Elasticity Engineering

Hong-Linh Truong
Faculty of Informatics, TU Wien

hong-linh.truong@tuwien.ac.at
http://www.infosys.tuwien.ac.at/staff/truong
@linhsolar
What is elasticity?

What is elastic computing?
Views on Elasticity

1. Demand elasticity
   Elastic demands from consumers

2. Output elasticity
   Multiple outputs with different price and quality

3. Input elasticity
   Elastic data inputs, e.g., deal with opportunistic data

4. Elastic pricing and quality models associated resources
Multi-dimensional Elasticity

SOCloud 2017
Multi-dimensional Elasticity

- Multi-dimensional elasticity
  - Resources, quality, and costs
- Elasticity in hybrid systems of human-based, things-based and software-based computing resources
  - Software, things and human capabilities as computing resources in multi-clouds
- End-to-end approach
  - the whole system and subsystems
  - Single provider and multiple providers
BUT HOW DO WE CARRY OUT ELASTICITY ENGINEERING?
Elasticity in slices of IoT, Network functions and cloud resources

Application example

IOT Cloud Applications

LightWeighted Analytics and Control

Large-scale Data Analytics

IOT Cloud Systems – the software layer

The edge – IoT units

Sensors → Gateways

Sensor data

The cloud – cloud services

Load Balancer → EventHandling Web Service → Message-oriented Middleware

Sensor data

Near-Realtime Data Processing

NoSQL BigData

„IOT + Network functions + Clouds“
Tasks in Elasticity engineering

- Service Developer
  - Easy to program elasticity requirements

- Service Owner
  - Reduced time to market, easy to reconfigure

- Infrastructure Provider
  - Several owners, developers and providers from different organizations

- Designing and programming software-defined elastic services
- Automatic deployment and configuration
- Elasticity monitoring and analysis
- Elasticity Control

- Service Owner
  - Reduced resources overprovisioning
  - Maintains service’s performance while reducing cost

- Service Developer
  - Easy to understand service’s elasticity boundaries

SOCloud 2017
Fundamental building blocks for the elasticity

- Conceptualizing and modeling elastic objects (and their instances) and execution environments
  - Diverse types of artifacts and their runtime in a similar manner

- Defining and capturing elasticity primitive operations associated with elastic objects and environments

- Recommending and Programming elastic objects
  - A service system can be built from elastic objects

- Runtime deploying, control, and monitoring techniques for elastic objects
Elasticity Detection

Elasticity Trend/Prediction Func

Elasticity Space Func (from when to when?)

Change point detection algs

Trend 1

Trend 2

Alessio Gambi, Daniel Moldovan, Georgiana Copil, Hong Linh Truong, Schahram Dustdar: On estimating actuation delays in elastic computing systems. SEAMS 2013: 33-42
Mapping Services Structures to Elasticity Metrics

Cloud service and elasticity metrics
- Service C1
- Service C3
- Service Group
- Service C2
- Web Service

Runtime application instantiation and runtime monitoring metrics
- Cloud Service
  - Unit 2
  - Unit 1
  - Unit 3
- VM1
- VM2
- VM3
- Virtual Cluster
- Cloud Provider

IO calls
Cost per IO
VM cost per Hour
CPU Usage
Latency
CPU usage
Network IOs
Aggregated
Cost

cost per user
average response time
availability
cost
Multi-level monitoring and analysis

Cloud Service
Service Topology
Service Unit

runsOn

Virtual Cluster
Virtual Machine

Monitoring Snapshot

determinedFrom

determinedBy

Elasticity Boundary

Elasticity Space

Elasticity Pathway

Elasticity Space Function

Elasticity Pathway Function

Note: Pathway=Trend
Several possible functions for determining Elasticity Space and Trend/Prediction

- for different types of service and elasticity behaviors

Scopes in service structure

- Cloud Service System
- Service Topology
- Service Unit
- Service Function

Elastic test frameworks

Common functions/user-defined functions

Apply different elasticity detection and prediction functions

Benchmarks

Machine learning

SOCloud 2017
Analysis detects problems but predefined strategies do not always work!

Changing elasticity specifications at runtime without stoping services
Put things together – Flows

„High-level but complete view“

Hong-Linh Truong, Schahram Dustdar, Frank Leymann, Towards the Realization of Multi-dimensional Elasticity for Distributed Cloud Systems (Submitted version), Cloud Forward Conference 2016, Elsevier Science Procedia Computer Science, 18-20 October 2016, Madrid, Spain
Summary

 Multi-dimensional elasticity
   Key concepts atop IoT, edge systems and clouds

 Elasticity engineering across platforms
   Complex problems need software, things and people in a single system (but composed from multiple subsystems)
   coordinating elasticity across platforms

 End-to-end elasticity toolsets
   Detection, monitoring, analysis and control
   runtime elasticity techniques for dealing with diverse types of services
   There will be no single one
Thanks for your attention

Hong-Linh Truong
Faculty of Informatics, TU Wien

hong-linh.truong@tuwien.ac.at
http://www.infosys.tuwien.ac.at/staff/truong