

コールドストレージGeo-Replicationシステムにおける プロキシサーバの構築および性能評価

研究背景

- データセンタのデータ管理コストが年々増加
- アクセス頻度の低い“コールドデータ”がストレージ利用効率悪化を招く
- 階層化ストレージを用い、データをアクセス頻度に応じたストレージに保存するで、ストレージコストを削減できる
- データ損失を防ぐため、遠隔地にデータの複製を保存する“Geo-Replication”が行われる

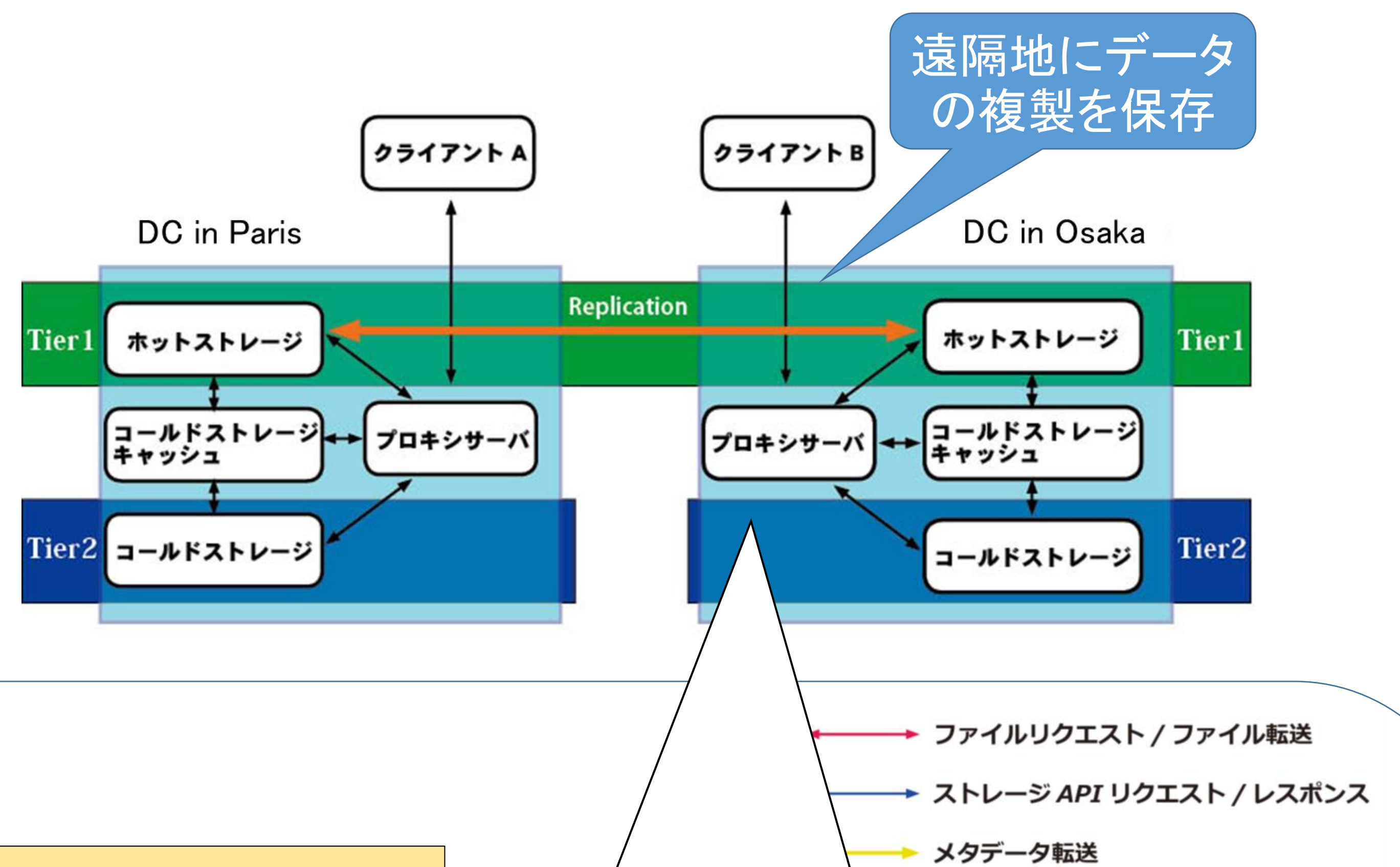
研究目的

- 階層化ストレージとGeo-Replicationを組み合わせたシステム(コールドストレージGeo-Replicationシステム)を用いて、データのアクセス頻度に応じて適切なストレージに保存し、データ管理によって生じるコストを削減する。また、データ損失のリスクを減らし、データの可用性を高める

システムアーキテクチャ

ホットストレージ	アクセス頻度の高いデータを保存する。アップロードされたデータは一度ホットストレージへ保存され、Geo-Replicationによって複製される
コールドストレージキャッシュ	コールドストレージへのデータ書き込みの際、データをキャッシュする
コールドストレージ	アクセス頻度の低いデータを保存するストレージ。ビット単価が低く消費電力が少ない

コールドストレージGeo-Replicationシステムアーキテクチャ

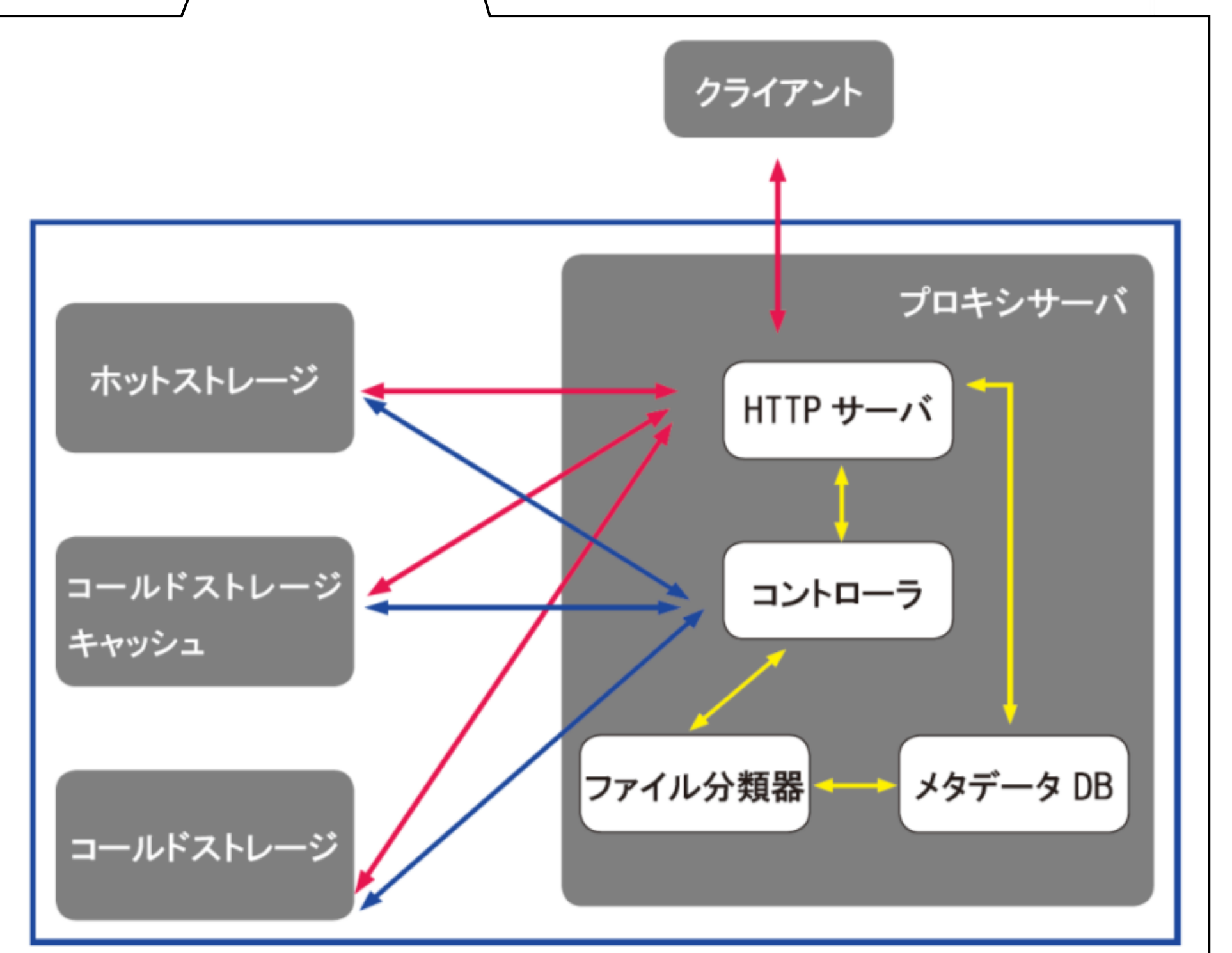


プロキシサーバ

プロキシサーバの役割

- クライアントとストレージ間のデータ転送
- ホットデータとコールドデータにデータを分類
- データを適切なストレージに保存するためのデータ転送命令の発行

- HTTPサーバ:** クライアントとストレージ間の通信を仲介するリバースプロキシの役割を果たす
- コントローラ:** データを適切なストレージに保存するために、データ転送命令を発行する
- ファイル分類器:** データをホットデータとコールドデータに分類する
- メタデータDB:** ファイルを分類する際に参照するメタデータ用データベース



プロキシサーバのシステムアーキテクチャ

プロキシサーバの性能評価

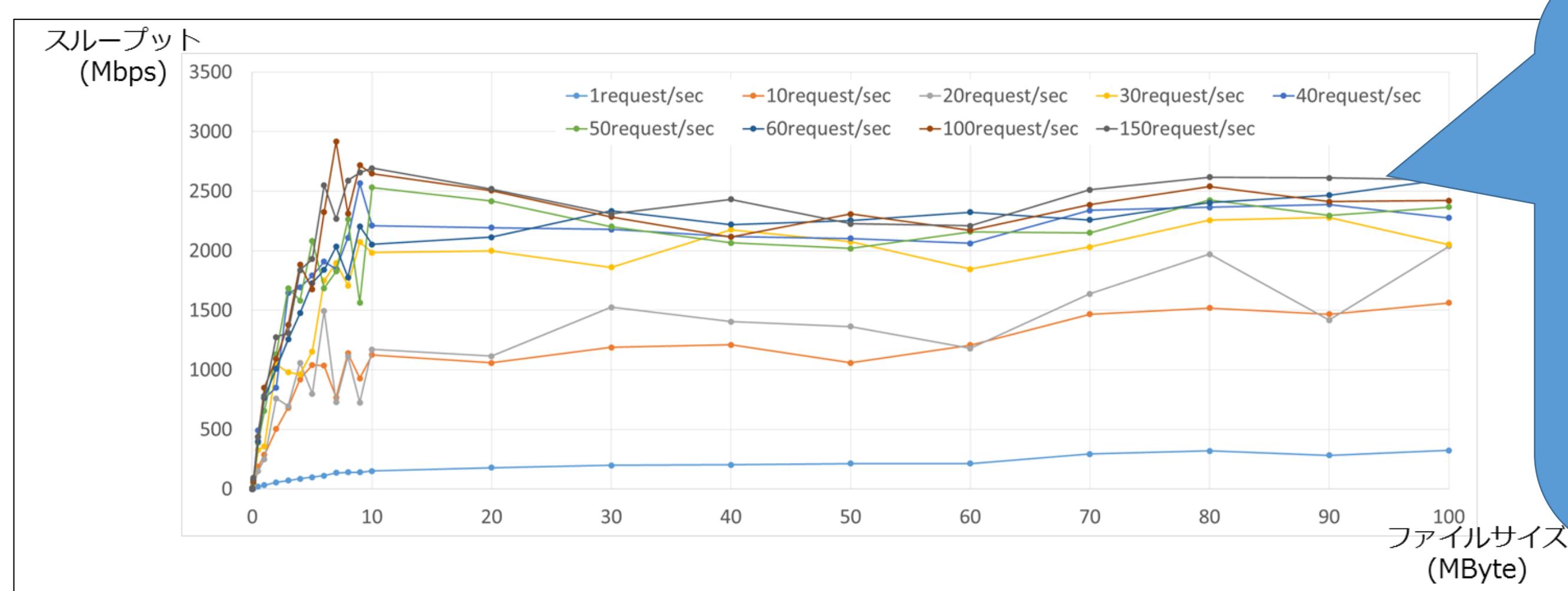
- 評価方法**
評価パラメータを変更しながらクライアントからホットストレージへデータをアップロード
- 評価パラメータ**
 - リクエストレート n (requests/sec)
 - ファイルサイズ f (MBytes)
- 評価指標**

スループット (Mbps)

$$T = \frac{nf}{\max_{1 \leq i \leq n} p_i}$$

(p_i はデータ i のアップロードにかかった時間)

リクエスト頻度/ファイルサイズに対するスループット



スループットがほぼ一定になっている。ホットストレージのI/O性能もしくはプロキシサーバがボトルネックになっていると考えられる

今後の課題

- 性能評価のボトルネックの原因説明
- 実運用を想定した様々な負荷テストの実行