Future Generation Computer Systems ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

A computational model to support in-network data analysis in federated ecosystems

Ali Reza Zamani [a], Mengsong Zou [a], Javier Diaz-Montes [a,*], Ioan Petri [b], Omer Rana [b], Manish Parashar [a]

[a] *Rutgers Discovery Informatics Institute, Rutgers University, USA*
[b] *School of Computer Science & Informatics, Cardiff University, UK*

HIGHLIGHTS

- We present a model that leverages software-defined networks to opportunistically exploit the latent computational capabilities located along the data path.
- Our model is able to use alternative methods of computation when our primary method cannot be used due to SLA constraints. In this paper we use a neuronal network approximation model as alternative to EnergyPlus.
- The description of two algorithms that introduces how the neuronal network model is trained and how neuronal network jobs are deployed across in-transit resources.
- A new set of experiments to validate and evaluate the effect that our new strategy has in the energy optimization of smart buildings.

ARTICLE INFO

Article history:
Received 1 July 2016
Received in revised form 8 February 2017
Accepted 21 May 2017
Available online xxxx

Keywords:
Software-defined networks
In-transit
Smart buildings
Cloud federation
CometCloud

ABSTRACT

Software-defined networks (SDNs) have proven to be an efficacious tool for undertaking complex data analysis and manipulation within data intensive applications. SDN technology allows us to separate the data path from the control path, enabling in-network processing capabilities to be supported as data is migrated across the network. We propose to leverage software-defined networking (SDN) to gain control over the data transport service with the purpose of dynamically establishing data routes such that we can opportunistically exploit the latent computational capabilities located along the network path. This strategy allows us to minimize waiting times at the destination data center and to cope with spikes in demand for computational capability. We validate our approach using a smart building application in a multi-cloud infrastructure. Results show how the in-transit processing strategy increases the computational capabilities of the infrastructure and influences the percentage of job completion without significantly impacting costs and overheads.

1. Introduction

There has been recent interest in moving away from centralized, large-scale data centers to a more distributed multi-cloud setting (as demonstrated by significant interest in cloud federation and interoperability efforts). Such a multi-cloud environment is often formed by a network of smaller virtualized infrastructure runtime nodes with an unstructured architecture. On the other hand, network providers are increasingly becoming potential sources of general purpose computation. They are minimizing the amount of network-specialized hardware hosted in their data centers and moving towards the use of commodity hardware. This strategy follows state of the art networking approaches, such as Software-defined networking (SDN) and Network Functions Virtualization (NFV). Software-defined networking (SDN) in particular is an approach devised to simplify network management through abstraction of lower-level functionality. Specifically, SDN separates control plane (where to send data) from data plane (data forwarding functions). This enables the software-based control plane to be run on commodity servers and to leverage the latest-generation of processors, which are faster than embedded-class processors in most switches [1]. On the other hand, NFV goes a step further and extends the as-a-service cloud model to offer networking functions on-demand using virtualization techniques. The key reason for using virtual machines (VMs) is the possibility of elastically scaling functions by simply adding or removing VMs

---

* Corresponding author.
E-mail address: javier.diazmontes@gmail.com (J. Diaz-Montes).

http://dx.doi.org/10.1016/j.future.2017.05.032
0167-739X/© 2017 Elsevier B.V. All rights reserved.

Please cite this article in press as: A.R. Zamani, et al., A computational model to support in-network data analysis in federated ecosystems, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.05.032.

based on data workload characteristics. This approach promises, as the cloud, a reduction in capital expenses and fast delivery of new functionality. As in the case of SDN, this approach is also implemented on commodity hardware [2].

Data centers managed and operated by network providers form a significant part of the current Internet infrastructure, as there is a large number of such data centers that are almost ubiquitous across the world. These data centers may not be as powerful as computational data centers, hosted by cloud providers or traditional high performance computing (HPC) providers. However, their ubiquity and the fact that we have to necessarily use them when moving data over the Internet, make them a very interesting source of pervasive computing at the edge of the network. Understanding how the availability of commodity servers within such “network data centers” can contribute towards data processing would enable an effective way to extend the boundaries of a cloud system — from a high end, often localized data center, to multiple distributed data centers that can process data while it is in transit from source to destination.

In this paper we propose a model to leverage the use of computational capabilities within such network data centers to offer general computation services co-located with network services. In this way, we can use more efficiently the resources of these network data centers while providing an extra source of revenue for those who operate and manage them. Hence, spare capacity within such network data centers can be more efficiently utilized and monetized. We extend the network controller capabilities to not only offer information about the network topology but also to identify sources of computation. We envision an ensemble of network data centers that can optimize the data routes based on flows and offer *in-transit* computational capabilities.

The rest of the paper is organized as follows. Section 2 presents our motivating use case. Section 3 presents our in-transit computational model, followed by the proposed in-transit optimization strategies in Section 4. Section 5 defines the problem of allocating workload using our computational model. Section 6 presents the implementation of our in-transit computational model. Section 7 describes evaluation and results. Section 8 collects the related work. Finally, Section 9 presents the conclusions and ongoing activities.

2. Motivating use case

An instrumented built environment, which can consist of single/multiple buildings (homes, office buildings, sports facilities, etc.), provides a useful scenario to validate the use of in-transit analysis capabilities. Depending on the number of sensors within a single building, the frequency at which data is captured from such sensors and the particular data analysis objective (e.g. reduce energy consumption, improve efficiency of HVAC (heating, ventilation and air condition) function, improve comfort levels based on occupancy, etc.), the computational capability requirements can vary significantly. In some instances such data is often analyzed off-line (in batch mode) to enable improvements in building design or to support long term facilities management. In other instances (evidenced by recent use of such instrumented environments), real time analysis needs to be carried out (over intervals of 15–30 min generally) to enable better energy efficiency and use of such infrastructure. When multiple such buildings are considered (e.g. within a business park, University campus or a housing association), the overall computational requirement can increase considerably.

In order to maintain a comfortable living environment, it is often necessary to consider multiple objectives that may have conflicting targets, e.g. minimum energy consumption, minimum CO_2 emission, or maximum comfort level. Optimization for the building operation stage (which can also include facilities management) requires different approach compared to the building design stage, e.g., some key design variables can no longer be changed (to find the most optimum solutions for design). It needs to take the as-built building environment to find the optimum solutions either against single or multi-objectives. To provide practical real time decision making in building energy management based on real time monitored data, it is necessary to develop a ’behavior’ of a building energy system by using various simulation tools. During the process, domain experts are often involved in order to identify the main use cases and scenarios with associated input parameters and feasible outputs. In the modeling process, different relevant components have to be assessed and calibrated iteratively, and the developed building energy simulation model is then executed (as the calculation engine) within a generic optimization program. In this work we execute multiple EnergyPlus[1] instances, a software that requires significant computational resources to run, with different input parameter ranges. (See Fig. 1.)

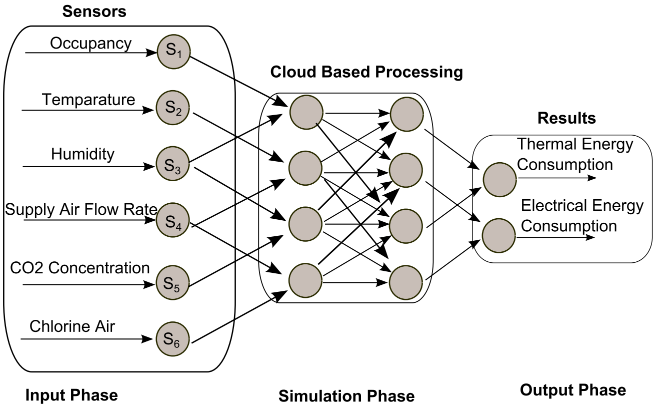

Fig. 1. Energy optimization scenario.

Various types of sensors are used to monitor energy efficiency levels within a building, such as: (i) solid-state meters for accurate usage levels, (ii) environmental sensors for measuring temperature, relative humidity (RH), carbon monoxide (CO), and carbon dioxide (CO_2) levels, (iii) temperature measurements using both mechanical (e.g., thermally expanding metallic coils) and electrical means (e.g., thermistors, metallic resistance temperature detectors (RTD), thermocouples, digital P-n junctions, infrared thermocouples) to provide sufficient accuracy. When dealing with large buildings such as sports facilities, the accuracy of these sensors is often questionable, largely because of the significant drift that occurs after initial calibration. In some buildings, there are specific requirements for sensors when monitoring CO_2 concentration, air flow, humidity, etc. and these sensors are more expensive to use and deploy.

We use sensor data from the *SportE*^{2} project pilot called FIDIA [3], a public sports building facility, located in Rome, Italy. *SportE*^{2} is a research project co-financed by the European Commission FP7 programme under the domain of Information Communication Technologies and Energy Efficient Buildings. This project focuses on developing energy efficient products and services dedicated to needs and unique characteristics of sporting facilities. The building has metering capability to determine consumption of electricity, gas, biomass, water and thermal energy. This data can be accessed through a specialist interface and recorded for analysis. The sub-metering of thermal and electrical consumption within grouped zones (gym/fitness and swimming pool is also provided along with “comfort” monitoring by functional area: gym, fitness

---

[1] http://apps1.eere.energy.gov/buildings/energyplus/.

Please cite this article in press as: A.R. Zamani, et al., A computational model to support in-network data analysis in federated ecosystems, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.05.032.