseteq}) \le d(V_{\subseteq})$。

证明：注意，E_{swp} 中的每条边都对应于 I^* 中的一次交换，而 E_{pot} 中的每条弧都对应于I^*中没有的一次交换。反设，在$(V_{\subseteq}) > d(V_{\subseteq})$中成立，并移除所有与边相对应的交换（$i$，$i+1$）$\in I^*$，这些边属于

$$E^{\text{swp}}(V_{\subseteq}) = \{\{v,v'\} : \{v,v'\} \in E_{\text{swp}}, \{v,v'\} \cap V_{\subseteq} \neq \varnothing\},$$

它们是至少与 $V_{\subseteq}$中的一个节点关联的边。然后，我们添加与入弧对应的交换，这些入弧属于

$$E^{\text{pot}}(V_{\subseteq}) = \{(v',v) : (v',v) \in E_{\text{pot}}, v' \notin V_{\subseteq}, v \in V_{\subseteq}\}$$

从 $V_{\subseteq}$ 外部的节点指向 $V_{\subseteq}$ 中的节点。接下来我们证明，每个新增的交换都会使总差异增加 2。注意，根据 E_{pot} 中弧的定义，每个与弧 $(v_{k_2}, v_{k_1}) \in E^{\text{pot}}(V_{\subseteq})$ 对应的交换 (i, $i+1$)，都会增加伴随集合 (S'_{2k_2-1}, S'_{2k_2}) 的差异。因此，只需证明：对于任意 $v_{k_1} \in V_{\subseteq}$，移除与 $E^{\text{swp}}(V_{\subseteq})$ 中的边对应的交换后，同一个交换 $(i, i+1) \in E^{\text{pot}}(V_{\subseteq})$ 会使集合 (S'_{2k_1-1}, S'_{2k_1}) 的集合差异增加 $+1$ 。否则，将存在两条弧 (v_{k_2}, v_{k_1})、$(v_{k_3}, v_{k_1}) \in E^{\text{pot}}(V_{\subseteq})$（$k_2$ 和 k_3 可以相同），其对应的交换 $(i, i+1)$ 和 $(j, j+1)$ 在移除并添加交换后，分别使集合 (S'_{2k_1-1}, S'_{2k_1}) 的集合差异增加 $+1$ 和 -1 。这意味着，在移除并添加交换之前，交换 $(i, i+1)$ 和 $(j, j+1)$ 分别使集合 (S'_{2k_1-1}, S'_{2k_1}) 的集合差异增加+1 和 -1 ，或反之。交换 $(i, i+1)$ 或 $(j, j+1)$ 在移除并添加交换之前会增加集合 (S'_{2k_1-1}, S'_{2k_1}) 的集合差异，因此本应包含在 I^* 中（注意，该交换与 I^* 中的交换没有共同元素），这与 $(i, i+1)$、$(j, j+1) \notin I^*$ 的事实矛盾。因此，每个新增的交换都会使集合总差异增加 2。此外，新增的交换彼此之间，以及它们与 I^* 中的其他交换之间，都没有共同元素，因为两个共享元素的交换不可能都使总差异增加 2。

因此，新增的 $\text{in}(V_{\subseteq})$ 次交换使总差异增加 $2*\text{in}(V_{\subseteq})$，而移除的交换使总差异减少 $2*d(V_{\subseteq})$。由于 $\text{in}(V_{\subseteq}) > d(V_{\subseteq})$，这意味着移除并添加交换后，集合总差异增加了，这与极大性的假设相矛盾。

由 I^* 导致的集合总差异。因此，$\text{in}(V_{\subseteq}) \le d(V_{\subseteq})$。 ■

现在，我们证明当 $|E^j_{\text{swp}}|+1-|V^j{}_{\text{swp}}| > 0$ 时，(11) 成立。反设在 swp 时，(11 $|E^j_{\text{swp}}|+$) 不成立。 $1-|V^j| > 0$ $V^j > |E^j_{\text{swp}}|$ 1, 那么，in(swp) 。根据命题，这不可能，因为 $d(V^j{}_{\text{swp}}) = |E^j_{\text{swp}}|$。

剩下要证明的是，在该情形下 (11 $|E^j_{\text{swp}}|+1-$) 成立，即证明 $|V^j_{\text{swp}}| = 0$

$$\text{in}(V^j_{\text{swp}}) - \text{out}(V^j_{\text{swp}}) \le |V^j_{\text{swp}}| - 4 \quad (12)$$

对 $j \in [J]$ 和 $|E^j_{\text{swp}}| + 1 - |V^j_{\text{swp}}| = 0$。在证明之前，我们注意到，伴随集合 (S_{2i-1}, S_{2i})、$i \in [t]$ 中包含的四个元素属于以下“类型”之一：

1) 类型 1：对于某个 $b \ge 1$，$\{a, a+b, a+b+1, a+2b+1\}$，其中 $\{S_{2i-1}, S_{2i}\} = \{\{a, a+2b+1\}, \{a+b, a+b+1\}\}$；2) 类型 2：对于某个 $b, c > 1$，$\{a, a+b, a+b+c, a+2b+c\}$，且 $a \ge 1$，其中 $\{S_{2i-1}, S_{2i}\} = \{\{a, a+2b+c\}, \{a+b, a+b+c\}\}$；3) 类型3：对于某个 $a, b \ge 1$，$\{a, a+1, a+1+b, a+2+b\}$，其中 $\{S_{2i-1}, S_{2i}\} = \{\{a, a+2+b\}, \{a+1, a+1+b\}\}$。

为证明这四个元素 $\{\ell_1, \ell_2, \ell_3, \ell_4\} \in S_{2i-1} \cup S_{2i}$、$i \in [t]$ 属于上述三种类型，我们将这四个元素排列为 $\ell_1 < \ell_2 < \ell_3 < \ell_4$，其中 $i \in [t]$ 任意。注意，根据 (S_{2i-1}, S_{2i}) 的平衡性质，我们有 $\ell_2 - \ell_1 = \ell_4 - \ell_3$。以下情况恰有一种成立：(1) $\ell_2 - \ell_1 = 1$，属于类型3；(2) $\ell_2 - \ell_1 > 1$ 和 $\ell_3 - \ell_2 = 1$，属于类型 1；(3) $\ell_2 - \ell_1 > 1$ 和 $\ell_3 - \ell_2 > 1$，属于类型 2。

我们注意到，当$|E^j_{\text{swp}}|+1-|V^j\text{swp}| = 0$时，可以排除类型 3 伴随集；也就是说，连通分量 G^j_{swp}无环。对于任意类型 3 伴随集(S_{2i-1}, S_{2i})，可以验证，从 $\{(\ell_1 - 1, \ell_1)$、$(\ell_2, \ell_2 + 1)$、$(\ell_3 - 1, \ell_3)$、$(\ell_4, \ell_4 + 1)\}$中，至多允许进行 2 次与 $S_{2i-1} \cup S_{2i} = \{\ell_1, \ell_2, \ell_3, \ell_4\}$ 仅共享一个 a 元素的交换。这是因为，不可能有三次交换都分别使(S_{2i-1}, S_{2i})的差异增加 1。此外，除了在伴随集(S_{2i_1-1}, S_{2i_1})与其他伴随集之间进行交换外，还总能使用交换$(a, a+1)$或$(a+b+1, a+2+b)$，进一步使(S_{2i_1-1}, S_{2i_1})的集合差异增加 2。这意味着，如果(S_{2i_1-1}, S_{2i_1})属于类型 3，那么 V^jswp 中的 v_{i_1} 应当存在自环，这与 G^j_{swp} 无环的假设矛盾。

接下来，我们在以下命题中刻画类型 1 和类型 2 的伴随集合的一些性质。证明见附录。

命题 2. 对于任意 $i \in [t]$，令 (S_{2i-1}, S_{2i}) 为类型 j, $j \in [2]$ 的伴随集合。那么，

$$d(v_i) + d_{out}(v_i) = j + 2. \quad (13)$$

现在我们可以证明 (12) 了。由命题 1 可得 $d_{\text{in}}(v_i) \le d(v_i)$，因为 $d_{\text{in}}(v_i)$ =in(v_i)。由命题2, $d_{\text{out}}\ v_i \ge 3 - d\ v_i$ $d_{\text{in}}\ v_i - d_{\text{out}}\ v_i \le$ 可得 ( ) ( )。因此，( ) ( ) ( )。对所有swp求和，得到 $2d\ v_i - 3$ $v_i \in V^j$

$$d_{\text{in}}(V^j_{\text{swp}}) - d_{\text{out}}(V^j_{\text{swp}}) \le 2 \sum_{v_i \in V^j_{\text{swp}}} d(v_i) - 3|V^j_{\text{swp}}|$$

$$
\begin{aligned}
&= 4|E^j_{\text{swp}}| - 3|V^j_{\text{swp}}| \\
&= |V^j_{\text{swp}}| - 4,
\end{aligned}
$$

这意味着 (12) 成立，从而完成证明。

参考文献

[1] C. Pan、R. Gabrys、X. Liu、C. Colbourn 和 O. Milenkovic，“动态分布式存储中的平衡且抗交换交易”，载于*2022*年*IEEE*国际信息论研讨会 *(ISIT)*。IEEE，2022年，第2385–2390页。[2] H. Dau 和 O. Milenkovic，“分布式存储中用于访问均衡的最大最小和Steiner系统”，*SIAM*离散数学杂志，第32卷，第3期，第1644–1671页，2018年。[3] Z. Liu、Z. Bai、Z. Liu、X. Li、C. Kim、V. Braverman、X. Jin 和 I. Stoica，“Distcache：面向大规模存储系统的可证明负载均衡与分布式缓存。”载于*FAST*，第19卷，2019年，第143–157页。[4] Y. M. Chee、C. J. Colbourn、H. Dau、R. Gabrys、A. C. Ling、D. Lusi 和 O. Milenkovic，“通过部分Steiner系统标记实现存储系统中的访问均衡”，设计、编码与密码学，第88卷，第11期，第2361–2376页，2020年。[5] S. E. Anderson、A. Johnston、G. Joshi、G. L. Matthews、C. Mayer 和 E. Soljanin，“擦除码存储中内容访问的服务率区域”，载于*2018*年*IEEE*信息论研讨会 *(ITW)*。IEEE，2018年，第1–5页。[6] M. Aktaş、G. Joshi、S. Kadhe、F. Kazemi 和 E. Soljanin，“服务率区域：编码分布式系统设计的新视角”，*IEEE*信息论汇刊，第67卷，7963，2021年12月，7940 no. pp. – 。[7] C. J. Colbourn，“Steiner系统的流行度区块标记”，载于*2020*年代数与组合编码理论会议 *(ACCT)*。IEEE，2020年，第41–46页。[8] ——，“面向数据流行度的平等Steiner三元系统”，设计、编码与密码学，第89卷，第2373–2395页，2021年。[9] A. Hedayat，“t-设计的权衡理论”，载于编码理论与设计理论：第二部分，设计理论，1990年，第101页。[10] E. J. Billington，“组合交易：近期成果综述”，《设计*2002*：进一步的计算与构造设计理论》，第47–67页，2003年。[11] H. Hwang，“关于 (v, k, t) 交易的结构”，统计规划与推断杂志，第13卷，第179–191页，1986年。

附录

命题 2 的证明

注意，涉及类型 1 伴随集元素的交换可以分为两组，$\{(a-1,\ a),\ (a+b+1,\ a+b+2),\ (a+2b,\ a+2b+1)\}$ 以及 $\{(a,\ a+1),\ (a+b-1,\ a+b),\ (a+2b+1,\ a+2b+2)\}$，其中同一组中的交换可以同时对伴随集 $(S_{2i-1},\ S_{2i})$ 的差异产生影响 $+1$ ，而不同组中的交换则无法同时对集合差异产生影响 $+1$ 。因此，I^* 不能包含来自两组的交换。当

$$
\begin{aligned}
d(v_i) = |I^* \cap (&\{(a-1,a),(a+b+1,a+b+2),\\
&(a+2b,a+2b+1)\} \cup \{(a,a+1),(a+b-1,a+b),\\
&(a+2b+1,a+2b+2)\})| = 0,
\end{aligned}
$$

也就是说，当 v_i 是 G_{swp} 中的孤立节点时，其中一组交换对应于 E_{pot} 中从节点 v_i 出发的三条出弧，这意味着 $d_{\text{out}}(v_i) = 3$。当 $d(v_i) \neq 0$ 时，不失一般性，我们可以假设

$$
I^* \cap \{(a,a+1),(a+b-1,a+b),(a+2b+1,a+2b+2)\} = \varnothing.
$$

然后，

$$
d(v_i) = |I^* \cap \{(a-1,a),(a+b+1,a+b+2),(a+2b,a+2b+1)\}|.
$$

$\{(a-1,\ a),\ (a+b+1,\ a+b+2),\ (a+2b,\ a+2b+1)\} \setminus I^*$ 中的交换对应于从 v_i 在 E_{pot} 中出发的 $d_{\text{out}}(v_i) = 3 - d(v_i)$ 条出弧。在任一情况下，我们都有 $d(v_i) + d_{\text{out}}(v_i) = 3$。

类似地，我们将涉及类型 2 伴随集元素的交换分为两组，$\{(a-1,a),(a+b,a+b+1),(a+b+c,a+b+c+1),(a+2b+c-1,a+2b+c)\}$以及 $\{(a,a+1),(a+b-1,a+b),(a+b+c-1,a+b+c),(a+2b+c-1,a+2b+c)\}$。当 $d(v_i) + d_{\text{out}}(v_i) = 4$ 时，可以采用与前文类似的论证。因此，$d(v_i) + d_{\text{out}}(v_i) = j + 2$。