Topology and Network Resources Discovery Protocol for Content-Aware Networks

Vlad Poenaru, Eugen Borcoci, Marius Zotea
Telecommunication Department
University POLITEHNICA of Bucharest
Bucharest, Romania
vlad.poenaru@elcom.pub.ro, eugen.borcoci@elcom.pub.ro, marius.zotea@elcom.pub.ro

***Abstract* —Cooperation of several network providers to offer a common infrastructure to support overlay virtualized networks is of interest in the context of content/information orientation of the current and Future Internet. This paper is part of a work on management framework of a complex system, aiming to construct multi-domain overlay Virtual Content Aware Networks (VCAN), QoS enabled. A protocol (TNRDP). is proposed, evaluated and basically implemented, running between domain managers, to discover inter-domain topology (represented with different degrees of abstraction) and available network resources of different IP domains. The scalability and efficiency is preliminary analyzed.**

Keywords — Multi-domain, Topology and Resources Discovery, Content-Aware Networking, Management, Future Internet.

I. INTRODUCTION

The current, and especially Future Internet, has a strong content/information orientation, including multimedia flows distribution, [1-3]. Customizing the media transport can be done by creating virtualized *Content Aware Networks* (VCAN) on top of the IP level and coupling the network layer more strongly with the upper layers by having *Network Aware Applications* (NAA). The Content Awareness (CA) novel concept means that new intelligent routers will process and forward the data, based on *content type* recognition or, even more, treating the data objects based on their *name* and not based on *location address*, [4], [5]. The VCANs can be constructed based on virtualization techniques [6], agreed to be used to overcome the ossification of the current Internet [1], [2]. The VCANs should be finally mapped onto networking infrastructures, while respecting the requirements of Service Providers exploiting these VCANs.

A complex system architecture, CAN/NAA oriented, is proposed in ALICANTE European FP7 ICT research project, "Media Ecosystem Deployment Through Ubiquitous Content-Aware Network Environments", [7]. It works on top of multi-domain IP networks, to offer services for different business actors playing roles of consumers and/or providers. The architecture defines several cooperating environments: *User (UE), Service (SE)* and *Network (NE).* The UE contains the End-Users (EU) terminals; SE contains High Level Service Providers (SP) and Content Providers (CP). The NE contains a novel entity CAN Provider (CANP) to manage and offer VCANs and traditional Network Providers (NP/ISP) - managing the network at IP level. On demand of SP, the CANP (represented by CAN Managers, each being associated to a network domain) creates unicast or multicast VCANs (QoS enabled) over multi-domain, multi-provider IP networks. VCANs are realized as parallel planes as in [8], but additionally being content aware. The network resources are provided by the NPs. They are managed quasi-statically (provisioning phase) and also dynamically (during delivery phase) by using media flows adaptation. The management is based on dynamic Service Level Agreements/Specifications (SLA/SLSs) negotiated and concluded between providers (e.g., SP, CANP). In the Data Plane, content/service description information (metadata) can also be inserted in the media flow packets by the Content Servers, then recognized and treated appropriately by the intelligent routers of the VCAN (called Media Aware Network Elements – MANE).

This paper proposes a *topology and network resource discovery protocol (TNRDP)* running between the CAN Managers, to collect inter-domain and abstracted intra-domain information, to support the VCAN mapping algorithms onto network resources. Section II presents samples of related work. Section III summarizes the overall ALICANTE architecture and VCAN management. Section IV presents assumptions, requirements and main characteristics of the TNRDP protocol. Implementation is summarized in Section V. Specific aspects, evaluation and optimizations are discussed in Section VI. Section VII contains some conclusions and future work outline.

II. RELATED WORK

The TNDRP belongs to the management framework, aiming to construct QoS enabled VCANs, , over several independent but interconnected network domains..Given that VCAN solutions are based on the overlay concept, inter-domain QoS peering and routing are of interest, based on the overlay network ideas [9], [10]. An overlay network is defined, which first, abstracts each domain with a node, represented by the domain resource manager, or more detailed with several nodes represented by the egress routers from that domain. There exist protocols to transport QoS and other information between nodes and, based on this information, QoS routing algorithms can choose some QoS capable path.

In [9], a Virtual Topology (VT) is defined by a set of virtual links that map the current link state of the domain without showing internal details of the physical network

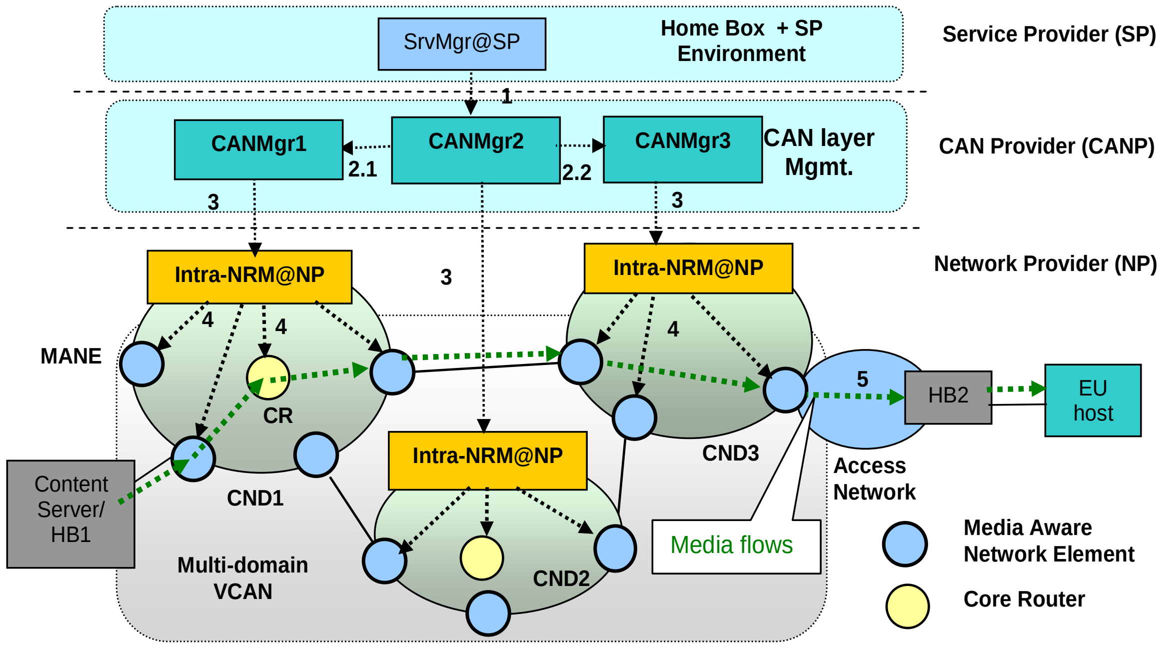

Figure 1. Example of network infrastructure (three domains) and management actions

Notations: SP - Service Provider; HB - Home Box; SrvMgr@SP – Service Manager at SP; CANMgr- Content Aware Network Manager at CAN Layer; Intra-NRM@NP – Intra-domain Network Resource Manager at Network Provider; CND – Core Network Domain; MANE – Media Aware Network Element; CR – Core Router; EU – End User Terminal

topology. Then *Push* and *Pull* models for building the VT at each node are considered and analyzed.

In the *Push* model each AS advertises its VT to their neighbor ASes. This model is suited for small topologies. In the *Pull* model, the VT is requested when needed, and only from the ASes situated along the path between given source and destinations; the path itself is determined using BGP. Also, the solution in [10] is based on BGP protocol. However we need a protocol to run between managers, so, although some ideas of [9] and [10] are valuable, our context is different. So, in Section IV we will develop our solution adapted to our system architecture to find overlay topology.

The system architecture is briefly presented in Section III to prepare extraction of TNDRP requirements. It is similar to Software Defined Networking, [11], [12]: evolutionary architecture; Control Plane / Data Plane separation; network control intelligence can be (logically) centralized in SW - based SDN controllers, which maintain a global view of the network; execution of the Control Plane infrastructure SW can be done on general purpose HW; the intelligent layers are decoupled from specific networking HW. Actually the Alicante CAN Manager and Intra-domain Network Resource Manager of the network Provider constitute together the “controller” defined in SDN.

III. System Architecture and VCAN Management

A. System Architecture

Figure 1 presents a simplified partial view on the ALICANTE architecture, with emphasis on the CAN layer and management interaction. The network contains several Core Network Domains (CND), belonging to NPs (they can be also seen as Autonomous Systems - AS) and access networks (AN). The ANs are out of scope of VCANs. One *CAN Manager* (CANMgr) exists for each IP domain to assure the consistency of VCAN planning, provisioning, advertisement, offering, negotiation, installation and exploitation. Each domain has an *Intra-domain Network Resource Manager* (Intra-NRM), as the ultimate authority configuring the network nodes. The CAN layer cooperates with some local entities called Home Boxes (HB) and SE by offering them CAN services. Details on this architecture are found in [7], [13], [14].

B. VCAN Management

The actions 1, 2, 3, 4 (Figure 1) are a simplified set performed in M&C Plane in order to negotiate, agree and install a VCAN:

1- Request (SLA/SLS) for a VCAN from SP - to an *initiator iCAN Manager* (e.g., CANMgr2);

2.1, 2.2 – “Horizontal” negotiations between the iCAN Manager and others, associated to different CNDs – for multi-domain VCAN mapping and construction;

3- "Vertical" requests from each CANMgr to its Intra-NRM, to install VCAN;

4- Commands given by each Intra-NRM (involved in the multi-domain VCAN) to configure its MANE and Core Routers (install ingress policies, and egress rules).

The number 5 action represents a data flow, transported on a unicast path over several network domains inside the VCAN, from Content server (or HB if it generates content).

The SP VCAN requests are expressed in the SLS. The SP knows the edge points of this VCAN, i.e., the MANEs IDs where different sets of HB will be connected. The requested VCAN belonging to a given QoS class, should be mapped onto real network topologies, while considering available transport resources of one or several network domains. The solution adopted (for unicast VCANs) was to run on the initiator CAN Manager a combined algorithm capable of computing QoS constrained paths, making logical resource reservation, and finally mapping the VCAN onto such resources [14]. The algorithm provides a global optimum and is appropriate for VCANs mapped onto MPLS paths. The latter are pre-provisioned by each Intra-NRM and then offered as available to the associated CAN Manager.

The *VCAN mapping problem is:* given a resource availability graph and a *Traffic Demand Matrix (TDM)*, how to map it onto a real graph while respecting the minimum bandwidth constraints and also optimize the resource usage.

The mapping algorithm needs input data at the iCAN Manager level, i.e. an "international" graph, plus link capacities, where the intra-domain paths are abstracted as traffic trunks.

The TNDRP should collect the inter-domain graph information(it is assumed that each CAN Manager has knowledge on its domain and inter-domain links with neighbors). In such a way, the VCAN initiator CAN Manager can learn the inter-domain *Overlay Network Topologies (inter-ONT)* and inter-domain available link capacities, and can map VCANs onto network resources. At domain level, a summary only of the intra-domain topology is uploaded by each Intra-NRM to the CAN Manager, represented in an abstract way by sets of virtual links (called *Traffic Trunks*), belonging to a given QoS class. This information is called *Resource Availability Matrix (RAM*). Each domain is free to upload whatever RAM it wants, depending on its own policy. This is important from a business point of view, given that it preserves each CND independence.

A partial description of the VCAN mapping algorithm is given in Figure 2, in order to derive the requirements of the TNRDP protocol. Assume that SP issues a request for *VCAN (I1, O1, O2, O3)* to CANMgr2 (this is the VCAN initiator). Here the generic *Ix, Oy* represent inputs and respective output points in the VCAN. Actually these will be mapped onto MANEs interfaces. Note that SP does not know which (if they exist) some transit domains are needed for the requested VCAN. It is not its job but the initiator CAN Manager must determine the best transit domains.

In order to run the mapping algorithm, the initiator iCANMgr should discover (viaTNRDP) the complete inter-domain and intra-domain abstracted summarized graphs (within a given region of the world).

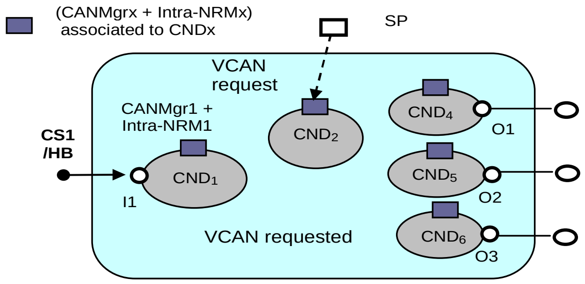

Figure 2. Example of a requested VCAN (I1, O1, O2, O3)

Figure 3 shows as an example the context associated to the VCAN request addressed to CANMgr2 playing the role of iCANMgr (see Figure 2). After running the TNRDP iCANMgr2 will know the inter-domain graph and the RAM abstractions of each Core Network Domain (CND1 … CND6). Then iCANMgr runs the VCAN mapping algorithm and determines all CNDs involved in the requested VCAN. Then it splits the SLS parameters, thus preparing new requests for each of these CNDs and negotiates *with each CAN Manager* involved, to agree and reserve resources for the VCAN. The main TNRDP design decision is on how a CAN Manager can obtain the topology and resources information? One can use several approaches.

On demand variant: the iCANMgr knows initially only a pure inter-domain graph (where each domain is abstracted with a node). Then it can determine (first step) via a routing algorithm, which domains could be candidates to participate to the VCAN. Then iCANMgr asks from them their RAMs and obtains the complete abstracted graph as in Figure 3. On this graph iCANMGr applies the VCAN mapping algorithm (second step). While being hierarchical, this solution does not provide a global optimum, given that the intra-domain RAMs are not known in the first step.

Proactive variant: the TNRDP works periodically or in event-triggered mode. Every iCANMgr has in any moment the image of the graph as presented in Figure 3, then it applies the VCAN mapping algorithm and finds the best paths by using an appropriate metric and constraints expressed in the TDM.

The VCAN mapping algorithm is described in [14]. Its details are out of scope of this paper. Shortly, the cost of a link (i,j) in the ONT can be *C(i,j) = Breq/Bij= Breq/Bavail*, where Bij is the available bandwidth on this link and *Breq* is the bandwidth requested for that link. The metric is additive, so one can apply a modified Dijkstra algorithm to compute the *Shortest Path Trees (SPT)*, i.e one tree for each ingress node where the traffic flows will enter. A basic simpler metric of 1/Bij as an additive link metric can be used.

This paper considers the proactive variant of TNRDP. The advantages are similar to other proactive protocols: it provides a global optimum of the paths; at any moment each CAN Manager is ready to receive VCAN requests and to serve them based on updated information produced by TNRDP. This is similar to the routes computation in

 ISBN: 978-1-61208-315-5

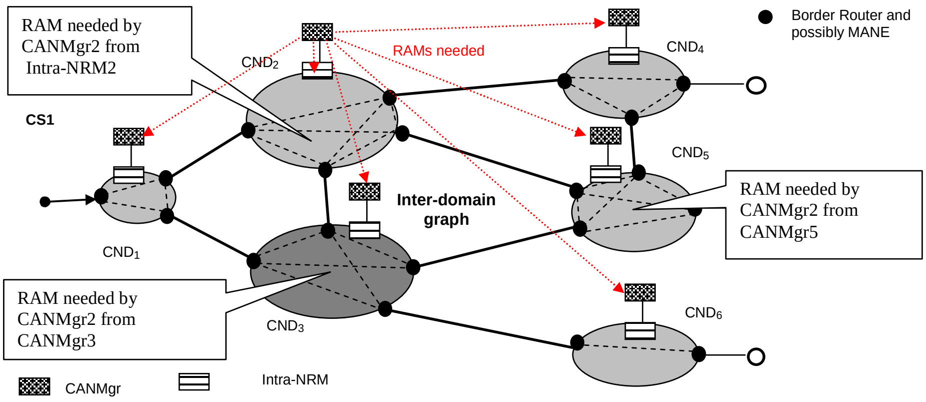

Figure 3. The graph information necessary to be known by CANMgr2 in order to be able to map the VCAN requested in Figure 1

proactive style and then their usage for forwarding the data packets. Actually, TNRDP decouples the topology and resources discovery process from the actual VCAN resource reservation and allocation processes. However, different from the routing protocols context, the VCAN reservation means a *stateful* approach. So, this solution creates some problems, e.g., a decision on VCAN-k mapping is taken at iCANMgr_k, based on the known graph at that instant, while some other SPs VCAN requests can arrive at some other CAN Managers, for resource overlapping with those used by the iCANMgr_k.

Another problem of the proactive TNRDP can be related to the signaling overhead, due to broadcast style, similar to OSPF working mode. These aspects will be discussed in more depth in Section VI.

IV. Topology and Network Resources Discovery Protocol

A. *Assumptions*

TNRDP is an application layer protocol.The folowing assumptions are considered valid for the context of TNRDP:

- Each Intra-NRM knows its physical connections and internal paths and uploads its RAM information to its CAN Manager by using a mechanism external to TNRDP. The RAM can be expressed in the most simple form as a set of tuples : *{(Router_in, Router_out, Capacity) (.. .)..}.*
- The CANMgrs exchange TNRDP messages with other CANMgrs.
- TNRDP does not solve itself security issues; reliable transport is supposed over TCP.
- The Managers Identities are statically known, by each manager in a given region of a larger network. Also the associated domains IDs (e.g., AS numbers) are known, or at least the ASes of neighbours.
- A simplifying assumption for the basic TNRDP version is that all domains are VCAN capable.
- VCAN Mapping algorithm itself is out of TNRDP scope.

B. *TNRDP Requirements*

The essential TNRDP requirements resulted from the previous sections are:

- To accommodate any number of CAN Managers;
- Scalability in terms of signalling traffic overhead;
- To serve any CANMgr to find out the topological and resources graph in a given network region;
- To support parallelism, i.e., it should leverage topological and resurces changes that can appear in several places in the same time;
- To disseminate in a stable way the topology and capacity information when changes appear in some domains.
- Run in “soft realtime” conditions, given that VCANs are established in aggregated way and not per-path. Also they are established and terminated by SPs not so frequently (tens-hundreds, at most thousands per day).

The main TNRDP design decisions are described below:

- Any CANMgr communicates directly (via TCP) only to its neighbors; two managers are defined as neighbors if their associated network domains have at least an inter-domain link. Before exchanging NSA (*“Network State Advertisments”)* a CANMgr must establish logical connection to its neighbors;
- Each CANMgr receiving a NSA message, aggregates this information to its own ONT and then broadcasts a new NSA to its neighbors;
- TNRDP is a simple stateful protocol. The Finite State Machine states of a given CANMgr related to one of its neighbors are: *Listen, Waiting for connection or disconnection confirmation, Connected* (NSAs can be exchanged);
- Any CANMgr may initiate Connect/Disconnect;
- If a CANMgr receives a Disconnect message form a neighbor, it will exclude from its known graph all information related to its links to that neighbor. If the disconnected domain is a leaf one, then the CANMgr observing this disconnection will also erase all information related to that disconnected domain;

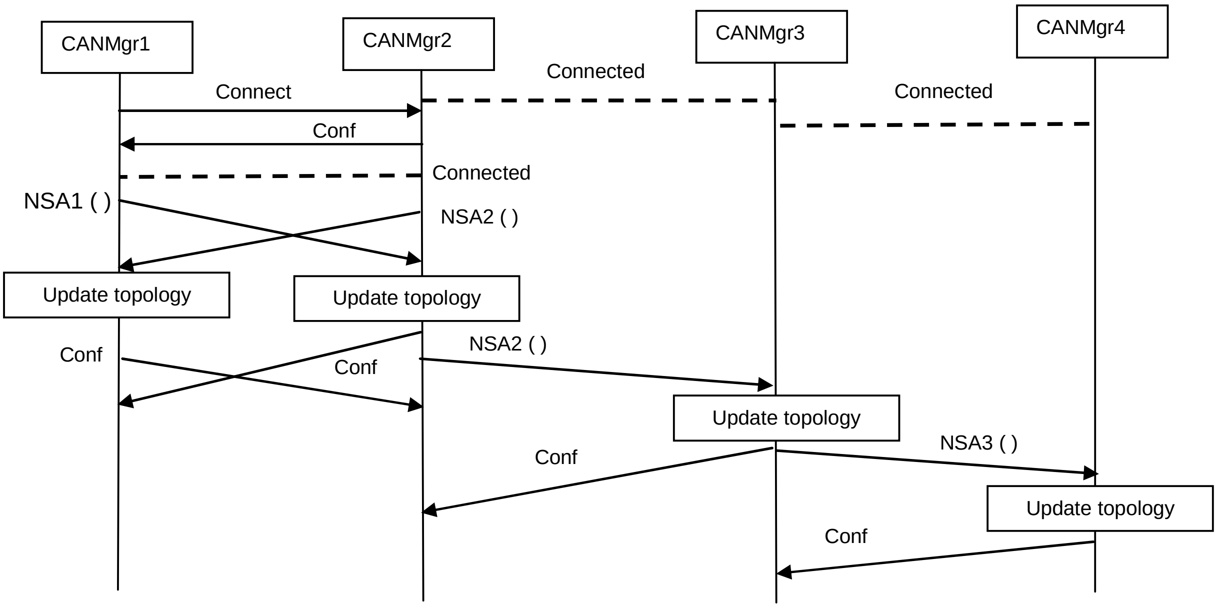

Figure 4. Message Sequence Chart example; Neighbors (CANMgr1-CANMgr2, CANMgr2-CANMgr3, CANMgr3-CANMgr4)

- NSA messages can be exchanged in triggered event style or periodically, in asynchronous mode (this simple assumption could be modified in the section related to optimizations)
- "Hello" messages will maintain the connection alive.

C. TNRDP main characteristics and basic operation

The TNRDP messages types are:

Connect – to initiate connection between two neighbors;

Disconnect – to ask the end of a connection to a neighbor; the NSA exchange will be stopped;

Network State Advertisments – broadcasted to neighbors to update the graphs. It contains parts of the global graphs, known by the sender;

Confirmation- sent in unicast mode as confirmation to "active messages";

Error- to signal synthax or semantic errors;

Hello - to maintain/confirm the connection.

V. IMPLEMENTATION

A *Java basic implementation* has been realized, as *proof of concept* and for preliminary evaluation/validation of TNRDP. The program has five classes:

Runner – is the starting point of the program. Here a configuration file representing the managers set data, topology and resources information, etc., is loaded as input data;

SimFactory – performs configuration file analysis and generates the topology;

SimHost – generates the manager instances and processes the states and mesasages of the managers;

Graph – contains the graph and also the operations to be performed on the graph;

The message format is the following: *Type, Seq_no, Src_Mgr_Id, Dest_Mgr_Id, Data_length, Data*, where the fields semantics is straightforward. The sequence numbers are inserted by each sender and incremented at each new message. They allow to control the pairs message/ confirmation and can also be used in mimimizing the amount of control traffic.

Figure 4 presents an example of success simple scenario in which the neighbors are: CANMgr1 - CANMgr2, CANMgr2 - CANMgr3, CANMgr3 - CANMgr4. It is supposed that connections already exist: 2-3 and 3-4. Then CANMgr1 wants to join the "community" and initiates a connection to CANMgr2. After connection is established the two partners exchange topology and resources information, via NSA1 and NSA2 messages. After updating its Database, CANMgr2 informs CANMgr3 about changes and CANMgr3 performs a similar action for CANMgr4. Here it is supposed an asynchronous mode, event-triggered style for communication.

Message – executes transmission of messages between the managers.

Each manager has three message queues: *input_Q*, *processing_Q*, and *output_Q*. The program runs in each (iterative) step all managers participating to topology. Each manager processes all messages received from other managers in the previous step, makes updates and sends messages to the output queue to be sent to other managers in the next iteration.

The topological graph and resource information has been represented in matrix form, considering all uni-directional logical pipes (intra-domain or inter-domain segments) of the graph. The simple entry in such a matrix could be *(Input_i, Output_j, Available_Bandwidth).*

VI. SPECIFIC VCAN-RELATED ASPECTS, EVALUATION RESULTS AND OPTIMIZATIONS

The TNRDP protocol runs in a specific environment given by the ALICANTE architecture. The full implementation should be compliant with CAN architectural requirements. One major TNRDP aspect is related to the *stateful* characteristic of a VCAN (i.e. working based on logical reservation and then resource allocation following QoS constrained paths determination). From this point of view the requirements are harder than in the classic routing protocols (OSPF, BGP); there, a router is free to make fast path changes (for routing and consequently – forwarding) if new conditions arise. In our case, a first precaution is that VCAN mapping algorithm is atomically performed at a given iCANMgr; i.e., the TNRDP is not allowed to change the topology and resource information until the VCAN mapping calculation is ended at this iCANMgr. However, this computation is done relatively fast (see [1]) in comparison with communication delays between CAN Managers. Another issue is that several SPs can concurrently request different VCANs, to different CAN Managers. Therefore, some competition for the same resources could appear, given the fact that two iCANMgrs could possibly compute and try to reserve the same path segments in un-coordinated way. This requires an additional confirmation/negotiation (after the initial VCAN mapping has been computed), done in hub style between iCANMgr and the others ones CAN Managers involved, but this step (i.e., unicast communication between iCANMgr and the other CAN Managers) existed anyway in the original approach of VCAN negotiation..

Scalability of TNRDP in terms of number of domains to be accommodated and their dimension is of interest. It is estimated that a VCAN-capable region could involve tens or at most hundreds (less probable) of CAN Managers. On the other side the system architecture supposed that each Intra-domain Network Resource Manager is aware of its paths and resources. Therefore such issue is not a bottleneck of the TNRDP protocol.

The proactive style implies broadcasting NSAs to neighbors and further in an area. The TNRDP messages overhead is of interest. The total number of messages following a given event depends on the number of domains and on the topology. On the JAVA implementation the following numbers have been measured, for a linear network topology. The complexity order is $O(D^2)$, where D is the number of domains, as it can be seen in Table I.

A general approximate estimation (for an arbitrary topology) of the message numbers can be done, considering that we have D domains (managers) and each manager has a connection order of n (it has to broadcast to n other domains a given message), while the diameter of the region is d (measured in number of domains). The overhead traffic produced by one event would be $OVH = O(2*D*n*d)$ messages. If we include confirmation then for $D = 100$, $n = 10$, $d=5$, we get $OVH = 10000$. A reduction can be achieved if after receiving a mesasage from a neighbor, a manager will wait (based on a timer) some time (e.g., ~100ms) for other possible messages to be received , then aggegating them and re-transmiting its new NSA.

TABLE I. COMMUNICATION COMPLEXITY

No. of domains	5	10	20	50	100	250
No. of messages	24	99	399	2499	9999	62499

VII. CONCLUSIONS AND FUTURE WORK

This work proposed a novel *topology and network resource discovery protocol* (TNRDP) usable between several domain managers, aiming to produce information to serve *Virtual Content Aware Networks* mapping onto network resources in an optimal way. Solutions analysis is done in the architectural context of a complex system aiming to media distribution. Proof of concept preliminary implementation has been done and a summary of performance analysis and optimization measures. Currently a full version of this protocol is in development in the framework of the ALICANTE project.

VIII. AKNOWLEDGEMENTS

This work was supported partially by the EC in the context of the ALICANTE project (FP7-ICT-248652) and partially by the project POSDRU/89/76909.

REFERENCES

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

[2] C. Baladrón, "User-Centric Future Internet and Telecommunication Services", in: G. Tselentis, et. al. (eds.), Towards the Future Internet, IOS Press, 2009, pp. 217-226.

[3] J.Turner and D. Taylor, "Diversifying the Internet," Proc. GLOBECOM '05, vol. 2, St. Louis, USA, Nov./Dec. 2005, pp. 760-765.

[4] 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.

[5] V. Jacobson et al., "Networking Named Content," CoNEXT '09, New York, NY, 2009, pp. 1–12.

[6] 4WARD, "A clean-slate approach for Future Internet", http://www.4ward-project.eu/.

[7] G. Xilouris, et al., FP7 ICT project, Deliverable D2.1: "ALICANTE Overall System and Components Definition and Specification", September 2011, http://www.ict-ALICANTE.eu/

[8] M. Boucadair, et al., "A Framework for End-to-End Service Differentiation: Network Planes and Parallel Internets", IEEE Communications Magazine, Sept. 2007, pp. 134-143.

[9] Fabio L. Verdi and Maurıcio F. Magalhaes "Using Virtualization to Provide Interdomain QoS-enabled Routing", Journal of Networks, April 2007, pp. 23-32.

[10] Z. Li, P. Mohapatra, and C. Chuah, Virtual Multi-Homing: On the Feasibility of Combining Overlay Routing with BGP Routing, University of California at DavisTechnical Report: CSE-2005-2, 2005.

[11] T. Koponen, et al., Architecting for Innovation. SIGCOMM CCR, 41(3), 2011.

[12] M. Casado, T. Koponen, S. Shenker, and A. Tootoonchian, Fabric: A Retrospective on Evolving SDN. In Proc. Of HotSDN, August 2012.

[13] E. Borcoci, and R. Iorga, “A Management Architecture for a Multi-domain Content-Aware Network” TEMU 2010, July 2010, Crete, http://www.temu.gr/2010/.

[14] R. Miruta and E. Borcoci (UPB) "Optimization of Overlay QoS Constrained Routing and Mapping Algorithm for Virtual Content Aware networks" ICNS 2013 - Lisbon, Portugal. http://www.thinkmind.org/index.php?view=article&articleid=icns_2013_4_30_10174.