Review Open Call 3 - “Medium Experiments”

IntelligentNFVAutoscaler

Alexandros Ntitoras

Modio Computing

4th Fed4FIRE+ Engineering Conference

8-10 October 2018, Bruges, Belgium

Outline

- Experiment description.
- Project results.
- Business impact.
- Feedback.

Experiment description (1/3)

CONCEPT AND OBJECTIVES

- **Concept**: Add ‘predictive’ functionality to Kubernetes autoscaler in order to proactively scale-in/out resources.
- **Objectives**:

  Leverage CPU load forecasts for Kubernetes predictive autoscaling.

  Comparison of Kubernetes stock and predictive autoscaling policies with various WebRTC load patterns.

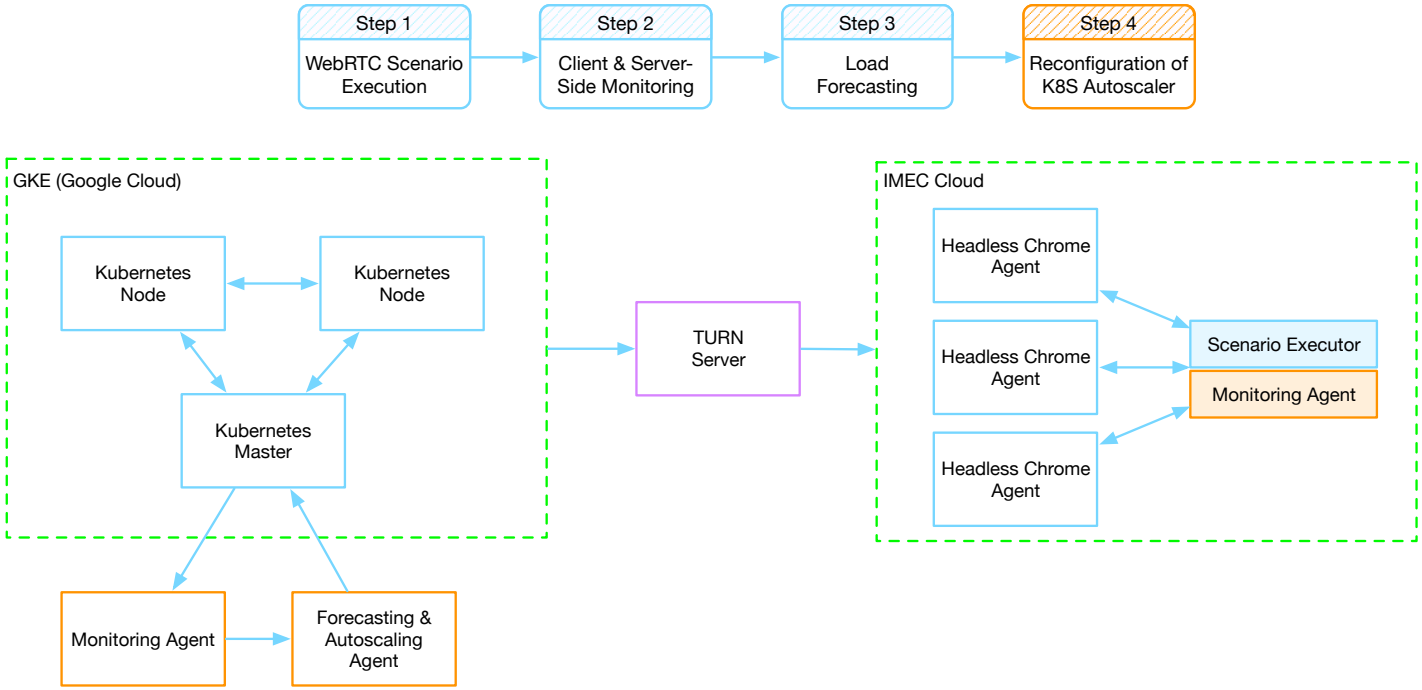
Experiment description (2/3)

BACKGROUND AND MOTIVATION

- **Background**: We have validated that our predictive autoscaling approach outperforms OpenBaton.
- **Motivation**: Market our predictive autoscaling approach for NFV 5G environments.

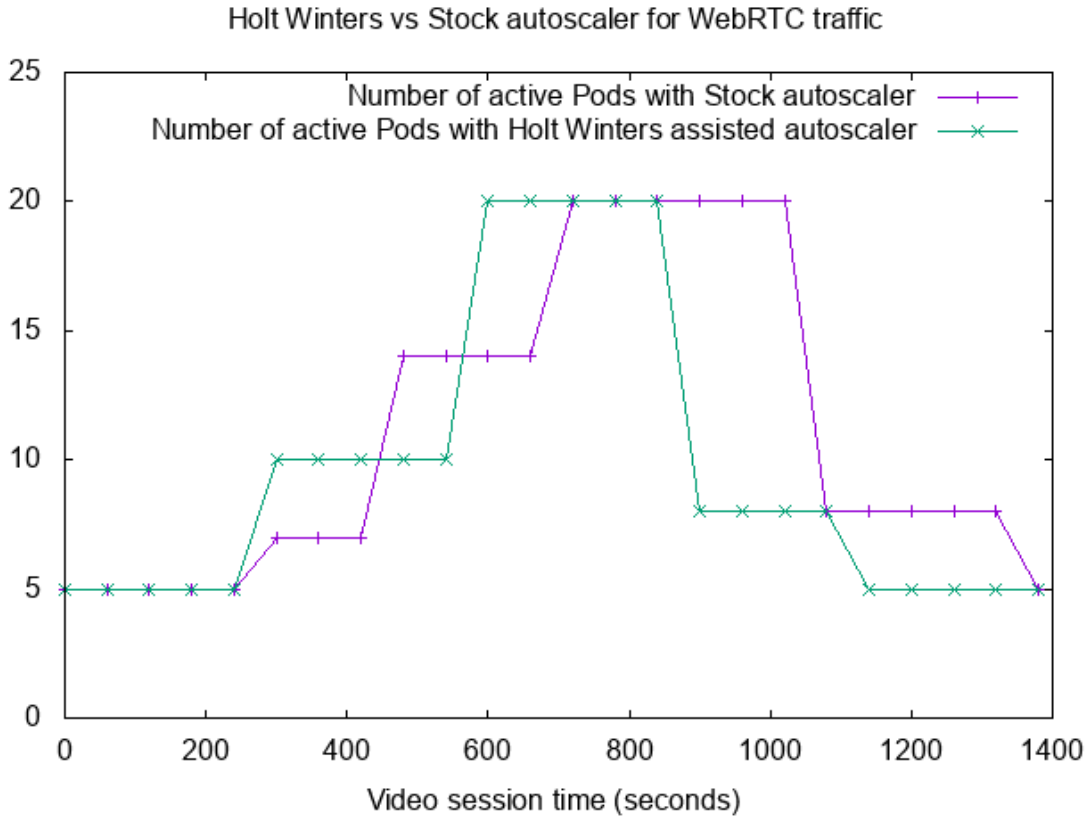
Experiment description (3/3)

EXPERIMENT SET-UP



Project results (1/3)

MEASUREMENTS



Project results (2/3)

MEASUREMENTS

- 20-30% energy saving through proactive shutdown of idle Pods during the ***scale-in*** phase.
- 5-15% QoS improvement in terms of ‘good sessions’ during the ***scale-out*** phase.
- Holt Winters has performed better than i) ARIMA and ii) Recurrent Neural Networks in the majority of experiments.

Project results (3/3)

LESSONS LEARNT

- Predictive autoscaling outperforms default Kubernetes autoscaler.
- Energy saving and QoS improvement of client services are two important dimensions of our Value Proposition and both have been validated over Fed4FIRE+.
- Plenty of room for experimentation with additional services, load patterns and machine learning algorithms.
- Same methodology can be applied to 5G use cases demanding intelligent resource provisioning within an NFV cloud.

Business impact (1/4)

IMPACT ON MODIO'S BUSINESS

- This experiment allowed us to move **from** a small-scale Proof of Concept demonstration over OpenBaton **to** an enhanced implementation and demonstrator for Kubernetes.
- Given that there is no competitor in the market, this project directly helps us in our VC raising activities to develop a Minimum Viable Product that will enter the 5G NFV market.

Business impact (2/4)

HOW FED4FIRE+ HELPED US

- The project's measurable outcomes (energy saving upon scale-in, QoS improvement upon scale-out) have helped us in strengthening our Value Proposition (VP).

- Leveraging the demonstrator and its obtained outcomes, we will soon be contacting VCs to raise backing for developing our MVP.

- The obtained measurements shall be included in our patent for predictive autoscaling for Kubernetes.

Business impact (3/4)

VALUE PERCEIVED

- Our primary goal is not to market our solution as a standalone software product, but through a licensing scheme. Discussions with potential clients are still needed to derive an estimated perceived business value, based on our experiment's outcomes:
    - Energy saving for the operator;
    - Better services offered to the end user.

Business impact (4/4)

WHY WE CAME TO FED4FIRE+

- To compile a demonstrator which we shall use for further VC backing.
- Approval from a recognised H2020 Future Internet project carries weight in the industrial community.
- Prior experience from our WiSHFUL open call project where we worked with the iMinds testbed.

Feedback (1/4)

USED RESOURCES AND TOOLS

- jFed tool and RSpec files for the provisioning of client resources.
- Virtual Wall Virtual Machines running multiple concurrent headless Chrome WebRTC sessions.
- Kubernetes cluster provisioned through Rspec for initial testing of the Kurento Kubernetes deployment and metrics monitoring.

Feedback (2/4)

ADDED VALUE OF FED4FIRE+

- Offered us additional resources, technical, financial support and reduced total experiment cost.
- Automated and repeatable deployments through jFed/RSpec allowed for faster experimentation setup time thus allowed more experiments to be carried out, overall boosting productivity.

Feedback (3/4)

ADDED VALUE OF OUR EXPERIMENT TO FED4FIRE+

- We provide a reference implementation of predictive autoscaling.
- Our core features are not tied to any specific algorithm.
- Our autoscaling agent for Kubernetes can reused by consequent experimenters wishing to explore predictive NFV resource management.

Feedback (4/4)

TESTBED ISSUES DURING EXPERIMENTATION

- Complexities related to IPv6 & ICE support in Kubernetes led to the provisioning of an external STUN/TURN server and Kubernetes cluster.


Co-funded by the European Union


Co-funded by the Swiss Confederation

This project has received funding from the European Union's Horizon 2020 research and innovation programme, which is co-funded by the European Commission and the Swiss State Secretariat for Education, Research and Innovation, under grant agreement No 732638.

WWW.FED4FIRE.EU