Multi-controller Scalability in Multi-domain Content Aware Networks Management

Eugen Borcoci, Mihai Constantinescu, Marius Vochin
University POLITEHNICA of Bucharest
Bucharest, Romania
emails: eugen.borcoci@elcom.pub.ro
mihai.constantinescu@elcom.pub.ro, marius.vochin@elcom.pub.ro

***Abstract* — This paper studies the management system scalability properties of a networked media eco-system. The system offers guaranteed Quality of Services (QoS) multimedia delivery, over multiple domain networks, based on creation of data plane slices named Virtual Content Aware Networks (VCAN). The VCANs are realized under control of a management plane, by centralized per network domain "controllers" - cooperating in order to span the VCANs over multiple IP domains. Scalability is an issue- similar to the multi-controller problem, in emerging Software Defined Networking (SDN) technologies. The management system architecture considered in this paper has been previously defined. This work provides a simulation model and results concerning the multi-controller communications. It is shown that SDN-like control approach is conveniently feasible in a multi-domain environment.**

Keywords* — *Content-Aware Networking; Software Defined Networking, Multi-domain; Management; Resource provisioning; Future Internet.

I. INTRODUCTION

The architectural solutions for the Future Internet constitute hot research topics today. It is recognized that traditional Internet has fundamental architectural limitations, and also ossification if compared to the current needs and considering its global extension [1], [2].

A significant trend in Internet is the information/content-centric orientation; consequently, significant changes in services and networking have been recently proposed, including modifications of the basic architectural principles. The revolutionary approaches are often referred to as *Information/Content-Centric Networking (ICN/CCN)*, [3], [4]. In parallel, evolutionary (or incremental) solution emerged, as Content-Awareness at Network layer (CAN) and Network-Awareness at Applications layers (NAA). This approach can create a powerful cross-layer optimisation loop between the transport and applications and services.

A new trend, targeting to achieve more flexibility in networking is *Software Defined Networking* (SDN) architecture and its associate OpenFlow protocol [5], [6], [7] where the control plane and data planes are decoupled and the network intelligence is more centralized, thus offering a better and flexible/programmable control of the resources.

The European FP7 ICT research project, *"Media Ecosystem Deployment Through Ubiquitous Content-Aware Network Environments"*, ALICANTE, [12], adopted the NAA/CAN approach, to define, design, and implement a Media Ecosystem spanning multiple network domains. This work considers as input the ALICANTE management architecture [14], which is similar to SDN with respect to the distribution of the main management and control functions among several controllers. Communication between controllers is necessary in order to accomplish multi-domain tasks.

This paper studies by simulation, based on Extended Finite State Machines (EFSM) model, some scalability aspects of the signalling protocol for multi-controller communication.

Section II presents some related work. Section III describes the management architecture. Section IV defines the inter-controller communication. Section V introduces the simulation model and Section VI describes the simulation results. Conclusion, open issues, and future work are shortly outlined in the Section VII.

II. RELATED WORK

The ICN approaches are very promising, but raise some research and deployment challenges like the degree of preservation of the classic transport (TCP/IP) layering principles, naming and addressing, content-based routing and forwarding, management and control framework, in-network caching, etc.

In SDN, the network intelligence is more centralized, thus offering a better and flexible control of the resources, quality of services, etc., due to the possibility to have an overall image of the system in the control plane and by allowing programmability of the network resources. The operators will get more freedom and speed in developing their services, without waiting long time for new releases of vendors' networking equipment. Although it seems to be very attractive, e.g., for data centers but not only, SDN exposes also many research challenges and open issues, both from architectural and from deployment point of view. Degree of centralization and relationship with scalability and reliability are examples. An extension of the SDN concepts is proposed in so-called *Software Defined (Internet) Architecture* [9], where the idea is also to decouple the architecture from infrastructure as to lower the barriers to architectural evolution. The SDIA approach tries to exploit SDN concepts but also traditional technologies (e.g., MPLS, software forwarding, etc.) in order to obtain evolvable architectures. SDN and SDIA are still evolutionary in contrast with "clean slate" ones, which are disruptive.

Currently there are concerns about SDN performance, scalability, and resiliency [8], [11], the main source for these problems being the centralization concept. It is clear that a central controller will have a limited processing capacity and the solution will not scale as the network grows (increase the number of switches, flows, bandwidth, etc.). Apart from the obvious solution to increase the controller performance, the second idea is to define a SDN multi-controller architecture. However, SDN still wants a consistent centralized logical view upon the network; this creates the need for controllers to cooperate and synchronize their data bases in order to provide together a consistent view at network level. Work in progress is developed at IETF towards constructing an inter-controller, [12]. While the vertical protocols between Control and Data Plane have seen significant progress by specifying Open Flow versions, [7] and implementing several types of controllers, [6], [10], the inter-controller cooperation and scalability issues are still open research issue.

ALICANTE architecture has considered, from the beginning, the scalability requirements in its management and control specification. These aspects will be more developed in the next section.

III. MANAGEMENT SYSTEM ARCHITECTURE

In ALICANTE architecture, [12], [13], [14], several cooperating environments are defined containing business actors: User Environment (UE), containing the End-Users; Service Environment (SE), containing SPs and Content Providers (CP); Network Environment (NE), where we find the new CAN Provider and the traditional Network Providers (NP) managing the network elements, in the traditional way at IP level. The "environment", is a generic grouping of functions working for a common goal and which possibly vertically span one or more several architectural (sub-) layers. Between actors, dynamic Service Level Agreements (SLA) can be established. A novel business entity, CAN Provider was defined, to which several SPs can independently ask customizable Virtual Content Aware networks, and then use them. Network Providers can cooperate to VCAN construction but preserving their independency in resource allocation. Flexible connectivity services have been achieved offering: Fully/partially/un-managed services.

The Internet is sliced by creation on demand logically isolated VCANs, realised as parallel logical planes based on light virtualisation (in the Data Plane only), and optimising inter and intra-domain mapping of VCANs, onto several domain network resources.

The architecture supports both V/H integration SLAs for several level of guarantees. Distributed M&C (each domain has its Intra-NRM and an associated CAN Manager) assures large scale provisioning. The VCANs are flexible supporting: unicast, multicast, broadcast, P2P and combinations with different levels of QoS/QoE, availability, etc. End Users and their Home-Boxes can ultimately benefit from CAN/NAA features by using VCANs.

Services Providers only ask VCANs and use them; they are not burdened with tasks to construct them. The architecture assures QoS/QoE optimization based on: CAN/NAA interaction; Cooperation between resource provisioning (SLA) and media flow adaptation; Hierarchical monitoring at CAN and network layers cooperating with the upper layers.

The ALICANTE architecture is conceptually similar to recently proposed SDN, although not following full SDN specifications (Fig. 1).

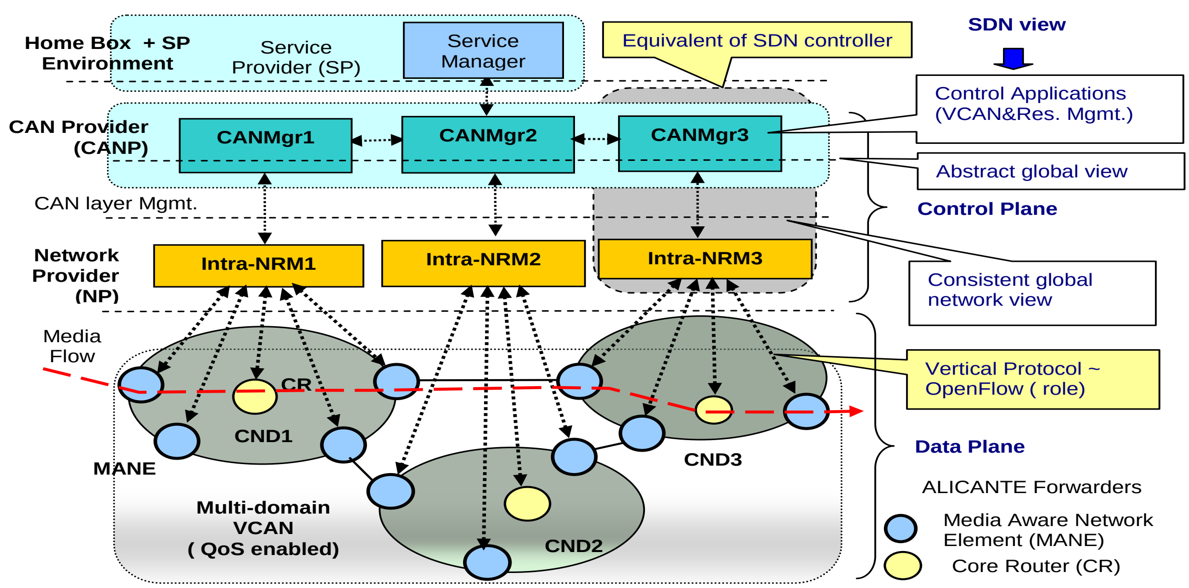

Figure 1. ALICANTE partially centralized management architecture and equivalence with SDN

Notations: SP – Service Provider; CANP - CAN Provider; NP – Network Provider; CND - Core Network Domain; CANMgr - Content Aware Network Manager; Intra-NRM – Intra-domain Network Resource Manager;

Both architectures are evolutionary and can be seamlessly developed. The Control Plane and Data Plane are separated. Note that Control Plane in SDN terminology is here actually Management and Control Plane. The QoS constrained routing, resource allocation, admission control and VCAN mapping are included in the CAN Manager. The "virtualization" of the network is performed by Intra-NRM, which hides the characteristics of MPLS technology by delivering to the CAN Managers an image of abstract matrix of connectivity logical pipes.

In SDN the network intelligence is (logically) centralized in SW -based SDN controllers, which maintain a global view of the network: maintain, control and program Data Plane state from a central entity. In our case [*CAN Manager + Intra-domain Network Resource Manager*] play together the role of an SDN controller for a network domain, controlling the MANE edge routers and interior core routers. Actually, we have a multi-domain logical network governed by several "SDN controllers" – which cooperates for resource management and routing. However, the degree of centralization is configurable in ALICANTE by defining the placement of CAN Managers and the sets of routers to be controlled.

In both SDN and this architecture, the Control Plane SW is executed on general purpose HW. The decoupling of the control with respect to specific networking HW is realized**:** the MANE and core routers are viewed by the upper CAN layer in abstract way.

Data Plane is programmable: all configurations for MANE and Core routers are determined in CAN and Network M&C and downloaded in the routers. ALICANTE architecture defines the control for a whole network (and not for single network devices): at CAN Manager level there exists an overall image on the static and dynamic characteristics of all VCANs; at Intra-NRM level there is a full control on the network domain associated with that Intra-NRM.

In SDN and our case also, the network appears to the applications and policy engines as a single logical switch. In our case, the network appears at higher layers as a set of parallel planes VCANs. This simplified network abstraction can be efficiently programmed, given that the VCANs are seen at abstract way; they can be planned and provisioned independently of the network technology.

IV. INTER-DOMAIN MANAGEMENT COMMUNICATIONS

One CANMgr (belonging to CANP) is the initiator of VCAN construction, at request of an SP. The VCANs asked should be mapped onto real multi-domain network topology, while respecting some QoS constraints. This provisioning is done through negotiations [14], performed between CAN Managers associated to each network domain. If necessary, the initiator communicates with other CANMgrs, to finally agree a reservation and then a real allocation (i.e., installation in the network routers) of network resources necessary for a VCAN. A CAN Planning entity inside each CANMGr runs a combined algorithm doing QoS constrained routing, VCAN mapping and resource logical reservation. In this set of actions, it is supposed that the initiator CANMgr knows the inter-domain topology at an overlay level and also a summary of each network domain topology, in terms of abstract trunks (e.g. *{ingress, egress, bandwidth, QoS class, ..}*). This knowledge is delivered by an additional discovery. service. Previous papers, of the [13], developed and implemented the combined VCAN mapping algorithm.

The overall system flexibility and scalability essentially depends on its Management and Control. For VCAN planning, provisioning and exploitation: it was adopted per-domain partially centralized solution; this avoids full-centralized VCAN management (non-scalable), but allowing a coherent per-domain management. However, the initiator CAN Manager, like in SDN approach, has the overall consistent image of a multi-domain VCAN.

There is *no per-flow signaling* between CAN Managers. The VCAN SP–CANP negotiation is performed per VCAN, described in terms of aggregated traffic trunks. The SP negotiates its VCAN(s) with a single CAN Manager irrespective, if it wants a single or a multi-domain spanned VCAN.

A hierarchical overlay solution is applied for inter-domain peering and routing where each CAN Manager knows its inter-domain connections. The CAN Manager initiating a multi-domain VCAN is the coordinator of this hierarchy, without having to know details on each domain VCAN resources. The monitoring at CAN layer and network layer is performed at an aggregated level.

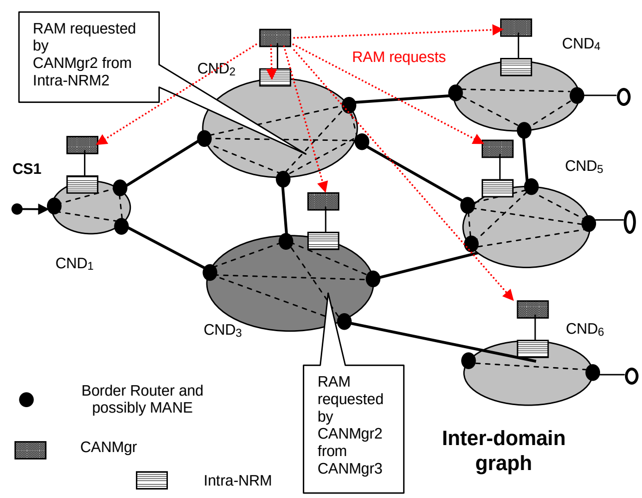

Figure 2. CAN Manager 2 issues RAM requests from each CANMgr involved in a VCAN

Fig. 2 shows an example of inter CAN Managers inter-controller signaling (i.e., inter-controllers in SDN terminology) in which the initiator CAN Manager 2 asked (in hub style) the other involved CAN Mangers (3, 4, 5, 6) to deliver to it their Resource Availability Matrices. Based on the received information, the initiator performs the VCAN mapping.

V. SIMULATION MODEL

The simulation objective of this paper is the evaluation of the M&C signaling overhead related to the negotiation activities between SP, CAN Managers, Intra_NRMs when the number of network domains and CAN Managers is a variable parameter.

Given the complexity of the M&C subsystem a simulation study has been developed. Real Time Developer Studio is a Specification and Description Language simulator, developed by PRAGMADEV. It comes in two versions: SDL and Specification and Description Language - Real Time (SDL-RT) [15].

SDL-RT is ITU standard SDL (based on Extended Finite State Machine Model - EFSM) extended with real time concepts. It is object oriented, has graphical language, allows modeling real-time features, combining dynamic and static representations, supporting classical real time concepts, extended to distributed systems, based on standard languages. It retains the graphical abstraction brought by SDL while keeping the precision of traditional techniques in real-time and embedded software development and making simpler the re-use of legacy code by using natively the C language. In SDL-RT, the C language is used to define and manipulate data. The ALICANTE management simulation model consists in: one Service Provider; N x CAN Managers, N x Intra_NRMs, where N is variable (1, ...16).

The specific target is to evaluate the time spent from the instant when an SP issues a VCAN request to an initiator CAN Manager, until the final confirmation of the VCAN installation is obtained by SP. The SP can choose any CAN_Mgr, based on their proximity and involvement in the requested VCAN. The chosen CAN_Mgr, named afterward the initiating CAN_Mgr, will interrogate the inter domain database about the network capability of the others domains, performs the inter-domain mapping algorithm, and will communicate with each CAN_Mgr identified by the inter-domain mapping algorithm as being involved in the requested VCAN. Note that the simulation model assumes parallelism in communication process from the initiator CAN Manager to different others (in “Hub” style). This is an important feature and design decision, assuring the scalability of negotiation process.

Fig. 3 describes the system processes. It contains the global variables, the instance of each class and the other blocks involved.

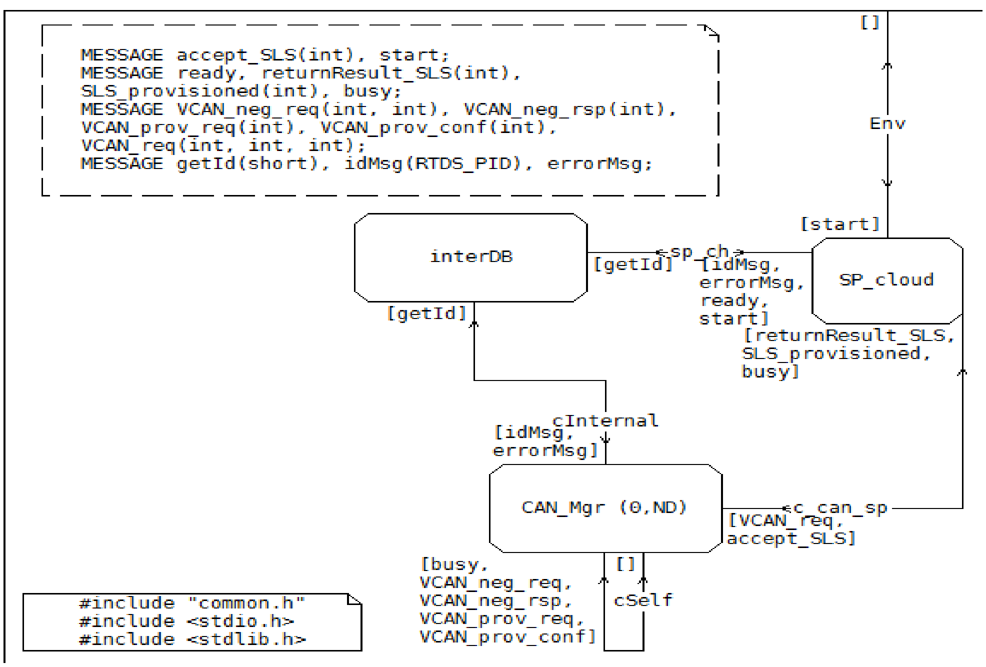

Figure 3. The system model used in RTDS simulations

The system consists of an *interDB* block, (used in simulation only, corresponding to an interdomain database that contains inter-domain network topology), a SP_cloud block, associated with the SP/CP requestors, and ND (Number of Domains) CAN_Mgr(s).

The work [14] describes in Message Sequence Chart (Fig. 4) forms the details of the signaling process between an initiator CAN Manager, and other CAN Managers involved in constructing a multiple domain VCAN. Here a simplified description is done.

The initiating CAN_Mgr send a *VCAN_neg_req* to each of corresponding CANMgr and enter into *"negociating"* state. Each corresponding CANMgr check its own capabilities (intra-domain mapping algorithm), respond to initiating CANMgr with a *VCAN_neg_rsp* message, and transits to *"waiting_for_acceptance_ext"* state. The initiating CAN_Mgr waits for all corresponding CAN_Mgr to respond, integrate the response, send that integrate response by a return_result_SLS message to SP_cloud and wait for a decision. That message indicates to SP that all requested resources are available and could be provisioned.

The SP analyzes the response and sends a provision request to the initiating CAN_Mgr, using the message accept_SLS. The initiating CAN_Mgr sends a provision request message, VCAN_prov_req, to each corresponding CAN_Mgr and waits for their response.

VI. Simulation Results

The simulations are focused on identifying the system behavior, and to determine a quantitative and qualitative estimation of the signaling time.

Being a real time simulator, the RTDS SDL-RT uses the internal PC clock to estimate the time for each task/process from the system. Therefore, the results are defined in "ticks", which are relative time units.

The simulation model just simulates the time consumed by the inter-domain and intra-domain mapping algorithms, but it does not actually compute that algorithms. However, the result of the mapping algorithm, the chosen CAN_Mgr and the Round Trip Time (RTT) delay between two communicating CAN_Mgr are introduced in simulator using a configuration file; its data is shown in Table 1.

TABLE 1. VARIATION IN RESPONSE TIME FROM DIFFERENT CAN_MANAGERS TO INITIATOR

	Average RTT=100	Average RTT=200	Average RTT=300	Average RTT=400	Average RTT=500
CAN_Mgr	RTT var	RTT var	RTT var	RTT var	RTT var
1	50	50	50	50	150
2	120	220	420	420	420
3	100	300	500	500	600
4	130	230	230	630	630
5	63	63	163	163	363
6	137	337	537	537	837
7	82	182	382	382	482
8	118	218	218	518	518
9	96	96	96	96	96
10	104	304	504	504	704
11	51	251	351	551	551
12	149	149	149	649	749
13	70	70	270	270	470
14	130	330	330	430	630
15	81	81	81	181	281
16	119	319	519	519	519
17	100	200	300	400	500

Each simulation uses one column from that table. Random variations have been introduced to emulate a real situation where the CAN Mangers are placed in different network domains and communicates via Internet.

While the simulation model uses an abstract time clock, in the experiments done, we can evaluate a time unit comparable to 1ms.

Two sets of data are used: one with a fixed (constant) RTT value for each corresponding CAN_Mgr, and one with the same average value as the constant RTT, but with a big dispersion. Both sets of data are shown in the Table 1.

The simulations are performed on two different machines, (i.e. named "Processor-1" - low power, and "Processor_2" - high power (Table 2).

TABLE 2 PC CONFIGURATION

PC configuration	Windows Experience Index	
	Processor-1	Processor-2
Processor	6	7.5
Memory(RAM)	5.9	7.8
Primary hard disk	5.9	7.9

The computing difference between the two PCs are just a qualitative criteria on evaluating the performance of a real CAN_Mgr machine when computing VCAN requests in ALICANTE environment.

Most of simulations are performed on a powerful PC, named "Processor_2". However, some of simulations are performed also on a slower PC, named "Procesor_1". These simulations allow a validation of results obtained from Processor_2". As expected, the results from "Procesor_1" have bigger relative time values compared with the values from "Processor_2", due processing time is shorter on powerful machine, as "Processor_2" (Table 2).

The range of RTT varies from 100 to 500 relative time units. Figures 4, 5 and 6 present the signaling time dependency (in relative time units) on the domains number implied in the requested VCAN, both from constant RTT and variable RTT.

The important result obtained and shown by the above diagrams is that the system is scalable versus the number of network domains. The simulation results shows that, even the start values are different, the signaling time has a tendency of convergence to 2500 value (~ms).

That convergence is explained by the fact that the signaling is made in parallel with all the CAN_Mgr involved, and the signaling time is depending on the domains numbers, the computation time spent on CAN_Mgr, and the RTT delay between each two communicating CAN_Mgr.

The simulator time unit can represent approximately one ms. This means that a single initiator CAN Manager could perform ~ 1200 VCAN requests per hour which is considered completely satisfactory into ALICANTE environment.

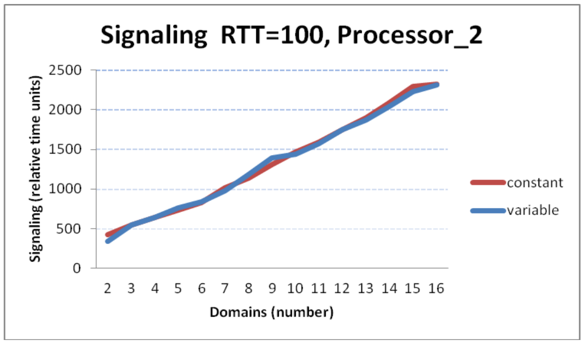

Figure 4. Signaling RTT=100, Processor 2

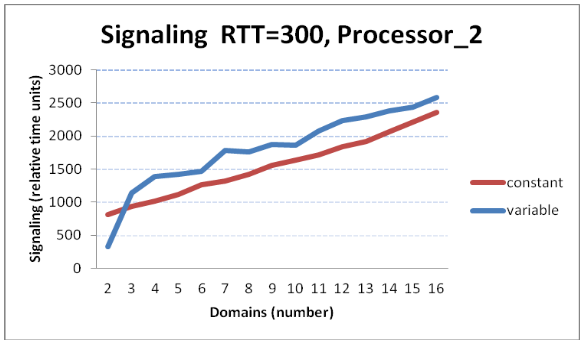

Figure 5. Signaling RTT=300, Processor_2

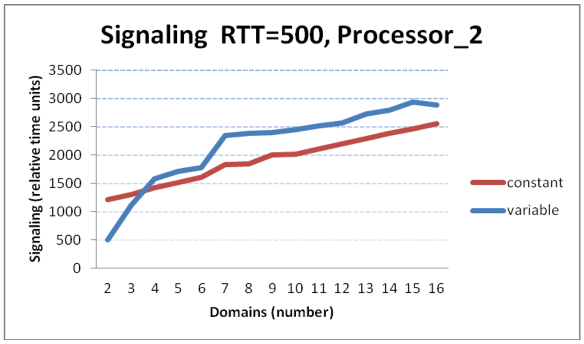

Figure 6. Signaling RTT=500, Processor_2

Fig. 7 shows a comparison between simulations on "Processor-1" and "Processor-2". It is clear that the system behavior is similar, only the convergence value is different (4500 for "Processor-1" and 2500 for "Processor-2"). Again, the convergence is present, and the difference on convergence value is the result of different computing time inside CAN_Mgr.

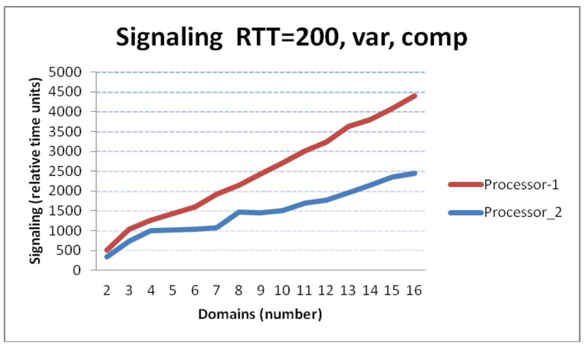

Figure 7. Signaling RTT=200, var, comparison

In order to have a validation from statistical point of view, a set of simulations were performed twice on "Processor-2", using different seeds. The simulation results are shown in Fig. 8. The convergence is present again, the small difference of the convergence value occurs due the seed influence on simulator internal algorithm, shown on the relative time units obtained on each simulation. That difference has a minor importance, as demonstrated by the several comparisons made based on the whole set of simulation results (Fig. 9, 10, and 11).

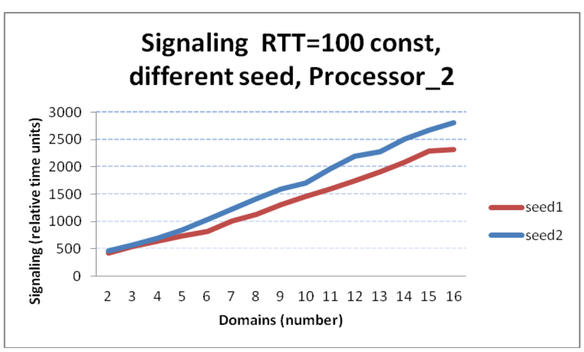

Figure 8. Signaling RTT=100 const, different seed, Processor_2

A convergence analysis is presented on Fig. 9, 10, and 11. The convergence is proved on both machines, for both fixed (constant RTT), and variable RTT.

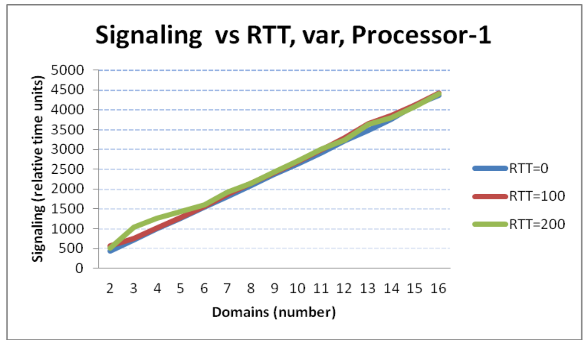

Figure 9. Signaling vs RTT, var, Processor_1

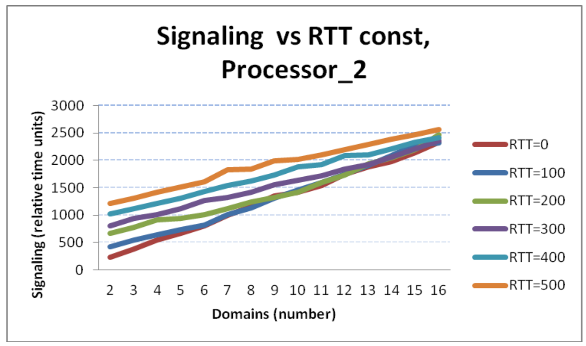

Figure 10. Signaling vs RTT const, Processor_2

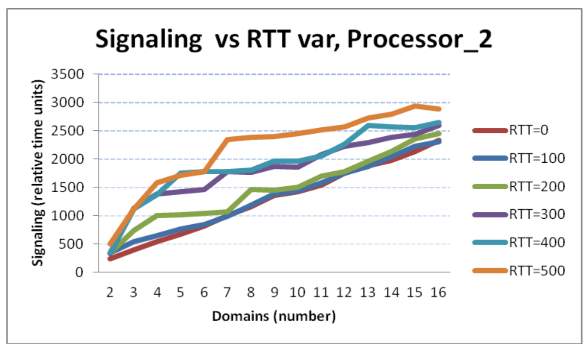

Figure 11. Signaling vs RTT var, Processor_2

Fig. 11 is the most interesting graph, showing that, despite big dispersion of RTT used, the signaling time is converging on the same value (2500) as for constant RTT. On that graph, the simulations used the same seed, but a range of average RTT from 0 to 500ms, with a high dispersion on each subset (same average RTT, different value for each RTT). Each RTT subset used in simulations is described in Table 1.

Moreover, the results from Fig.11 highlight that the RTT delay between each two communicating CAN_Mgr has a minor influence on the convergence value. According with the overall simulation results and comparing with the computing capability of CAN_Mgr and number of domains involved, that influence could be ignored.

VII. Conclusions

The management architecture of the ALICANTE system has been described, showing the similarity with SDN approach. Equivalence between an SDN controller and the pair {CAN Manager and Intra-domain Network Resource Manager} has been analyzed. Scalability problems appear in this multi-controller environment.

A simulation model based on EFSM model has been constructed.

It was found that signaling time dependency on the domains number implied in the requested VCAN is converging to an approximately constant value, being determined by the speed of the processor equivalent to the CAN_Mgr + IntraNRM used for simulation and by the number of domains involved.

That convergence is not influenced by the communication delay between the communicating CAN_Mgr.

The signaling in ALICANTE system is growing slowly and linearly with the number of domains involved, validating ALICANTE management system scalability.

Further work should evaluate the capacity of one controller to command a given number of network elements (routers) by using a vertical protocol (similar to OpenFlow).

Acknowledgment

This work has been partially supported by the FP7 European research project ALICANTE (FP7-ICT-248652) and partially by the projects POSDRU/107/1.5/S/76909.

References

[1] J. Schönwälder, M. Fouquet, G. D. Rodosek, and I. Hochstatter (2009), "Future Internet = Content + Services + Management", IEEE Communications Magazine, vol. 47, no. 7, Jul. 2009, pp. 27-33.

[2] T. Anderson, L. Peterson, S. Shenker, and J. Turner, "Overcoming the Internet Impasse through Virtualization", Computer, vol. 38, no. 4, Apr. 2005, pp. 34–41.

[3] J. Choi, J. Han, E. Cho, T. Kwon, and Y. Choi, "A Survey on Content-Oriented Networking for Efficient Content Delivery", IEEE Communications Magazine, March 2011.

[4] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking Named Content," CoNEXT '09, New York, NY, 2009, pp. 1–12.

[5] N. McKeown, et. al., "OpenFlow: Enabling Innovation în Campus Networks", March 2008, - http://www.openflow.org/documents/openflow-wp-latest.pdf.

[6] B. A. A. Nunes, M. Mendonca, X. N. Nguyen, K. Obraczka, and Thierry Turletti, "A Survey of Software-Defined Networking: Past, Present, and Future of Programmable Networks", ian. 2014, http://hal.inria.fr/hal-00825087

[7] OpenFlow Switch Specification, V 1.3.0 (Wire Protocol 0x04), June 25, 2012, https://www.opennetworking.org/images/stories/downloads/sdn-resources/onf-specifications/openflow/openflow-spec-v1.3.0.pdf

[8] S. H. Yeganeh, A. Tootoonchian, and Y. Ganjali, "On Scalability of Software-Defined Networking", IEEE Communications Magazine, February 2013, pp.136-141.

[9] B. Raghavan, T. Koponen, A. Ghodsi, M. Casado, S. Ratnasamy, and S. Shenker, "Software-Defined Internet Architecture: Decoupling Architecture from Infrastructure", HotNets-XI Proceedings of the 11th ACM Workshop on Hot Topics in Networks, 2012, pp. 43-48, doi: 10.1145/2390231.2390239.

[10] A. Tavakkoli, M. Casado, and S. Shenker, "Applying NOX to the Datacenter," Proc. ACM HotNets-VIII Wksp., 2009.

[11] H. Yin, H. Xie, T. Tsou, D. Lopez, P. Aranda, and R. Sidi, draft-yin-sdn-sdni-00.txt, "SDNi: A Message Exchange Protocol for Software Defined Networks (SDNS) across Multiple Domains", June, 2012

[12] FP7 ICT project, "MediA Ecosystem Deployment Through Ubiquitous Content-Aware Network Environments", ALICANTE, No248652, "D2.1: ALICANTE Overall System and Components Definition and Specifications", http://www.ict-alicante.eu/.

[13] E. Borcoci, R. Miruta, and S. Obreja, "Multi-domain Virtual Content-Aware Networks Mapping on Network Resources", EUSIPCO Conference, 27-31 August 2012, Bucharest, http://www.eusipco2012.org/home.php

[14] C. Cernat, E. Borcoci, and V. Poenaru, "SLA Framework Development for Content Aware Networks Resource Provisioning", AICT 2013 Conference, http://www.thinkmind.org/index.php?view=article&articleid=aict_2013_8_30_10084

[15] http://www.sdl-rt.org/, Apr 23, 2013