malloc

- mechanism
  - block metadata
    - header
      - boundary tag
        - coalesce support
      - header fields
        - block size
        - in-use bit
        - flags
          - prev-in-use
          - alignment
    - footer
      - boundary tag
  - memory block
    - block
      - coalescing of free blocks
        - immediate
          - fast
            - double-linked list
        - delayed
          - speed optimization
      - block splitting
        - splitting threshold
      - reuse
        - end of block
          - block tail reuse
          - exact-fit reuse
        - free block selection strategy
          - list order
          - optimal fit
          - half fit
          - ordering
            - size-ordered
            - address-ordered
            - random-fit
  - arena
    - end of heap
      - top chunk / wilderness block
        - high watermark
    - multiple arenas
- policy
  - locale
    - free memory
    - kernel memory
    - userspace
    - NUMA node (implied)
  - type
    - solid state drive
    - RAM
    - storage-based RAM
      - swap
      - remote memory system (disaggregate memory)
  - free list policy
    - ordering strategy
    - per-size vs global lists
    - immediate vs. deferred reuse
  - allocator design
    - lookup table
      - high-mark, end block of heap
      - free subpool
        - local coalescing
        - per-arena or global structures
- allocation
  - binning
    - per-run binning
    - size bins (per-size class)
  - size-class
    - fixed-sized
    - rounded-size
      - power of two
      - alignment constraint
  - run
    - block pool for a single size-class
    - managed
      - free list
      - bitmap
  - layout
    - arena
      - array slot
        - tcache indexing
      - linked list
        - circular doubly-linked list
        - chain-style list
    - page/chunk management
      - large vs. small chunk
      - page alignment and mapping
      - multiple page size (huge page)
    - cache locality
      - NUMA-aware placement
      - cache-line alignment
      - prefetching hints
    - thread-local storage
      - per-thread arenas
      - thread cache (tcache)
    - fragmentation management
      - slab allocation layout
      - run-based layout
    - locking strategy
      - lock-free vs. locked in arena or bin
      - fine-grained locking per-arena or size-class
    - memory tagging or protection
      - guard page
      - metadata isolation
    - metadata storage
      - inline metadata in header
      - separate metadata pool
    - allocator-specific structure
      - bitmaps
      - hash table
      - tree
        - radix
- strategy
  - fit and reuse
    - ordering
      - address-ordered fit
        - first-fit
          - first-fit decreasing
        - next-fit
          - roving point
      - sized-ordered fit
        - worst-fit
        - best-fit
          - best-fit decreasing
      - random-ordered fit
        - random-fit
    - fit type
      - first-fit
      - best-fit
      - next-fit
        - roving poiint
      - worst-fit
      - exact-fit
      - optimal-fit
      - random-fit
    - segregated free list
      - segregated fit [PSC71]
        - exact list
          - indexed fit
            - fast fit
            - bitmapped fit
          - bitmapped fit
      - segregated storage
        - sequential
          - first-fit
            - address-ordered fit
            - first-fit decreasing
            - next-fit
              - roving point
        - simple-fit
          - worst-fit
          - best-fit
            - best-fit decreasing
      - strict-size class
        - strict size classes w/ rounding
          - buddy system
            - conventional binary buddy
            - Fibonacci buddy
              - generalized Fibonacci buddy
            - weighted buddy
            - double buddy
        - strict-size classes w/ range lists
          - first-fit decreasing
          - best-fit decreasing
  - coalescing & splitting
    - coalescing
      - immediate
        - fast
          - double-linked list
      - delay
        - speed optimization
    - splitting
      - block-splitting
        - splitting threshold
  - frequency (reuse timeline)
    - allocation request
    - free() calls
  - performance pattern
    - live predetermination
      - sizing streaming
        - uniform size increment
        - exponential sizing
      - ordering
    - process behavior
      - phase behavior
      - statistical distribution
        - statistical delay process
          - random-decay
            - sudden death, group of
  - design objective
    - portability
    - compatibility
    - space efficiency
    - time performance
      - plateau
      - slow ramp
        - forward
        - backward
      - fast peak
    - error detection
      - anomalies
        - coalescing bug
        - reuse bug
  - allocation
    - binning
      - per-run
      - size bin
    - size-class
      - fixed-size
      - rounded-size
      - indexed size
    - run
      - block pool for a size-class
      - free list managed
    - layout
      - arena
        - array slot
        - linked list
          - circular doubly-linked list
          - chain-style list

Presented with xmind