

15-NA00

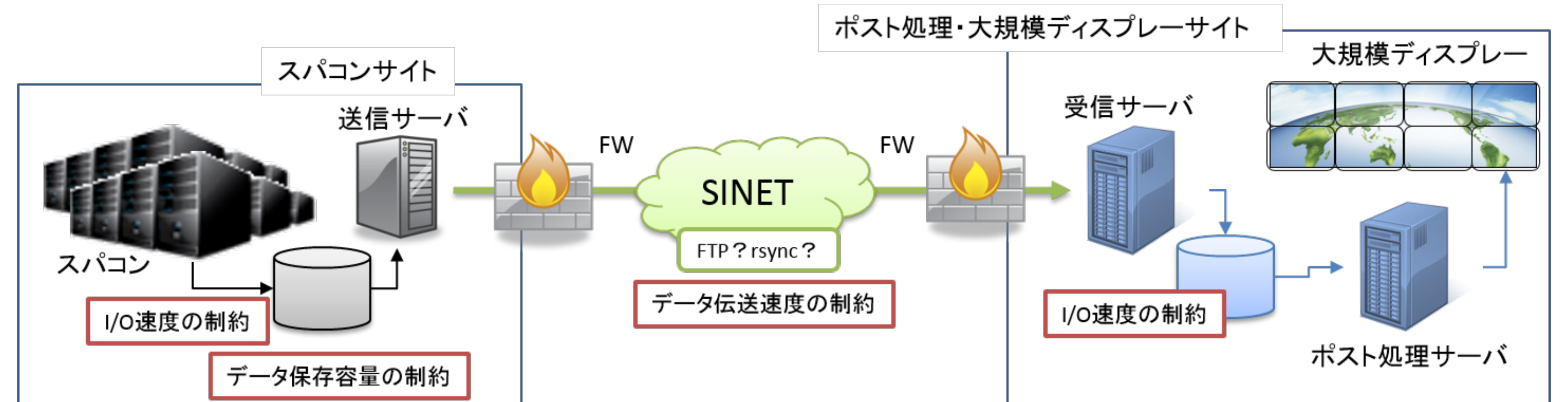
村田健史 (独立行政法人情報通信研究機構)

クラウドを活用したビッグデータポスト処理環境実現のためのデータ伝送実験



情報通信研究機構 統合データシステム研究開発室 長屋嘉明
 京都大学 学術メディアセンター 深沢圭一郎
 大阪大学 サイバーメディアセンター 伊達進・木戸善之
 名古屋大学 情報基盤センター 荻野正雄
 九州大学 情報基盤研究開発センター 南里豪志
 筑波大学 計算科学研究センター 建部修見
 愛媛大学 大学院医学研究科 木村映善

これまでの問題点

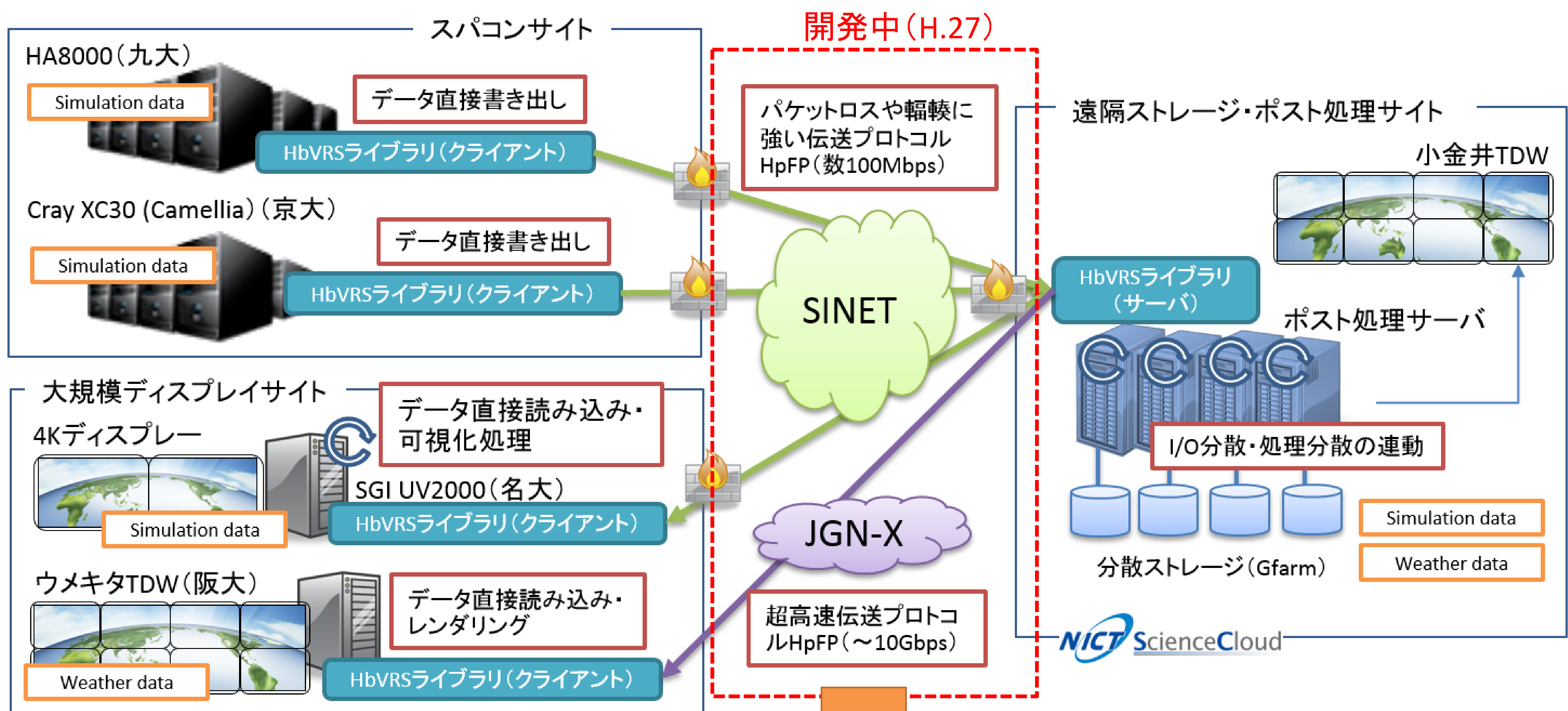


研究目的

現在、計算科学基盤であるスーパーコンピュータ(以下スパコン)による計算規模は大規模化しており、その結果出力されるビッグデータのポスト処理環境の重要性が指摘されている。大規模スーパーコンピュータにより生成されたビッグデータのポスト処理を高速かつ低遅延で処理するためには、スパコンからデータを直接書き出し、即時にポスト処理システムからも参照できる遠隔ストレージが有効である。遠隔高速ストレージは古くから研究開発が進められてきた。マルチポイント-マルチポイントの並列ストレージシステムの高速化の研究(例えば、Henschel et al., SC12, 2012)は多くあるが、シングルクライアントから利用できる遠隔高速ストレージはほとんど存在しない。これは、実用性の高いデータ伝送プロトコルが存在しない、ストレージ側のI/O速度のボトルネック、データ伝送プロトコルとストレージシステムを効果的に組み合わせることが容易ではないことなどが理由である。本研究では、スパコンの計算ノードから直接データを遠隔ストレージに書き出したり、遠隔ストレージから直接データを読み込んでポスト処理を行ったりするためのツール(後述のHbVRSツール)を開発する。このHbVRSクライアントを京都大学、九州大学のスパコンおよび大阪大学、名古屋大学のポスト処理環境・可視化環境に実装する。NICTサイエンスクラウドが構築・運用しているGfarmによる分散ストレージシステムにHbVRSサーバを実装し、これを遠隔ストレージシステムとして利用する。これらをもとに、実データを用いて実環境実験を行う。スパコンからの遠隔ストレージ利用では、32~128コアで1ステップあたり100GB程度のデータファイルを数時間に一度書き出す仕組みを実現する。また、L2環境下のポスト処理では、可視化サーバからバンド幅(1Gbpsまたは10Gbps)でのデータファイル読み出しを達成する。

研究概要

本申請では、理想環境での動作検証実験ではなく、実利用システムでの実現を目標としている。したがって、最終目標はスパコンユーザが遠隔ストレージ・ポスト処理環境をローカル環境のように使いこなし、さらに得られた可視化データを遠隔地の大規模ディスプレイに表示するなどの実際のアプリケーションを実現するインフラを提供することである。そのために、本研究では実際の計算科学研究者(スパコンユーザ)が実際に利用している環境(スパコンおよびポスト処理)において高速仮想遠隔ストレージの実験を行う。具体的には、京都大学や九州大学のスパコンで実際に掲載されているシミュレーションコードを用いてのデータ書き出しを行う。また、遠隔ストレージに保存されている実際の気象データやシミュレーションデータを本研究で開発するツールを用いてポスト処理や可視化を行う。これにより、SINET上での遅延、輻輳やパケットロス、各大学のFirewall、スパコンから遠隔ストレージまでのルーティング、データ書き出しとデータ保存、およびポスト処理のコンカレント処理など、実際にシステムを利用する際に想定される問題を一つずつ解決することで実用化に近づくことが期待される。



これまでの成果

特許申請準備中のためポスター会場で口頭でご説明します。おおむね順調に目標を達成しています(H.27年5月現在)。

- ①通信プロトコルの仕組み、②通信プロトコル開発状況、③室内実験結果、④ファイアウォール越え通信実験結果、⑤分散ストレージシステムへの組み込み実験結果