

ITEM #229 - UnalignedAND-BTP Fish-Control Backbone - Exactness Guards, Recall Modes, and Anti-Leak Contracts

Conversation : UnalignedAND-BTP 算法加速

20260111

Authors: Sizhe Tan & GPT-Obot

ITEM #229 - UnalignedAND-BTP Fish-Control Backbone

Exactness Guards, Recall Modes, and Anti-Leak Contracts

Category

DBM-COT · Metric Space Intelligence · UnalignedAND-BTP · Fish-Control Backbone

Status

Backbone / Architecture-level (stable direction anchor)

0. Motivation & Positioning

UnalignedAND-BTP (BucketTreeOfPermutation) 是 DBM Metric Distance 的核心内核算法，用于解决：

- 2A：两个复杂 Metric Space 点（Graph / Sequence / Image Starmap）的最优或 Top-K 匹配；
- 2B：由若干已知小结构 组合/预测一个大型结构（如生物分子、复杂模式拼装）。

该算法计算复杂度高、结构嵌套深，任何工程级加速都极易引入**静默漏解**（silent miss）。

因此，本 ITEM 不讨论具体优化技巧，而是固定一套“**鱼控级**”骨架：

先确保不漏解（Exactness Guards），再允许可控降本（Recall Modes），
用 Contracts 把所有潜在逻辑漏洞显式化并锁住。

1. Two Canonical Operating Modes（两种规范运行形态）

1.1 EXACT Fish-Control Mode（不漏解优先）

目标：在算力允许范围内，保证 Top-K / 最优解的完备性。

原则：

- 任何 **REJECT / 剪枝** 必须来自可证明的下界（LB）；
- Anchor / CCC / Quality 只能排序或 DEFER，不能拍脑袋排除；
- 慢可以，漏不行。

典型适用：

- 算法验证阶段
 - 论文级结果
 - 结构预测/科学计算的“最终答案”阶段
-

1.2 RECALL-Controlled Fish-Control Mode（吞吐优先）

目标：显著降本提速，允许漏解，但必须 可观测、可量化、可回补。

原则：

- Anchor + CCC Gate 可直接 REJECT 低质量候选；
- 必须输出 coverage / miss-risk / reject-reason 统计；
- 必须保留 DEFER / 回补通道。

典型适用：

- 大规模扫描
- 实时系统
- 预筛 / warm-up 阶段

2. Anti-Leak Philosophy (反漏解设计哲学)

所有加速都可能出错，因此系统必须具备三道硬兜底阀门：

1. Coverage Guard (切片层)

- 搜索域必须被 core 并集覆盖；
- $halo \geq Reach(A)$ ，防止边界漏解。

2. Candidate Guard (召回层)

- EXACT 模式下，Anchor 不得成为排除条件（除非 must-hit 有证明）。

3. Filter Guard (过滤层)

- EXACT 模式下，CCC/质量指标 只能 DEFER，不能 REJECT；
- 只有 $LB > threshold$ 才允许 REJECT。

这三道阀门的作用是：

宁可慢、宁可候选多，也不允许静默漏掉有效解。

3. Core Concepts (统一抽象)

3.1 Reach / Span 统一尺度

为避免“ $3 \times \text{size}$ ”的歧义，统一使用：

- **Reach(A)**：A 的匹配影响半径（steps / lags / hops / patch radius）
- **SpanNeed(A)**：A 完整可验证匹配所需的最小跨度
- **CoreSpan(P)**：Piece 核心区的有效跨度

不漏解条件：

$\text{CoreSpan}(P) \geq \text{SpanNeed}(A)$
 $\text{HaloWidth}(P) \geq \text{Reach}(A)$

3.2 CCC as Evidence, not Magic

CCC（Common Concept Core）不是“神秘判定器”，而是：

- 一种 **Occurrence** 级的证据压缩；
 - 用于 下界估计、排序、解释链；
 - 在 EXACT 模式下 不得单独作为排除依据。
-

4. The 8-Stage Fish-Control Pipeline

Overview

UnalignedAND-BTP 的所有加速与控制，最终被摊平成 **8 段线性流水线**。

Mermaid: 8-Stage Pipeline

flowchart TD

```
A[1. Compute Reach(A) / SpanNeed(A)]
B[2. Generate Pieces with Core + Halo]
C[3. Anchor Search → Occurrences]
D[4. OccurrenceCccGate\n(ACCEPT / DEFER / REJECT)]
E[5. Piece Merge / Union]
```

```
F[6. Piece-level Cheap LB Filter]
G[7. UnalignedAND-BTP\n(Anytime + LB/UB)]
H[8. MatchKey / Conflict\n/ SafePeel & Iterate]

A --> B
B --> C
C --> D
D --> E
E --> F
F --> G
G --> H
H -->|residual| B
```

5. Contract Table（反漏解契约总表）

#	Contract	Scope	Core Guarantee	Leak Risk if Violated
C1	PieceWithHalo	切片	Core 覆盖 + Halo \geq Reach(A) \Rightarrow 不因边界漏解	高（跨边界真解丢失）
C2	AnchorSet	召回	EXACT 模式下 Anchor 只排序不排除	中-高
C3	OccurrenceCccGate	过滤	EXACT：仅 LB 可 REJECT； RECALL：需风险标注	高（静默漏解）
C4	MatchKey / Conflict / SafePeel	组合	去重确定性 + 冲突显式化 + 仅 core 内可剥离	中
C5	BtpSolverBounds	内核	LB 可证明、单调；支持 Top-K 截断	极高（错误剪枝）

Contract 不是优化，而是“鱼控护栏”：
违反 Contract 的代码，哪怕跑得快，也不被允许进入主干。

6. Engineering Guidance（工程落点）

- 模块化：8 段流水线必须可单测、可统计、可日志化；

- **EXACT 默认**：新模块先在 EXACT 模式下通过，再谈 RECALL；
 - **DEFER 通道**：任何“不确定但可能有解”的候选必须有去处；
 - **可逆调试**：Peel / Remove 必须可回放（日志或版本化）。
-

7. Roadmap & Evolution

- Phase-1：Toy Sequence / Graph 在 EXACT 模式下跑通（证明不漏）；
 - Phase-2：引入 CCC-based LB 加强，提升 EXACT 性能；
 - Phase-3：RECALL-Controlled 大规模部署，DEFER 回补闭环；
 - Phase-4：跨域应用（生物结构、复杂图、长期时间序列）。
-

8. Closing Remark（鱼控宣言）

我们不是在“钓一条鱼”，也不是“撒一张网”。

我们做的是鱼控：

先控制生态、边界与演化方向，

再谈速度、规模与收益。

ITEM #229 作为 UnalignedAND-BTP 的方向盘与骨架，

其价值不在于“算得多快”，

而在于“永远知道自己有没有漏掉真正的鱼”。
