Write Amplification Reduction in Flash-Based SSDs Through Extent-Based Temperature Identification

Mansour Shafaei   Peter Desnoyers   Jim Fitzpatrick

Outline

- Background
- Related Work
- Extent-Based Temperature Identification
  - Evaluations
- Conclusion

Background

- Performance of flash-based storage dominated by
  - Cleaning costs (Write Amplification)
    - The number of internal copies required before erasing blocks

- WA reduction
  - Selection (Non-predictive)
    - Cost-Benefit based block selection
  - Placement (Predictive)
    - Hot/Cold data segregation
      - High performance
      - Requires write frequency/recency tracking

Data Temperature Identification Schemes

- Implicit
  - Length heuristic
    - Short host writes → shorter-lived data
      - Inaccuracy issues

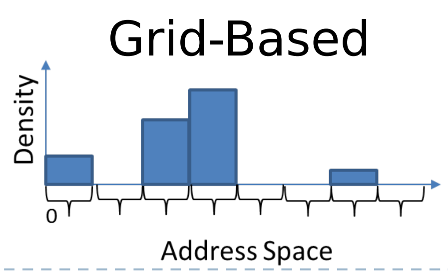
- Explicit
  - Per LBA data access tracking
    - EWMA
      - Requiring excessive resources
  - Approximated
    - E.g., Multiple Hash Functions (MHF) (Hsieh et al.) and Multiple Bloom Filters (MBFs) (Park et al.)

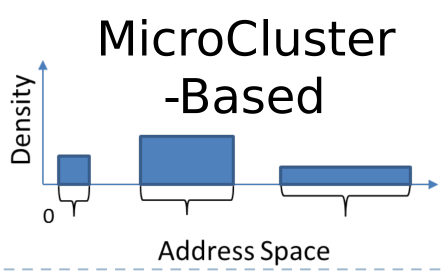
Density-based clustering

- Data temperature identification can be considered as a density-based stream clustering problem
  - Well-studied subject in the AI domain

- Definition
  - Finding arbitrary shape clusters based on the density of data in a given address space
    - Forming dense clusters separated by sparse areas

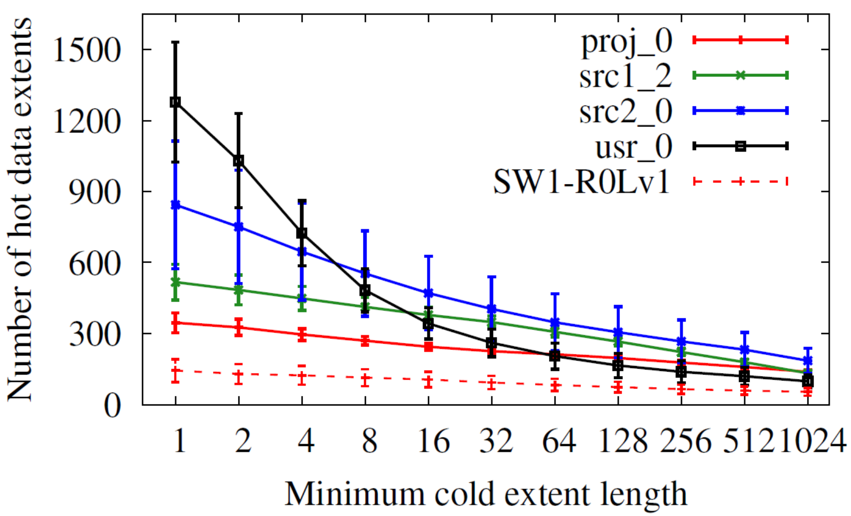
- Types

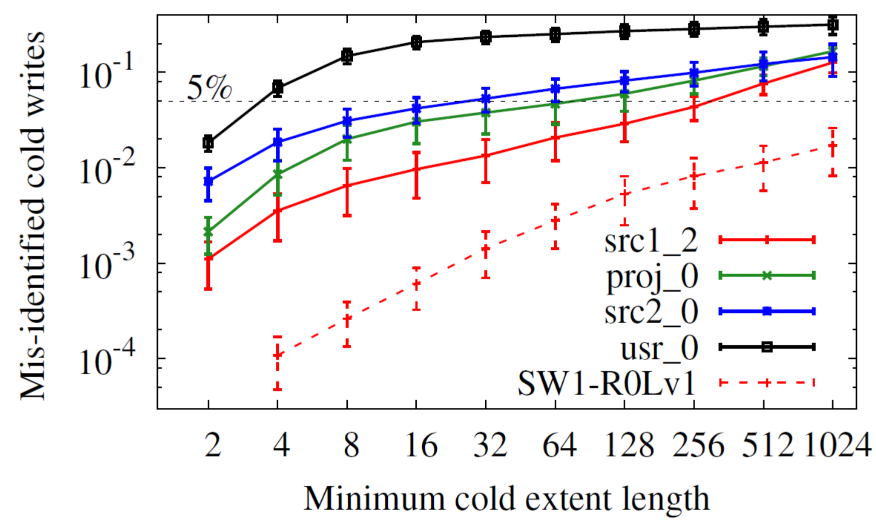




Applicability of Density-Based Clustering to Storage Workloads

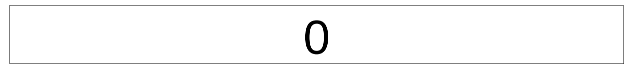
- Running trace driven test to measure:
  - Number of extents needed for accurate identification
  - Accuracy loss by merging hot extents separated by small amounts of cold data





Extent-Based Temperature Identification (ETI)

- Beginning with a single extent covering the entire address space



- Extent and counter updates based on incoming writes
  - If boundaries of corresponding extent match incoming write's offset and length:
    - Increments the counter

  - Otherwise, if # of extents < max. # of allowed extents:
    - Creates new extents accordingly and increments the counter values

  - Else:
    - Increments the extent counters fully or partially(>α%) overwritten by the write

ETI (cont.)

- Periodic aging
  - Divide all counters by two and round off
  - Merge all adjacent hot extents with equal counter values
  - Merge all adjacent cold (counter < β) extents

ETI Overhead

- More complex
  - Requiring an ordered search structure vs. Bloom filter
- Per write operation execution
  - Vs. existing scheme's per block operations
  - Up to 96% reduction among available traces
- Optimized search by taking advantage of spatial locality
  - Beginning each search at the location of latest accessed extent

Experimental Setup

- Environment
  - FTLSim
    - A high-level SSD simulator written in C and Python
    - Page-map FTL
    - 32K of 128-page blocks ~ 16 GB
    - 7% spare factor
    - A global greedy cleaning algorithm with hot and cold data segregation

- Metric
  - Write Amplification

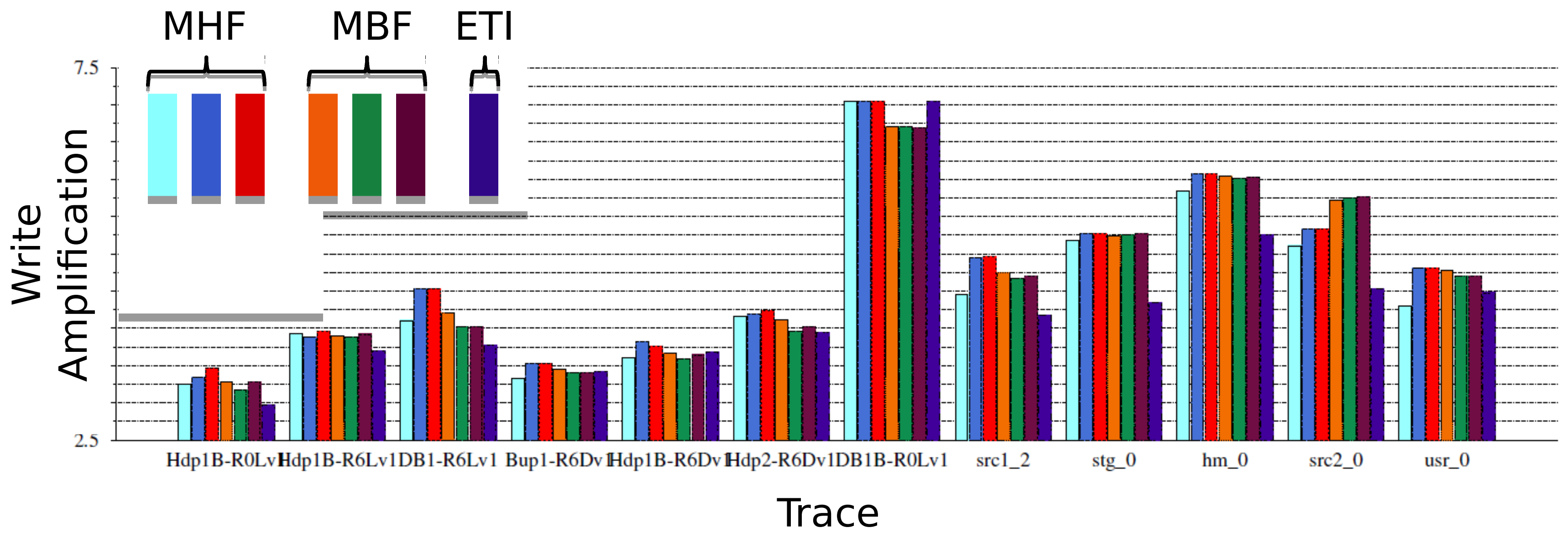
Experimental Setup (Cont.)

- Schemes
  - ETI (max. # of extents=2048)
  - MHF (table size=2048, CRC HFs)
  - MBF (BF size=2048, CRC HFs)

- Workload
  - Selected MSR traces
  - A series of traces provided by an industrial partner
    - One RAID-0 and two RAID-6 variants
    - Two backup applications, a database, a Hadoop HDFS cluster, and an unspecified software application

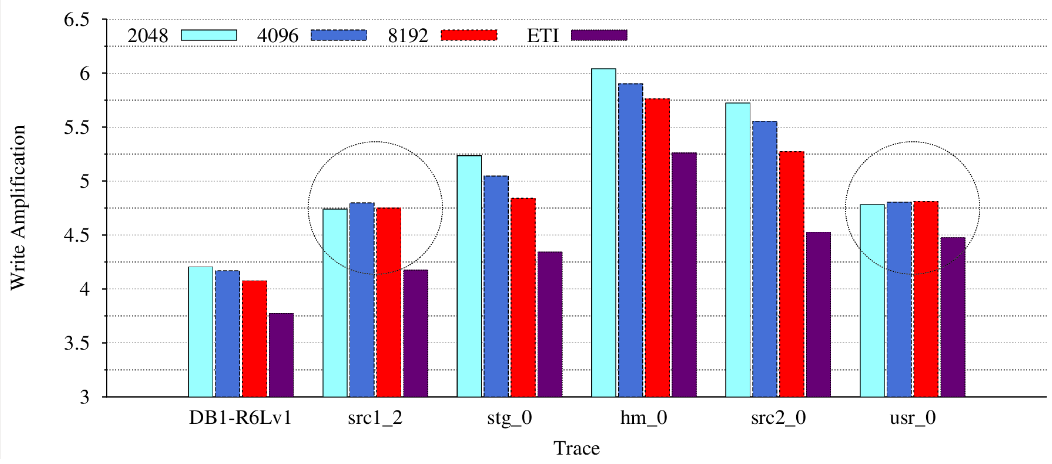
Performance Comparison

- Using ETI
  - Up to 21% WA reduction
  - Never worse than MHF
  - Only 4% worse in one single trace compared to MBF



Performance Comparison (cont.)

- Dedicating more memory resources to the implemented hash-based schemes for a more fair comparison
  - MBFs with twice and 4 times larger BF sizes
  - Modest improvements in four traces
    - Significantly less than what achieved by ETI
  - Almost no improvement in two other traces!



Conclusion

- We propose Extent-based Temperature Identification (ETI) as a one-dimensional density-based clustering scheme to get used combined with a dual frontier, page-mapped Greedy translation layer.
- We measure by write amplification reduction unlike previous work which used identification accuracy.
- Results show by up to 21% WA reductions over both Multiple Hash Functions and Multiple Bloom Filters schemes.