Centre for Advanced Studies, IBM Toronto Lab

Scalable Adaptive Web Services

Marin Litoiu, Centre for Advanced Studies, IBM Canada
Mircea Mihaescu, IBM US
Dan Ionescu, Bogdan Solomon, University of Ottawa

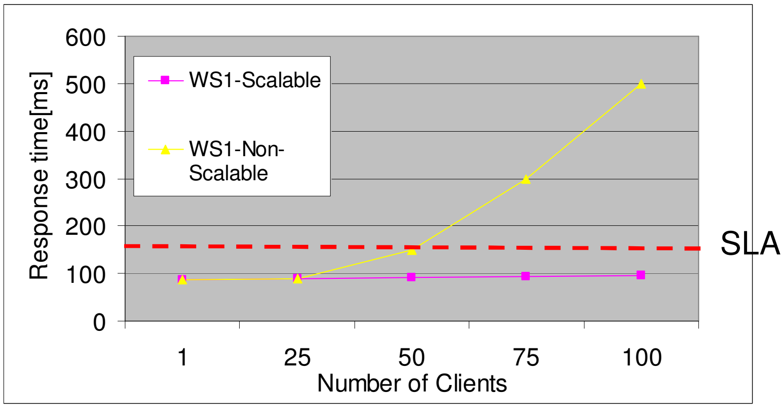
SDSOA 2008 May 11, 2008

Content

- **Introduction**
- **Feedback Loops**
- **Challenges**

Centre for Advanced Studies, IBM Toronto Lab

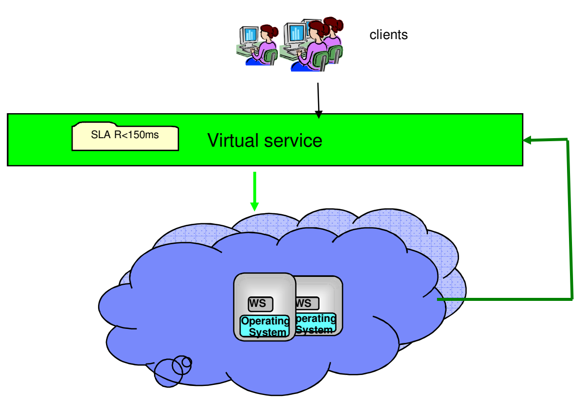
Scalable and Adaptive Web Services(1)



SDSOA 2008 May 11, 2008

Centre for Advanced Studies, IBM Toronto Lab

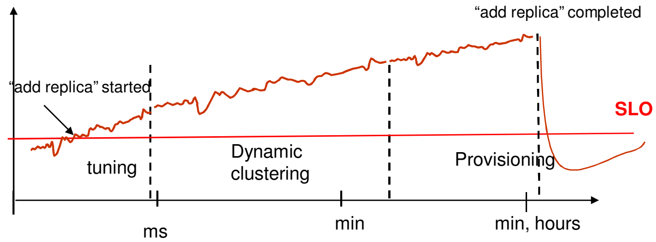
Scalable and Adaptive Web Services(2)



SDSOA 2008 May 11, 2008

Challenges…doing it predictive

Performance, actions



- There is a delay between measurement and the end of change execution
  - tuning (ex: change no of threads) can be done in ms
  - provisioning can be done in s, min, hours…
- Without prediction, the adjustments might come too late
  - breaches of SLA, loss of customers…

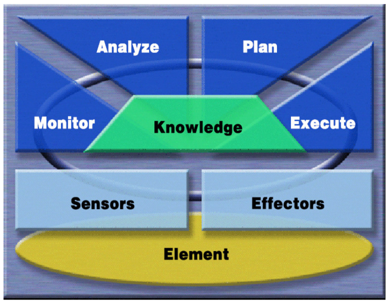
More challenges..

- **Ensure stability**
  – Do not overreact to transients
  – How much to change? – Need a model of the system
- **Minimum intrusiveness**
  – How much instrumentation is enough?

Feedback Loops

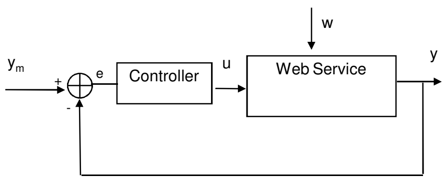
Various feedback loops proposed

- Performance Model Based Adaptive Control
- Linearized Feedback Control
- Autonomic Robust Control (this presentation)



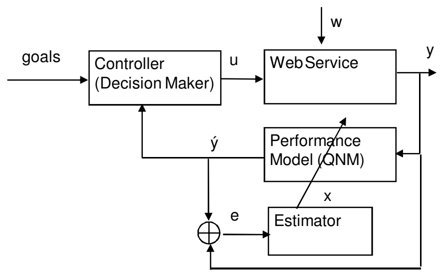
Linearized Feedback Control

- **A linear system describes the system to be controlled and the controller equations are derived based on a well-developed theory.**
  - The controller acts directly on the error *e* by integrating, derivating and/or amplifying it.
  - The main advantage of this approach is that is very fast and deals with non-steady states.
  - The main disadvantage is the static nature of the controller. It is assumed that the model of the service is static – its usefulness is limited.



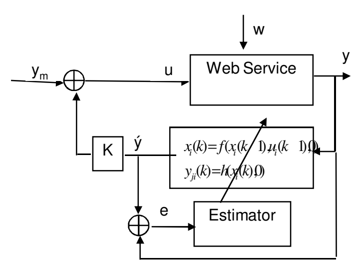
Performance Model Based Adaptive Control

- **Explicit queuing performance model of the system is created and maintained at run time.**
  - Model is maintained by an Estimator, which, based on the output of the system (*y*) and output of model (*ý*) computes new parameters (*x*) for the model.
  - The model and the estimator work in a feedback loop.
  - The role of controller is to achieve the performance *goals*, which are expressed as service level agreements.
  - The queuing models capture steady-state system behaviour – appropriate for controlling long term trends in the service and therefore is very appropriate for provisioning.



Autonomic Robust Control

- **Extended linearized feedback architecture**
  - The system is assumed non-linear
  - The model has to cope with ever changing model parameters and even with unknown and un-modeled states and disturbances.
  - The structure and an initial value for parameters are constructed off-line, through established system identification methods
  - Once the model is obtained, an estimator and its associated controller will keep the QoS automatically in line with the SLA.



Challenges

- **Large variety of workloads.**
  - Very likely, a feedback loop addressed one type of workloads
  - There is not yet a clear understanding when to switch between the schemes.
- **The composition of controllers for aggregated autonomic web services.**
  - Most control schemes address individual web services
  - No theory on how to combine multiple autonomic loops
    - and what is the effect of aggregation on the overall quality of control?
- **Availability of the measurement points**
  - how do we measure a the service that runs in an inaccessible administration domain?

Centre for Advanced Studies, IBM Toronto Lab

Conclusions

- **Hybrid technique has been introduced for model identification**
  - Initially the model parameters have been calculated using the RPE method using an online procedure.
  - The method was applied considering that the model parameters are piecewise constant, and that on the interval on which the output samples were collected the system model does not change.
  - The Akaike information model was then applied to obtain the best order of the model which can represent the real process.
  - A robust estimator was built considering the initial model obtained.
- **The results obtained led to considering a process model which contains a known and modeled part and an uncertain and un-modeled part as well.**
  - The un-modeled part was represented by applying the first variation (variation calculus) to the functions considered as the known part of the model.
  - This procedure allowed in turn building robust state estimators insensitive to uncertainties in parameter and model structure.

