Cronus: Computer Vision-based Machine Intelligent Hybrid Memory Management

Thaleia Dimitra Doudali (IMDEA Software Institute)*
Ada Gavrilovska (Georgia Tech)

@ MEMSYS 2022

* Work done while at Georgia Tech (PhD).
Hybrid Memory Management is Complex

Dynamic Data Movements!

It is a complex decision mix to manage the data allocated across memories.

E.g., Which / How much / Where / When to move data?

Why do we need more intelligent systems?

- ZBs
- Application data sizes
- Complex data access patterns
- Exploded system parameter space

Hard to balance
Machine Intelligent Hybrid Memory Management

The Vision.

Applications

System-level Resource Manager

Data Access Monitoring

ML training

Re-train?

prediction

yes

no

Trained Models

Resource Management

practical

Heterogeneous Hardware

Hybrid Memory
Machine Intelligent Hybrid Memory Management

Laying the grounds for the practical integration of ML.
**System Design of Kleio**

**Kleio** is a hybrid memory page scheduler with machine intelligence. [Best Paper Award Finalist at HPDC 2019.]

---

1. **Page Access Monitoring**
   - Page Selector
     - small subset
     - Pages for ML
     - bigger subset
     - Pages for History

2. **Page Hotness Prediction**
   - ML-based predictions (Per page RNN models)
   - History-based predictions
   - Page Hotness Prediction

3. **Page Migration Selection**
   - Calculate hot vs. cold pages

---

**Kleio** extends existing lightweight hybrid memory management with the necessary amount of machine intelligence to boost application performance.

---

**Applications**

**System-level Resource Manager**

**Heterogeneous Hardware**

---

**Not all pages “need” ML.**
The Key(s) to a Practical and Efficient ML-based System Design

Apply ML **when** and **where** necessary.

- **Page Selector**
  - Priority
  - Page hotness predictions
  - Pool of Pages
  - History

- RNN

- **Memory Access Trace**
- **Hybrid Memory Configuration**

Apply ML on a small page subset.

- Foundations for practical use of ML.

Carefully select pages for ML.

- Application performance boost.

The page selection is not a lightweight process.

Can we accelerate the page selection process?
Can we accelerate the page selection process via image-based decisions?
Insight from Visualizing Pages Selected for ML

Neighboring pages that are part of distinct access patterns across time receive similar priority for ML.
Cronus: Image-based Page Selection Pipeline

1. **Image Creation:**
   1. 256x256 image size.
   2. Page priority = hotness x hotness variance.
   3. Perceptually uniform colormap.

2. **Pattern Detection** with automatic color thresholding.

3. **Page Selection** with reverse pixel-to-page mapping.

Open Source Codebase:
https://github.com/GTkernel/cronus-sim
Evaluation

**Objective:** Evaluate the effectiveness of the Page Selection.

**Comparison**
- **Kleio:** Performance-based Page Selection.
- **Cronus:** Image-based Page Selection in viridis colormap.
- **Grayscale:** Image-based Page Selection in grayscale colormap.
- **Viz-Black:** Image-based Page Selection in black-and-white.
- **Thres-Elbow-Benefit:** Analytical Page Selection with thresholds.

**Goals** (compared to Kleio)
- ~ Similar Page Selection Quality.
- ~ Similar Application Performance.
- ↓ Reduced Page Selection times.
Evaluation

Page Selection Quality.

Cronus $\sim$ Kleio

Application Performance.

Cronus $\sim$ Kleio

Cronus makes an image-based high quality page selection that delivers similar performance to Kleio.
Evaluation

Page Selection times.

Cronus drastically reduces by $400x$ the page selection times, down to few seconds.

Remaining Challenges.

Larger Workloads
- More Patterns.
- Harder to Visualize.
Summary of Cronus

Takeaways:

- **An image is worth a thousand.. lines of code.**
  - Image processing and computer vision methods can unlock new opportunities in reducing system complexity and overheads.

- **It is all about the image color and metadata.**
  - A perceptionally uniform colormap best captures the most effective page ordering.
  - The metadata enable a standalone image-based pipeline with no need to store huge raw data (memory access trace).

---

Greek Trivia: According to the ancient Greek mythology, Cronus (Kronos) was the King of the Titans and the god of time.

Open Source Codebase: [https://github.com/GTkernel/cronus-sim](https://github.com/GTkernel/cronus-sim)