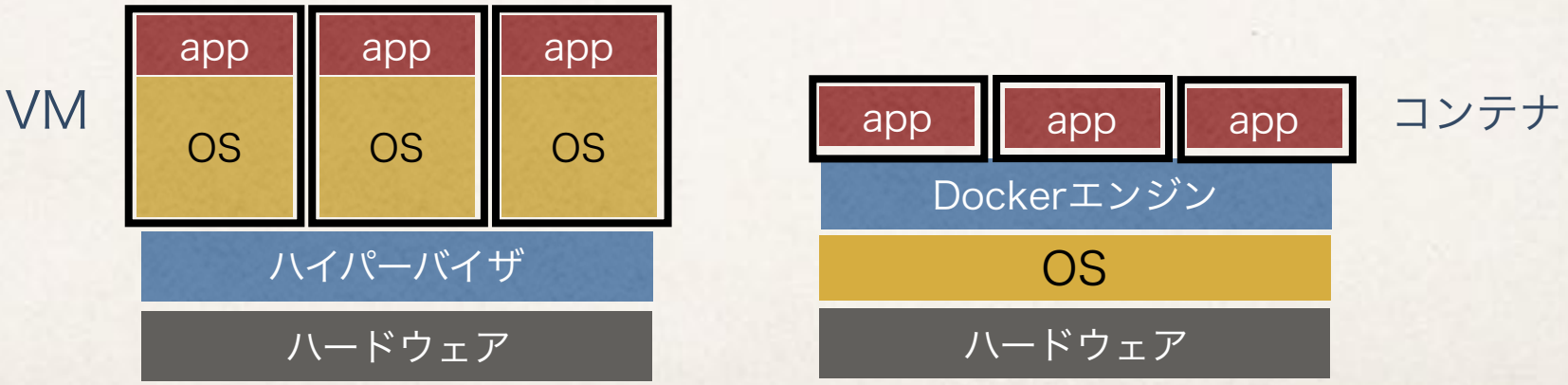
OverlayFSを用いたコンテナに対するサービス妨害攻撃の防止

九州工業大学
佐藤寛文　光来健一

コンテナ型仮想化

- アプリケーションのための仮想実行環境を提供
  - 例：Docker
  - 計算機全体を仮想化するハイパーバイザ型仮想化より軽量に動作
    - 高速に起動することができる
    - 少ないリソースしか必要としない

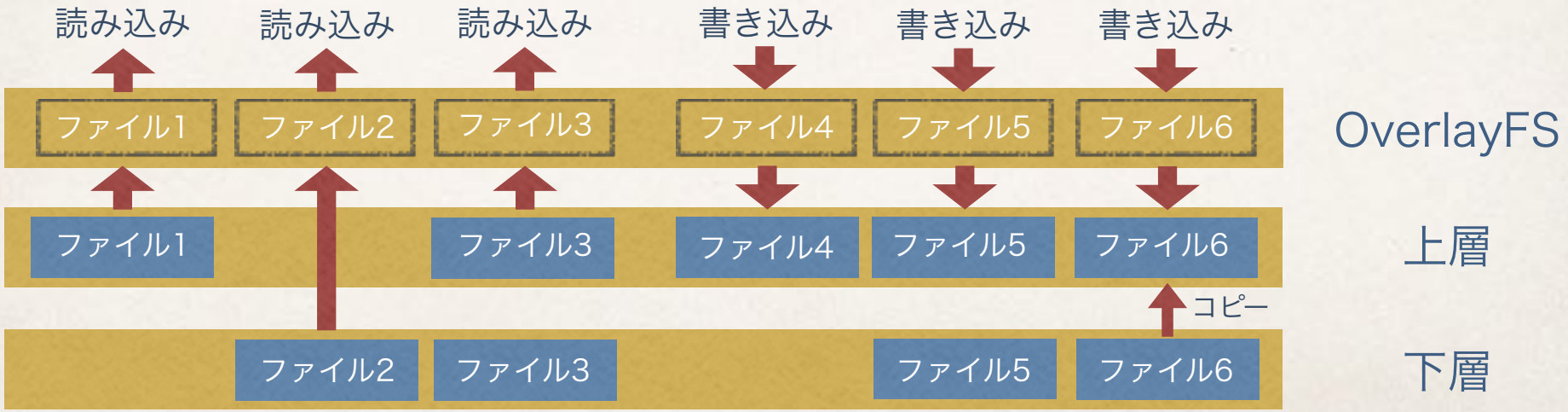


コンテナのディスクイメージ

- コンテナが用いるファイルを格納した仮想ディスク
  - ベースイメージを複数のコンテナで共有
  - コンテナごとに差分イメージを用意
    - 個別のアプリケーション、ログなどを格納
- OverlayFSを用いて重ね合わせる
  - 下層にベースイメージ、上層に差分イメージ

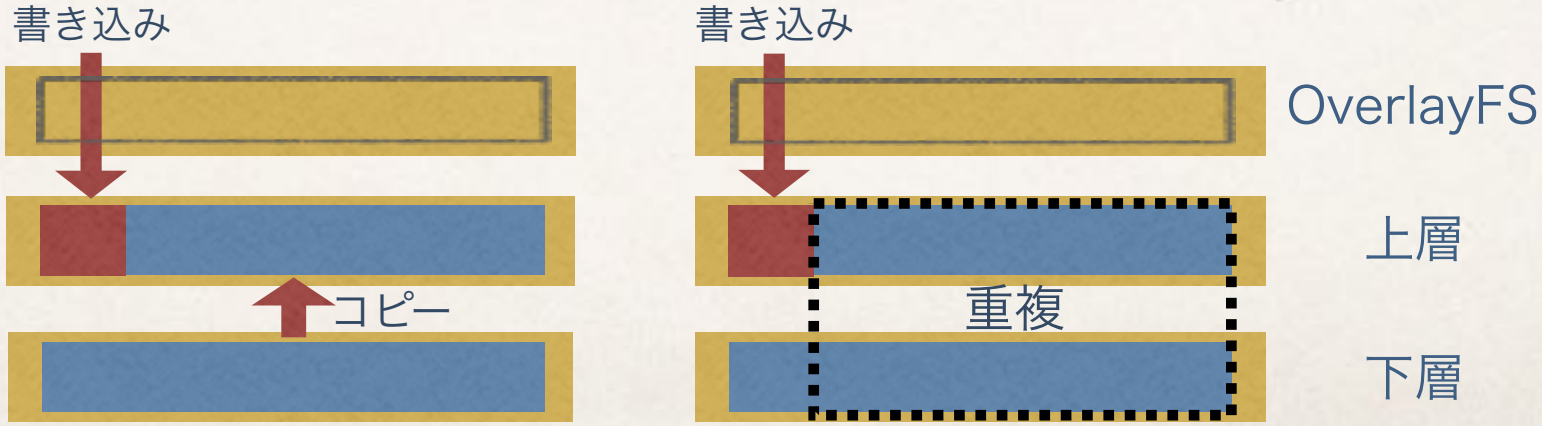
OverlayFSの読み書き処理

- ファイルの読み込み
  - 上層にあれば上層から、なければ下層から読み込む
- ファイルへの書き込み
  - 上層にあれば上層に書き込む
  - 上層になければコピーオンライトを行う
    - 下層のファイルを上層にコピーし、上層に書き込み



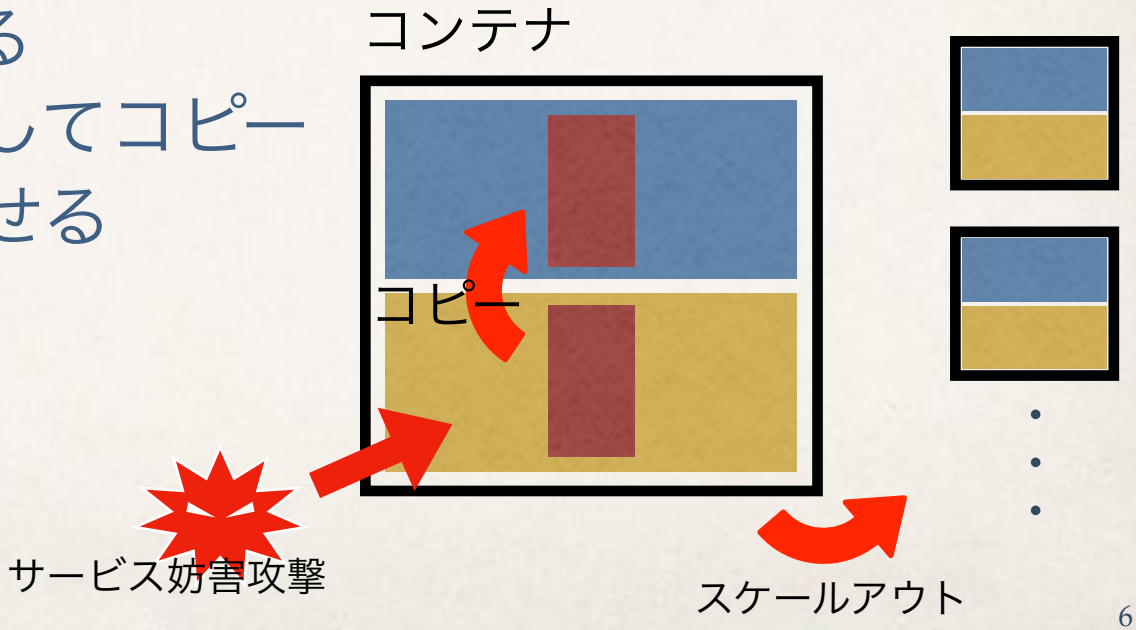
OverlayFSの問題

- 下層のファイルを最初に書き換える際にコピーオンライトのオーバヘッドが大きい
  - ファイル全体が一括でコピーされる
  - コピーが完了するまでコンテナが停止
    - コピーはシステム全体の性能にも影響を与える
  - ディスク容量を圧迫
    - 下層と上層でほとんど同じファイルを持つことになる



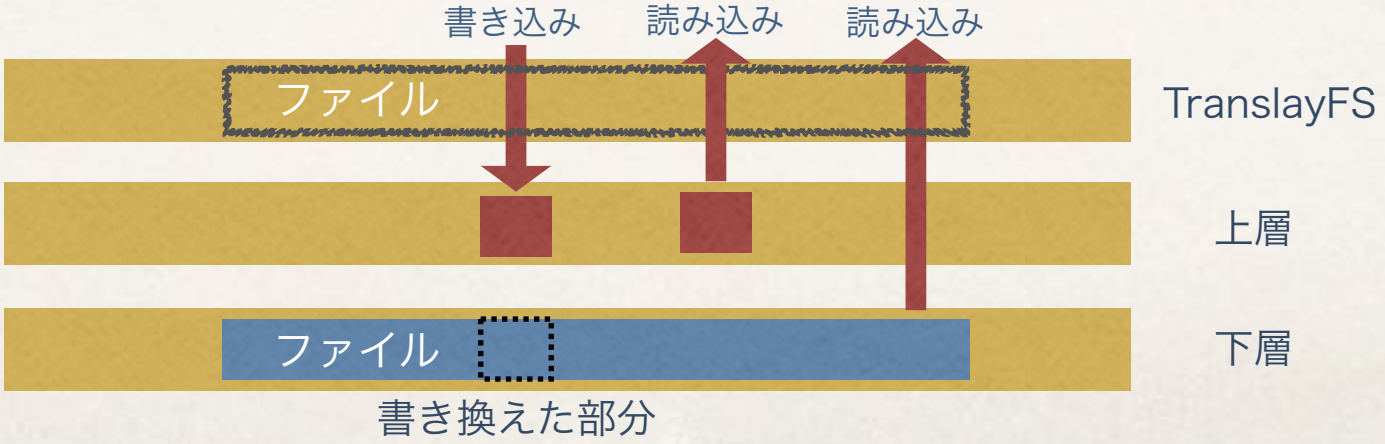
コンテナに対するサービス妨害攻撃

- コピーオンライトを意図的に発生させることでコンテナを一時的に停止させることができる
  - ファイルサイズに比例した時間だけ停止
- コンテナがオートスケールする場合に被害が拡大
  - 大量のリクエストを送信してスケールアウトさせる
  - 新しいコンテナに対してコピーオンライトを発生させる
  - これを繰り返す



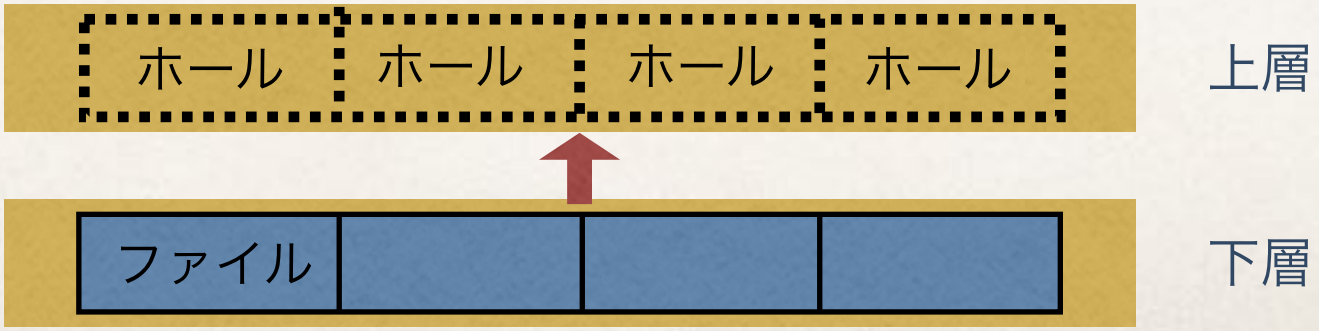
提案：TranslayFS

- OverlayFSを改良し、コピーオンライトのオーバヘッドを削減したファイルシステム
  - 上層には書き換えたデータのみを保持
    - それ以外のデータは下層から読み込む
  - ディスク容量を節約
  - コンテナに対するサービス妨害攻撃を防止
    - コンテナを長時間停止させることはできなくなる



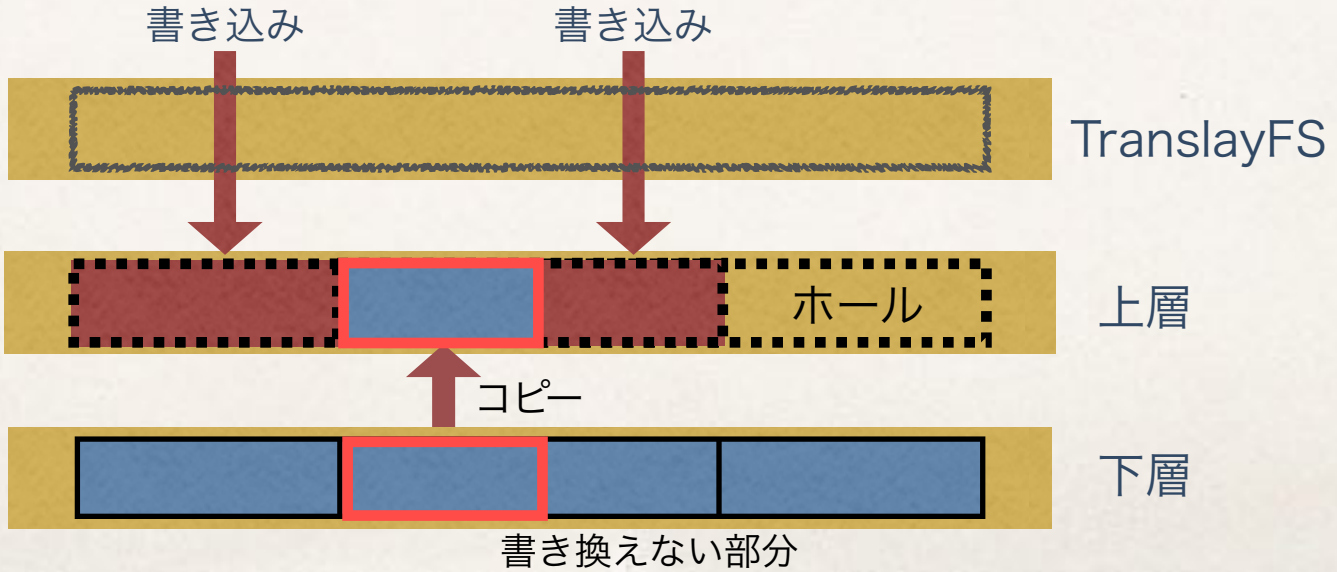
スパースファイルの活用

- ファイルの変更部分を管理するのは煩雑
  - データベースを用いると性能低下の恐れ
- スパースファイルを作成して効率よく管理
  - スパースファイルとは？
    - 実際のデータを持たないファイル
    - ホールと呼ばれる空のブロックから成る
    - ディスク容量をほとんど消費しない
  - ファイルに対して初めて書き込みが行われた時に作成



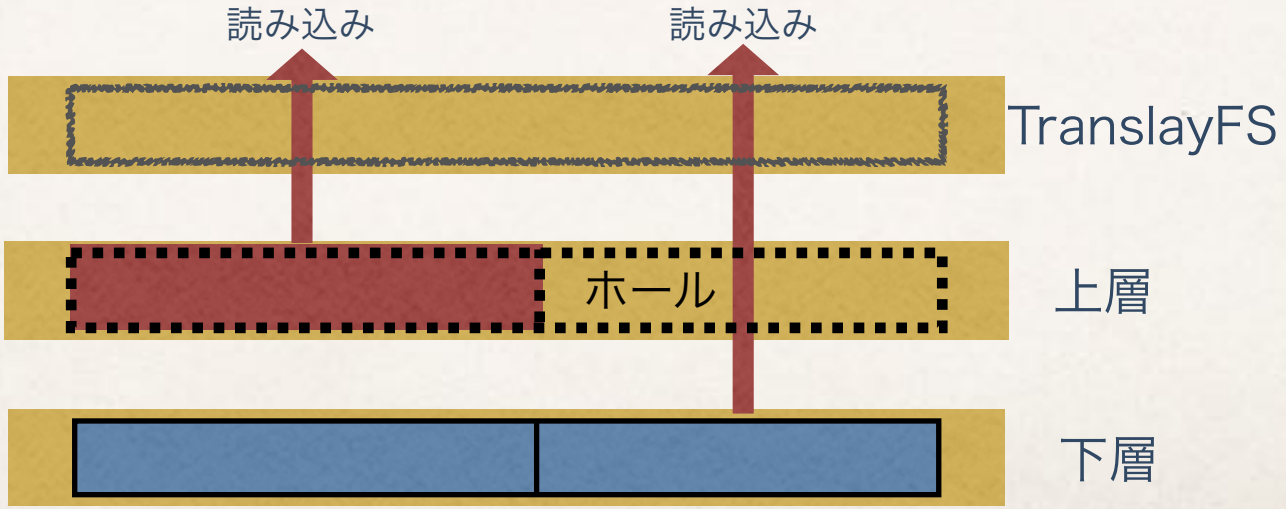
TranslayFSの書き込み処理

- 上層のスパースファイルにブロック単位で書き込む
  - 書き込んだブロックだけが実際のデータを含む
    - その分だけディスク容量を消費
  - 下層のブロックの一部だけを書き換える場合は、書き換えない部分だけを下層からコピー
    - コピーを最小限に抑えることができる



TranslayFSの読み込み処理

- 上層のスパースファイルまたは下層のファイルから読み込む
  - 上層の対象ブロックがデータを含んでいれば上層から
  - 上層の対象ブロックがホールなら下層から
  - データはブロック単位で書き込まれているため、どちらかのブロックだけ読み込めばよい

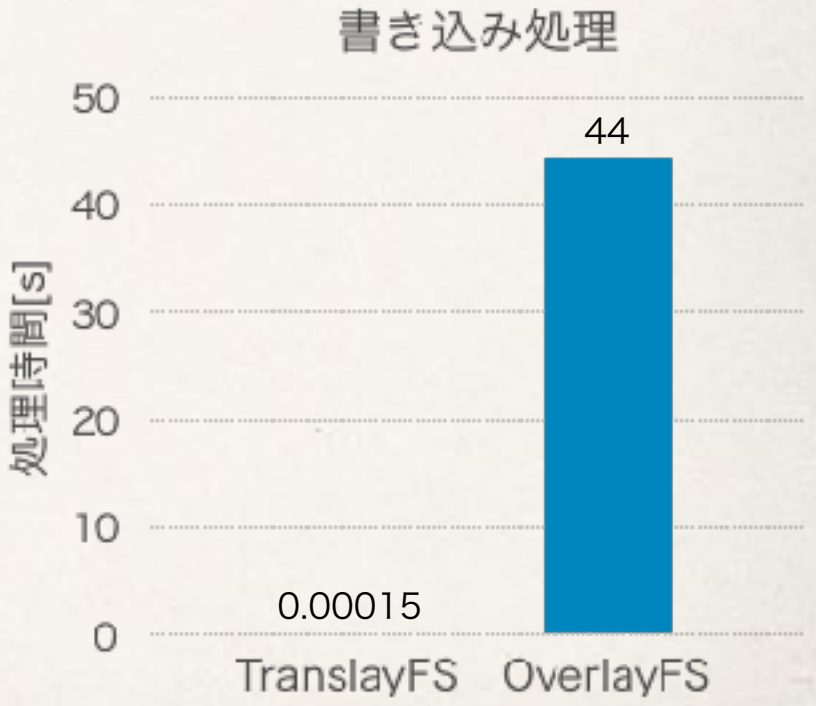


実験

- 目的
  - OverlayFSとの読み書き性能の比較
  - 他のストレージドライバとの読み書き性能の比較
    - AUFS、ZFS、devicemapper
- 実験環境
  - CPU：Intel Core i7-3770 CPU @3.40GHz
  - メモリ：8GB
  - ハードディスク：SATA3 HDD
  - OS：Linux 4.4.0

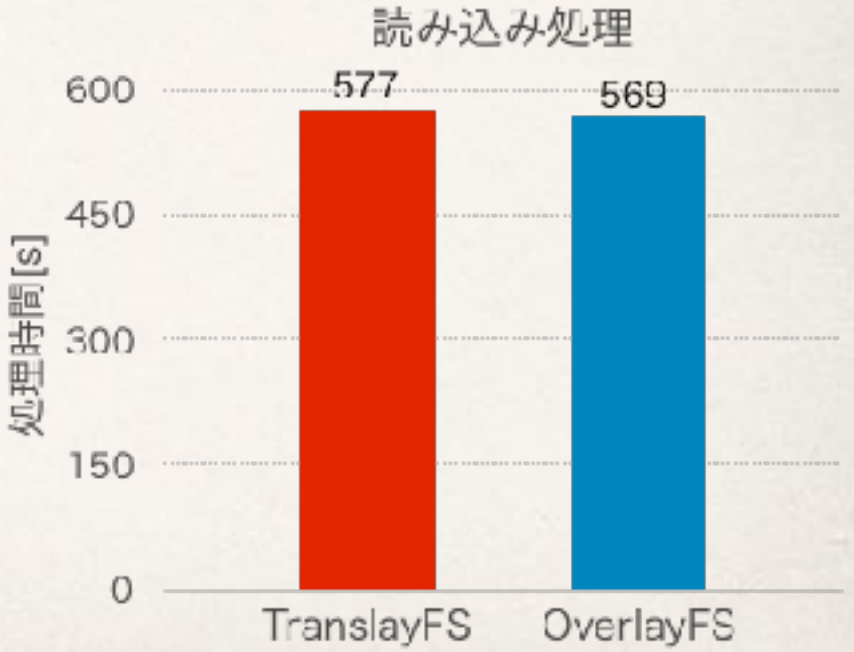
書き込み性能の比較

- 10GBの巨大なファイルに1バイトの書き込みを行う時間を測定
  - 10GBのファイルを下層に作成
- 実験結果
  - OverlayFSでは44秒かかった
    - 10GBのファイルがコピーされた
  - TranslayFSでは0.15ミリ秒で完了
    - スパースファイルが作成され、4KBだけ書き込まれた



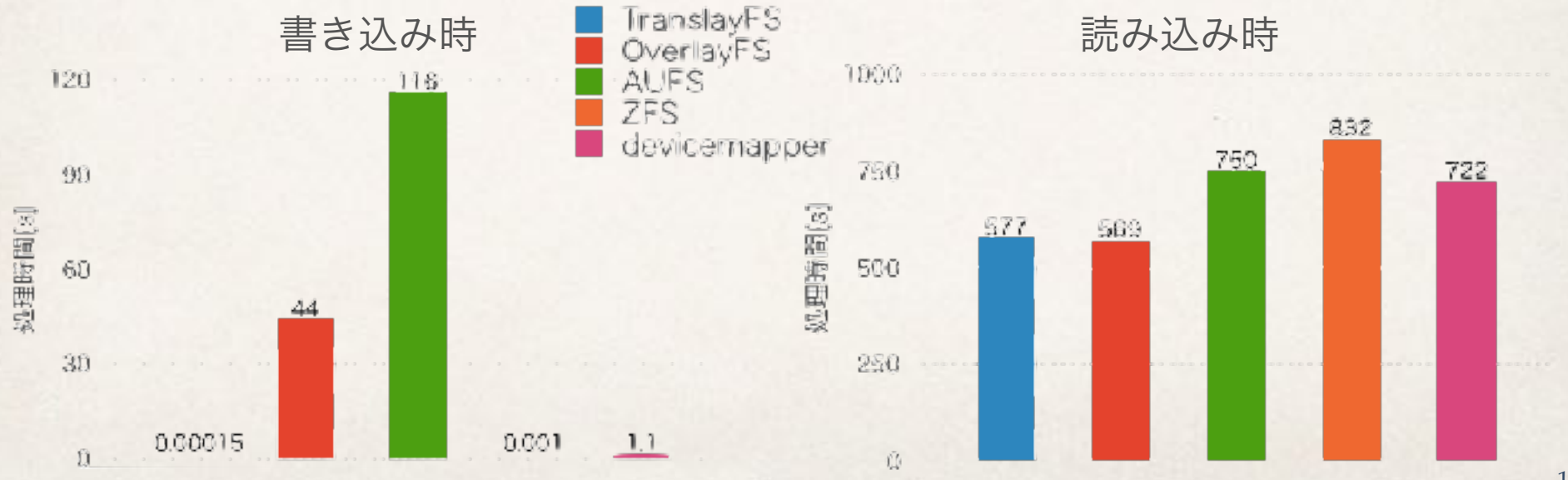
読み込み性能の比較

- 1バイトの書き込みを行った10GBのファイル全体を読み込む時間を測定
- 実験結果
  - TranslayFSは1.3%の性能低下
    - ブロック単位で処理するオーバヘッドなど
    - ホールを検出するオーバヘッド
  - 書き込み性能の向上を考えれば許容範囲内



様々なストレージドライバとの比較

- 他のストレージドライバについて同様の読み書きを行う時間を測定
  - AUFSは書き込み時間がOverlayFSよりも長い
  - ZFSは書き込み時間は短いが、読み込み時間が長い
  - devicemapperは読み書き時間が少しずつ長い



関連研究

- Unionファイルシステム（OverlayFS等）
  - 複数のファイルシステムを重ねるファイルシステム
  - ファイル単位でコピーオンライトを行うオーバヘッドが大きい
- ZFS［LLNL］、Btrfs［Oracle］
  - 差分管理などを提供する高機能なファイルシステム
    - ブロック単位でコピーオンライトを行う
  - 性能や安定性に問題がある
- devicemapper［Red Hat］
  - 差分管理を提供するブロックデバイス
  - ファイルシステムより扱いにくい

まとめ

- OverlayFSを改良し、コピーオンライトのオーバーヘッドを削減したTranslayFSを提案
  - 上層には書き換えたデータのみを保持
  - スパースファイルを用いることで効率よく実現
  - コンテナに対するサービス妨害攻撃を防止
- ファイルに対する最初の書き込み性能が大幅に向上
  - 読み込みでは1.3%の性能低下が見られた
  - 他のどのストレージドライバよりも高速
- 今後の課題
  - ブロックの一部だけに書き込まれた際の処理の実装
  - ファイル読み書き以外の性能の測定