Revenue Models for Streaming Applications over Shared Clouds

Rafael Tolosana-Calasanz, José A. Bañares, Congduc Pham and **Omer F. Rana**

{rafaelt, banares}@unizar.es, congduc.pham@univ-pau.fr, o.f.rana@cs.cardiff.ac.uk

Universidad de Zaragoza
Zaragoza, Spain

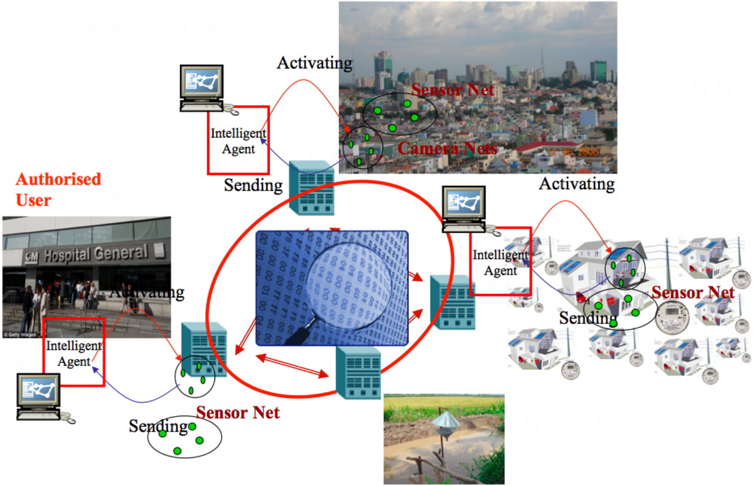
Université de Pau
Pau, France

Cardiff University
Cardiff, United Kingdom

July 14^{th}, 2012

C4BB4C 2012 – ISPA 2012
Madrid, Spain

Sensor Data Aggregation



Zaragoza

Approach & Focus

Adaptive infrastructure for sensor data analysis

- **Multiple** concurrent data streams with **SLA**
- **Variable** properties: rate and data types; various processing models
- Support for **in-transit** analysis, enforcing **QoS**
- Support for **admission** control & flow **isolation** at each node
- In case of QoS violation, **penalisation**

Approach & Focus

Adaptive infrastructure for sensor data analysis

- **Multiple** concurrent data streams with **SLA**
- **Variable** properties: rate and data types; various processing models
- Support for **in-transit** analysis, enforcing **QoS**
- Support for **admission** control & flow **isolation** at each node
- In case of QoS violation, **penalisation**

Key focus

- Understanding **revenue models** for in-transit analysis
- Understanding the impact on **faults** on such a revenue model

Outline

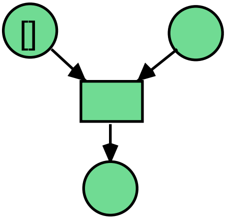
1. Background
   - Petri nets
   - Reference nets
2. In-transit Analysis
3. System Architecture & Revenue Model
   - Token Bucket
   - Autonomic Data Transfer Service
4. Evaluation
5. Conclusions and Future Work

Outline

Petri nets

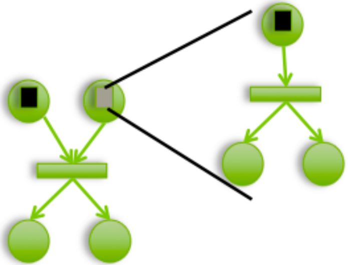
Characteristics

- directed bipartite **graph**
- **2** types of **nodes**: places and transitions
- **arcs**: place-transition, transition-place
- **tokens**: move on the graph
- **static** structural nature



Universidad Zaragoza
1542
CARDIFF UNIVERSITY PRIFYSGOL CAERDYD
UNIVERSITÉ DE PAU ET DES PAYS DE L'ADOUR

Reference nets



Characteristics

- **tokens** can be nets – dynamic hierarchies of Petri nets
- **Java** inscriptions & **Renew** interpreter
- we can build executable rapid prototype models – concurrency

Outline

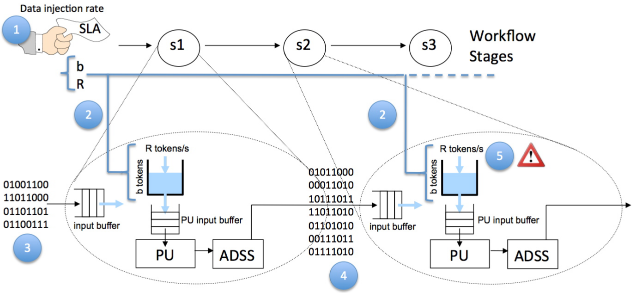
In-transit Analysis

Characteristics

- Perform **partial/full** processing of data from source to destination
- Benefit from availability of **slack** in the network – i.e. availability of excess capacity at processing nodes
- Useful to support: filtering, statistical analysis (min, max, avg) over a window size – i.e. common (often repeated) operations
- Same operation available at **multiple nodes** – location of analysis not important

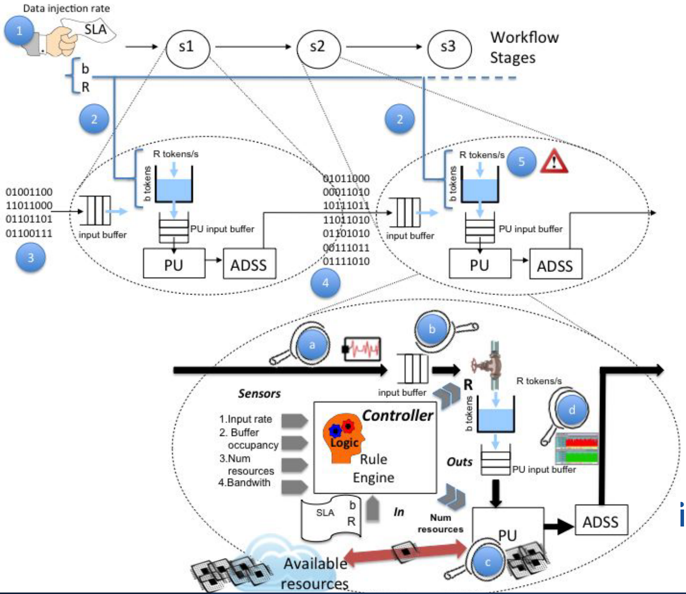
Outline

System Architecture

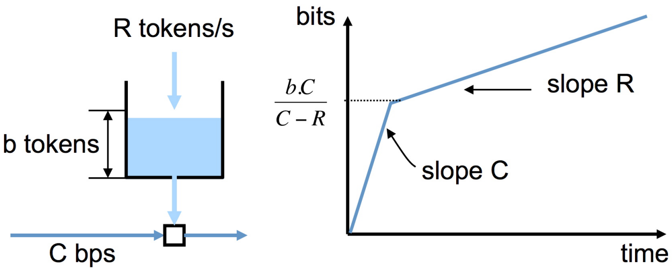


- **3** key components / node: Token Bucket, Processing Unit & output streaming
- Each component provides various **tunable parameters** – these can be externally modified

System Architecture

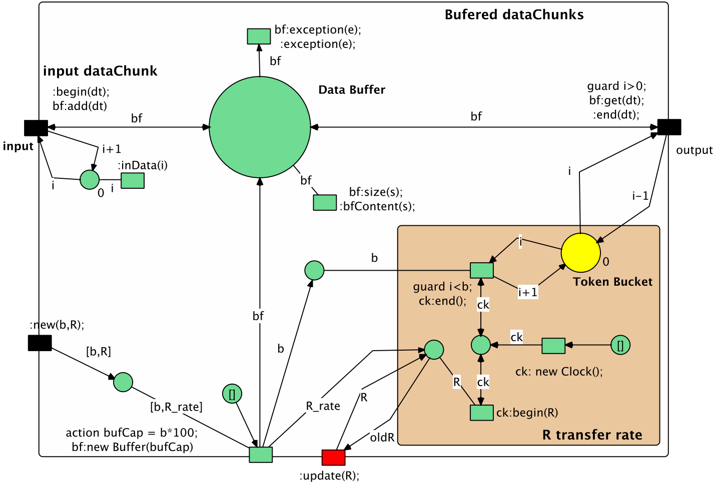


Token Bucket Behaviour



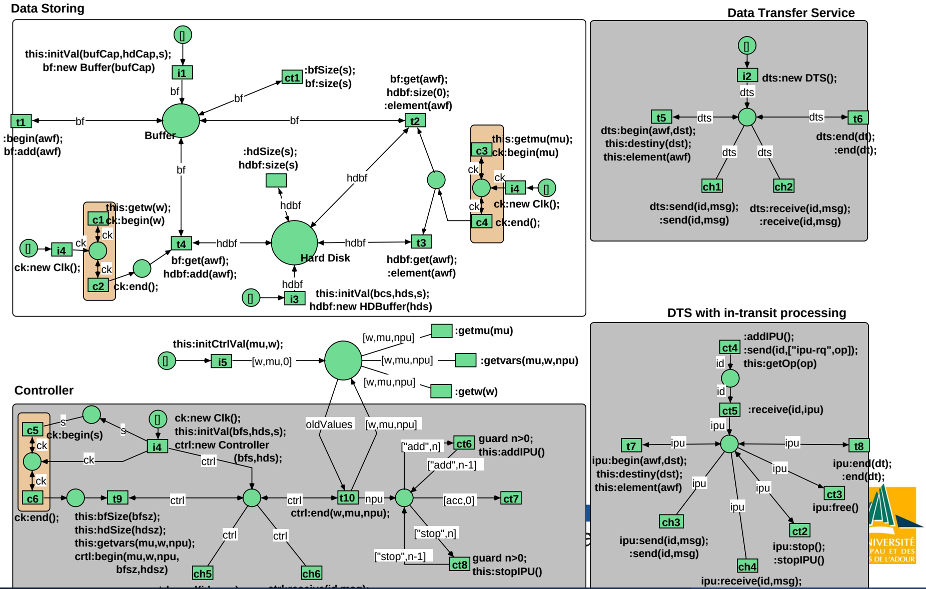
- Two key parameters of interest: R and b.
- Behaviour is dictated by changes in these two parameters.

Token Bucket & Processing Units



1542
Zaragoza
PRIFYSGOL CAERDYDD
UNIVERSITÉ DE PAU ET DES PAYS DE L'ADOUR

Autonomic Data Transfer Service



Revenue Model for a Cloud Provider

Elements of the Model

- **Revenue**: prices charged to clients $\sum_{a=1}^{n}\sum_{b=1}^{m} Pr(O_{ab})$
- **Cost**: for performing such operations $\sum_{a=1}^{n}\sum_{b=1}^{m} c(O_{ab})$
- **Penalisation** in case of QoS violation for client a: $PSLA_a$

Revenue Model for a Cloud Provider

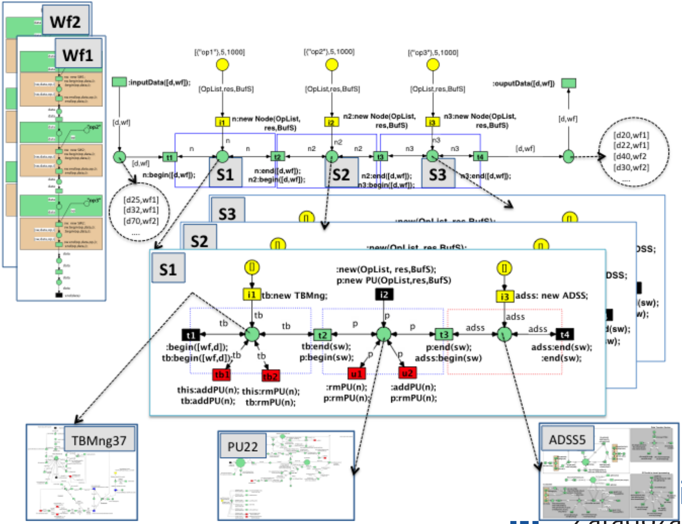
Elements of the Model

- **Revenue**: prices charged to clients $\sum_{a=1}^{n} \sum_{b=1}^{m} Pr(O_{ab})$
- **Cost**: for performing such operations $\sum_{a=1}^{n} \sum_{b=1}^{m} c(O_{ab})$
- **Penalisation** in case of QoS violation for client a: $PSLA_a$

Minimizing the cost

- $\sum_{a=1}^{n} \sum_{b=1}^{m} Pr(O_{ab}) - min(\sum_{a=1}^{n} \sum_{b=1}^{m} c(O_{ab}), \sum_{a=1}^{n} PSLA_a)$

Renew Model



idad Zaragoza 1542

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

UNIVERSITÉ DE PAU ET DES PAYS DE L'ADOUR

Dynamic Resource Provisioning

Conditions

- In the event of failure, **actions**: i) increase of resources, ii) assume penalisation
- In case of failure
  - there is an **increase** in execution time
  - resources are not released as expected – other flows are affected
- The action that minimises cost will be taken

Dynamic Resource Provisioning

Conditions

- In the event of failure, **actions**: i) increase of resources, ii) assume penalisation
- In case of failure
  - there is an **increase** in execution time
  - resources are not released as expected – other flows are affected
- The action that minimises cost will be taken

Model extension

- initial estimated NumRes: $\hat{NumRes_j} = \sum_{i=1}^{n} R_i/\hat{\delta_{ij}}$
- real-time estimated NumRes: $\hat{NumRes_j} = \sum_{i=1}^{n} R_i/\delta_{ij}$

Outline

Experiments

Outline

Conclusions and Future Work

Conclusions

- in-transit processing of **multiple** data streams over a **shared** (elastic) infrastructure
- dynamic Token Bucket (admission control): support of **variable bursts**
- **elastic Processing Unit**: add / reduce computational resources
- Autonomic Data Transfer Service: **adaptive** transfers
- Analyse **Revenue** models in the presence of faulty resources
  - two actions: i) add more resources, ii) assume penalisations
  - the action that minimises cost will be taken

Conclusions and Future Work

Conclusions

- in-transit processing of **multiple** data streams over a **shared** (elastic) infrastructure
- dynamic Token Bucket (admission control): support of **variable bursts**
- **elastic Processing Unit**: add / reduce computational resources
- Autonomic Data Transfer Service: **adaptive** transfers
- Analyse **Revenue** models in the presence of faulty resources
  - two actions: i) add more resources, ii) assume penalisations
  - the action that minimises cost will be taken

Future Work

- Currently validating in an Electrical Vehicles scenario
- Considering bursts, variable income data rates
- Trying to minimise the number of resources: Green computing

Revenue Models for Streaming Applications over Shared Clouds

Rafael Tolosana-Calasanz, José A. Bañares, Congduc Pham and **Omer F. Rana**

{rafaelt, banares}@unizar.es, congduc.pham@univ-pau.fr, o.f.rana@cs.cardiff.ac.uk

Universidad de Zaragoza
Zaragoza, Spain

Université de Pau
Pau, France

Cardiff University
Cardiff, United Kingdom

July 14^{th}, 2012

C4BB4C 2012 – ISPA 2012
Madrid, Spain