基于异或的隐私保护码优化研究

金星彤 李鹏 王刚 刘晓光 李忠伟

（南开大学计算机与控制工程学院，天津 300350）

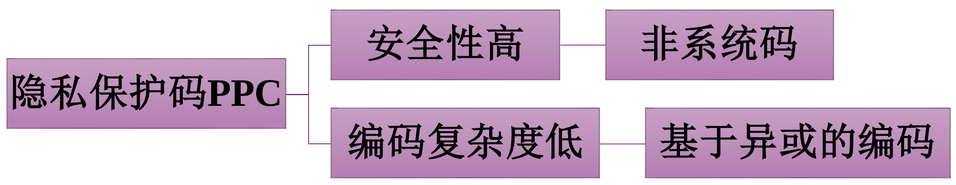
•研究背景

- 海量数据→云存储行业兴起
- 单云存储系统：安全隐患
  - 数据完整性、数据保密性、服务可用性、厂商锁定
- 多云存储系统：隐私保护编码
  - 很大程度上解决以上问题
  - 应用于移动平台：编码效率低

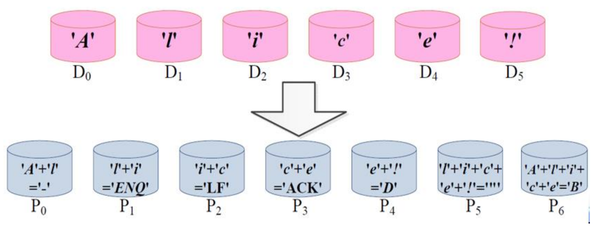
•研究内容与贡献

- 基于最优调度次序优化编码效率
  - 搜索最优调度：减少编码异或次数
  - 设计基于最优调度的编码算法
- 基于SIMD 技术并行优化编码效率

•隐私保护编码介绍



基于异或的
非系统纠删码



•基于异或的纠删码优化思想

- *合法调度遵循准则*
  - 前c个元素是X中的元素：

    $s_i = x_i, (i < c)$，即前c个元素组成（$c \times c$）的单位矩阵
  - 任何非X中的元素s_k必须遵循：

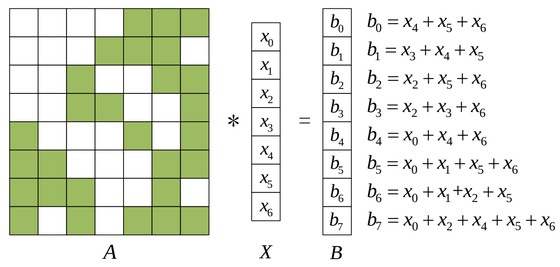
    $s_k = s_i + s_j, (i < j < k)$，即$s_k$必须由它之前的元素经过异或得到

- *搜索最优调度定理*
  - 设s_i，s_j和$s_k \in S$，其中$i < j < k-1$，$s_k = s_i + s_j$。那么存在一个合法调度S'，满足：

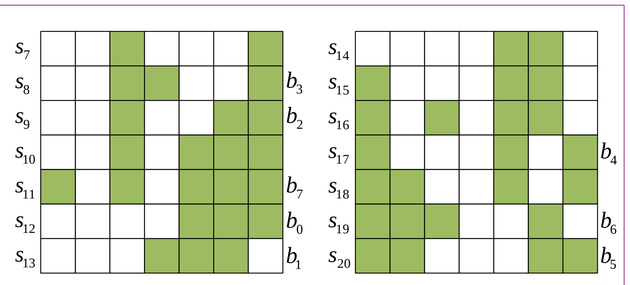
$$S' = \left\{ s'_x | s'_x = \begin{Bmatrix} s_x \ when \ x \le j \\ s_k \ when \ x = j+1 \\ s_{x-1} \ when \ j+1 < x \le k \\ s_x \ when \ x > k \end{Bmatrix} \right\}$$

•PPC编码优化策略

- 基于最优调度的优化
  - 搜索纠删码的最优调度
    - 枚举法：宽度优先算法BFS



PPC生成矩阵



一种优化调度

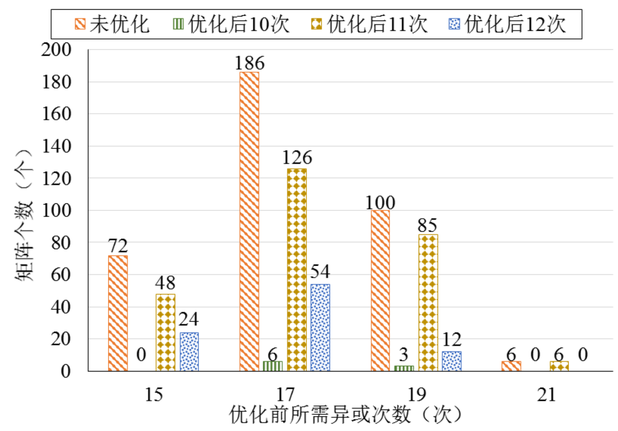
- 基于最优调度的编码
  - 平凡算法：编码过程即为由生成矩阵按行序对原始数据进行异或操作
  - 基于最优调度的编码：存储调度次序，在编码时按照调度顺序处理原始数据

- 基于SIMD技术的并行优化

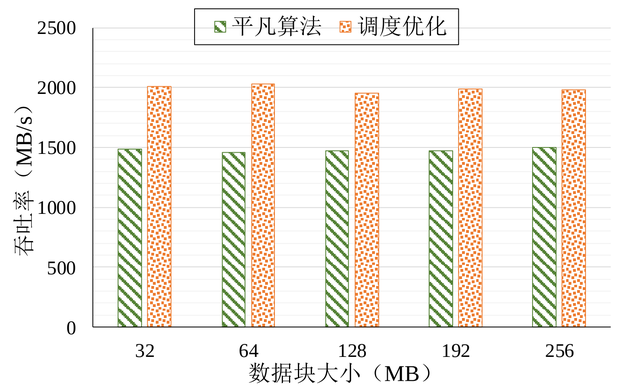
基于异或的PPC编码算法中含有大量异或运算：采用AVX2技术在基于最优调度次序的编码方案上进行进一步并行优化（对编码过程中的每一步异或操作进行并行化操作）

•实验结果

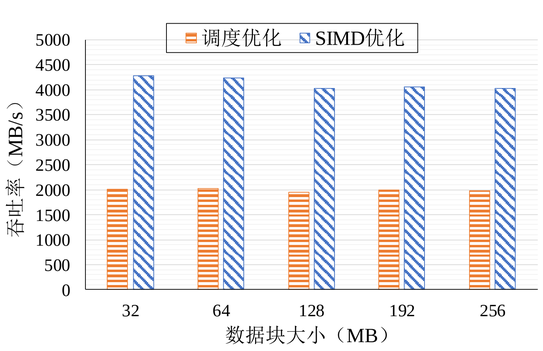
- 优化前后异或次数对比

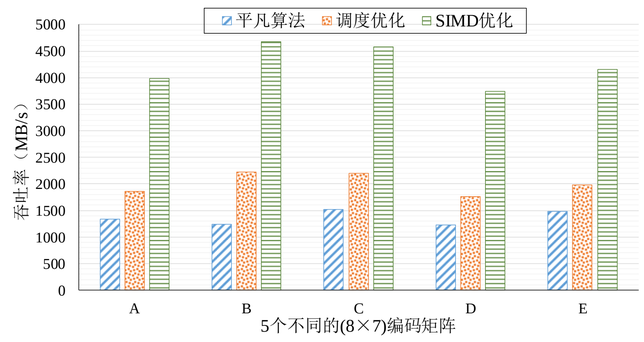


- 最优调度编码性能分析



- SIMD并行优化性能分析





•小结

- 本文从减少编码过程中异或次数以提高数据编码性能的角度出发，研究基于异或的非系统小纠删码优化问题。
- 实验结果表明，优化后异或次数比优化前平均减少了36.9%，编码性能提高了34.8%，进行SIMD优化后进一步提高了107.1%。
- 综上，对基于异或的隐私保护码的异或次序进行优化可以很大程度上提高数据编码性能，基于SIMD并行优化每次异或操作后编码性能明显提升。