Dynamic Compact Routing Project

Kick-off meeting - Jan 16, 2009
INRIA Sophia-Antipolis

Dimitri Papadimitriou

Alcatel-Lucent BELL NV

dimitri.papadimitriou@alcatel-lucent.be

Meeting Agenda

10h30-11h00: Introduction	- 30min (All)
11h00-12h00: Project overview, motivations and objectives	- 60min (Dimitri)
12h00-13h00: Technical Phase 1	- 60min (Dimitri)
13h00-14h00: Lunch	
14h00-15h00: Technical Phase 1	- 60min (Cyril)
15h00-17h00: Technical Phase 2	- 120min
17h00-17h45: Detailed work plan, phasing/milestones	- 45min
17h45-18h15: Wrap-up and Conclusions	- 30min

Introduction

Introduction

1. Scientific project

- Context: (Future) Internet
- Topic: Distributed Dynamic Routing
- Approach: Science vs Engineering

2. Round Table

- Partners presentation/background
- Partners expectations

3. Administrative issues - if any remaining

Project Motivations and Objectives

Problem Statement

The Internet routing system is facing challenges in terms of

1. **Scalability**
2. Routing system dynamics: stability and convergence
3. Security

Main reasons:

- Resulting from its expansion, the Internet routing system has to cope with a growing number of sites, routes, and Autonomous Systems (with increasing meshedness but steady average AS path length)

→ **Increasing number of RT entries whereas shortest path routing scales ~ n log(n)**

- User/site addressing vs network addressing (overload of IP address space usage): topology independent address prefix allocation that impedes prefix aggregation

→ **Contribute BGP routing system instability (→ sustain higher dynamicity)**

- Existing solutions to mobility, site multi-homing, and inter-domain TE (using address prefix de-aggregation) exacerbate the limitations of the current routing system

→ **Routing system must not only scale with increasing network size/number of hosts but also with growing set of constraints and functionality**

Problem Statement

Impacts:

- User vs network addressing space (<-> overload of IP addressing space usage) - impacts TCP and other transport layer protocols/end-to-end communication
- **Sub-linear scalability of routing system wrt to number nodes ideally ~ log(n) - note: today scaling of routing system (shortest-path routing) ~ n log (n)**
- **Routing scalability not dissociable from routing system dynamics (stability and convergence properties)**

Root Causes:

Cause 1: Topology vs aggregation

- Originally, host addresses assignment based on network topological location
- Conditions to achieve efficient address aggregation and relatively small routing tables (tradeoff routing information aggregation vs granularity) are not met
- Deterioration root causes: increased AS meshedness, host mobility (Mobile IP), site multi-homing (~25% of sites), traffic-engineering

⇒ Super-linear growth of Routing Table (RT) even if network itself would not be growing (***routing protocol must not only scale with increasing network size !***)

Cause 2: Inter-domain routing protocol (BGP)

- Protocol implementation specifics: may be circumvented
- Protocol architecture: BGP is a path-vector protocol (eliminates DV count-to-infinity problem) but subject to Path exploration that affects convergence time:

Theoretical convergence time: upper bound ~ O(N!) and lower bound = W[(N-3) x MRAI timer]

Observed convergence time: (Max_AS-Path - Min_AS-Path) x MRAI timer

- Protocol usage: policy-based routing (- no policy distribution)

→ inter-AS oscillations (policy conflicts: local preferences over shortest path selection)

→ intra-AS oscillations (MED-induced oscillations)

BGP scalability and convergence problems

Scaling of routing algorithm: Routing Table (RT) size growth rate > linear (super-linear)

1. Routing engine system resource consumption ⇒ *cost growth rate* ~ 1.2-1.3/2 years

   - Routing space size

↑ #routing table entries ⇒ ↑ memory

↑ #routing table entries ⇒ ↑ processing and searching (lookup)

   - Number of peering adjacencies between routers

↑ #peering adjacencies ⇒ ↑ memory (due to dynamics associated with routing information exchanges)

2. Exacerbates BGP convergence time

   - BGP convergence time is limited by access speeds of DRAM (used for RIB storage)
     - DRAM capacity growth rate: ~4x every 3.3 years (**faster than Moore's law**)
     - DRAM access speed growth rate: ~1.2x every 2 years
   - BGP convergence time degradation rate (estimation):

$$\frac{\textit{RT growth rate } [1.25\text{-}1.3]}{\text{DRAM access speed growth rate } [1.1]} \sim 10\% \text{ per year}$$

Note: speed limitations can be absorbed using parallelism

Alternatives: Solution space

BGP improvements

- BGP multi-path
- Fast re-routing
- AS-path limit (diameter)
- Route cause notification

Compact Routing

- Name dependent schemes: e.g. TZ scheme, BC scheme
- Name independent schemes: e.g. Abraham scheme

… as of today none can efficiently deal with topology dynamics such as the Internet (dynamic routing)

Hybrid routing protocols

- Combination of LS/PV: Hybrid Link-state Path-vector (HLP)
- Combination of LS/DV: LVA

Others

- Loc/ID separation (host-based: SHIM6, HIP - router-based: LISP, GSE)
- User-controlled multi-path routing (elimination)
- Geographical routing
- Polymorphic routing

Objective

Routing problem space:

- **Alternative 1 (evolutionary)**: BGP re-considered (is it possible ?) or new candidate protocol like HLP - but no improvement possible on RT size scale from aggregation
- **Alternative 2 (disruptive)**: topology-dependent **compact routing** on locators or move directly to topology-independent compact routing (same worst case)

In both cases: how to account for topology dynamics ?

Bottom line:

- Routing requires coherent full-view (network graph topology or distance to destination) and support of topology dynamics $\Rightarrow$ timely routing updates
- Routing information exchange and its processing cost cannot grow slower than linearly on Internet

→ Challenge: ***compromise between routing scaling and dynamics***

Construct in polynomial time a compact routing scheme that minimizes the stretch bound for Internet-like graph while i) requiring only o(n) bits of routing information per node and ii) minimizing communication costs

Project Overview

Project Tasks

Tasks:

- **Task 1** (Specification and formal verification): Dynamic compact routing scheme formal specification and verification (analytical)
- **Task 2** (Experimentation): Dynamic compact routing scheme quantitative performance evaluation (in terms of number of routing table entries and memory size) on Internet-like graphs

Deliverables: to each task corresponds a specific deliverable

- Deliverable D1 for Task 1
- Deliverable D2 for Task 2

Project Timeline

Duration: 13 months (1st Mar. 2009 -> 31th March. 2010)

Timeline:

- **T0 (March 1st 2009)** : beginning of the study
- **T1 (T0+03** months) : meeting(*) on progress on Task 1, start preparation of Task 2
- **T2 (T0+06** months) : meeting(*) on progress on Task 1, start of Task 2

  draft version of D1 available
- **T3 (T0+07** months) : first final version of D1 available
- **T4 (T0+09** months) : meeting(*) on progress on Task 2
- **T5 (T0+12** months) : meeting(*) on progress on Task 2, draft version of D2 available
- **T6 (T0+13** months) : final version of D2 available (and presentation at Alcatel-Lucent Bell Antwerp of the global results)

Note:

- Ad-hoc Interim meeting and/or conference calls on progress of either on Task 1 or Task 2 can further complement this timeline
- At T6 (T0+13), deliverable D1 can be object of a revision based on the results obtained as part of Task 2

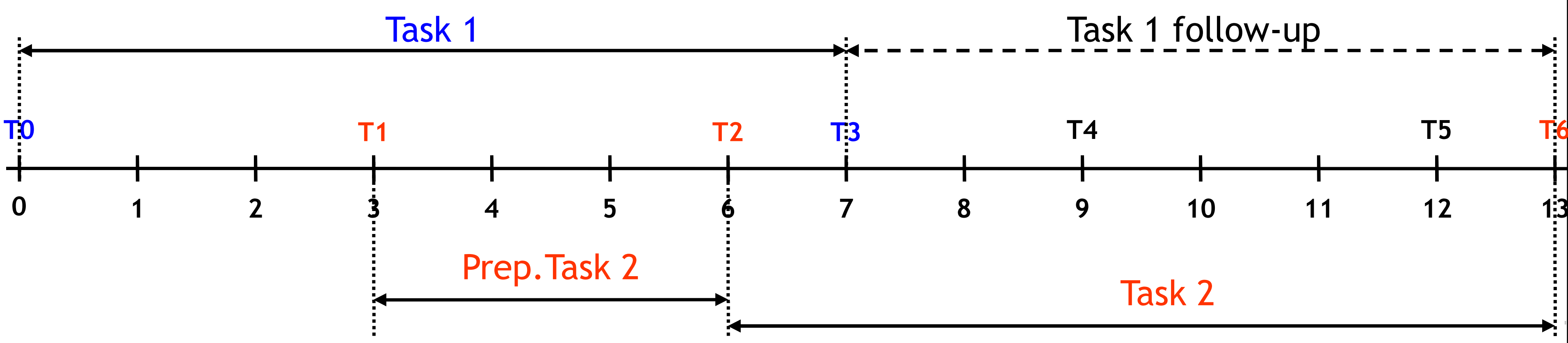
(*) all partners present

Project Timeline: Tasks

Timeline:

- **T0 (March 1st 2009)** : beginning of the study
- **T1 (T0+03** months) : meeting on progress on Task 1, start preparation of Task 2
- **T2 (T0+06** months) : meeting on progress on Task 1, start of Task 2

  draft version of D1 available
- **T3 (T0+07** months) : first final version of D1 available
- **T4 (T0+09** months) : meeting on progress on Task 2
- **T5 (T0+12** months) : meeting on progress on Task 2, draft version of D2 available
- **T6 (T0+13** months) : final version of D2 available

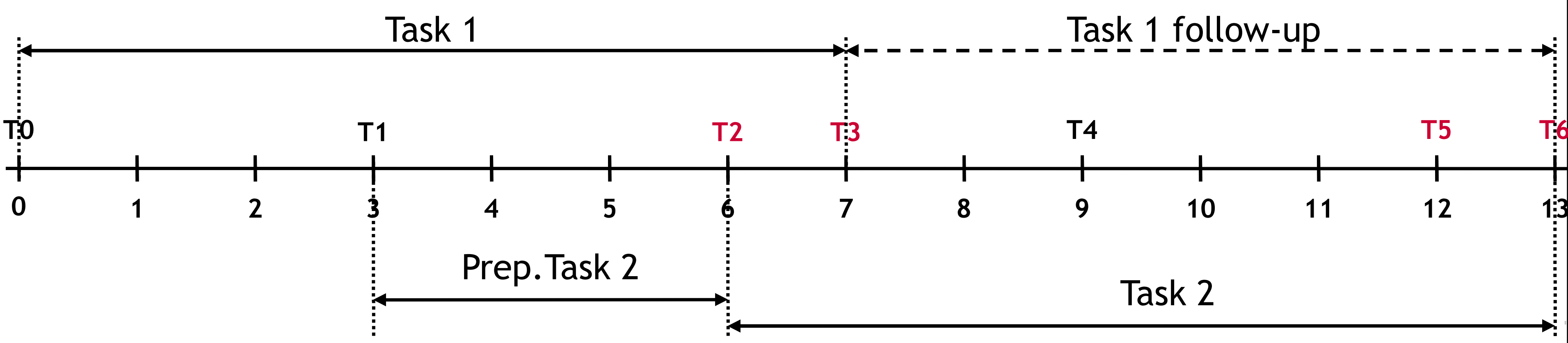


Project Timeline: Deliverables

Timeline:

- **T0 (March 1st 2009)** : beginning of the study
- **T1 (T0+03** months) : meeting(*) on progress on Task 1, start preparation of Task 2
- **T2 (T0+06** months) : meeting(*) on progress on Task 1, start of Task 2

  draft version of D1 available
- **T3 (T0+07** months) : **first final version of D1** available
- **T4 (T0+09** months) : meeting(*) on progress on Task 2
- **T5 (T0+12** months) : meeting(*) on progress on Task 2, **draft version of D2** available
- **T6 (T0+13** months) : **final version of D2** available

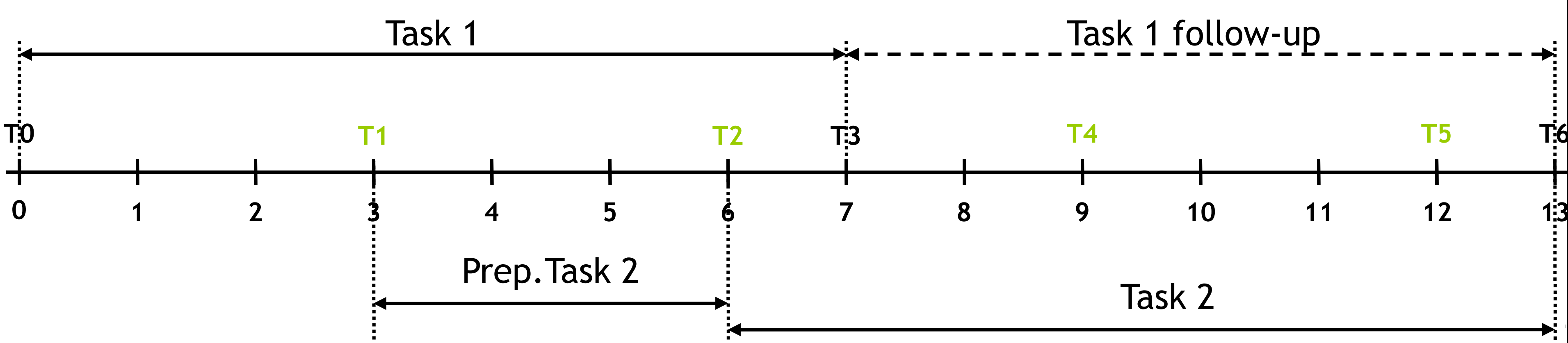


 (*) all partners present

Project Timeline: Meetings

Timeline:

- **T0 (March 1st 2009)** : beginning of the study
- **T1 (T0+03** months) : meeting(*) on progress on Task 1, start preparation of Task 2
- **T2 (T0+06** months) : meeting(*) on progress on Task 1, start of Task 2
  draft version of D1 available
- **T3 (T0+07** months) : first final version of D1 available
- **T4 (T0+09** months) : meeting(*) on progress on Task 2
- **T5 (T0+12** months) : meeting(*) on progress on Task 2, draft version of D2 available
- **T6 (T0+13** months) : final version of D2 available



(*) all partners present

Project Leadership

- **Task 1** technically leadership by Universite de Bordeaux

Duration Task 1: from T0 to T3

Follow-up during period from T3 to T6

- **Task 2** technically leadership by INRIA/Sophia-Antipolis (projet MASCOTTE)

Duration Task 2: from T1 to T2 (preparation), T2 to T6

Note: preparation phase can start earlier e.g. at T0

- *Both tasks are under the technical supervision of Alcatel-Lucent Bell*