

オーバーレイネットワーク技術の 非常時通信への適用に関する一検討

長谷川剛 (大阪大学)
亀井聡 (NTT)
村田正幸 (大阪大学)

研究の背景 (1)

- ネットワーク障害への対応
 - 冗長系(ハードウェア・ソフトウェア)の構築、障害発生時の切り替え
 - ・ 発生し得る障害を想定し、効率の良い手法を用いる
 - ・ コストと性能のトレードオフ
 - 多くは、単一障害を想定
 - ・ 同時には1つしか故障しない
- 大規模ネットワーク障害
 - 災害、テロ、ルータOSの不具合...
 - 複数機器の同時故障・面的な障害
 - ネットワーク接続性、重要通信の優先が最重要
 - 発生確率は小さいため、冗長系で対応するにはコストが大きい

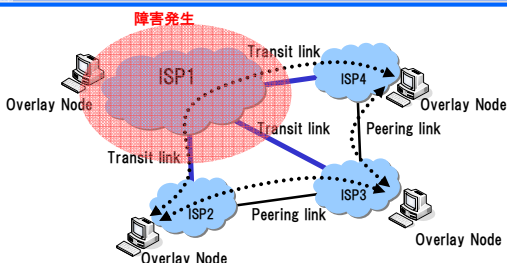
研究の背景 (2)

- Border Gateway Protocol (BGP)
 - インターネットのAS間ルーティングを司る
 - AS間リンクの経済的コスト、政治的思惑など、ネットワーク性能以外の事情を考慮することが可能
- 大規模ネットワーク障害への対応
 - BGPのルーティングテーブルの収束が遅くなる(数分~数時間)場合がある
 - ・ BGPは収束時間の理論的上限が存在しない
 - ・ 最小: ホップ数×固定時間で伝播
 - ・ フルメッシュ接続に近い場合、収束時間が大きくなる
 - 改善手法
 - ・ BGPそのもの、ルータの改変を必要とし、早期の展開は困難

研究の目的

- 大規模障害発生時の早期回復、高接続性の実現
 - オーバレイネットワークを利用
 - ・ IPネットワークの改変が(ほとんど)必要ない
 - ・ 早期のサービス展開が可能
- オーバレイルーティング技術の適用
 - オーバレイネットワークを用いてトラフィックの経路制御を行う
 - 大規模障害発生時のルーティングへ適用
 - ・ ユーザ性能重視ではなく、ネットワーク接続性重視
 - ・ 従来手法の欠点を改善し、適用可能なネットワーク規模を拡大

オーバレイルーティング (1)



- ISP1が落ちた時の、ISP2とISP4との到達性
 - BGPでは不可能(オペレータ同士の折衝が必要)
 - オーバレイルーティングを使うと、ISP3を経由した経路を自動的に利用可能
 - ・ アプリケーションレベルで動作するため、IPルーティングを超えたルーティングが可能

オーバレイルーティング (2)

- 参考: Resilient Overlay Network (RON) [13]
 - 参加ノード間でフルメッシュ論理ネットワークを構築
 - ・ 定期的な生死確認、ネットワーク性能の計測
 - ・ ノード間で情報交換を行う
 - ノード間のパケット転送を、直接行うか、他ノードを中継させるかを判断
 - ・ 2ホップまでの経路を用いる
 - ネットワーク障害を固定時間で全てのノードが把握可能
 - 計測、情報交換のオーバーヘッドが大きい
 - 50~100ノード程度までならば動作する

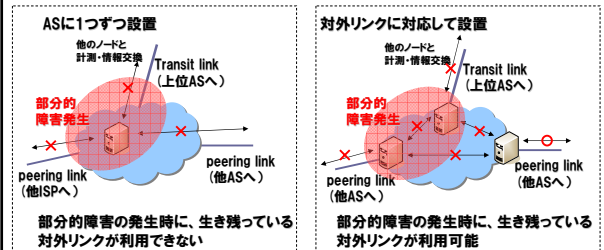
[13] D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris. "Resilient overlay networks." in *Proceedings of 18th ACM Symposium on Operating Systems Principles*, Oct. 2001.

提案手法の方針

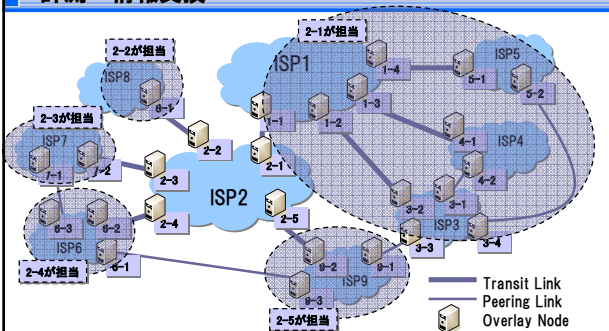
- フルメッシュ論理ネットワークを用いる
 - ネットワーク障害を短時間で発見するため
- オーバーレイノードの設置場所の検討
 - 参加するAS (ISP) のオペレータが戦略的に設置
 - ASトポロジが既知であることを利用
 - ASの部分的障害に対応する
- オーバーヘッドの削減
 - RONと同等の計測、情報交換を少ないオーバーヘッドで実現
- 3ホップ以上の経路も用いる
 - ネットワーク接続性を向上させるため

オーバーレイノードの設置

- 対外接続リンクに対応させてノードを設置
 - ネットワークの部分的障害に対応
- AS数に対してノード数が増える
 - 計測・情報交換のオーバーヘッドが増大



計測・情報交換



- IPルーティング情報を基に、担当範囲を分割
- 計測・情報交換のための通信回数を削減できる
 - 上の例では、125回から31回へ

到達性情報テーブル

		送信側ノード番号				
		1	...	i	...	N
受信側ノード番号	1					
	...					
	j			(接続性, タイムスタンプ)		
	...					
N						

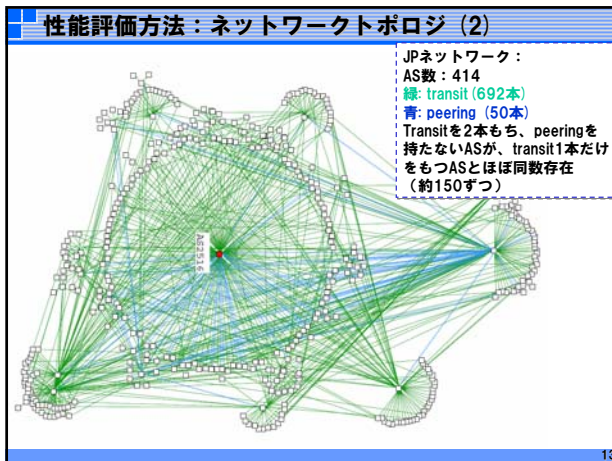
- 自分が計測した情報と、他ノードとの交換によって得た情報も交換する
- 交換によって得た情報のタイムスタンプを確認して自ノードのテーブルを更新

経路探索

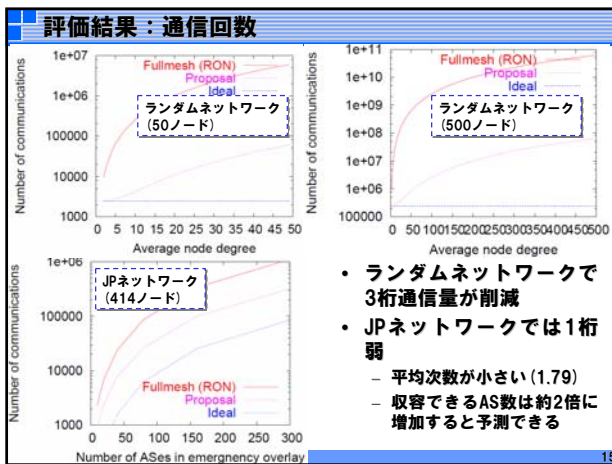
- 各ノードはダイクストラ法によって他ノードへの最短経路を算出
 - フルメッシュ (に近い) ネットワークなので計算時間は短い
- 障害発生から代替経路発見までの時間
 - 1ホップ、2ホップの伝播が必要な場合：固定時間で発見可能 (RONと同様)
 - 3ホップ以上伝播が必要な場合：最悪でも (ホップ数-1) × 固定時間で発見可能
 - 複数ASを経由する場合でも、直接パケットが届く相手には1ホップでの伝播が可能
 - (BGPでは経由するASホップ分だけ伝播に時間がかかる)

性能評価方法：ネットワークトポロジ (1)

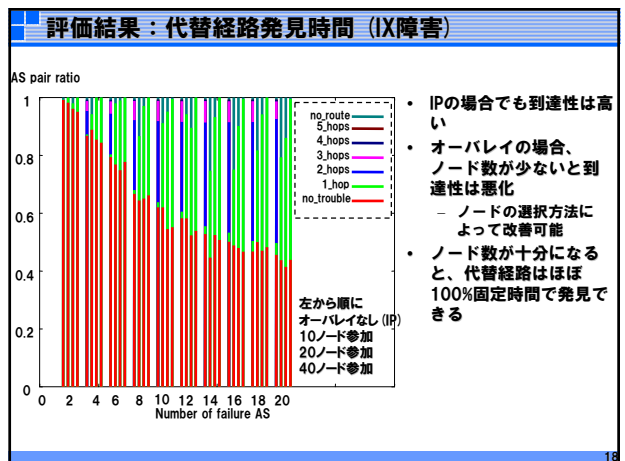
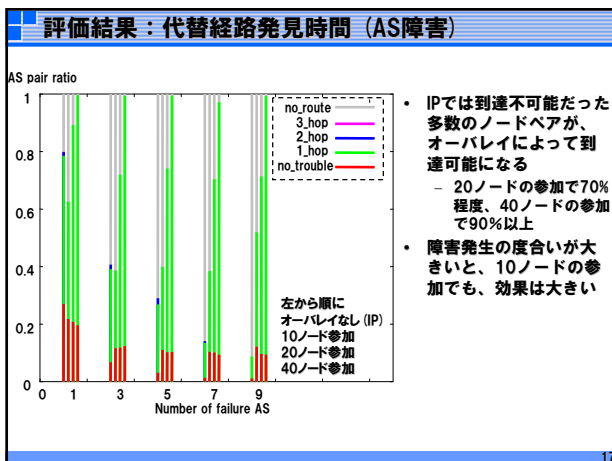
- ランダムネットワーク
 - ノード (AS) 数 50, 500
- CAIDA (<http://www.caida.org/>) が公開しているASネットワークトポロジ
 - AS間リンク情報
 - トランジットリンク・ピアリングリンクの区別
 - そのうちJPNIC管轄のASを抜き出したものを用いる (JPネットワーク)



- ### 性能評価方法：通信回数
- 評価指標：計測・情報交換に必要な通信回数
 - 各ASに1つずつオーバーレイノードを設置する場合
 - 通信量の観点からは理想的
 - 対外リンク毎にオーバーレイノードを設置する場合
 - 単純にフルメッシュで通信する場合
 - 提案する通信量削減手法を用いる場合



- ### 性能評価方法：代替経路発見時間
- 発生させるネットワーク障害
 - AS障害：選択したASおよび対外リンクが全て使えなくなる
 - IX障害：選択したAS群の間のリンクが全て使えなくなる
 - 共に、次数の大きいASから順に選択
 - 評価指標：代替経路発見時間
 - 全てのノードペア間の代替経路が発見されるまでの時間
 - IPの場合
 - 障害発生地点からのASホップ数×固定時間
 - 提案手法の場合
 - オーバーレイノードは、ピアリングリンクを持つASからランダムに選択
 - ネットワーク到達性の評価も行う
 - 代替経路が発見できないノードペアの割合



まとめと今後の課題

- **オーバレイルーティング技術を用いた非常時オーバレイネットワークの提案**
 - フルメッシュ論理ネットワークを構築
 - 効率的なノード間計測・情報交換
- **性能評価結果**
 - 通信回数を1-3桁改善
 - ネットワーク接続性の向上
 - 短時間での代替経路発見
- **今後の課題**
 - 他のネットワークを用いた評価
 - オーバレイルーティングへの切り替え方法の検討
 - 異なる方法を用いた非常時通信の提案

19