Optimizing the design of elastic content distribution systems for data-intensive scientific communities

Daniel Higuero

Advisors: J.Carretero, F. Isaila

Computer Science Department, Carlos III University, Spain

Universidad Carlos III de Madrid
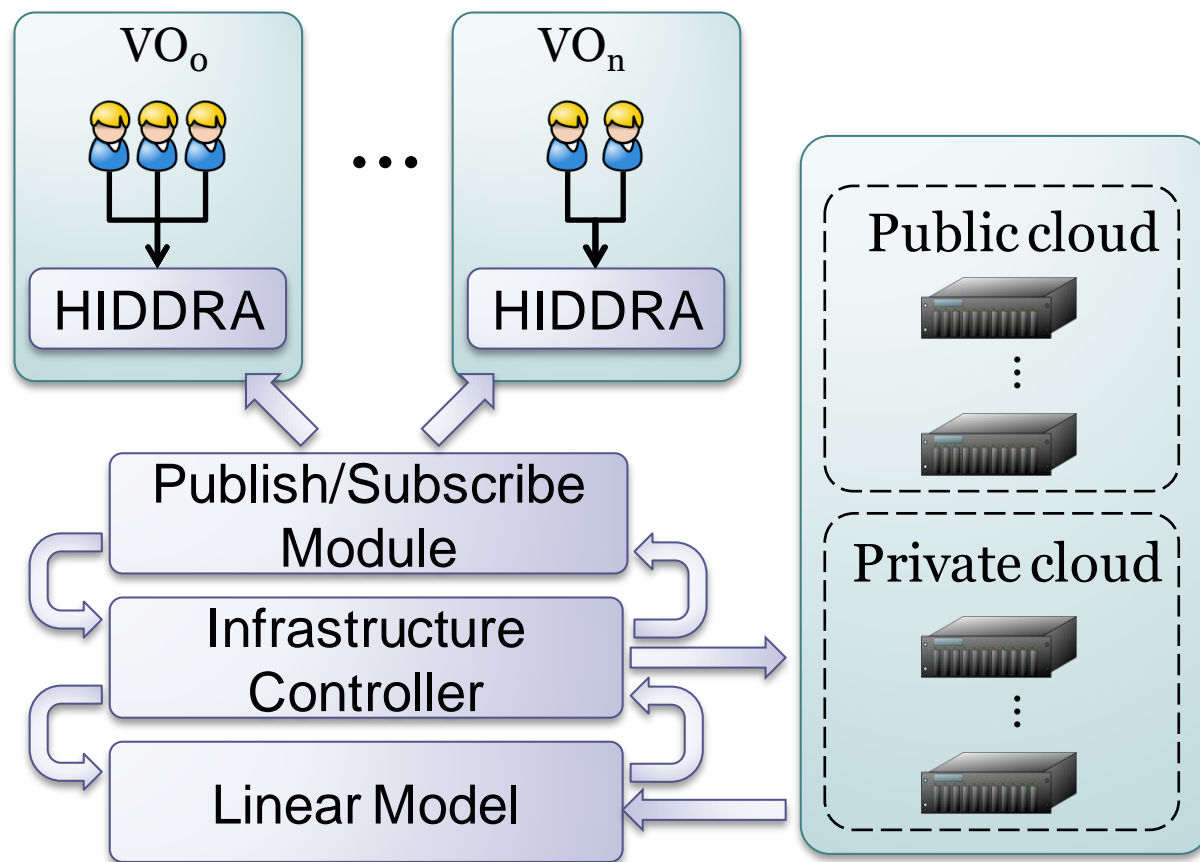
ARCOS

Motivation

- Data size is increasing
  - Web 2.0
  - Multimedia content
  - Scientific experiments
- Users are geographically distributed
- Networks do not evolve at the same speed
- Energy cost is increasing

How to improve data distribution and existing infrastructure utilization in content distribution networks?
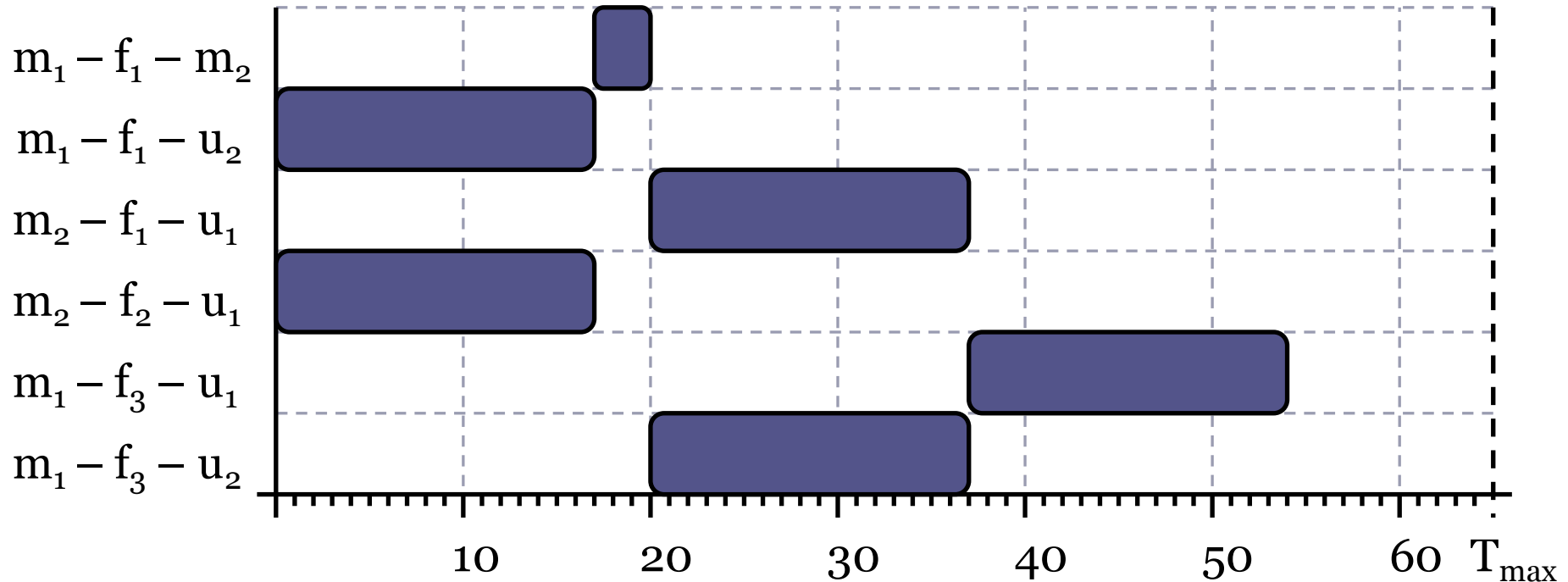
Thesis goals

- Leverage network analysis for improving data distribution
  - Understand the users
- Model the data distribution process and the supporting infrastructures using linear programming techniques
  - Understand the distribution process
- Elastic data distribution infrastructures
  - Cost-effective solutions that adapt to user needs

Architecture overview

Solution provided by the linear model

$(source - f_i - dest)$

Chart rows (top to bottom):
- $m_1 - f_1 - m_2$
- $m_1 - f_1 - u_2$
- $m_2 - f_1 - u_1$
- $m_2 - f_2 - u_1$
- $m_1 - f_3 - u_1$
- $m_1 - f_3 - u_2$

X-axis: 10, 20, 30, 40, 50, 60, T_{max}

ARCOS

Ongoing work

- Linear model optimization
  - Explore different solver configurations
  - Explore alternative heuristics
- Large scale evaluation using cloud infrastructures
  - Model real data distribution infrastructure
  - Deploy on Amazon or OpenStack
  - Evaluation of real use cases

Thanks for your attention!

Universidad
Carlos III de Madrid

ARCOS