Solving virtual network resource allocation problems using a Constraint Satisfaction Problem model

Fouad Guenane*, Pierre Yves Dumas*, Michele Nogueira†, Guy Pujolle*
*LIP6 - UPMC - Sorbonne Universités, France
†NR2 - Federal University of Parana, Brazil
Email: {fouad.guenane, pierre-yves.Dumas, guy.pujolle}@lip6.fr; michele@inf.ufpr.br

***Abstract*—We define virtualization as a set of techniques to run multiple operating systems on the same physical machine (physical node) sharing its resources. These technologies have a huge success because they improve safety, reliability and flexibility. Several studies in the field have been made by different teams, but the allocation of physical resources remains an open problem in the field of network virtualization. We identified this problem as a graph matching problem. We express and solve it like a Constraint Satisfaction Problem (CSP). Our CSP model includes Qos requirements and energy saving. To our knowledge, no previous work used CSP approach and combined Qos and Green aspects to allocate resources.**

***Index Terms*—Virtual networks, resources allocation, Constraint satisfaction Problem (CSP), Choco.**

I. Introduction

Virtualization has revolutionized the concept of network, new technologies like Cloud computing and virtual network have strengthened the concept of virtual server: this concept considers several servers which are logical entities installed on physical machines. The power of data centers can place thousands of virtual servers, which can migrate based on differents criteria: performance, power consumption, cost or other numerous parameters. Today networks are also at the stage of virtualization, several virtual routers can be installed on one physical router. Virtual routers can also migrate (transferring from physical node to another) based on a network criteria : cost, power, reliability, congestion control...etc. This notion of virtual routers simply extends to all network equipment such as switches, MPLS Label Switch Routers, box, gateways, SIP servers, PBXs, etc.

Network virtualization must provide a feeling of freedom of movement at all levels. Each virtual network must be free to implement its own topology to rout functions. Furthermore, its control protocols should be customized independently of the underlying physical network and other coexisting virtual networks.

Several studies in the field have been made by different teams [1][2][3][4][5][6]. However, the allocation of physical resources remains an open problem in the field of network virtualization [7] and the matching of the virtual functions to the physical resources is not yet optimized in terms of latency, architecture, quality of service, etc.).

This optimization problem is very complex due to many constraints, but we can propose a more optimal solution then the previous ones. Several studies in the field did not take into account all the properties of the allowance (discussed later). Some studies focused on the knowledge of virtual requests in advance and others were done in a centralized way.

In this project, we aim to develop a solution for the dynamic resource allocation problem by modeling it as a constraint satisfaction problem. We present in section-2 the related works which focuses on resources allocation problem. In section-3 we discuss and present our CSP model. Its implementation and our results are described in section-4. This implementation uses Choco, a java library that enables constraint programming within imperative programming. Finally, in section 5, we conclude.

II. Related Works

In a network virtualization environment, a number of virtual networks coexist on the same physical network. Each virtual network is composed of a subset of resources of the underlying physical network. A physical network (also called network substrate) belongs and operates by an infrastructure provider (InP) whose goal is to make a profit from the resources of the network by leasing its customers (service providers (SP) or suppliers of virtual networks (VNPs)). A virtual network consists of a set of virtual nodes, each hosted on a physical node, and a set of virtual links, each set in the physical path or a set of physical links. Virtual networks that share the same physical network are completely isolated from each other, they can be used to provide end-to-end without the need for application protocols or tools unified control. [7] The architecture of virtual networks has several characteristics as Coexistence: many virtual networks (possibly from different SP) coexist on the same physical infrastructure, Recursion: also known as nesting, refers to the ability of a virtual network to be implemented on another virtual network creating a hierarchy in the network virtualization environment and Inheritance: virtual networks in the upper levels of the hierarchy of recurrent competition (children) could inherit the properties of virtual networks in the lower levels (their parents).

All these characteristics met to assure: (i)flexibility: each virtual network is independent of the underlying physical infrastructure, (ii) manageability: Each virtual network can be managed independently with the ability to provide network administration tools for one or more VLANs, (iii) stability and convergence: the scope of any configuration errors in the underlying physical infrastructure should be limited in order to affect a minimum number of virtual networks on top of the physical infrastructure and (iv) heterogeneity: not only the underlying physical infrastructure could be composed of heterogeneous technologies (like optical and wireless networks), the virtual layer should be too.

A. Resources allocation problem

Creating a virtual network requires allocation of physical resources to its nodes and links in accordance with a number of properties. We represent a virtual network and a physical network by respectively two graphs, and we consider the allocation problem like a problem of mapping the virtual network over physical resources. Several mapping algorithms have already been proposed to allocate physical resources efficiently and we present them later in the document. The paper [2] is from our point of view the basement of the field, it defines the foundation of virtual network instantiations and provides a model of resource allocation problem. Its model and results inspired several later works. Indeed, [2] proposes rules for instantiating a virtual network. However, the process of creating virtual networks started with the virtualization of physical resources. Therefore, a group of virtual resources is created and represented by a virtual layer that implements the abstraction of physical available resources. This virtual network will be subject to three interrelated steps which are Resource description, Resource discovery and Resource providing.

The main objective is clear: the instantiation of the virtual network must be economical in terms of physical resources. Resource allocation must be done with the idea of optimizing these resources. Notion of boundary nodes and links has been introduced for the assignment of a maximal number of allocated resources. To date, no process for selecting limits is able to maximize gain (services) generated by instantiating the virtual network in line with the Trade Policy Provider. Functions must also take into account the overhead of CPU and bandwidth of nodes used in order to devoid a supplementary traffic overload. These goals do not take into account the flexibility and reliability of links and it must be reminded that assignment of physical network nodes for the virtual network without the violation of bandwidth constraints is NP-complete as defined in [2]. So the problem is very complex and calls for an efficient and scalable solution to optimally instantiate nodes.

B. Solution classification

In our research, we identified several algorithms that we classify as follows:

- **Centralized approach**: a central entity is responsible for mapping virtual networks to the physical network. It must maintain updated physical network information (resources) to make appropriate decisions to allocate resources. This approach could suffer from scalability problems. In addition, the communication between the central entity and the other nodes in the physical network (updated information on resources available) will generate extensive network overload.
- **Distributed approach**: to cope with the problems of the centralized approach, the process of resource allocation can be distributed over the entire or part of the physical nodes in InP. In general, each physical node involved in the allocation of resources used his local knowledge to this effect. Communication protocols and cooperation are needed to coordinate the process.

In addition of centralized and distributed approaches, [2] [3] present two other approaches. **Static approach** (without reconfiguration): does not allow any change in the assignment of resources during the lifetime of the virtual network and **Dynamic approach** (with reconfiguration): adaptive, it can change the allocation of resources in dependence applications, Qos and virtual network performance.

To our knowledge, there are around a little more than 25 algorithms for resource allocation in virtual networks. They are centralized or distributed, with or without re-configuration. We cannot describe everyone of them but [8] classify them to describes their approach and requirements. To resume this section, the fundamental problem of virtual network instantiation are the optimization of resource allocation offered by the network in accordance with physical constraints as well as compliance with the specifications imposed by the "Service Level Agreement" (SLA), including the Quality of Service (Qos). Complexity is added when the environmental aspect is taken into account. Indeed, the ability to save energy can be integrated into the algorithm as an additional constraint. Whatever the context is, any proposed allocation must support the desired traffic (required one).

III. Model and Architecture

Methods for solving a problem of graph matching can be divided into two classes [9] **the exact methods** and **the approximate methods**. (i)The exact method assumes that a sub-graph exists and the task is to find it. But in some situations, the data is altered and a perfect match can not be found. (ii) The approximate method finds a solution optimizing a matching objective function. In our study the different algorithms that are based on these methods (Ullmann, Nauty, Schmidt and Druffel) does not correspond perfectly to our problem. Indeed, the matching network must consider the capacity of nodes and links.

A. Matching graph

We model the physical network as a weighted undirected graph $G^s = (N^s, L^s, A^s_N, A^s_L)$ where $N^s andL^s$ respectively represent sets of nodes and links of the physical network, each node has properties A^s_N that we have identified as the CPU and the memory capacity, it is the same for links A^s_L whose main property is the ability in terms of bandwidth. We also take into account the existing roads between all nodes of the physical network and represent a matrix called P^s. The instantiation of the virtual network is characterized by a query that is also modeled by a weighted undirected graph, $G^v = (N^v, L^v, C_N^v, C_L^v)$, the pair ,$N^v and L^v$ represents the logical topology of the virtual network. The virtual network is characterized by constraints on virtual nodes and links that constitute C_N^v, C_L^v modelize respectively constraints of virtual node N and virtual link L. Physical resources allocation to the virtual network is the matching of G^V on a part of G^S, respecting the constraints of the virtual network: $f : G^V \rightarrow (N', P', R_N, R_L)$ The global allocation function which is an application for starting domain G^V, and arrival domain which consists of:

- $N' \subset N$: the set of allocated nodes
- $P' \subset S$: the set of paths, a path can be composed of multiple links,
- R_n, R_L: respectively the allocated resources of nodes and links.

The assignment problem of nodes and links are dependent and must be treated simultaneously:

- Nodes assignment: $f^N(N^V, C_N^v) \rightarrow (N', R_n)$
- Links assignment: $f^L(L^V, C_L^v) \rightarrow (P', R_L)$

Symbol	Comment
N^s	Set of physical nodes
L^s	Set of physical links
A_N^s	Physical node properties
A_L^s	Physical link properties
N^v	Set of virtual nodes
L^v	Set of virtual links

TABLE I: Model's symbols

Allocate a large number of virtual links on a physical link can be disturb a Qos (cut, bottleneck,...). A router that would have a high number of instantiated virtual machines are not appreciated it can be decreased its performances, this is the business experience that guides us in the use of some useful parameters that concern nodes and/or links in order to improve the performance of the physical network by passing the virtual networks that are mapped above and reduce the rate of packet loss, lowering the packet collision due to cross traffic flow in several physical links. The load has to be shared over the entire substrate (physical) network. We aim to be specific enough to manage the quality of service and the economy of energy. To achieve that, we identified two functions of utility depending on the SLA, so that we can model the service level agreement (SLA). This eventually allows to improve both provider utility and client utility.

B. Service Level Agreement

SLA is a contract between a service provider, in this case a physical network provider and a client (virtual network) [10]. The contract specifies what services the provider must provide to the customer and what penalties are suffered when a service is not provided. However, customer expectations in most market studies emphasize requirements [10] [11] as reliable measurement of quality of service (Qos), provision of the expected quality of service and optimizing the use of resources.

Client	Provider
Carrier Service Provider	Carrier Service Provider
Internet Service Provider	Carrier Service Provider
Enterprise/ operator	Internet Service Provider
User	Internet Service Provider

TABLE II: Interaction between SLA's Actors

In the field of virtualization, SLA can be produced between different protagonists, Table II shows the possible interactions between these actors for the implementation of a virtualization service. For example a Carrier Service Provider who has a very huge physical infrastructure can provide to an ISP or an operator's his physical infrastructure to instantiate virtual networks on demand, the ISP or operator can then propose this service (network virtualization) to its customers. The operational implementation of SLA in network technologies is reflected in the adequacy of the various parameters that are customized for client needs, these parameters are presented in table III.

P	Parameters	type
P1	Packet lost	Nonfunctional
P2	Delay	Nonfunctional
P3	Jitter	Nonfunctional
P4	Bandwidth	Functional
P5	CPU and Memory	Functional
P6	Topology	Functional
P7	Availability	Nonfunctional
P7	Admission Control	Functional

TABLE III: SLA's Parameters

We identify two types of parameters, functional and non-functional. Indeed, some properties such as topology and functional resources are representative of the needs and constraints that must be strictly respected, conversely nonfunctional properties correspond to those on Qos and availability.

A contract must take into account two aspects, the customer and the supplier satisfaction. A utility function must be set up. For this purpose, we use a classification of services and its correspondence with the SLA. We have adapted the proposed correspondence [10] to our needs. The following figure shows that:

We classify services. A client needs to match one or more classes. Our solution calculates the most optimal allocation based on the objective function that will be defined. The supplier shall implement all the relevant parameters of the SLA constraints. Then, the resource allocation considers these conditions. From there, we decompose the problem into two sub-problems.

C. Client utility

The client expresses his needs in the SLA. In our SLA model, his needs are expressed by several properties defined

	Packet lost	delay	jitter	Bandwidth	CPU memory	Topology	Availability	Admission Control
Voice	=	++	++	+	=	=	+	++
Videophone	=	++	+	++	=	=	+	++
Telephony	=	=	0	0	=	=	+	++
Multimedia	+	++	0	++	=	=	+	++
VOD	=	++	0	++	=	+	+	++
VPN	Depends about the encapsulated data				=	++	++	++
Real Time Data	++	++	0	+	=	=	+	++
Data (web, mail , e-Commerce)	++	+	0	0	=	=	+	++
Streaming	++	=	0	0	=	=	+	++

++	Very High Performance	=	Performance Best Effort
+	High performance	0	indifferent

Fig. 1: Mapping SLA Services

for several distinct services. The main task is to choose an allocation that meets the needs of the customer to the best of these properties. How can we compare allocations? How to decide which one is the most appropriate? We decided to base our decisions on **multi-criteria** techniques.

The multi-criteria analysis or, more accurately, multi-criteria decision-making methods, are relatively new techniques which are in progress. Some [12] support multi-criteria decision as an alternative to traditional optimization methods based on the definition of a single function. The interest of MCDA is to consider a set of criteria from different types (expressed in different units), without necessarily transform it. By the way they integrate any type of criteria, these procedures seem to better afford to move to a judicious compromise rather than optimum, often obsolete.

Example: the delay parameter and its importance for VoIP service and streaming service. It is clear that delay is critical for VOIP service, but at the same time it is less important for the streaming considered as an average important criterion. According to [9][10], we have ordered the nonfunctional parameters according to their importance by service and the result is presented in Table IV. this helps us to make a weight over each parameter by service to respond and offer the needed Qos.

Services Parameter	packet lost	Delay	Jitter	Availability
VOIP	3	1	2	4
Videophone	3	1	2	4
Telephony	3	2	4	1
Multimedia	2	1	4	3
VOD	3	1	4	2
VPN	3	1	4	2
Real time data	2	1	4	3
Streaming	1	3	4	2

TABLE IV: Importance of SLA's parameter by service

Our model will allow us to compare between two proposed allocations, it takes into account the functional aspects that are expressed in the SLA, as well as nonfunctional (which are related to Qos) and express precisely the client's needs.

D. Provider Utility

The interest of the physical network provider is the monetary gain it can get in instantiating a virtual network. We identify two types of gains, one relating to the SLA and trade agreements between him and the client, the other to the economic aspects of energy that induces budgetary savings.

a) SLA's gain: We have previously discussed and modeled trade relations between Client and Provider. It expresses the gains generated by the satisfaction of customer needs from the SLA, the allocation of resources which are deployed to answer to nonfunctional customers needs in terms of quality of service engenders a gain. It is subject to the ability of the operator to provide this quality for a certain part of times.

b) Energy saving: Today, on each computer entity (hardware), a provider should take into account the costs associated with its use, so the ecology and earnings can reach a common goal. Our model takes into consideration the energy saving and expresses this aspect, putting the physical node in standby if we don't really need to instantiate virtual node over it, then carrying about the Qos. It means that as soon as we can choose to instantiate a virtual node over two physical nodes, we instantiate it on the already used one.

IV. Implementation and Test

A constraint is simply a logical relation among several unknowns (or variables), each taking a value in a given field.The algorithm "backtrack" is a blind search of the solution by experiment sets of variable assignments to find a solution. The complexity of the backtracking algorithm depends on the number of solutions to trying, for a CSP the complexity evolved on an exponential manner depending on the number of variables and the size of their respective fields.

The implementation is based on the constraints defined with the java library *Choco* which also embed a constraint solver. This allows us to express the graph for each network separately and the Qos constraints that characterizes each one, taking into account the functional and nonfunctional characteristics of each network expressed in SLA. Our program, thanks to the CSP solver of Choco, returns the most optimal solution to our resource allocation problem.

For the implementation of our project, we have developed the CSP module. First, we express the characteristics of virtual networks in form of CSP. The CSP must take into account several points such as:

- Satisfying the functional constraints.
- Then satisfying the nonfunctional parameters to the best.
- Eventually saving energy by allocating virtual networks to physical networks with nodes already instantiate, and pausing the rest of the nodes.

Our solution is a centralized solution, so our CSP is aware of all of the physical network resources allocated or not, a cluster is set up with all the necessary informations. We define two types of constraints, Capacity-Constraint and Projection-Constraint, which represent respectively the functional and nonfunctional constraints.

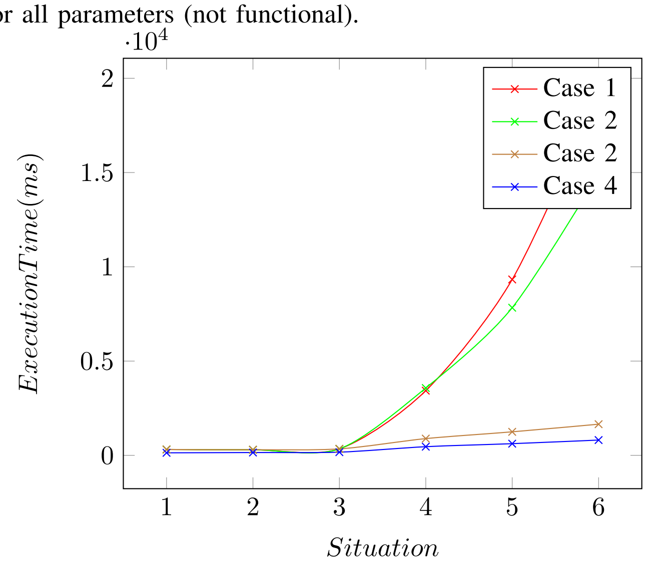
Previously, Fig. 2 presents different types of Qos, depending on the service provided by its virtual network. We propose a clear and simple method to express this difference and the importance of the order of the main nonfunctional parameters that must take into account a possible solution to provide the most optimal allocation. For this, each parameter has a weight. A weight depends on the service and its corresponding SLA.

To save energy, we joined a global score function which is equal to the score of each parameter multiplied by its weight. Functional Scoring: The score of the functional parameters is a global value that is equal to the number of physical routers allocated to whatever service or settings multiplied by their weight. Nonfunctional Scoring: The score of the Nonfunctional parameters is equal to the sum of the individual scores for each nonfunctional parameter multiplied by the value of its weight.

In conclusion, the global score is the aggregation of scores (functional and nonfunctional). Thus, to limit the abusive use of physical routers, an objective function has been implemented in order to minimize the global Score.

We present our tests in this section in order to validate our approach. These tests are based on the execution time. The tests were performed on a HP-ProBook 4530s machine with an Intel Core i5-2430M CPU 2.40 GH processor with a memory of 4 gigabytes and operating system Windows 7 Professional 64-bit. We chose to simulate generic physical routers with the same CPU and memory, but with a different rate on each link. For a virtual router, CPU and memory capacity are equal (between virtual routers) but bandwidths are different (for each virtual link).

We varied the number of constraints following this conditions. So, in the first case the weight of a service is the weight of a VOIP network. In the second one the weight of a service is 1 for all nonfunctional parameters. In the third case we apply a weight of a service equal to 0 for all nonfunctional parameters and in the final case the weight of a service is 0 for all parameters (not functional).



This figure allows us to have a clear view of the differences between the execution times. So, it is obvious that the case 1 has the greatest value, when our constraints are applied to the maximum for response to Qos needed by the virtual network instantiate service. This increases the time resolution because the search space is bigger.

The combination exposure of the solution and all set of research leads us to believe that our solution does not manage to scale. But some research areas are inaccessible because the constraints are not respected (because the algorithm operates without a backtrack down to the sub-tree when a constraint is violated), and other exploration strategies can be used to obtain better results.

We are in the process of testing the performance of the solution and the first results are encouraging and promising.

V. Conclusion

In this project, we realized a solution based on CSP to dynamically allocate resources in virtual networks with the possibility of reconfiguration. We identified the problem as a problem of graph matching. One of the methods to solve combination problems is to express it in a constraint solver. We used this mathematical tool to formulate the resource allocation problem, taking into account the Qos and energy saving requirements. The code was optimized and some advanced features of Choco where used. This resulted in a shorter execution time. It remains to find a solution to the problem of links assigning, trying to implement a load balancing between the virtual links (possibly machines) according to the SLA established by the provider.

References

[1] :A. Haider, R. Potter, et A. Nakao, « Challenges in resource allocation in network virtual-ization », in 20th ITC Specialist Seminar, 2009, vol. 20, p. 22.

[2] :Y. Zhu et M. Ammar, « Algorithms for assigning substrate network resources to virtual network components », in Proc. IEEE INFOCOM, 2006.

[3] : J. Lu et J. Turner, « Efficient mapping of virtual networks onto a shared substrate », Washington University, USA, Tech. Rep, 2006.

[4] : I. Houidi, W. Louati, et D. Zeghlache, « A Distributed and Autonomic Virtual Network Mapping Framework », in Autonomic and Autonomous Systems, International Conference on, Los Alamitos, CA, USA, 2008, vol. 0, p. 241-247.

[5] : M. Chowdhury, M.R. Rahman and R. Boutaba. « Vineyard: virtual network embedding algorithms with coordinated node and link mapping », in Networking, IEEE/ACM Transac-tion on 2012.

[6] : Farooq Butt, Nabeel and Chowdhury, Mosharaf and Boutaba, Raouf, «Topology-Awareness and Reoptimization Mechanism for Virtual Network Embedding ». Springer Berlin / Heidelberg 2010

[7] : N. M. M. K. Chowdhury and R.Boutaba, « Network virtualization: state of art and research challenges », Communications Magazine,2009.

[8] Belbekkouche, A., Hasan, M., & Karmouch, A. (2012). Resource discovery and allocation in network virtualization.

[9] :C. Desrosiers, « Le matching de graphes: une introduction ». Ecole Polytechnique de Montréal, mars-2006.

[10] : Long-Tae Park Kyungpook Nat. Univ., Taegu, « Management of service level agreements for multimedia Internet service using a utility model ». Communications Magazine, IEEE, 2001

[11] : Emmanuel Marilly, Olivier Martinot, Hélène Papini, Danny Goderis, « Service Level Agreements:a main challenge for Next Generation Networks» IEEE Communication, 2002.

[12] : Fuad Aleskerov, Denis Bouyssou, Bernard Monjardet, « Utility Maximization, Choice and Preference», Springer-Verlag Berlin Heidelberg, 2007