Extended Cloud Computing (EC$^2$)

Marin Litoiu (York)
John Chinneck, Murray Woodside (Carleton)
Ken Salem (Waterloo)
Core Team

- Marin Litoiu, York University
  - Adaptive and autonomic systems; software design and architecture; cloud computing; performance engineering
- John Chinneck, Carleton University
  - Optimization algorithms in software; feasibility and infeasibility in optimization; applied optimization
- Murray Woodside, Carleton University
  - Performance engineering of software architecture and design; design for high performance; scalability; design optimization
- Ken Salem, University of Waterloo
  - Query optimization; transaction management; data replication and synchronization; virtualization; storage systems; database access control; array databases
- Mike Smit (Post Doctoral Fellow), York University
- Students (TBD)
Extended Cloud: Two Tiers Architecture

- Many clients, many edges, one core
- The client interacts with the closest edge; heavy processing and storage delegated to the core;
Extended Cloud Platform

Platform as a Service (Core and Edge Services)
- Development services
- Deployment services
- Runtime services
- Runtime management services
Extended Cloud Computing

Adaptive Mgmt Framework for Extended Cloud

Optimization for Resource Adaptation

Storage & Data Services in Core/Edge of Extended Cloud

- Framework for Design & Analysis of Adaptive Cloud Mgmt
- Goal Mgmt in Hierarchical Adaptive Systems
- Runtime Supply Chain Mgmt
- Partitioning Computing & Storage Between Edge & Core
- Strategies & Layer Coordination in Cloud Computing
- Forecasting
- Two-Tier Storage Services
- Support for Bi-Directional Data Streams
- Mgmt of Limited Edge Resources
Adaptive Management Framework for Extended Cloud Environment

- Adaptive management is the main focus of this theme
  - Based on feedback loops: monitoring, analysis, planning and execution
- Objective 1: Framework for design and analysis of adaptive cloud management
  - How do we design a cloud feedback loop as a repetitive engineering process
  - How do we synthesize and analyze feedback loops
  - How do we decompose/compose edge and core feedback loops. Centralized vs decentralized control
- Objective 2: Goal management in hierarchical adaptive systems
  - Many goals: profit, user satisfaction, cost, QoS, QoE
  - Goals have sub-goals, and goals change
  - How how design, decompose and manage goals?
- Objective 3: Runtime supply chain management
  - A smart application needs resources at runtime
  - What are the policies that govern the acquisition of resources? How do we express them; What services are needed to broker resources?
A possible EC² feedback loop
Objective 1: Partitioning computing and storage between edge and core
- What can be partitioned and migrated at runtime?
- Static and dynamic optimization algorithms for resource allocation and migration
- Suboptimal algorithms

Objective 2: Strategies and layer coordination in cloud computing
- IaaS, PaaS, SaaS + Edge and Core
- Decomposition of global optimization in layers and tiers
- What data is available at each layer/partition?

Objective 3: Forecasting
- Predictive optimization: based on the forecasted values
- Models for clouds and applications
- Forecasting techniques
Storage and data services in the Extended Cloud

- **Objective 1: Two-Tier Storage Services**
  - storage services on edge and core to support
    - large unstructured objects (files), similar to Amazon S3
    - structured data (records), similar to Cassandra

- **Objective 2: Support for Bi-Directional Data Streams**
  - At runtime data flows in both directions: edge to core and core to edge
    - For sensor, video and audio data
  - Need to define abstractions for data stream storage

- **Objective 3: Management of Limited Edge Resources**
  - Runtime tradeoff analysis of different QoS levels
  - Dynamic negotiation between core and edge
  - Adaptive loops are essential
Extended Cloud Computing (EC²): Two Tiers Architecture
The EC² PaaS provides
- Development, deployment and management services for EC² smart applications

Project 2.1: Adaptive Management Framework for Extended Cloud Environment
- Framework for design and analysis of adaptive cloud management
- Goal management in hierarchical adaptive systems
- Runtime supply chain management

Project 2.2 Optimization for Resource Adaptation
- Partitioning computing and storage between edge and core
- Strategies and layer coordination in cloud computing
- Forecasting

Project 2.3 Storage and data services in the Extended Cloud
- Two-Tier Storage Services
- Support for Bi-Directional Data Streams
- Management of Limited Edge Resources

Interacts with Smart Applications & Smart Edge
SAVI Testbed & Theme Integration Activities

**Experiments in the Lab Year 1**
1. Create isolated apps that use video and mobility & require autonomic behaviour in the cloud & edge
2. Develop resource management for cloud based infrastructure
3. Examine edge opportunities for SAVI cluster, programmability, ASA
4. Theorize potential solutions from wireless/optical access between the apps and the smart edge
5. Build up the testbed theme - Isolate control plane - Identify authorization capabilities - Define SAVI-ized cluster requirements

**Experiments on Small Network Year 2**
- Demonstrate smart apps over virtual cloud infrastructure
- Demonstrate smart apps over virtual smart edge infrastructure (SAVI cluster)
- Demonstrate integrated WiFi/LTE over Fiber
- Interconnect two SAVI clusters

**Multinode Network Experiments Year 3**
- Demonstrate slices of wireless/optical access network
- Extend to proof-of-concept at scale testbed; Provide network slices; Demonstrate smart apps over virtual extended cloud infrastructure (integrated cloud & smart edge)

**Full Testbed Capability Year 4 & 5**
- Demonstrate slices of wireless/optical access network
- Buildout applications and virtual infrastructure to multiple locations
- Campus-wide test with wireless cells & optical backhaul

Demonstrate smart apps over virtual cloud infrastructure

Demonstrate slices of wireless/optical access network

Experiments on Small Network Year 2

Multinode Network Experiments Year 3

Full Testbed Capability Year 4 & 5

Experiments in the Lab Year 1