Functional Performance

Martin Thompson - @mjpt777

Are all memory operations equal?

Sequential Access

-

Average time in ns/op to sum all longs in a 1GB array?

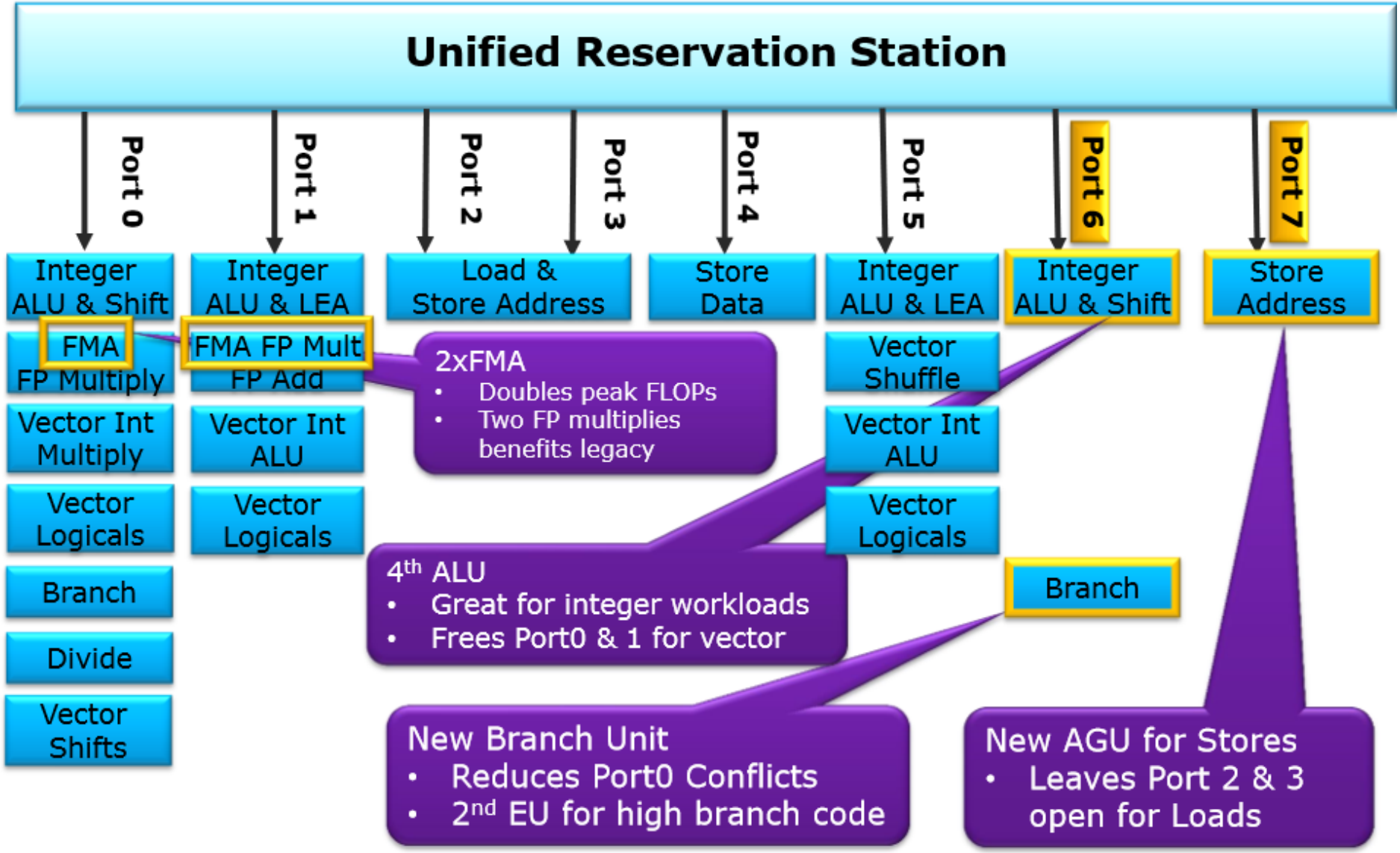
Access Pattern Benchmark

```
Benchmark                  Score     Error   Units
=================================================
sequential                 0.832  ± 0.006   ns/op
```

~1 ns/op

Really???
Less than 1ns per operation?

Haswell Execution Unit Overview



Access Pattern Benchmark

Benchmark	Score	Error	Units
sequential	0.832	± 0.006	ns/op
randomPage	2.703	± 0.025	ns/op
dependentRandomPage	7.102	± 0.326	ns/op
randomHeap	19.896	± 3.110	ns/op
dependentRandomHeap	89.516	± 4.573	ns/op

Access Pattern Benchmark

Benchmark	Score	Error	Units
sequential	0.832	± 0.006	ns/op
randomPage	2.703	± 0.025	ns/op
dependentRandomPage	7.102	± 0.326	ns/op
randomHeap	19.896	± 3.110	ns/op
dependentRandomHeap	89.516	± 4.573	ns/op

~90 ns/op

Some Fundamental Laws

$$r = s(2 - \rho) \; / \; 2(1 - \rho)$$

Queueing Theory

$$r = s(2 - \rho) / 2(1 - \rho)$$

r = mean response time

s = service time

ρ = utilisation

Queueing Theory

$$r = s(2 - \rho) / 2(1 - \rho)$$

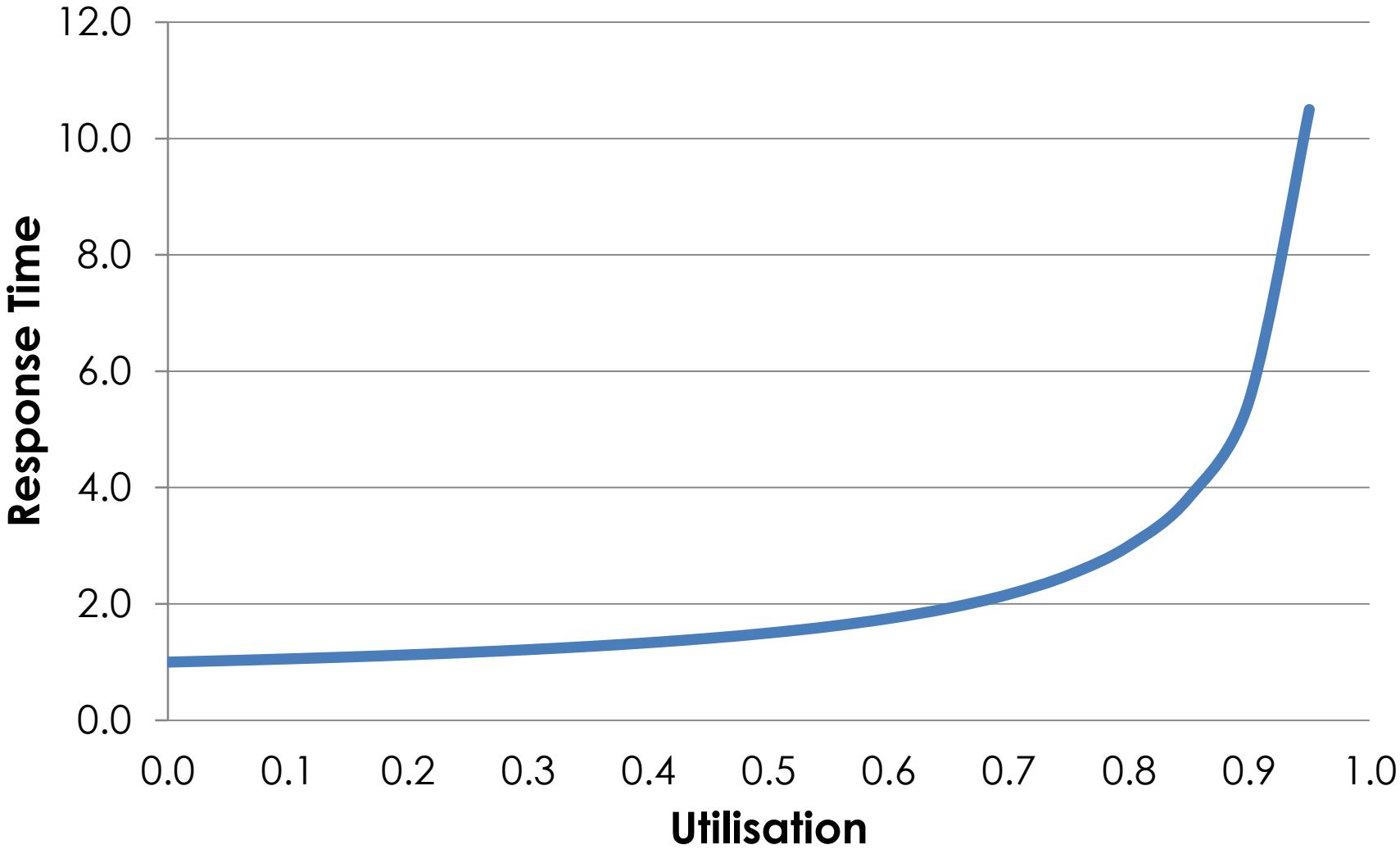
r = mean response time

s = service time

ρ = utilisation

Note: $\rho = \lambda * s$

Queueing Theory



$$C(N) = N / (1 + \alpha(N - 1) + ((\beta * N) * (N - 1)))$$

Universal Scalability Law (USL)

$$C(N) = N / (1 + \alpha(N - 1) + ((\beta * N) * (N - 1)))$$

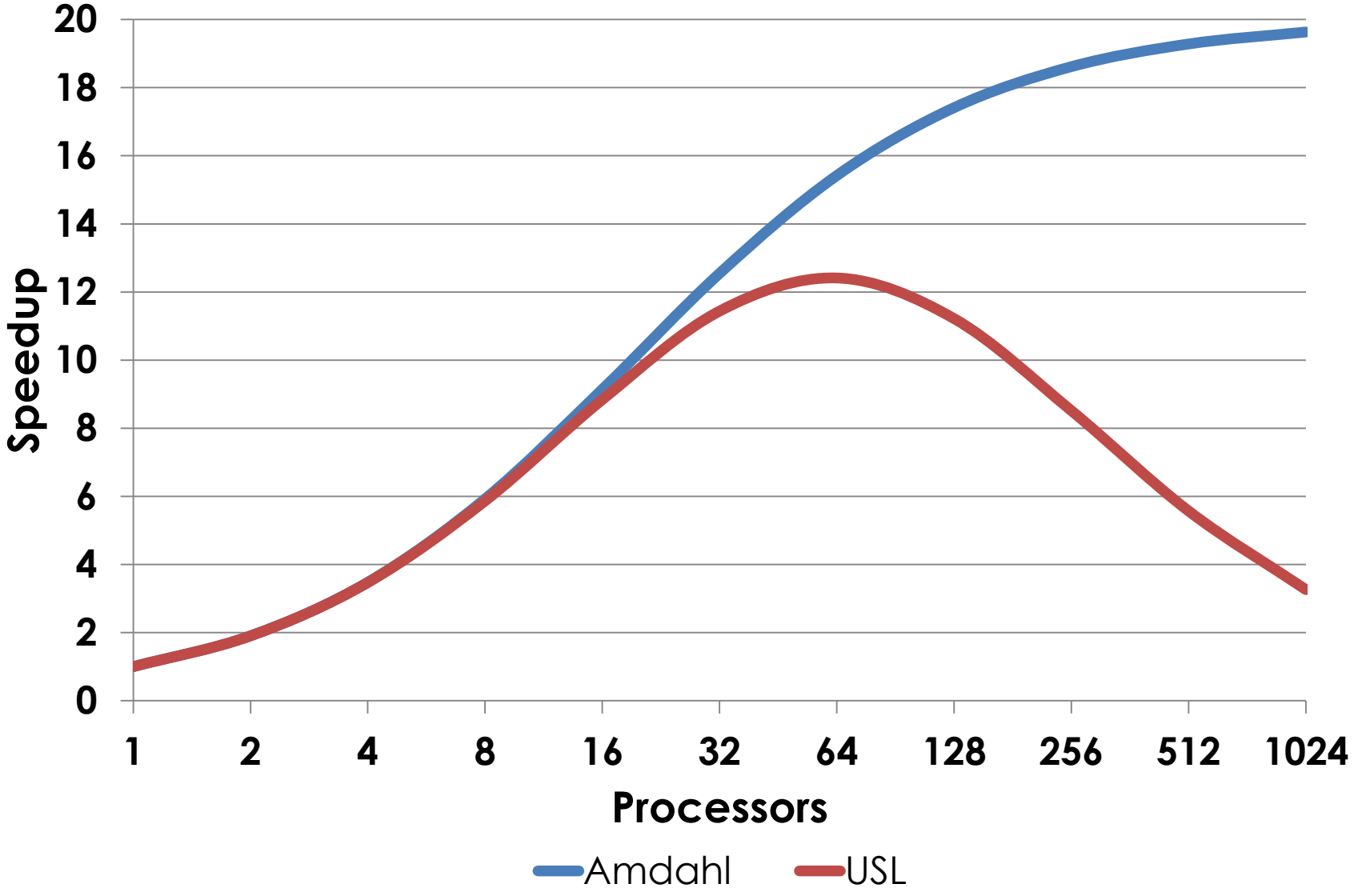
C = capacity or throughput

N = number of processors

α = ***contention*** penalty

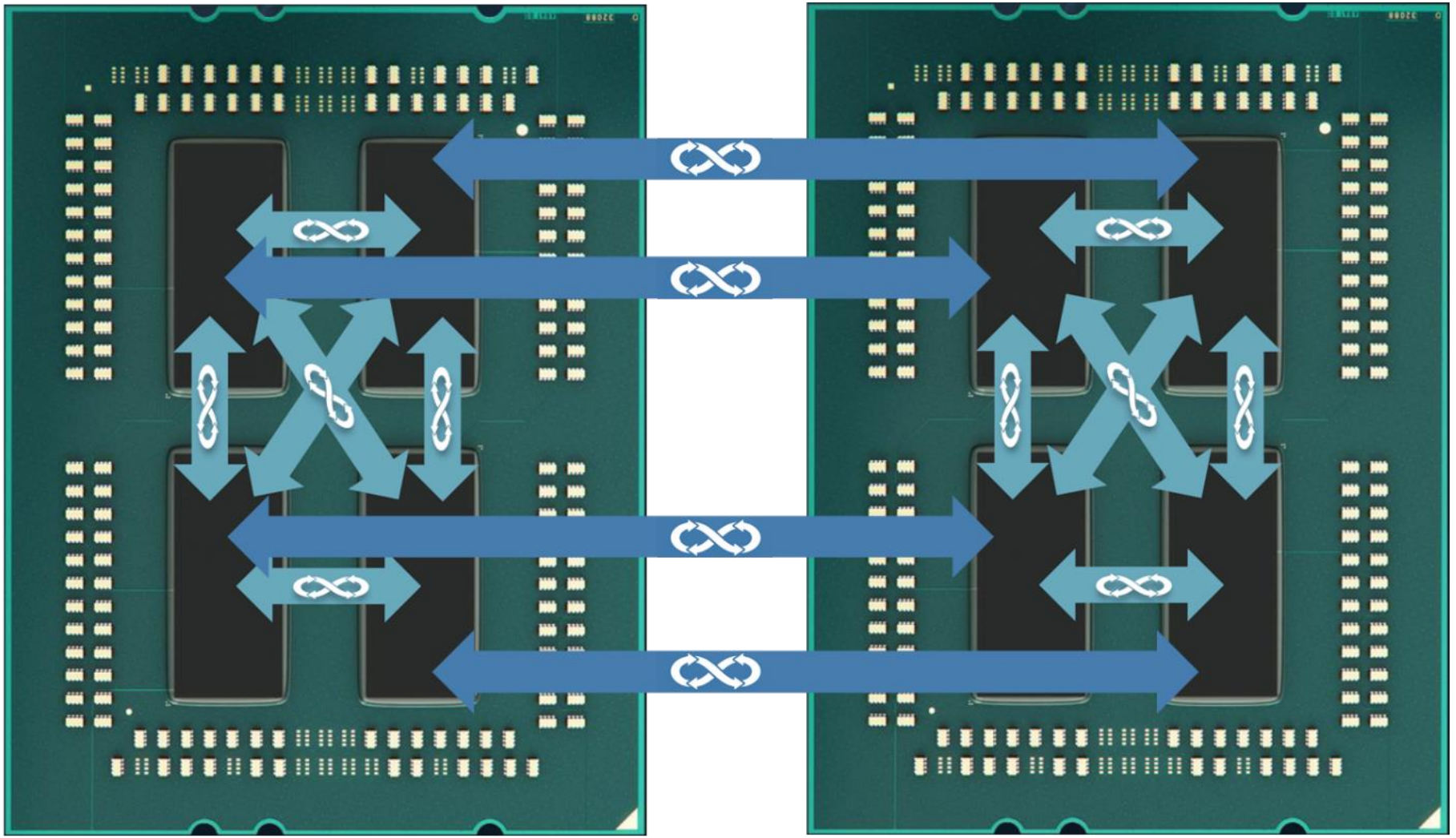
β = ***coherence*** penalty

Universal Scalability Law (USL)



GRADUATE

AMD EPYC Infinity Fabric



Source: AMD

Systems Engineering

Messaging

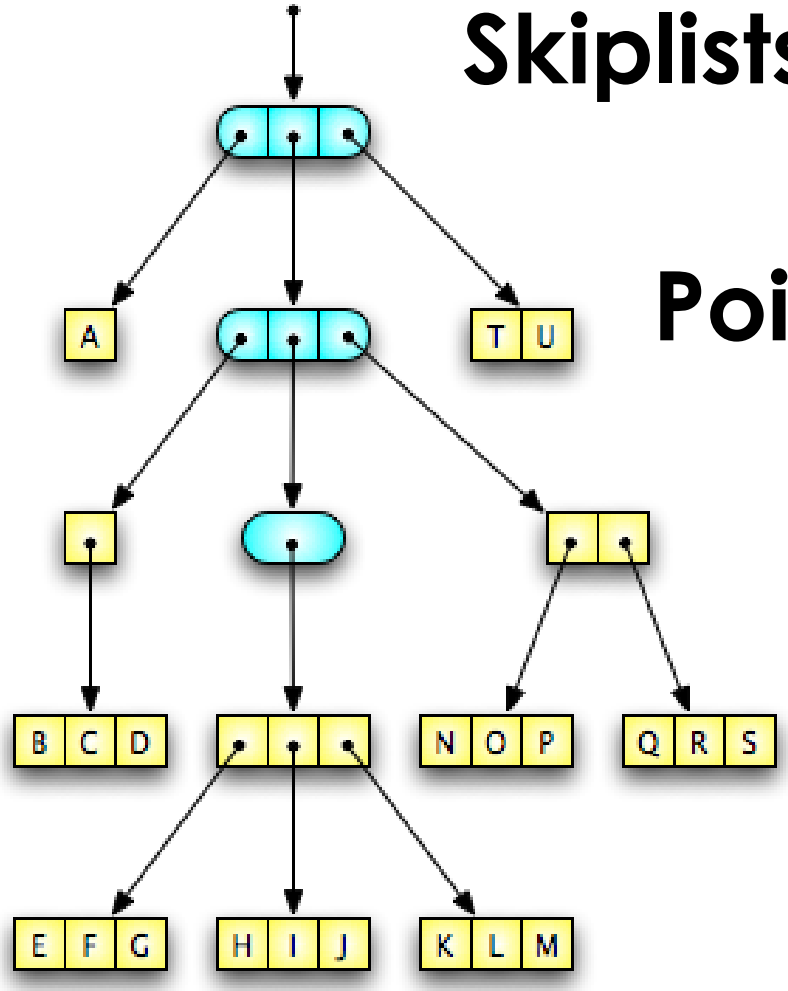
Concurrent
Distributed
Network & Protocol Understanding

What are the normal approaches?

Mechanical Sympathy

Skiplists, Trees, Scoreboards

Mechanical Sympathy



Skiplists, Trees, Scoreboards

->

Pointer chasing hell

Immutability ?

Concurrent Tree Updates

=>

Path Copy

Garbage Collection Hell

Functional data structures
are like sausages,
the more you see them being
made, the less you will sleep

CRDT ???

CRDT ???

Conflict-free Replicated Data Type

CRDT ???

Conflict-free Replicated Data Type

Operation Based

Commutative Replicated Data Type

CmRDT

CRDT ???

Conflict-free Replicated Data Type

Operation Based

Commutative Replicated Data Type

CmRDT

State Based

Convergent Replicated Data Type

CvRDT

Commutative Replicated Data Types

- **Operations are replicated**
  - **e.g. Add 10 to value, or -10**

Commutative Replicated Data Types

- Operations are replicated
  - e.g. Add 10 to value, or -10
- Operations are commutative
  - i.e. They can be applied in any order

Commutative Replicated Data Types

- Operations are replicated
  - e.g. Add 10 to value, or -10
- Operations are commutative
  - i.e. They can be applied in any order
- Operations are NOT idempotent
  - Network transport must be reliable

Convergent Replicated Data Types

- **State is replicated**
  - **i.e. Whole data structure is replicated**

Convergent Replicated Data Types

- **State is replicated**
  - **i.e. Whole data structure is replicated**

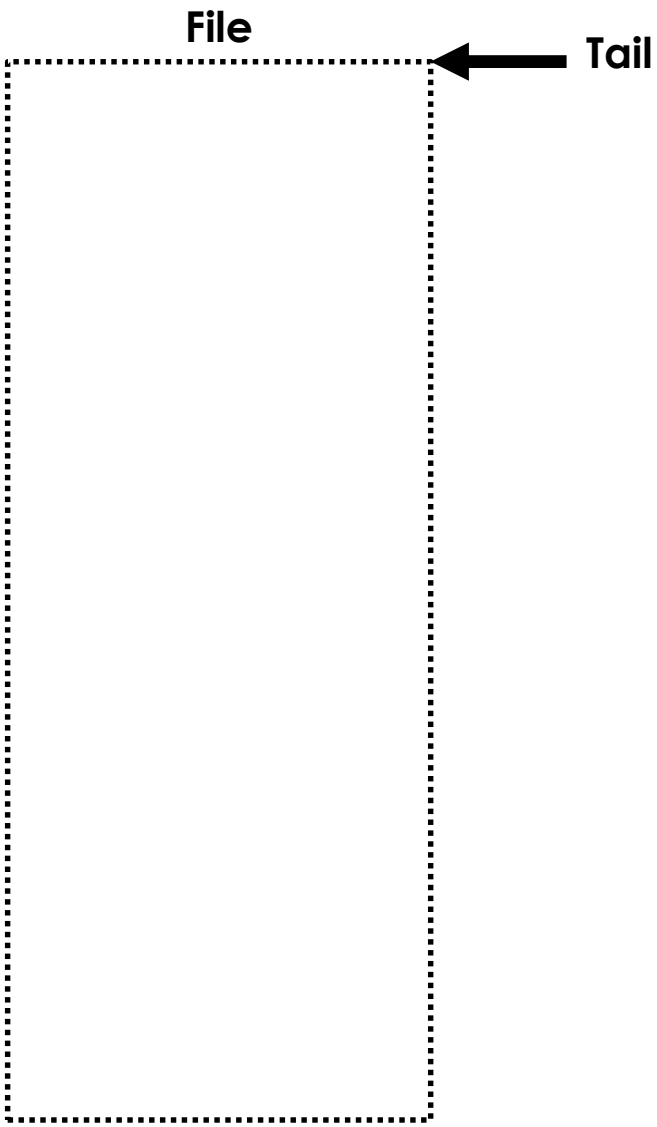
- **Updates must monotonically increase state**
  - **e.g. Append only, remove requires tombstones**

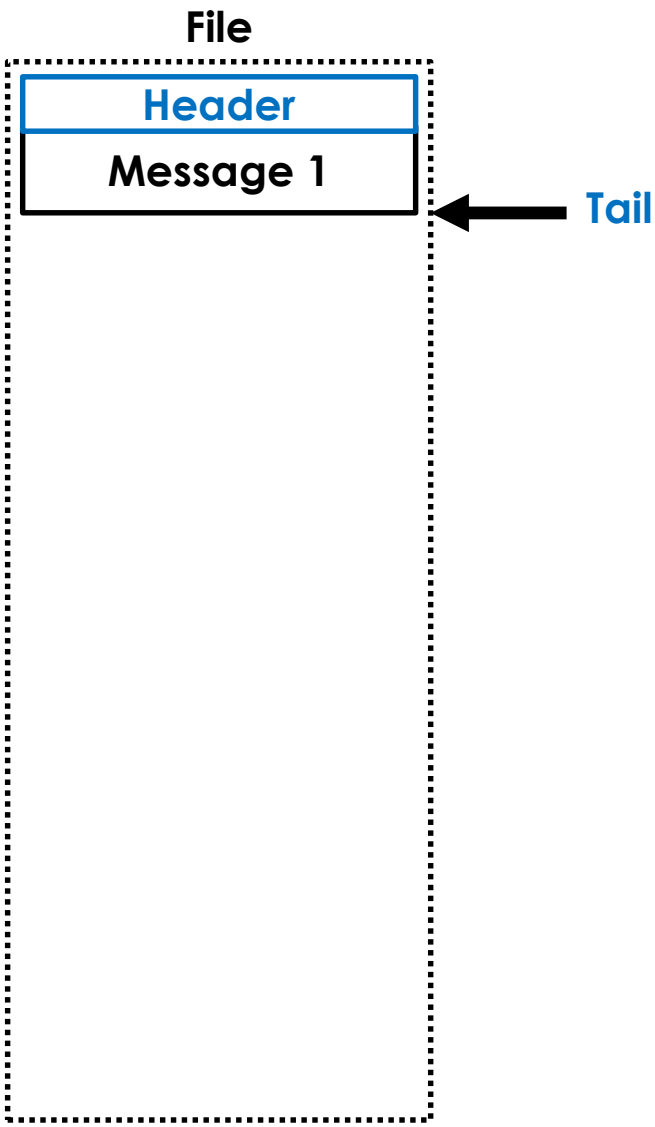
Convergent Replicated Data Types

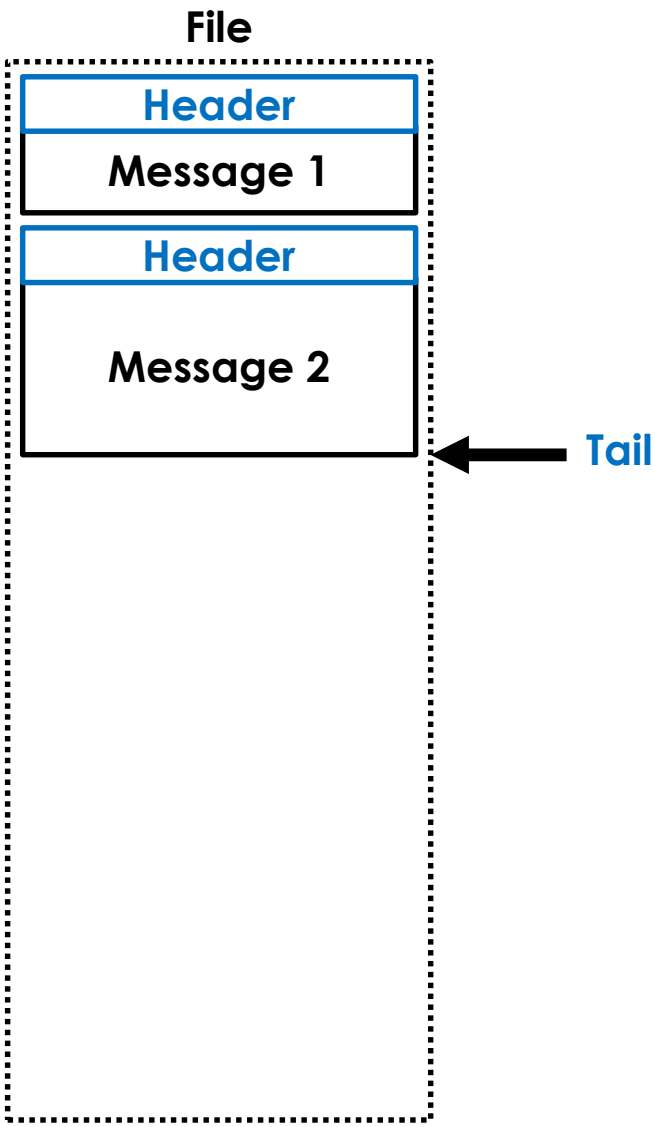
- State is replicated
  - i.e. Whole data structure is replicated
- Updates must monotonically increase state
  - e.g. Append only, remove requires tombstones
- Merge function must be associative, commutative, and idempotent
  - A reliable transport is not required
  - Some designs support deltas for efficiency
  - Resolving conflict can be difficult

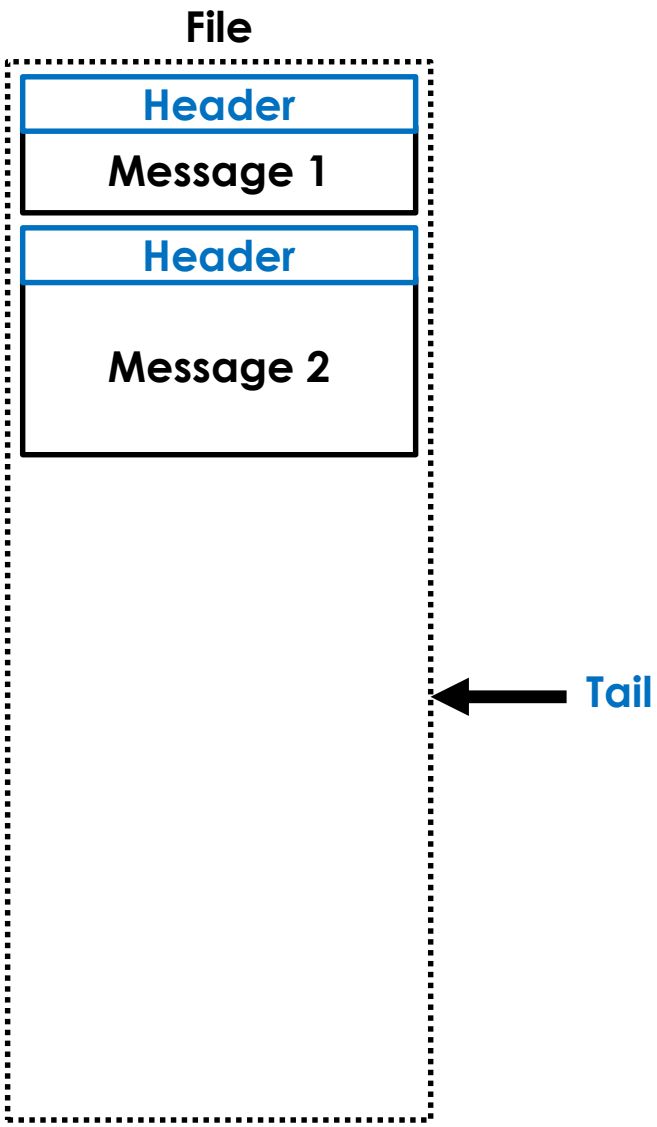
Aeron

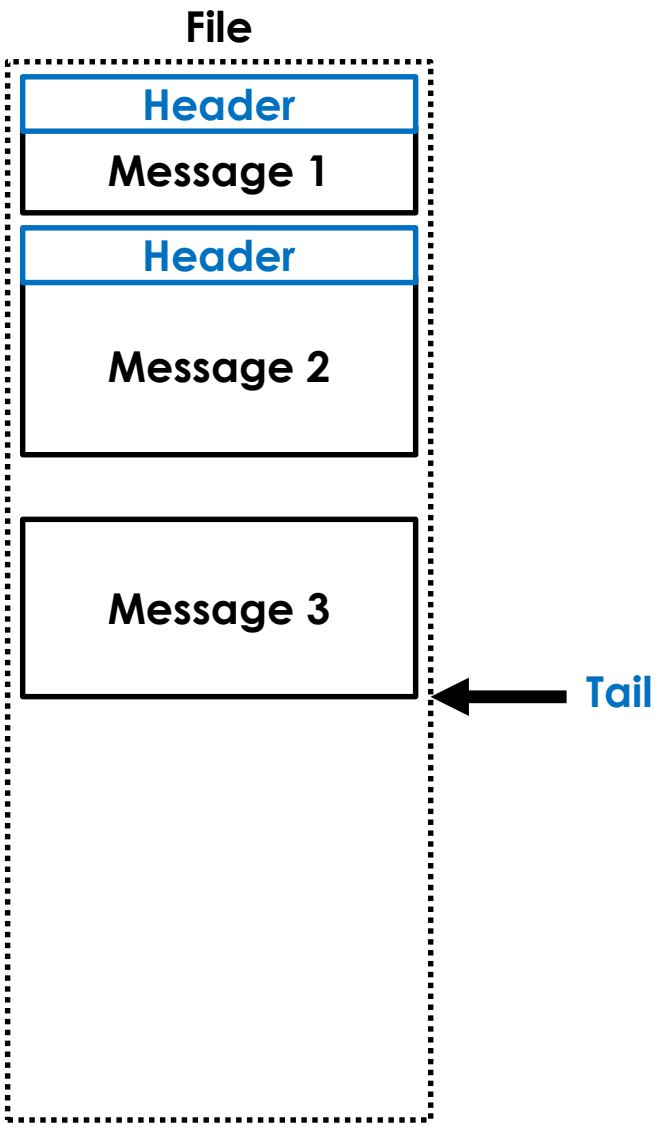
Log Buffer

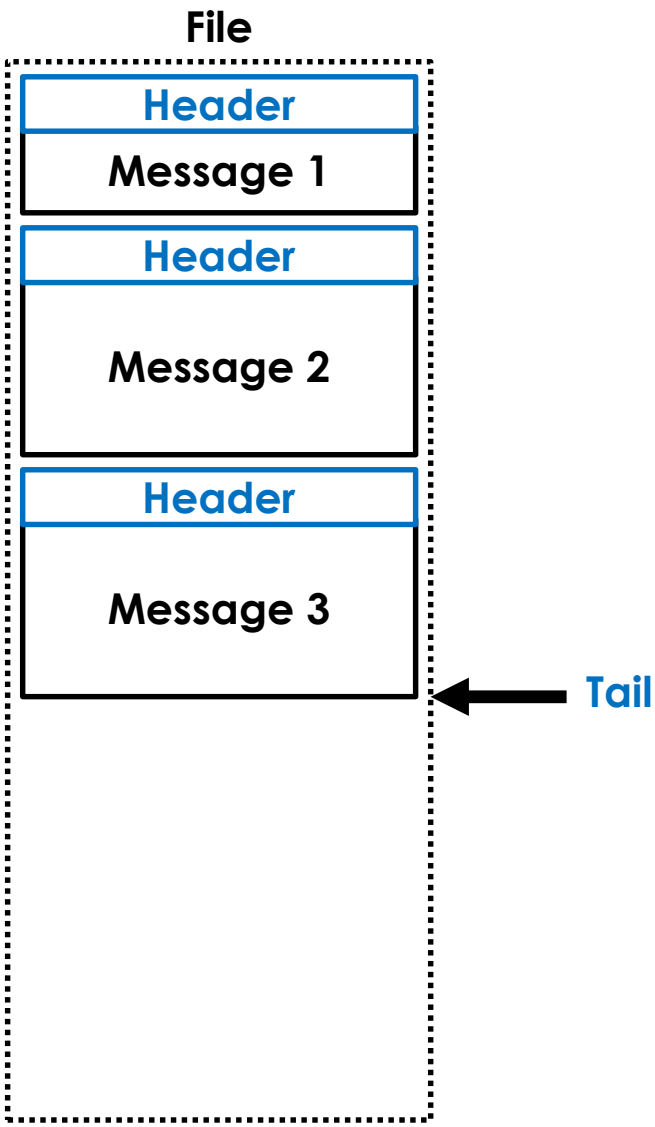

File
Tail


File
Header
Message 1
Tail


File
Header
Message 1
Header
Message 2
Tail


File
Header
Message 1
Header
Message 2
Tail

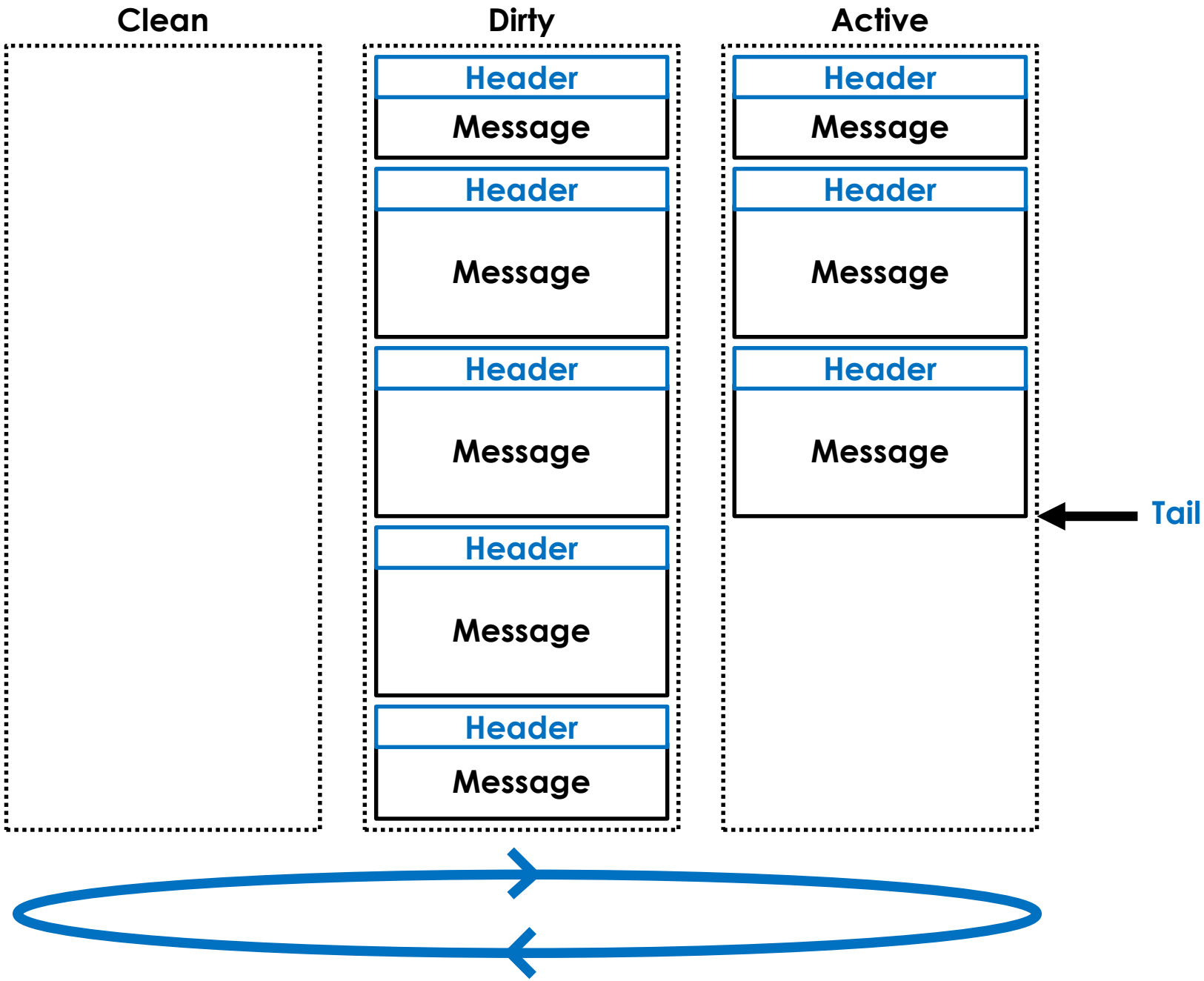

File
Header
Message 1
Header
Message 2
Message 3
Tail


File
Header
Message 1
Header
Message 2
Header
Message 3
Tail

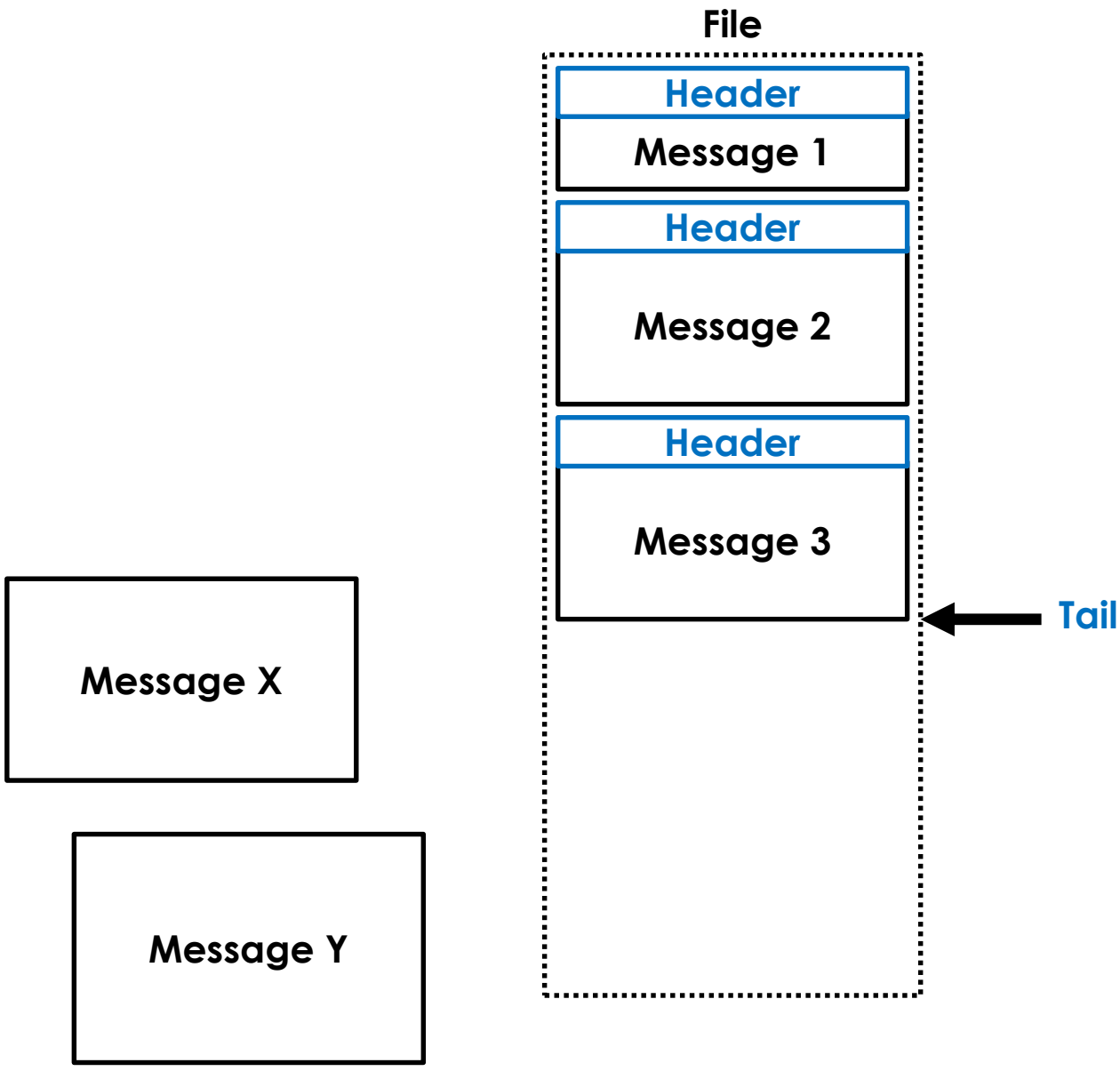
One big file that goes on forever?

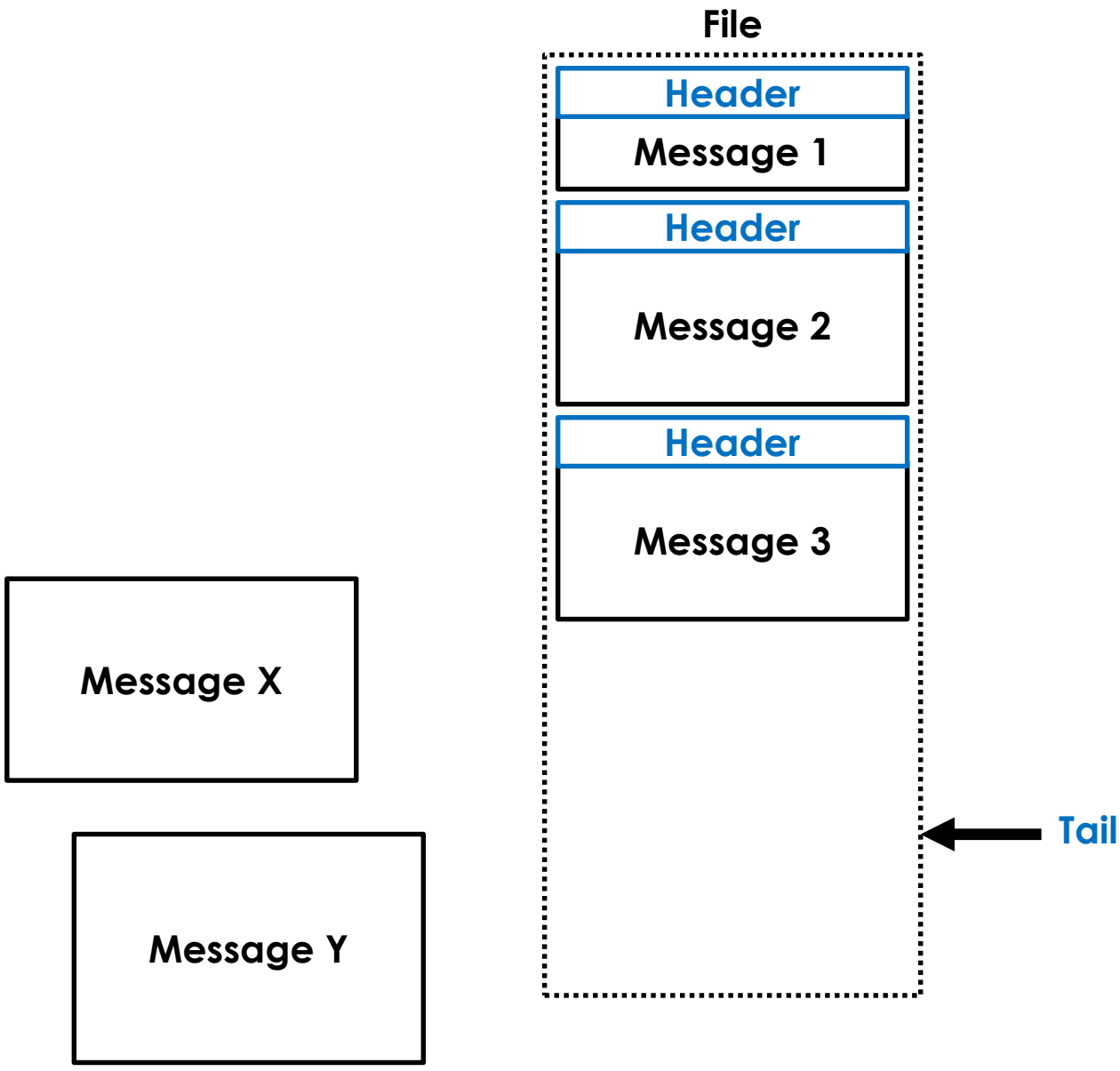
No!!!

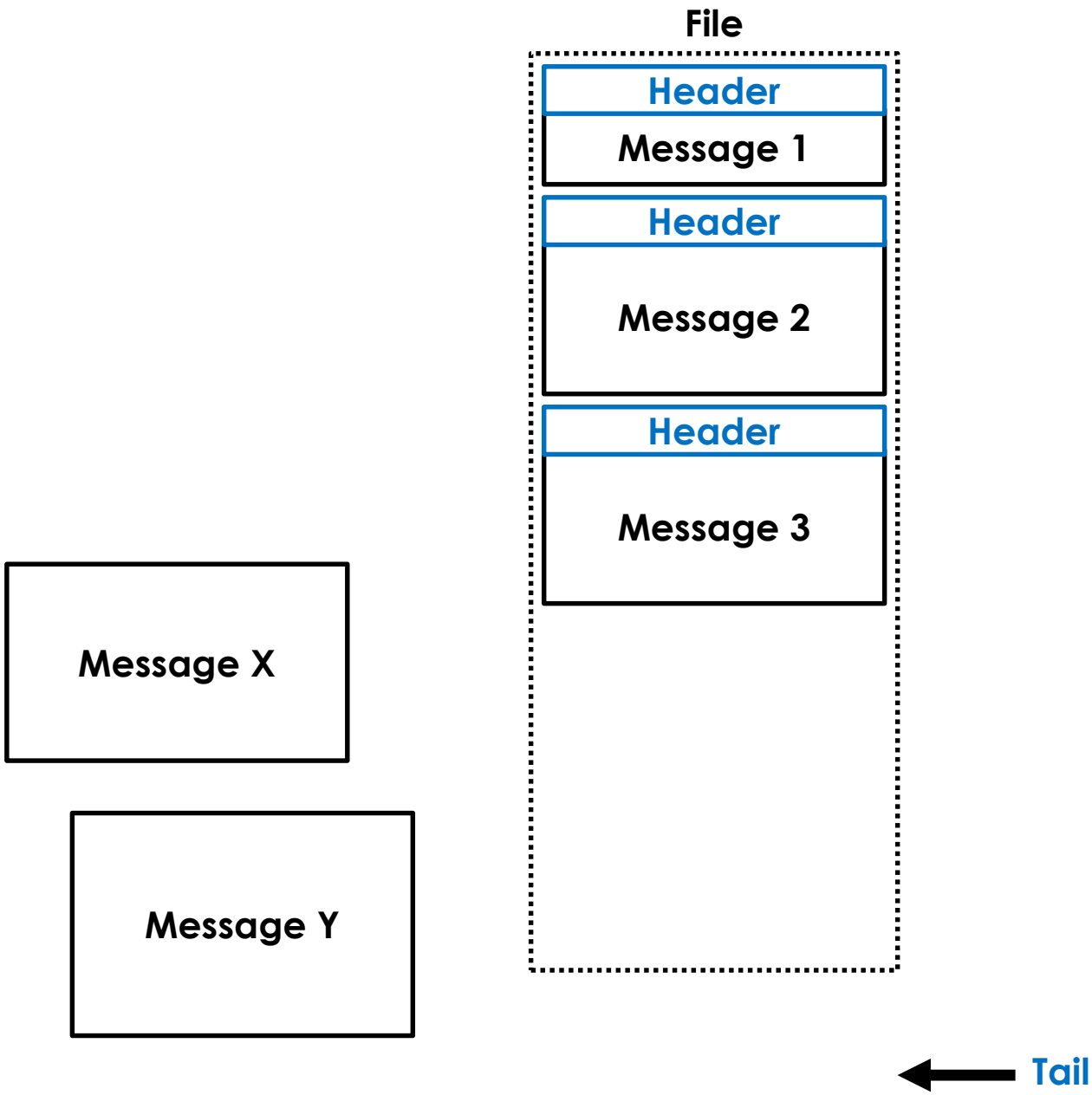
Page faults, page cache churn, VM pressure, ...

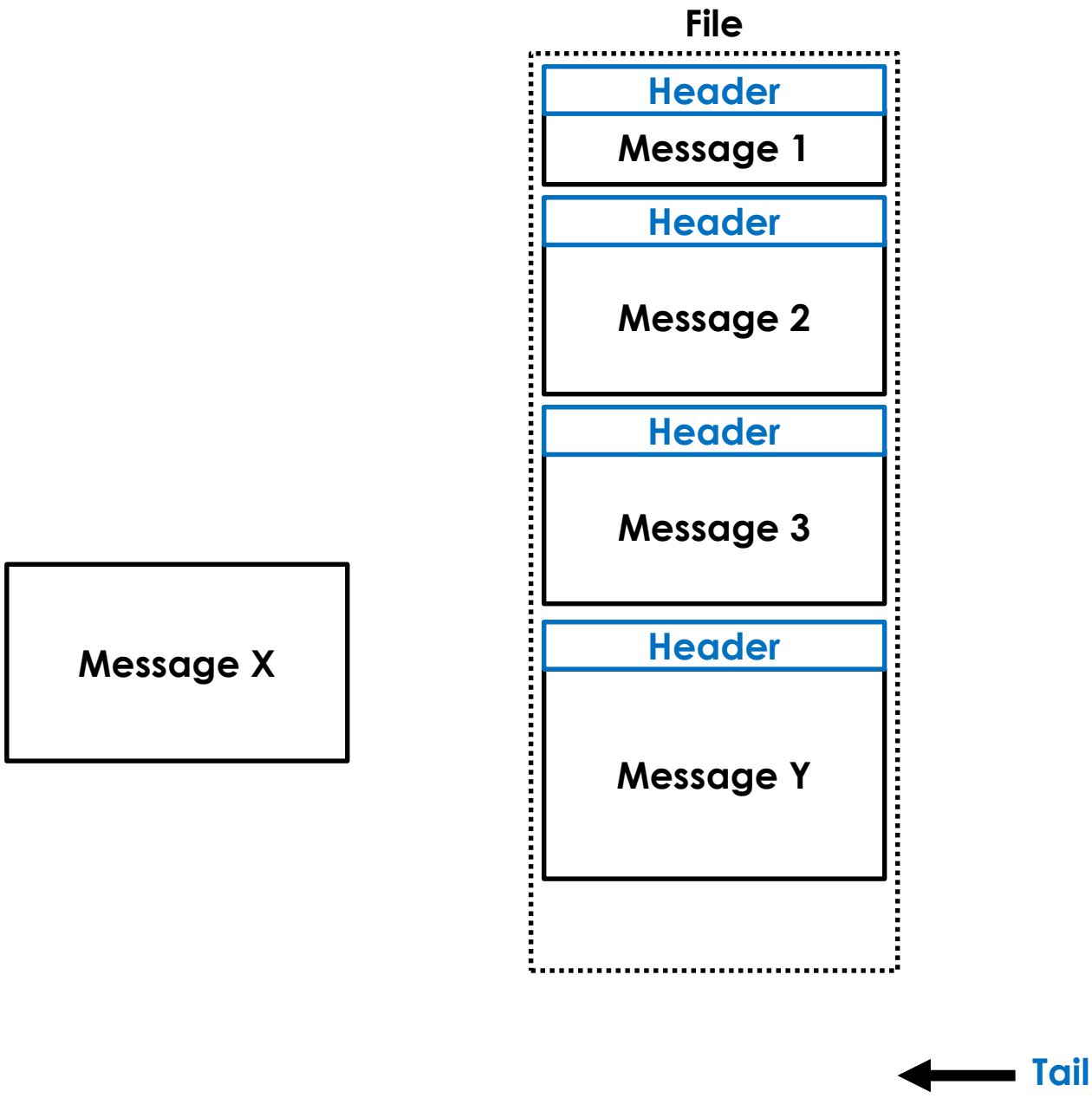

Clean
Dirty
Active
Header
Message
Header
Message
Header
Message
Header
Message
Header
Message
Header
Message
Header
Message
Tail

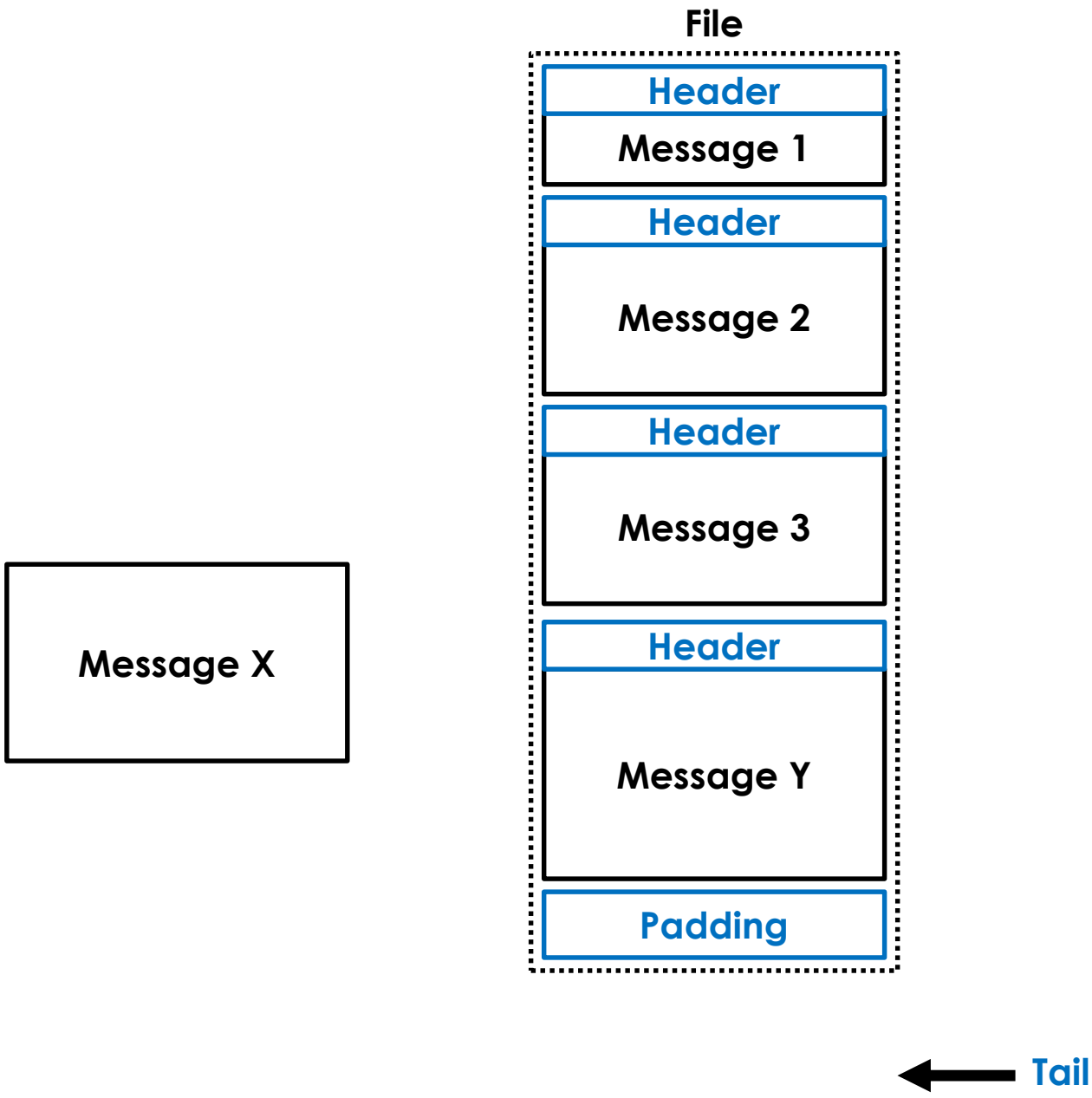
Concurrent Publication?

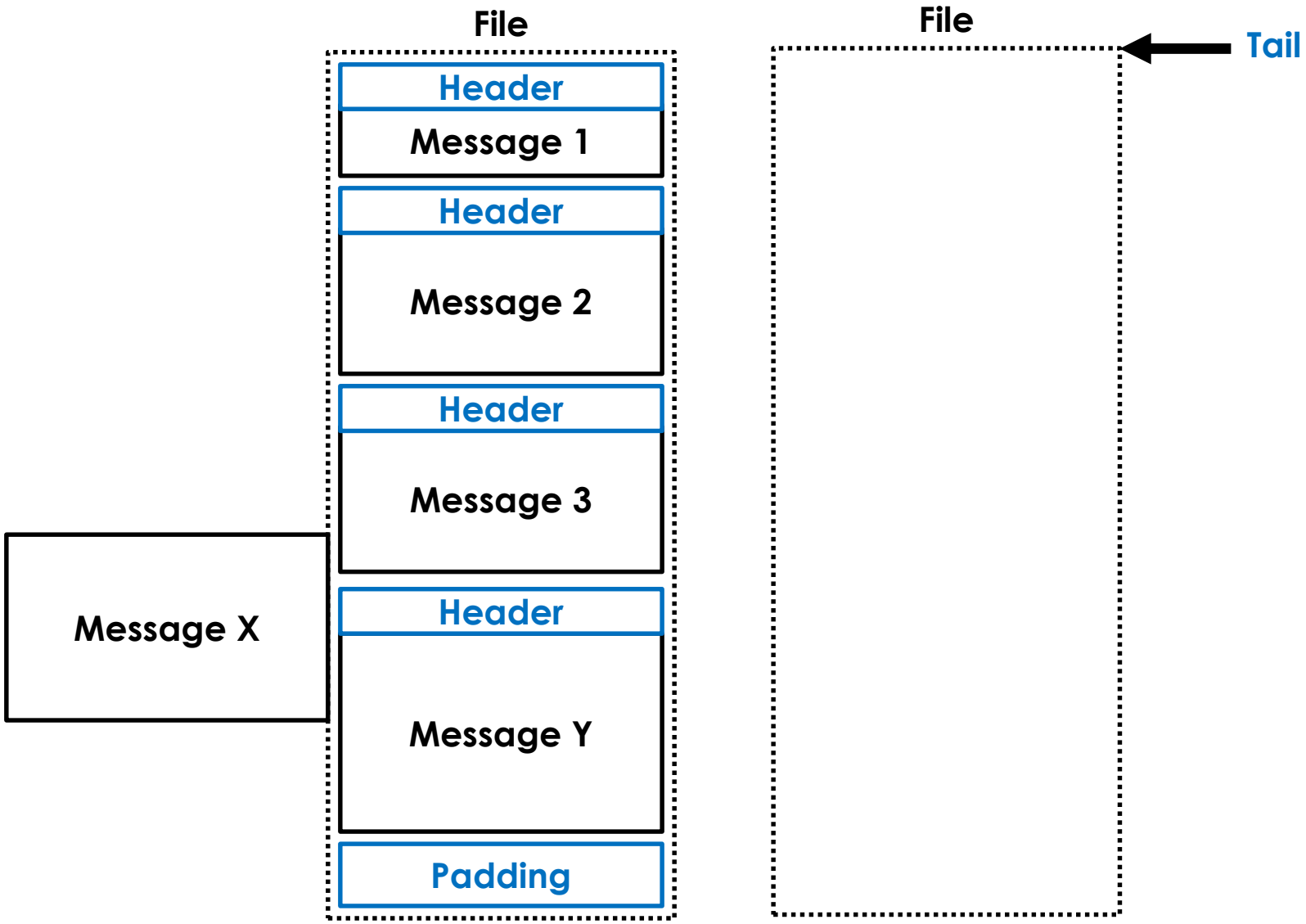

File
Header
Message 1
Header
Message 2
Header
Message 3
Tail
Message X
Message Y

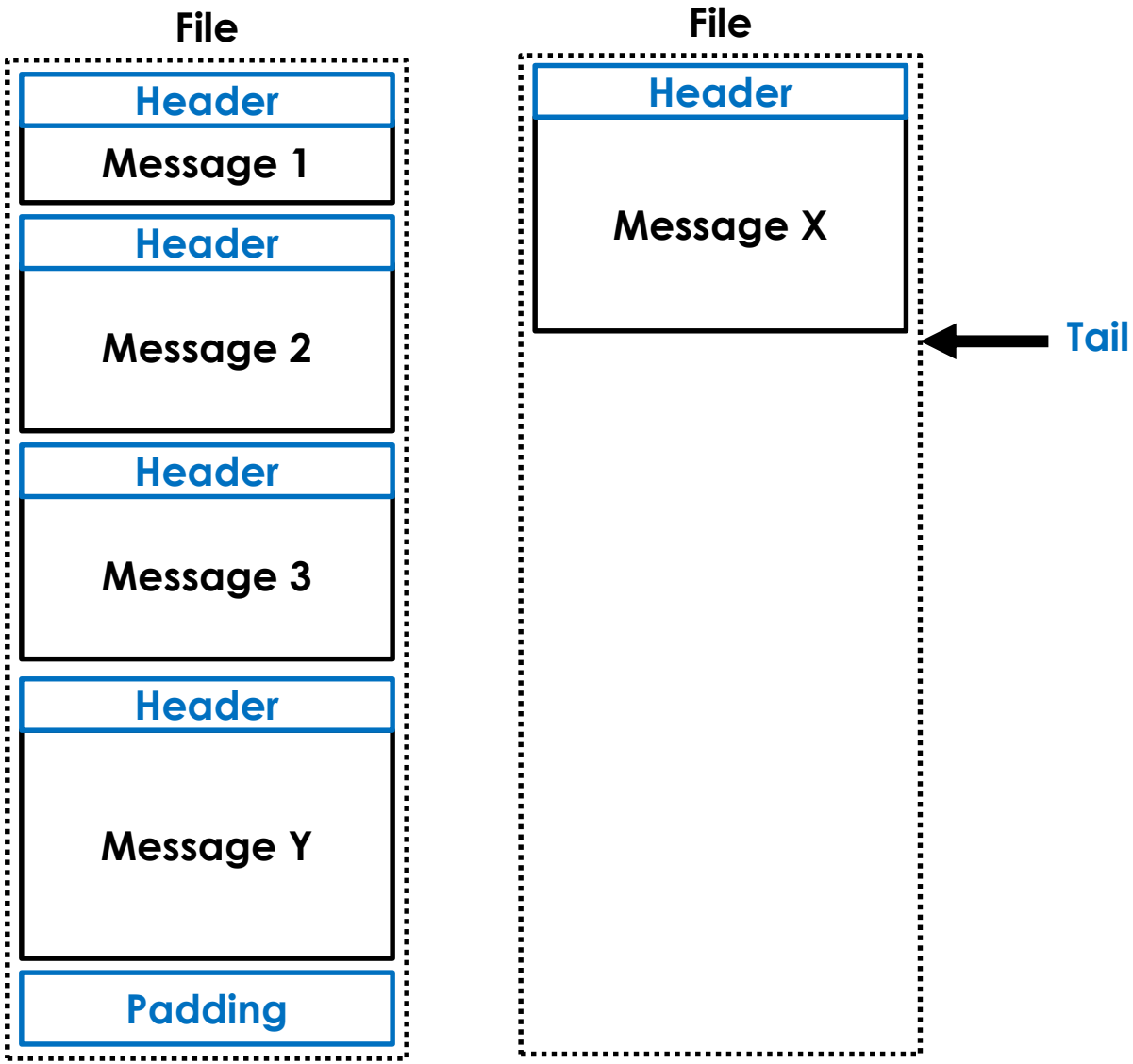

File
Header
Message 1
Header
Message 2
Header
Message 3
Message X
Message Y
Tail


File
Header
Message 1
Header
Message 2
Header
Message 3
Message X
Message Y
Tail


File
Header
Message 1
Header
Message 2
Header
Message 3
Header
Message Y
Message X
Tail

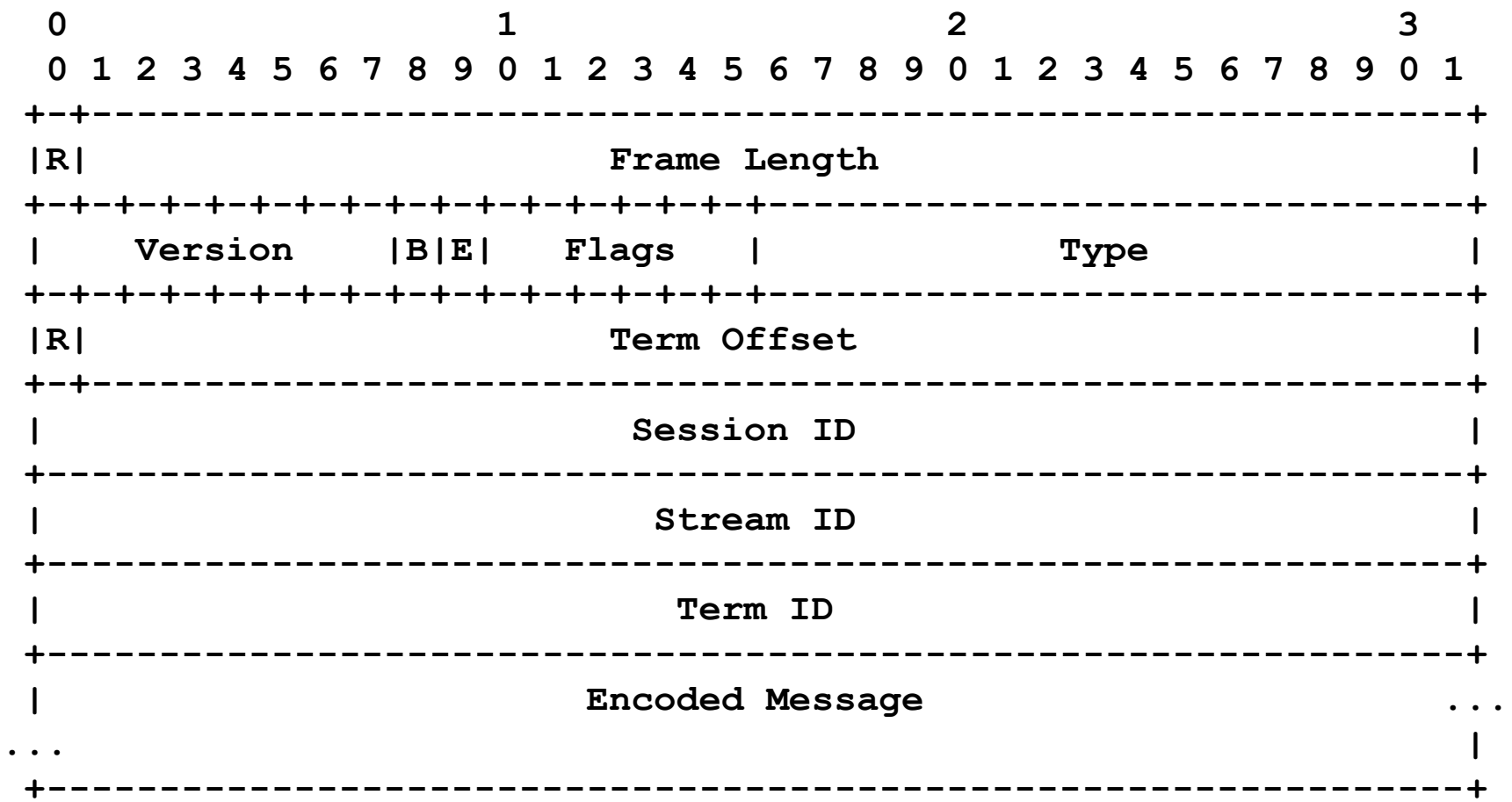

File
Header
Message 1
Header
Message 2
Header
Message 3
Header
Message Y
Padding
Message X
Tail


File
Header
Message 1
Header
Message 2
Header
Message 3
Message X
Header
Message Y
Padding
File
Tail


File
Header
Message 1
Header
Message 2
Header
Message 3
Header
Message Y
Padding
File
Header
Message X
Tail

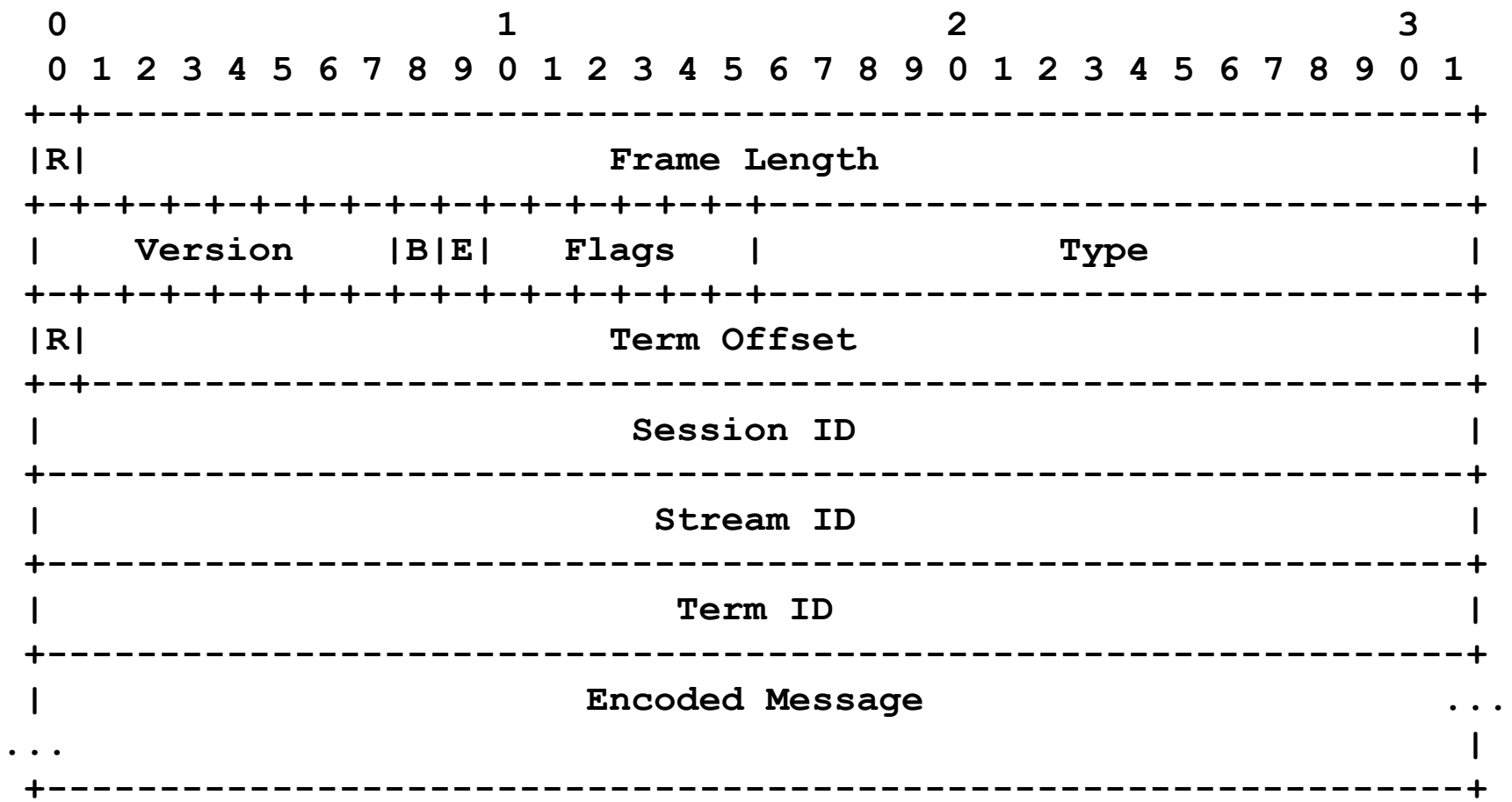
What's in a header?

Data Message Header

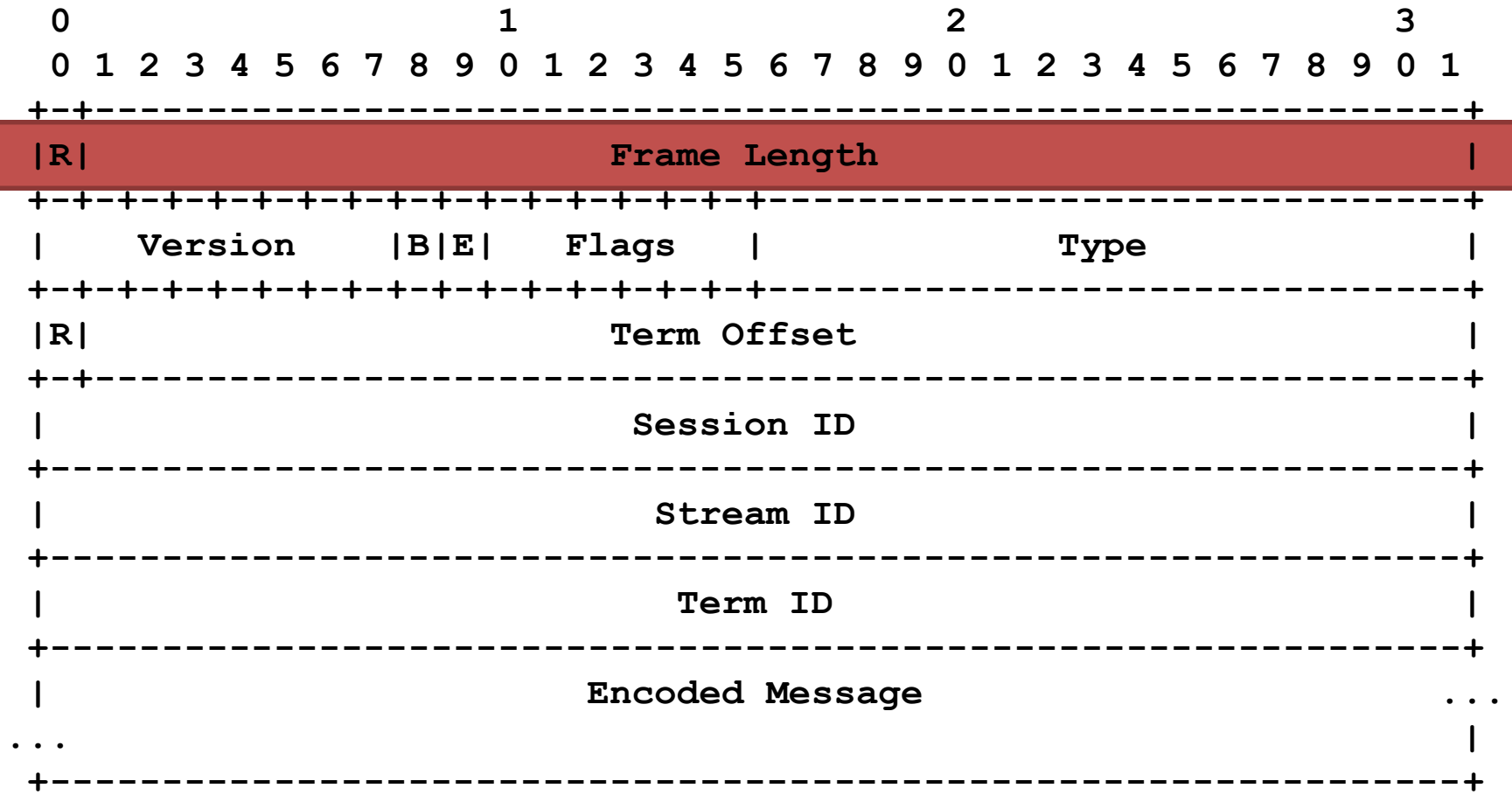


What if a Publisher dies mid operation?

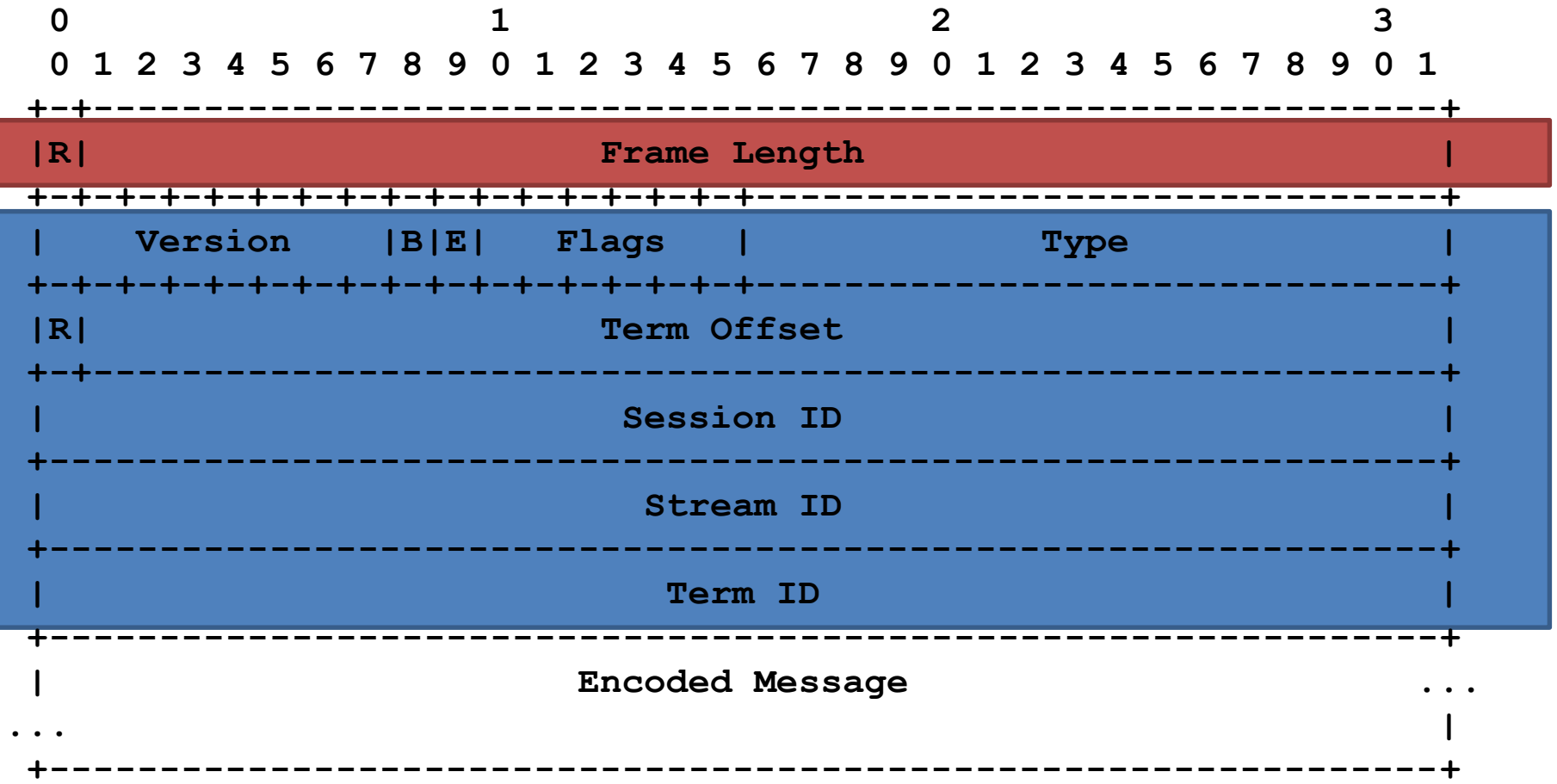
Data Message Header



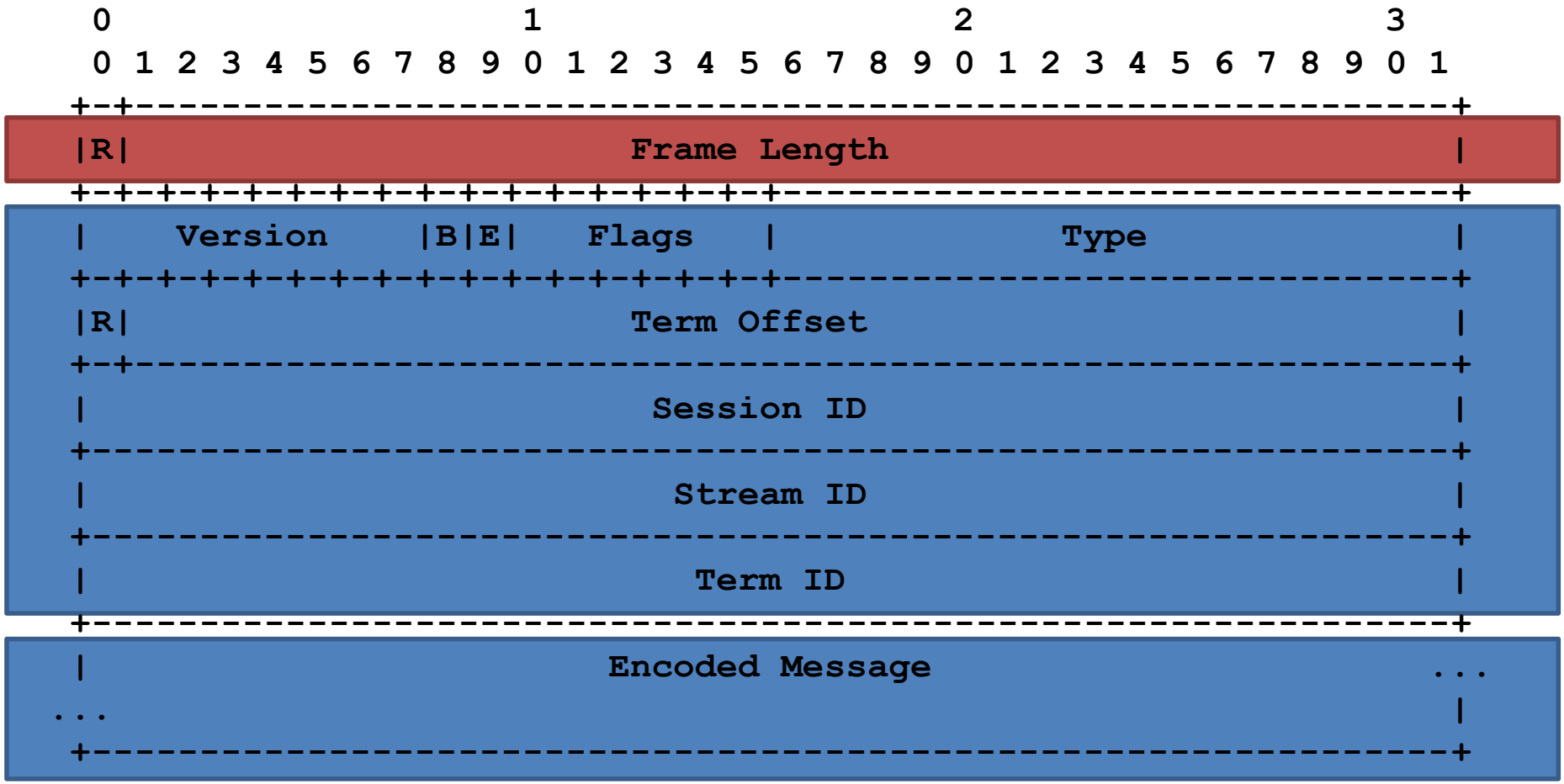
Data Message Header



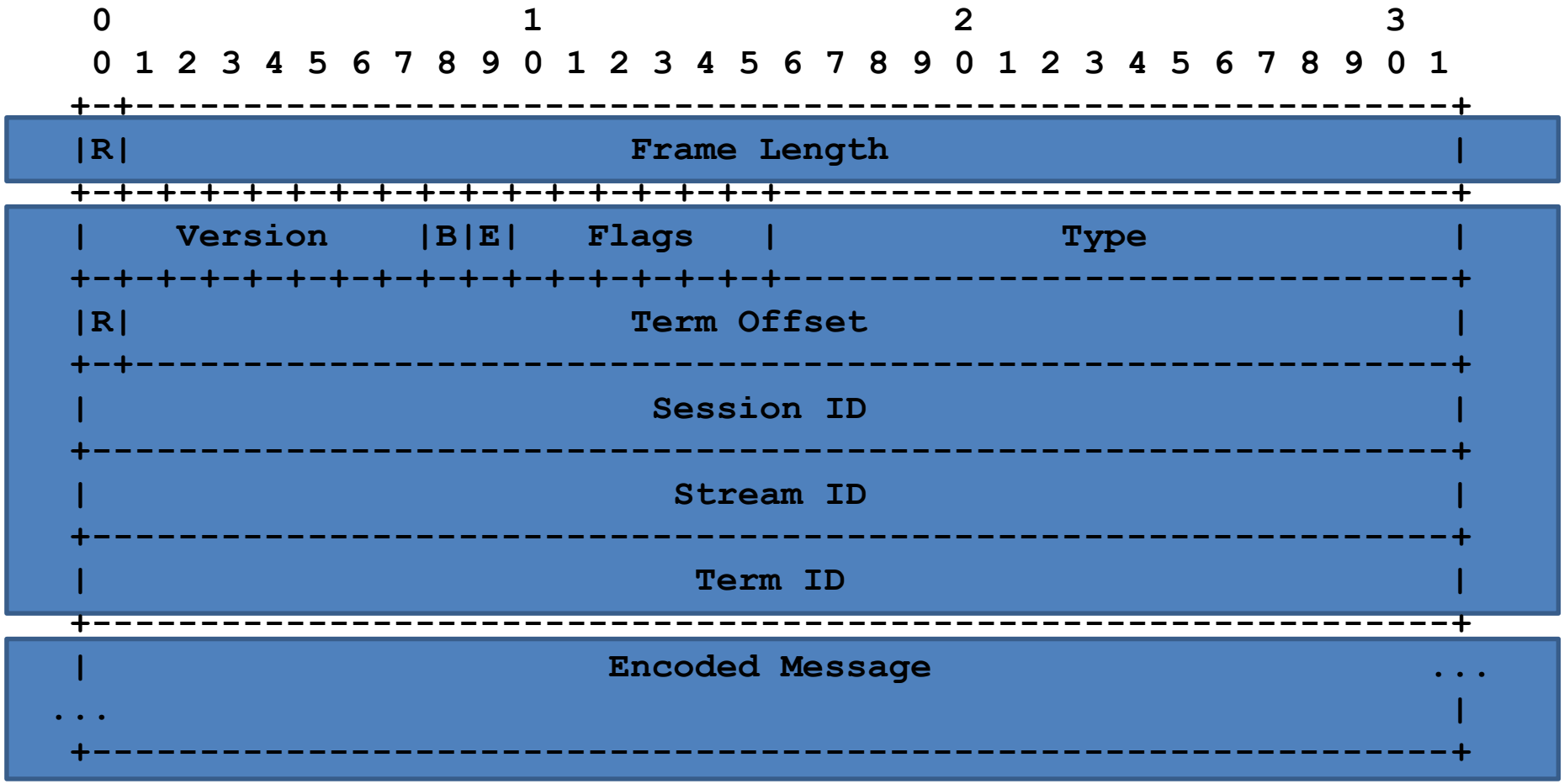
Data Message Header



Data Message Header

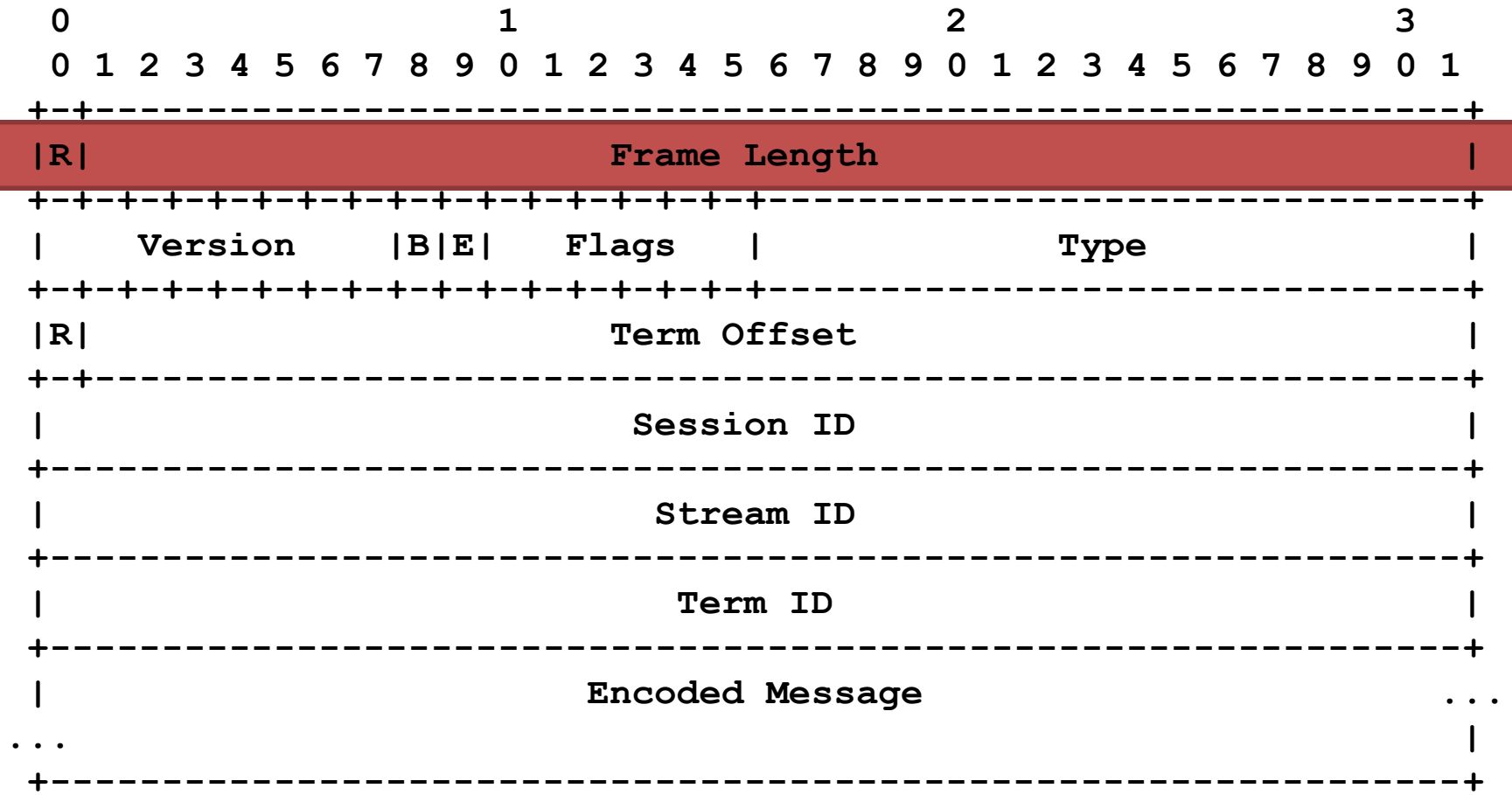


Data Message Header

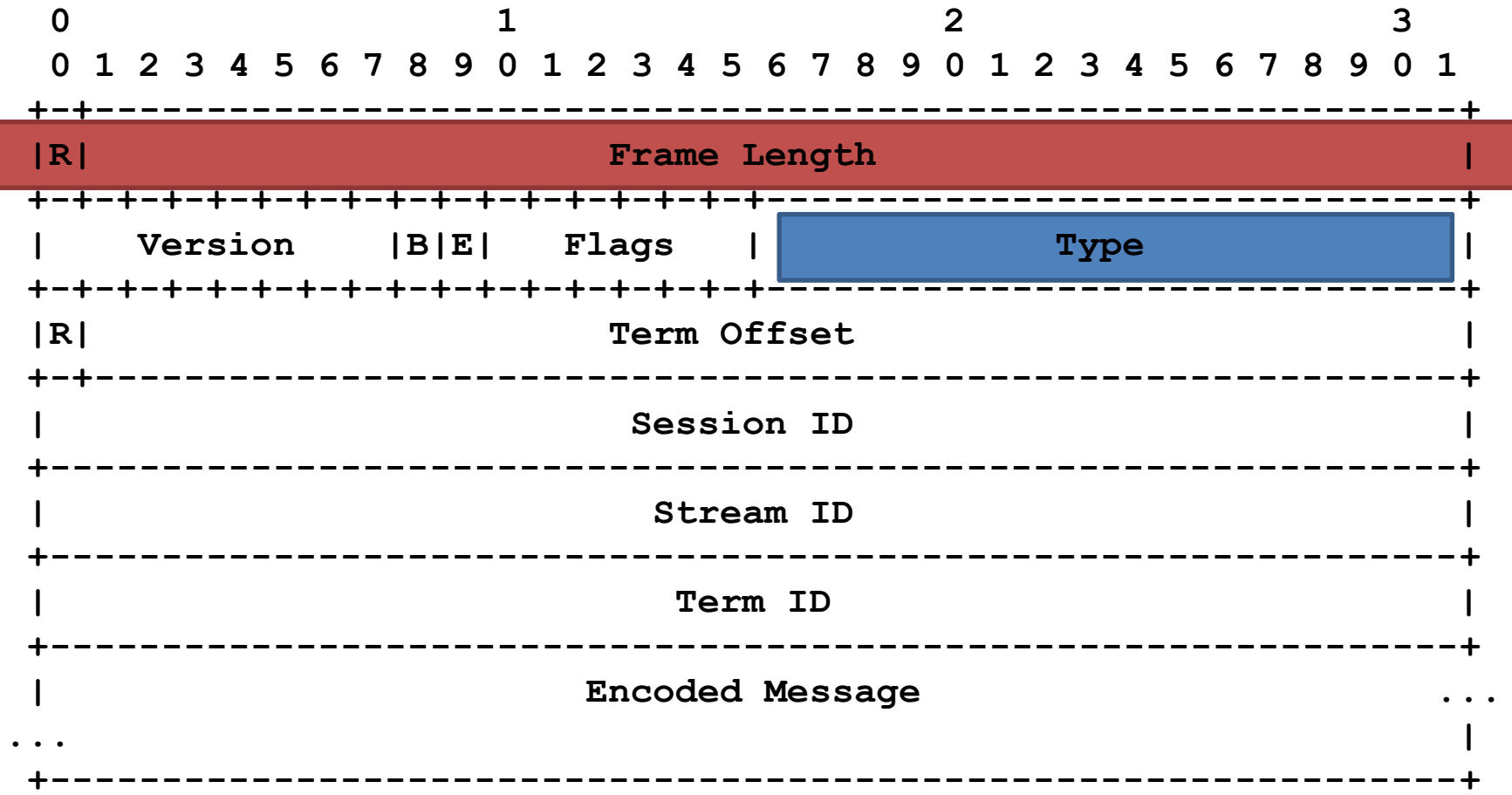


How to remove a failure?

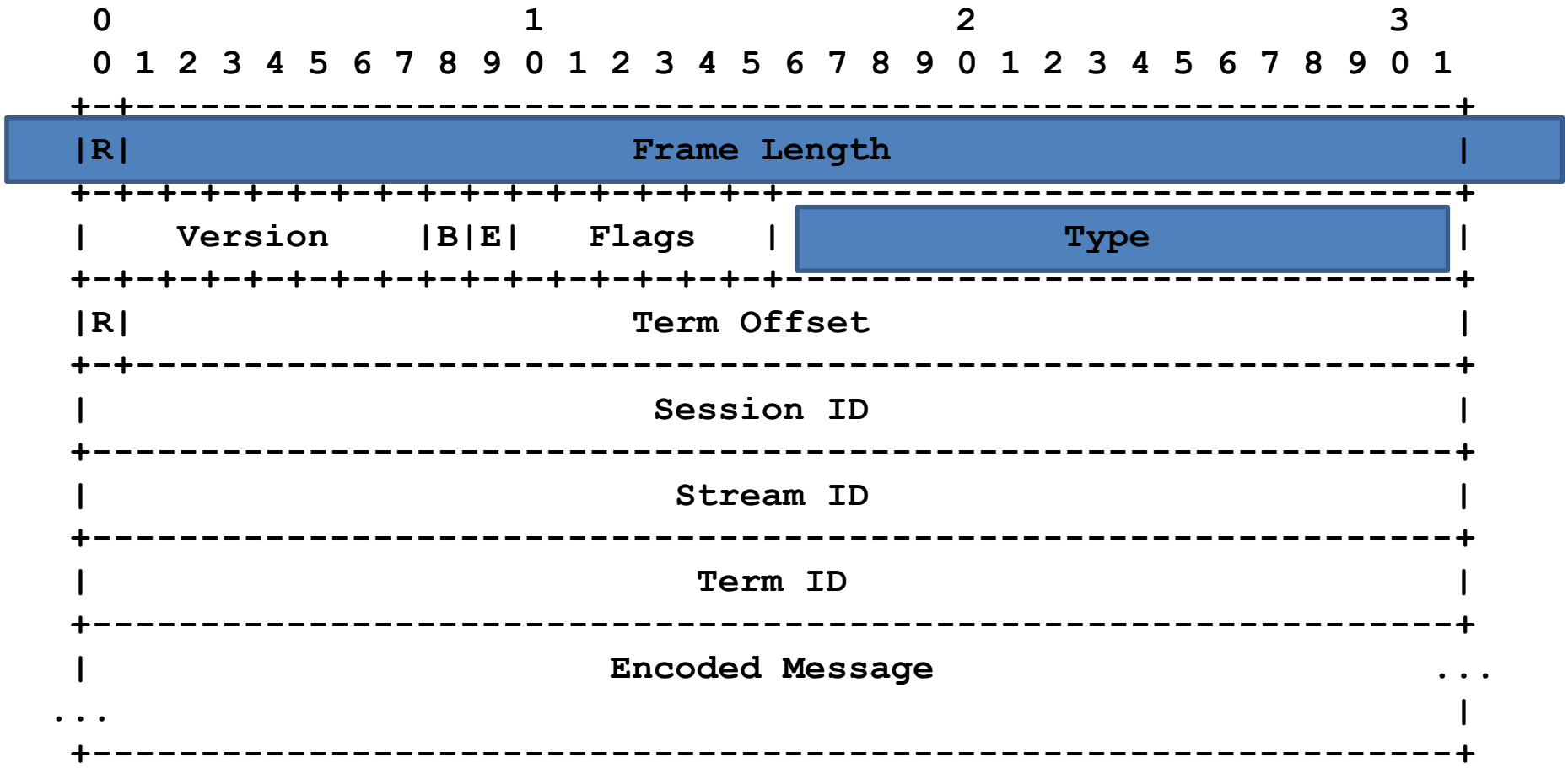
Data Message Header



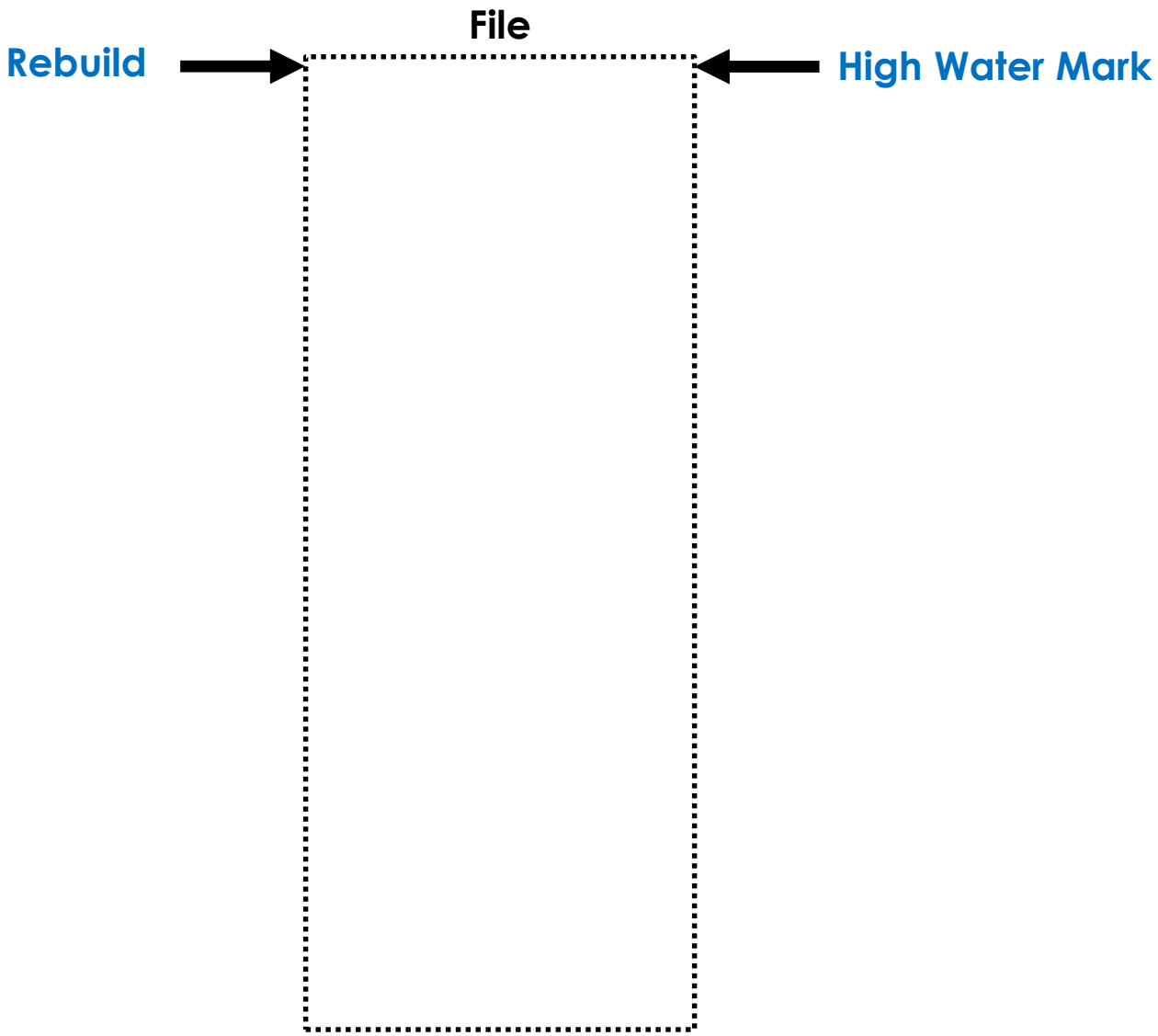
Data Message Header

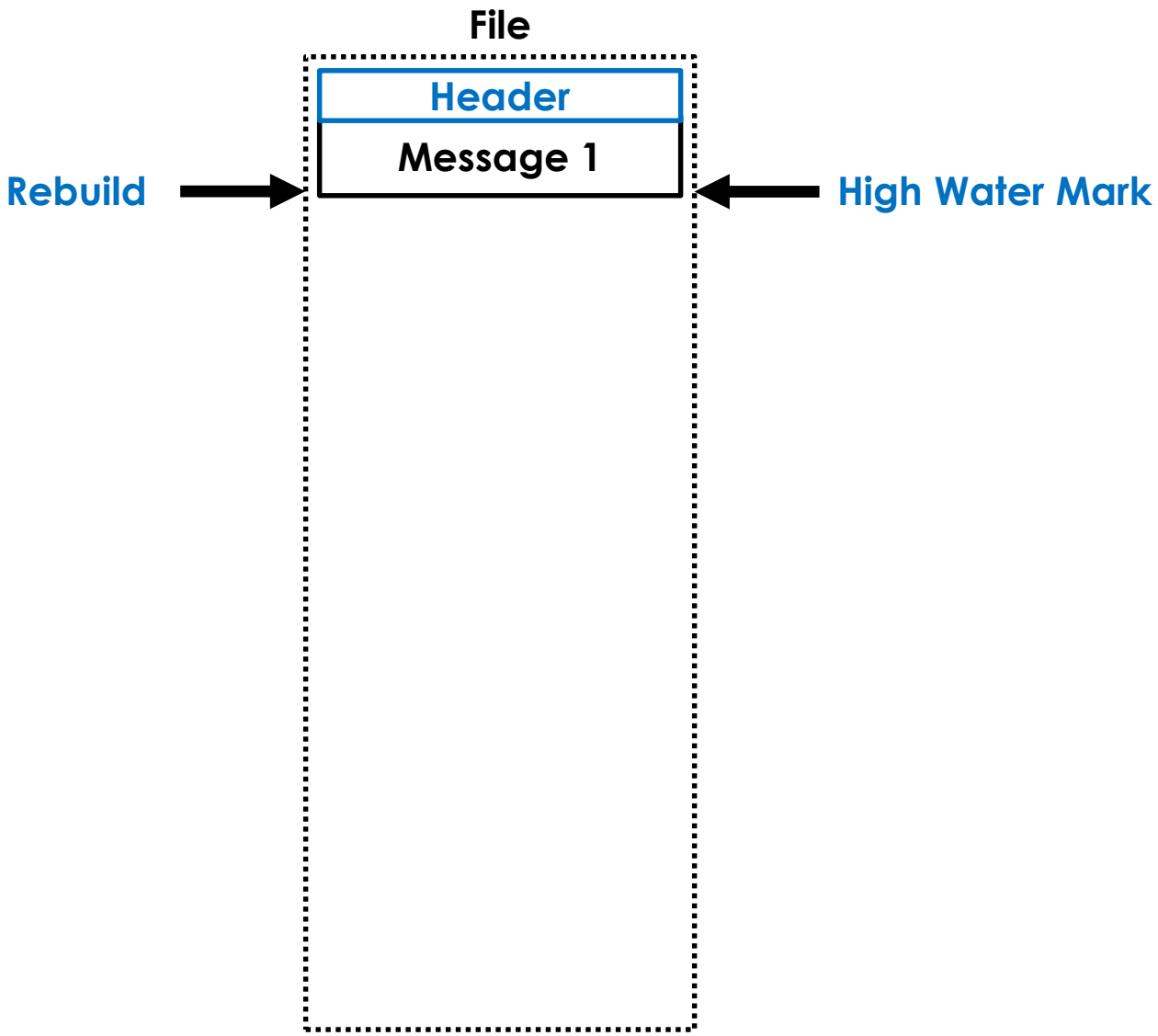


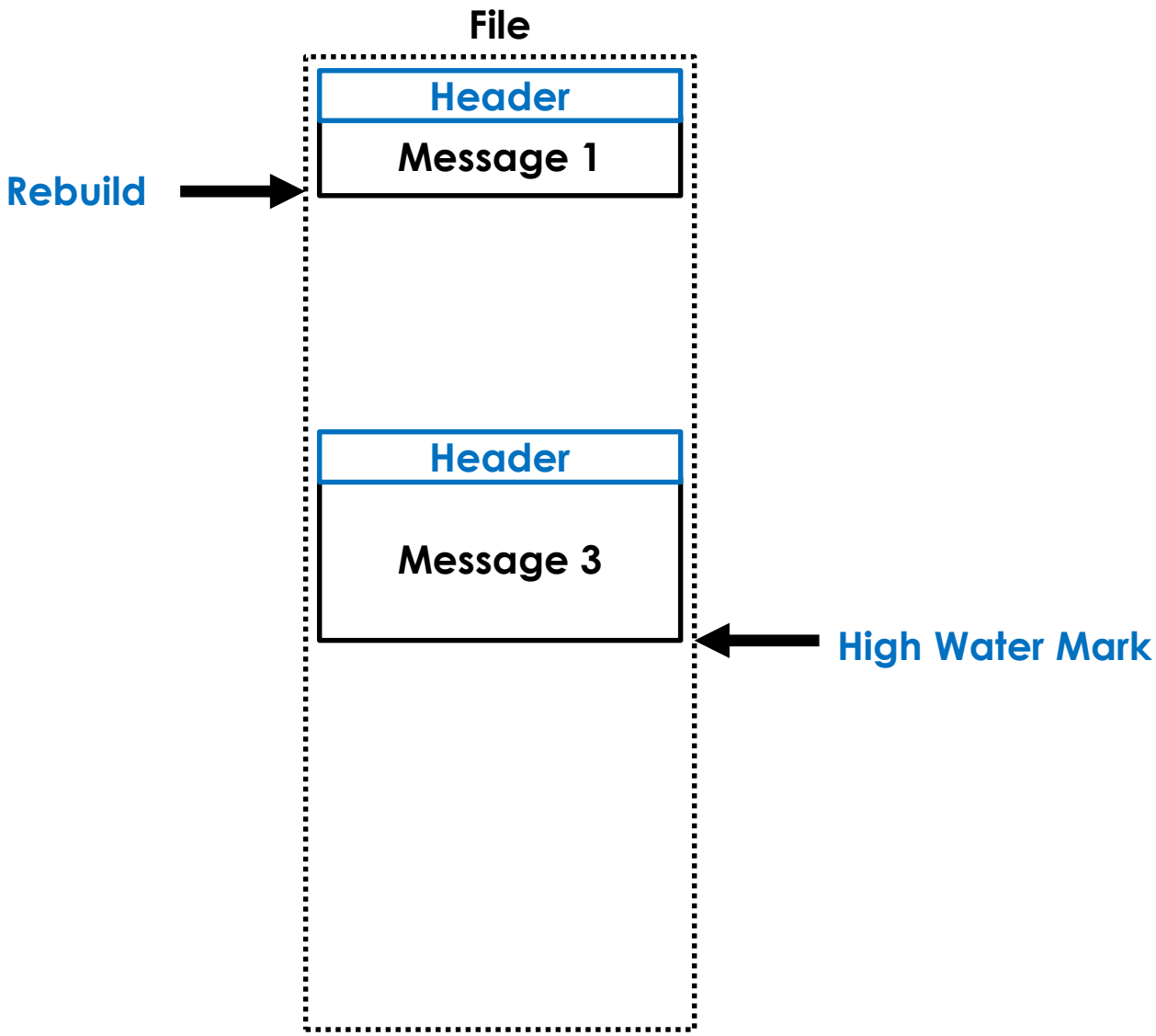
Data Message Header

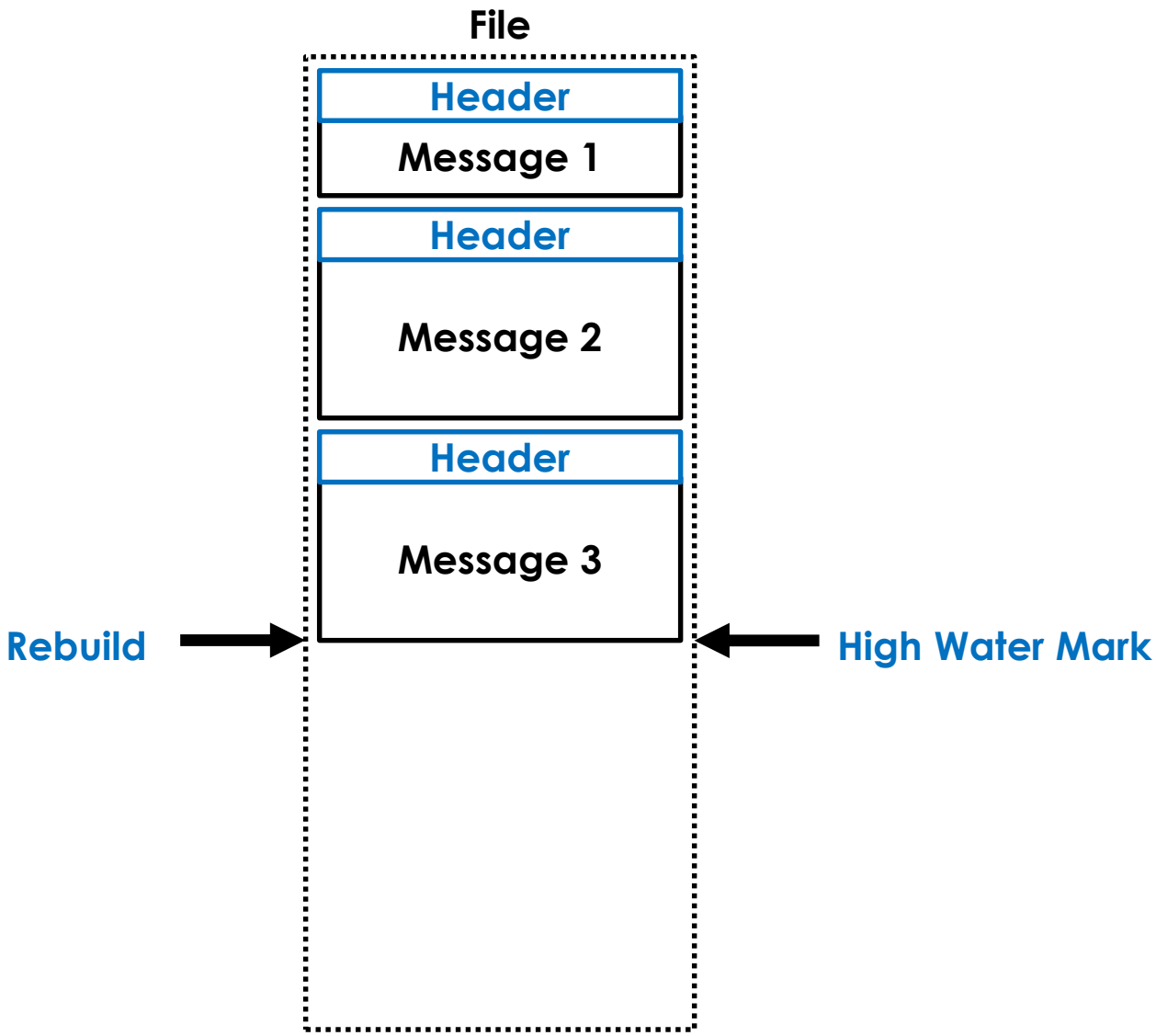


How do we replicate a log?


File
Rebuild
High Water Mark

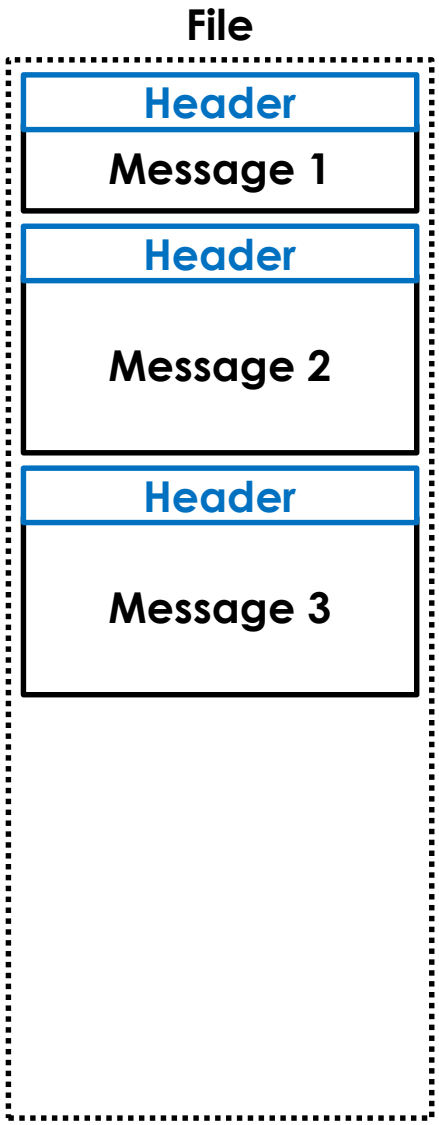

File
Header
Message 1
Rebuild
High Water Mark


File
Header
Message 1
Rebuild
Header
Message 3
High Water Mark


File
Header
Message 1
Header
Message 2
Header
Message 3
Rebuild
High Water Mark

What if a gap is not filled?

The Log Buffer has
Strong Eventual Consistency



Consider the implications of this memory layout

How do we know reason about the state of the system?

Publishers, *Senders*, *Receivers*, and *Subscribers* all keep position counters which are monotonic

In closing…

Shared Mutable State Is Evil ???

Single Writer

Shared Nothing

Build Protocols on Monotonically Increasing State

https://github.com/real-logic/aeron

Questions?

Twitter: @mjpt777

"Travel is fatal to prejudice, bigotry, and narrow-mindedness, and many of our people need it sorely on these accounts. Broad, wholesome, charitable views of men and things cannot be acquired by vegetating in one little corner of the earth all one's lifetime."

- Mark Twain