

複数サイトにまたがる仮想クラスタの管理機構

中田 秀基¹ 広淵 崇宏¹ 横井 威¹ 江原 忠士^{1,2} 谷村 勇輔¹ 小川 宏高¹ 関口 智嗣¹
1. 産業技術総合研究所 2. 数理技術先端研究所

背景

- 仮想クラスタによる管理コストの低減
- サイト内での資源共有によって可用性を向上
 - サイト間での資源共有はできない
- 大規模な仮想クラスタを構築することは難しい

目的

- 複数のサイトにまたがる仮想クラスタを可能にする管理システムを構築
- サイト間資源共有によるさらなる可用性の向上

仮想クラスタへの要請

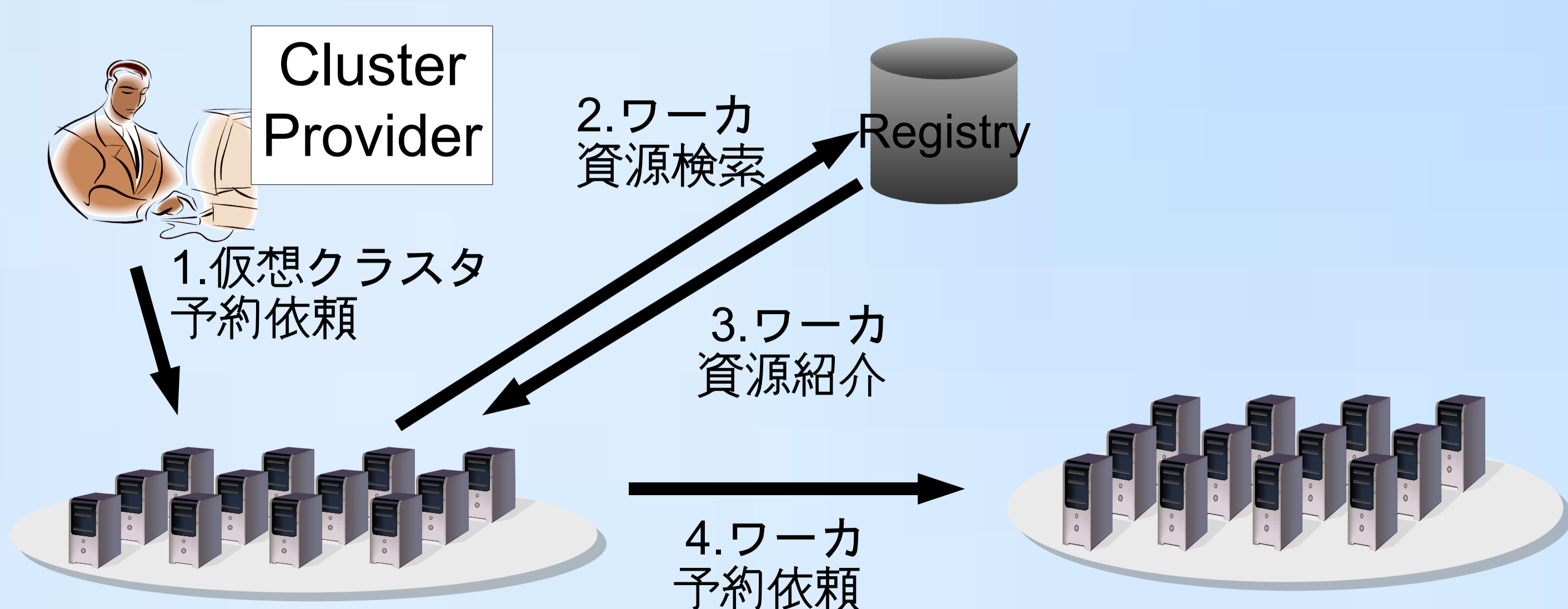
- シングルシステムイメージ
 - フラットなネットワーク
 - 共有ファイルシステム
 - 共有名前空間
- アプリケーション, 管理ユーティリティのインストール, コンフィギュレーション

提案仮想クラスタシステムの概要

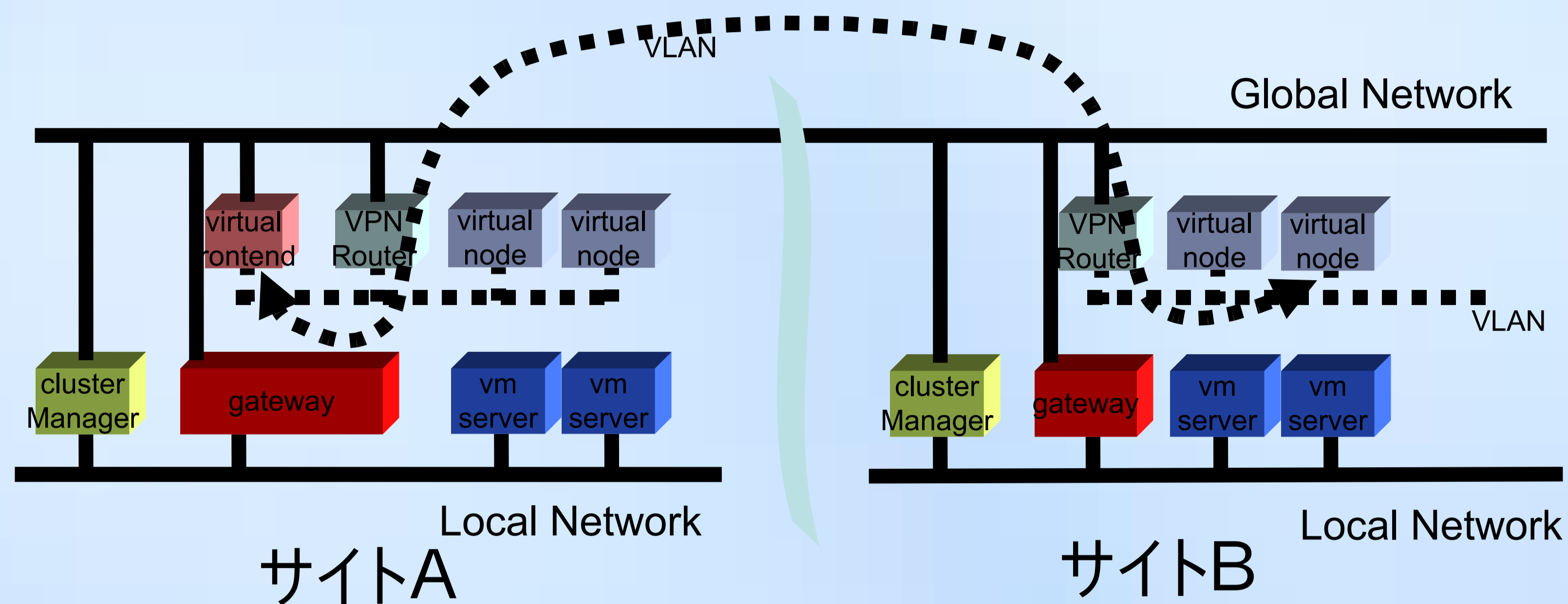
- 計算機, ネットワーク, ストレージの仮想化
 - 計算機 - VMware Server
 - ネットワーク - Tagged VLAN
 - ストレージ - LVM + iSCSI
- ソフトウェアの一体的デプロイ
 - NPACI Rocksを使用
 - 仮想ノード上に, 仮想フロントエンドノードを構築し, VLANで構築されたネットワーク上で各ノードをインストール
 - “Roll” によってソフトウェアを自動インストール
 - インストール時に各ノードをPXEブート, インストール

複数サイト対応の仮想クラスタ管理機構

- 各サイトが独立に仮想クラスタ構築機構を運用
- レポジトリサーバに登録
- 資源を必要とするサイトが, マスタとなりワーカーサイトを検索, 依頼
- REST風webインターフェイス



- ソフトウェアVPNで Tagged VLANをブリッジ接続
 - PXEブート, DHCPにも対応
 - Rocksには手を加えることなく利用可能



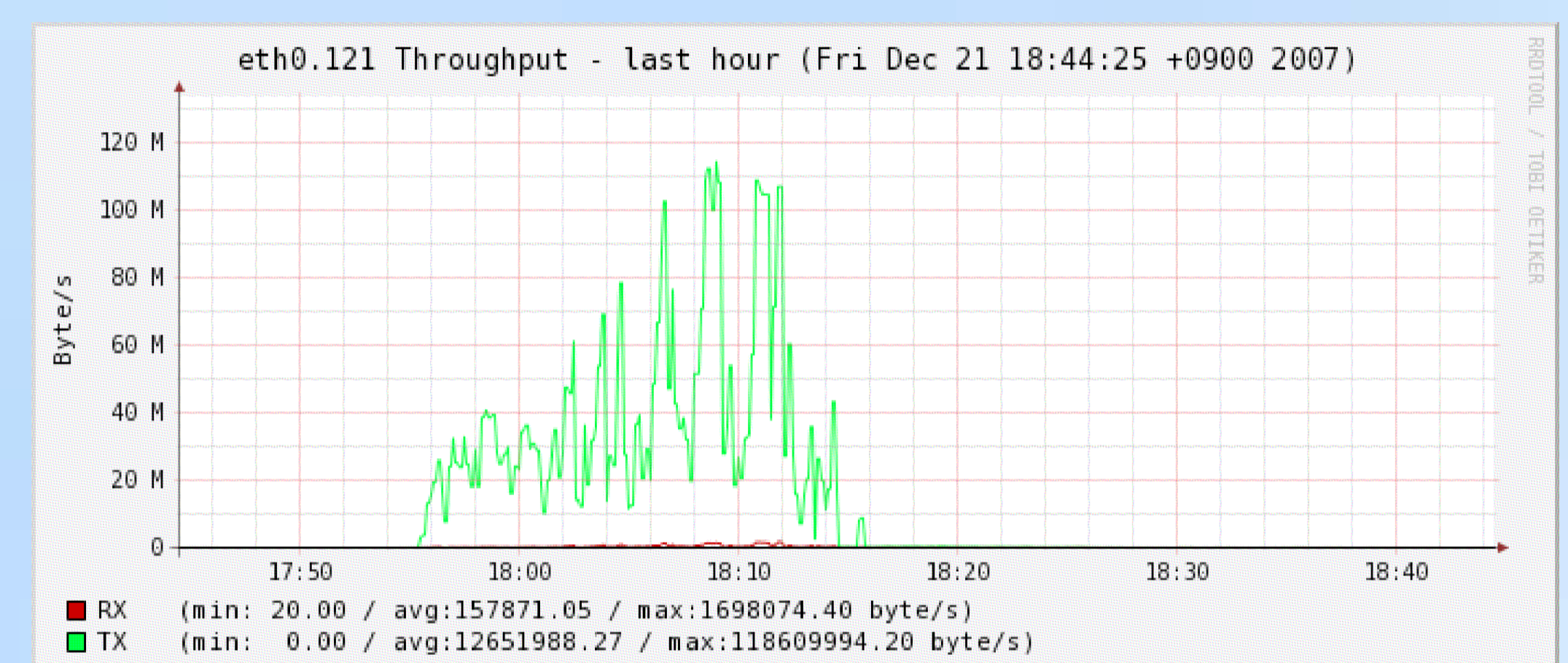
インストールパッケージ配布の最適化

- Rocks はbit torrent 的なパッケージ共有機構を持つ
- ディスクイメージのダウンロードはキャッシュされない
- トランスペアレントなHTTPキャッシュを導入

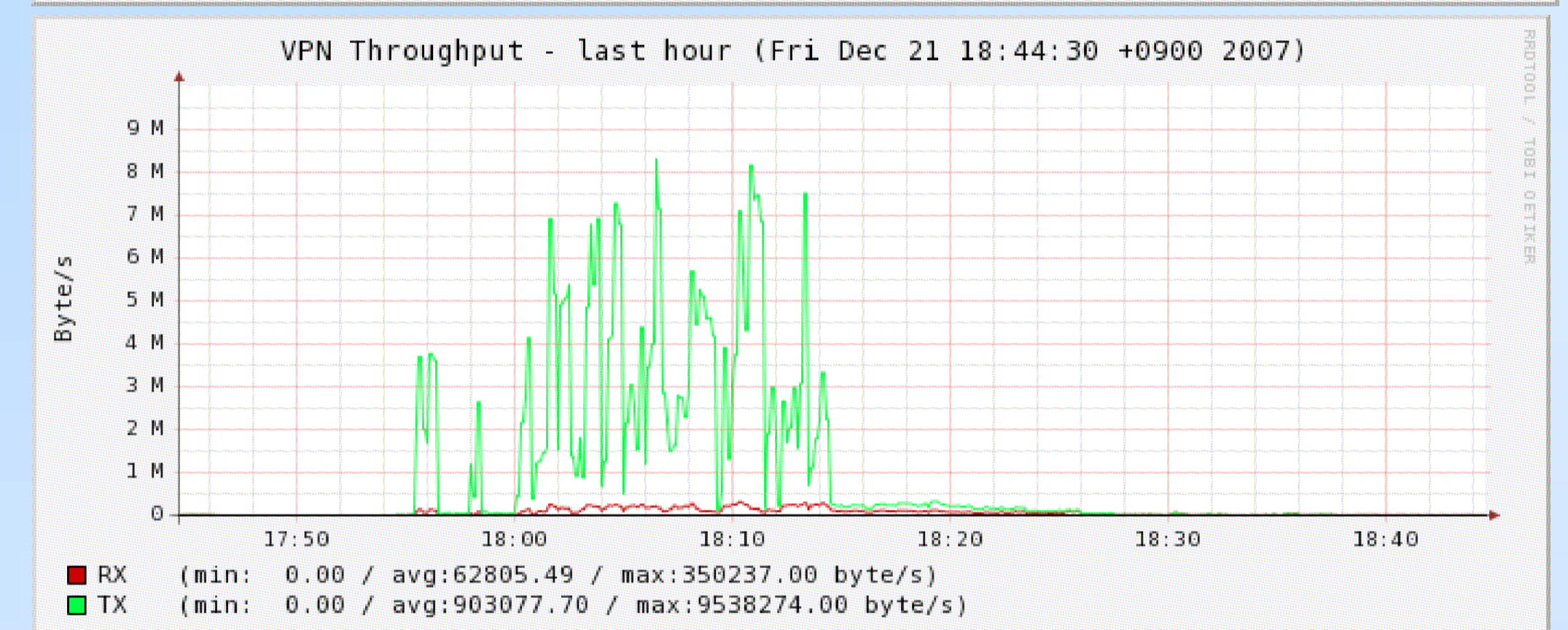
予備評価

- インストール時のサイト間トラフィックを測定

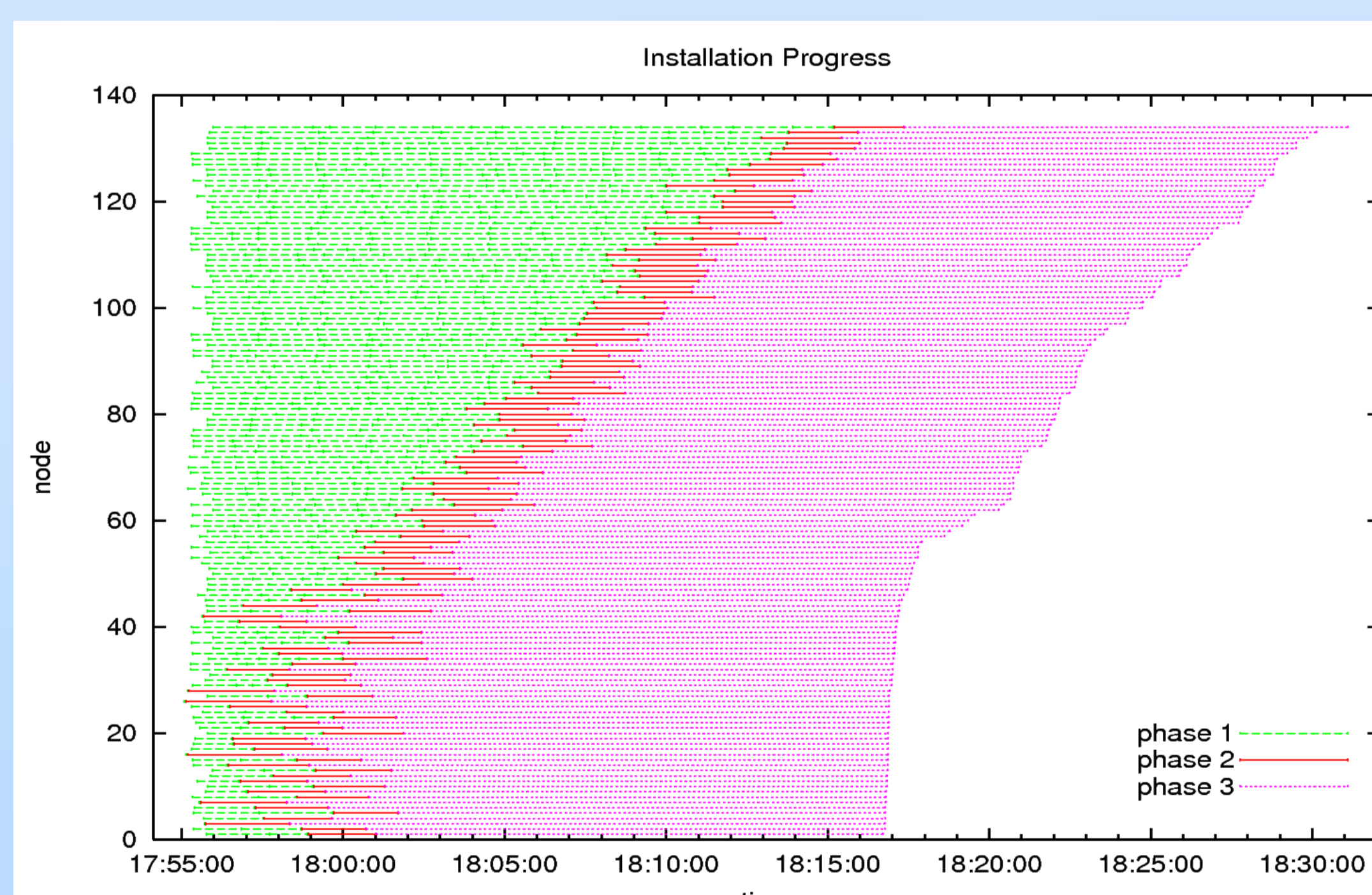
キャッシュなし相当
(キャッシュサーバの出口で計測)



キャッシュあり相当
(VPNの出口で計測)



130台
インストール時間の
内訳



結論

- 複数サイトにまたがる仮想クラスタ管理機構の提案
- 透過キャッシュを導入することで, インストール時のサイト間トラフィックを減少させることができることを確認

今後の課題

- インストール時間の削減
 - Rocks が用いるキックスタートファイルの事前生成
- 仮想クラスタ管理機構間インターフェイスのWSRF化