Γενικες Οδηγιες Αναληψης Θεματος Διπλωματικης Εργασιας

Κάθε πρόταση θέματος διπλωματικής εργασίας, για ανάληψη από φοιτητή του ΠΜΣ «ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ», που κατατίθεται για έγκριση από τη Γ.Σ. του τμήματος ΗΜΜΥ του ΕΛ.ΜΕ.ΠΑ, θα πρέπει να συμμορφώνεται με τον ακόλουθο οδηγό παρουσίασης και ανάλυσης υποψήφιου θέματος.

Γενικές Οδηγίες

1. Κάθε υποψήφιο θέμα θα πρέπει να καλύπτει επαρκώς τις ακόλουθες ενότητες:

    A) Διατύπωση του προβλήματος με αναφορά στις state-of-the-art τεχνολογίες

    B) Τρόπος αντιμετώπισης του προβλήματος και επιλογή εργαλείων ανάπτυξης

    C) Αναμενόμενα αποτελέσματα

2. Η μέγιστη έκταση κάθε υποψήφιου θέματος δε θα πρέπει να υπερβαίνει τις 3 σελίδες

Ακολούθως παρατίθενται δύο ενδεικτικά υποδείγματα παρουσίασης υποψήφιων θεμάτων.

HELLENIC MEDITERRANEAN UNIVERSITY

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

SURNAME XXXXXXXXXXXXXXXXXX	*NAME* XXXXXXXXXXXXX	*STUDENT NO.* XXX
PROGRAM OF STUDIES **INFORMATICS ENGINEERING**	*ACADEMIC UNIT / DEPARTMENT* **SCHOOL OF ENGINEERING/DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING**	

FINAL THESIS PROPOSAL FORM

PROPOSED THESIS TITLE

LoRaWAN device modelling based on NETCONF and YANG

FULL NAME OF THESIS SUPERVISOR

XXXXXXXXXXXXXXXXXXXXXXXXX

FULL NAME AND AFFILIATION OF THESIS EXAMINER A (if available)	*FULL NAME AND AFFILIATION OF THESIS EXAMINER B (if available)*
XXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXXXX

PROPOSED THESIS EXTENDED ABSTRACT (to be continued on page 2)

The large number of networking hardware vendors and their device-specific CLI scripts, makes the configuration of even the smallest network a colossal task. This has resulted in service providers and network operators to start adopting SDN principles like model-driven management [1]. Software Defined Networks, originally named OpenFlow networks, surfaced in 2008 at the University of California, Berkley, and Stanford University. The main idea was to enable academics to experiment on various vendor-specific hardware in a uniform manner. This led to the separation of network control and packet forwarding, what is known as, the control plane and the data plane respectively [2].

Generally, the control plane will consist of network applications, the controller platform and controller modules which interact with each other through the Northbound Interface. The data plane is made up of the data forwarding devices, for example, switches, routers etc. The interface between the data plane and the control plane is called the Southbound Interface. Southbound APIs and protocol plugins define the set of instructions that the forwarding devices utilise, and the protocol in which the control and data plane communicate with [3][4]. Some well-known APIs and protocol plugins include OpenFlow, Cisco's OpFlex, OVSDB, ForCES, SNMP, BGP and others. To the previous list, we can also add a configuration protocol called NETCONF.

The NETwork CONFiguration (NETCONF) protocol is used to manage, to retrieve, upload and manipulate configuration data of network devices. Compatible devices can be accessed through a straightforward API, enabling the transceiving of configuration data sets. NETCONF encodes its Remote Procedure Calls (RPCs) in a standardised XML schema over securely connected sessions. This means that

NETCONF does not necessarily need a controller, so a Northbound Interface high-level programming language can directly interact with low-level network devices using the NETCONF protocol. NETCONF can be divided into four abstract layers. The Secure Transport Layer, which is responsible for providing a secure connection route between the server and the client, the Message Layer, which is responsible for encoding the RPCs and the notifications in simple XML frames, the Operations Layer, a basic set of protocol operations called as RPCs in XML encoded format and the Content Layer. The latter two layers are covered by data models and protocol operations which are stated explicitly in a data modelling language called YANG[5].

YANG is used by the NETCONF protocol, RPCs and notifications to model its configuration data and its state data. YANG is not only used by the NETCONF protocol, but can be used by other protocols as well, like RESTCONF, CoAP, etc. YANG can use other encodings other than XML, for example in IETFs RFC7951 it is proposed that YANG encodes in JSON. YANG models data as trees in a hierarchically manner, where nodes have names and they have values or child nodes, a clear and short description of the nodes and interactions between nodes[6].

An indispensable part of networks is the Internet of Things. In 2017 the total number of active device connections was 16.4 billion with 36 percent of them being IoT devices. It is estimated that in 2025 the total number of active device connections will be 34.2 billion with 63 percent of them being IoT devices[7]. According to the European Research Cluster on the Internet of Things, IoT is a "… network infrastructure ... based on standard and interoperable communication protocols …"[8]. Network infrastructures can be Wireless Local Area Networks (WLAN), wired networks, cellular networks, Low-Power Wide Area Networks (LPWAN) and others.

Low-Power Wide Area Networks are characterised for their low energy consumption, their long-range connectivity and their low cost of deployment, and mainly designed for sensor networks powered by batteries. The low energy consumption is achieved because of low data rates reaching up to a few dozens of kbits per second, a few times a day, allowing the sensor node in being placed in sleep mode for most of its operational time. LPWANS are expected to grow from 25 million devices to 2 billion devices by 2025[7], with LoRaWANs being the market leader.

The technology behind the LoRaWAN communication protocol is its wireless, long-range communication link called LoRa. The LoRa physical layer is based on Chirp Spread Spectrum Modulation, like Frequency Shifting Key modulation in terms of power consumption, but with longer communication range. Longer communication range combined with the star network topology used by LoRaWANs, means that a single gateway can cover hundreds of square kilometres while preserving the battery life of the nodes.

In this thesis, a NETCONF/YANG module for managing LoRa based devices will be designed. Previous work on this subject has been presented as Internet-Drafts of the IETF[9][10]. The drafts expired in 2016 and were not approved as RFCs. Since then new LoRaWAN specifications have been announced by the LoRa Alliance and to the best of my knowledge updated versions of YANG modules, for LoRa devices, have not been designed. Additionally, to the above work, the designed modules will be implemented and tested on a LoRaWAN based network.

References

[1] J. Medved, R. Varga, A. Tkacik, and K. Gray, "OpenDaylight: Towards a Model-Driven SDN Controller architecture," in *Proceeding of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks 2014*, 2014, pp. 1–6.

[2] N. McKeown *et al.*, "OpenFlow," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 2, p. 69, Mar. 2008.

[3] D. Kreutz, F. M. V. Ramos, P. Esteves Verissimo, C. Esteve Rothenberg, S. Azodolmolky, and S.

Uhlig, "Software-Defined Networking: A Comprehensive Survey," *Proc. IEEE*, vol. 103, no. 1, pp. 14–76, Jan. 2015.

[4] M. Jarschel, T. Zinner, T. Hossfeld, P. Tran-Gia, and W. Kellerer, "Interfaces, attributes, and use cases: A compass for SDN," *IEEE Commun. Mag.*, vol. 52, no. 6, pp. 210–217, Jun. 2014.

[5] R. Enns, M. Bjorklund, J. Schoenwaelder, and A. Bierman, "Network Configuration Protocol (NETCONF) RFC 6241," *IETF*, 2011.

[6] M. Bjorklund, "YANG A Data Modelling Language for the Network Configuration Protocol (NETCONF) RFC 6020," *IETF*, 2010.

[7] K. L. Lueth, "State of the IoT 2018: Number of IoT devices now at 7B – Market accelerating," *iot-analytics.com*, 2018. [Online]. Available: https://iot-analytics.com/state-of-the-iot-update-q1-q2-2018-number-of-iot-devices-now-7b/.

[8] D. Gordon, J.-H. Hanne, M. Berchtold, A. A. N. Shirehjini, and M. Beigl, "Towards Collaborative Group Activity Recognition Using Mobile Devices," *Mob. Networks Appl.*, vol. 18, no. 3, pp. 326–340, Jun. 2013.

[9] A. Pelov, L. Toutain, Y. Delibie, and A. Minaburo, "draft-pelov-yang-lora-01," *IETF*, pp. 1–8, 2016.

[10] A. Minaburo and T. Bretagne, "draft-minaburo-6lpwa-yang-lora-01," *IETF*, pp. 1–7, 2016.

SIGNATURE (STUDENT)

SIGNATURE (THESIS SUPERVISOR)

DATE

SIGNATURE (ACADEMIC UNIT)

HELLENIC MEDITERRANEAN UNIVERSITY

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

SURNAME XXXXXXXXXXXXXXXXXX	NAME XXXXXXXXXXXXX	STUDENT NO. XXX
PROGRAM OF STUDIES **INFORMATICS ENGINEERING**	ACADEMIC UNIT / DEPARTMENT **SCHOOL OF ENGINEERING/DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING**	

FINAL THESIS PROPOSAL FORM

PROPOSED THESIS TITLE **Private cloud architectures and technologies for biomedical applications**	
FULL NAME OF THESIS SUPERVISOR XXXXXXXXXXXXXXXXXXXXXXXXX	
FULL NAME AND AFFILIATION OF THESIS EXAMINER A (if available) XXXXXXXXXXXXXXXXXXX	FULL NAME AND AFFILIATION OF THESIS EXAMINER B (if available) XXXXXXXXXXXXXXXXXXXXXXX

PROPOSED THESIS EXTENDED ABSTRACT (to be continued on page 2)

The last few years, cloud technology has matured in such an extent that has rapidly moved from the research laboratories to consumer products and commodity services. Cloud technology now offers hardware and software as on-demand services and consumers can benefit from services such as data storage and computational power, in a scalable and user-tailored model.

Cloud technology is particularly attractive due to the sharing of data at low cost capabilities for application domains with high requirements for storage and/or computational resources and/or requiring the aggregation of large amounts of intrinsically distributed heterogeneous data within complex workflows. Pioneered by Amazon's IaaS and EC213, Cloud Computing has become attractive for both enterprise and research. Most current available Cloud Computing facilities, particularly commercial public Cloud Computing (e.g. Amazon EC2, S314), are built on dedicated machines within the data center. These services show the benefits of Cloud Computing in the aspects of elastic, dynamic, unlimited storage and computation capacity.

Applications in the biomedical domain, such as clinical trial management, genome sequencing, pathway analysis, clinical imaging, as well as tumor, organ or functional modeling store and process huge amounts of data and are in need of such services, so they could greatly benefit of this technology uplift. However, concerns arise when considering employing them into scientific research, including security, ethical and legal issues, regulation and performance.

Private (community) clouds become very useful for biomedical scientific research since they allow for full control on data access, reliability and storage management. Rather than making investment on massive dedicated machines, a private cloud provides a more cost-effective solution by using existing heterogeneous computing facilities within an organization. Such a private cloud infrastructure harvests the underused computing resources within an organization, thus saving the cost of purchasing new servers and at the same time adhering to the ethico-legal requirements of the biomedical domain.

As a result, it is believed that, the nature of the application domain (i.e. biomedicine) demands a combination of public cloud and private cloud services where some critical data resides (or is processed) in an enterprise's private cloud while other data is stored and is accessible from a public cloud storage provider. With such approaches, i.e. hybrid cloud storage and processing, enterprises are able to mix and match cloud resources between local data center infrastructure and scalable, on-demand infrastructure. Hybrid cloud storage combines the advantages of scalability, reliability, rapid deployment and potential cost savings of public cloud storage with the security and full control of private cloud storage.

In this thesis we will analyze the special requirements of the current applications and research domains in the biomedical domain – such as the virtual physiological human and the Digital Patient initiatives. In parallel we will evaluate and compare existing cloud technologies and architectures with respect to their strengths, weaknesses and trade-offs. We will also focus on the implementation of a private (or hybrid) cloud architecture that aims to deal with the requirements of the biomedical domain, on selected use cases. Finally, we will evaluate the suitability, efficacy and competency of the proposed architecture in selected use cases.

In specific, we will analyze the state-of-art technologies in cloud computing and provide a private community cloud by harvesting existing IT facilities. The deployment of this private cloud will mainly use open source platforms, for example, Eucalyptus for low-level middleware and service management, and Hadoop for distributing computation across all of the computing nodes. Eucalyptus is an open source cloud computing and Infrastructure-as-a-Service (IaaS) platform for enabling private clouds. Eucalyptus takes a company’s existing IT infrastructure and adds a virtualization software layer with scalable, secure Web services that work together to create elastic pools with the ability to dynamically scale depending on the company’s needs. Eucalyptus also offers an Infrastructure-as-a-Service product in partnership with Amazon that delivers full compatibility with Amazon’s Web Services APIs. This makes it possible to efficiently extend a Eucalyptus private cloud to a hybrid cloud environment. The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer, so delivering a highly available service on top of a cluster of computers, each of which may be prone to failures.

During deployment and service development, we will particularly focus on distributed processing of biomedical data in the cloud. Specific data access and management as well as computation-oriented use cases will be defined. Indicative use cases are a) researchers who need small amounts of computing for long periods of time, for example, a bioinformatician may need to subscribe for a year or more to host a genomics tool to analyze specific cause-and-effect disease relationships; or b) a modeling scientist may require for large amounts of computing over a short period of time, i.e., burst computing. Such use cases will be taken into account when designing resource allocation (e.g. storage, compute power) management.

As a result, in the context of this thesis, we plan to a) Investigate how current cloud capabilities need to be enhanced to be used in the biomedical domain which has strict requirements for reliability, availability, performance and security; b) Select the most suitable from the variety of cloud offerings that are available in order to be used in the biomedical domain; c) develop a private cloud infrastructure – mainly using open source platforms; d) Investigate the extend, complexity and benefits of using private (community) clouds in the biomedical domain. Finally, we will also attempt to address the related legal issues in such a private cloud computing environment.

References

1. http://www.eucalyptus.com/

2. Vadim Truksha, Cloud platform comparison: CloudStack, Eucalyptus, vCloud Director and OpenStack, IT

infrastructure architect, Altoros Systems Inc., special to Network World

July 16, 2012, available at: http://www.networkworld.com/news/tech/2012/071612-cloud-platformcomparison-260923.html

3. Daniel Nurmi, et al.,The Eucalyptus Open-source Cloud-computing System, 9th IEEE/ACM International

Symposium on Cluster Computing and the Grid, 2009. Available at:

https://dspace.ist.utl.pt/bitstream/2295/584876/1/CCGrid2009_Eucalyptu

4. Doug Cutting, Hadoop and Big Data, Available at: http://www.cloudera.com/content/cloudera/en/whycloudera/hadoop-and-big-data.html

5. Haddop Overview, Available at: http://wiki.apache.org/hadoop/ProjectDescription

6. Use case study of Hive/Haddop, http://www.slideshare.net/evamtse/hive-user-group-presentation-fromnetflix-3182010-3483386 (accessed 26 May 2013)

SIGNATURE (STUDENT)

SIGNATURE (THESIS SUPERVISOR)

DATE

SIGNATURE (ACADEMIC UNIT)