An Method of Service Composition with Optimal Resources Allocation in Software-Defined Networking

Olena Tkachova, Ievgen Duravkin
Infocommunication engineering department
Kharkov National University of Radio Electronics
Kharkov, Ukraine
olena.tkachova@nure.ua, ievgen.duravkin@nure.ua

Hassan Mohamed Muhi-Aldeen
Telecommunication system department
Odessa National Polytechnic University
Odessa, Ukraine
hassan_mohammed88@yahoo.com

***Abstract* — The work is devoted to developing the complex methods of service composition and provision in Software-Defined Networking. The service provision mechanism in the multiservice network under Software-Defined Networking concept is analyzed. An overview of existing service provision methods that are widely used in distributed computing networks is proposed. The solution of the multicriteria task of services selection and composition with optimal resources allocation are proposed in the work. The model of complex service provision in Software-Defined Networking and algorithm of service composition with optimal resource allocation are also represented.**

Keywords — complex service; Software-Defined Networking; QoS criterion; resources allocation; service provision

I. Introduction

The Software-Defined Networking (SDN) is one of the more perspective technology of organization and management of modern info-communication networks. The main idea of SDN is separation the management functions from data transfer functions [1-3]. In accordance with the SDN concept, all management and monitoring functions are transferred to a separate central unit - the controller [3, 4]. The controller generates an optimal solution for services provision based on information about the network topology and current characteristics. Controller together with management modules forms the control system. SDN network infrastructure includes both physical and virtual elements. This allows significantly improve the performance and scalability of the network. Rapid deployment, flexible combination of multiple network functions on a single servers' platform, automatization of administrative procedures are the main advantages of SDN [5].

One of the main tasks of the control system is supporting the availability and reliability of services. The solution of the task usually based on structural network redundancy and resources reservation. Such approach leads to an unjustified increasing the cost of provided services without supporting necessary level of QoS. The developing the methods of complex services provision with resources optimization that allows supporting required QoS characteristics is a relevant task.

The analyses of service composition mechanisms and resource distribution effectivity and methods of increasing the quality of complex services in distributed network are proposed in work [6-9]. The work [6] is devoted to solving the problem of search services and the allocation of resources in order to minimize the total time of service provision. In [6, 7] introduced a new criterion for assessing the quality of services - processing requests delayed consideration, arising from the geographical remoteness of the resource from the consumer. The method based on genetic algorithm to choose the resources taking into account the geographical distribution. In [8] the resources allocation is based on QoS characteristics such as the response time, cost, availability, and reliability. The task is formulated as an optimization problem with constraints on the QoS metrics. Various schemes such as sequential execution, cyclical, parallel and selective of service composition are considered in [9]. The resulted service selection is based on such metrics as reliability and availability.

The above approaches are widely used in the management of a distributed network infrastructure. However, their direct application in the SDN is difficult for several reasons related to the high cost of network resources in the allocation process, lack of registration of computing elements, a limited number of analyzed QoS metrics and possible solutions of resources allocation. Thus, the solution of the multicriteria task of services selection and composition with optimal resources allocation is relevant. An effective solution to the task allows increasing availability and ensures the required quality of service in SDN. The model of complex services provision with a flexible choice of QoS criterions in SDN, and method of optimal resources allocation based on QoS requirements are proposed in the work.

II. The Model Of Services Provision In Software-Defined Networking

Multiservice networks, constructed under the SDN concept, often have a complex structure; computation and forwarding elements such network are geographically distributed. The controller executes the gathering and analyses of information about current network topology and network characteristics on the control plane. OpenFlow switches via OpenFlow Protocol [10, 11] provide the information about active network components. The controller has no complete map of the network infrastructure; the network map includes an only limited set of logical networks called VLANs (Virtual Local Area

978-1-5386-0983-5/17/$31.00 ©2017 IEEE

Network, VLAN). VLANs that are created dynamically and forming abstractions as overlay networks [3].

In such case the process of resources allocation for service provision includes the next step [1, 2, 12]:

(1) End user sends a request for service provision. Regarding Service Layer Agreement (SLA) each service has his own requirements to quality of service provision: response time, throughput, reliability, etc. The requirements are taken into account in Type of Services (ToS) fields.

(2) Requests from end users with different intensity arrive at the ports of OpenFlow switches. OpenFlow switch parses the request, analyses ToS, Source, and Destination fields in the header of the request and forwards the request to the Proxy/Web- server or computation node directly if IP address of those devices is matched as a destination. In case if any field was not mashed OpenFlow controller forward request to the controller.

(3) Controller performs a service orchestration regarding end user requirements. If service is complex and includes a set of atomic services the task scheduler analyses QoS requirements for each atomic servers, put the right order of atomic services and find the proper resource for service provision - the controller analyses the current network resources, determines what compute node can satisfy the users' requirements.

(3a) Find the logical network where the services are allocated. In the case of using the virtual resources, such characteristics as transparency interaction and interoperability between VM, good compatibility with hardware that impact to further availability of service. When the appropriate software is found the compatible hardware resources selected.

(3b) The hardware resources can be allocated both in the same LAN with the user or remotely LAN. The choice of LAN for service provision based on the requirements for quality of provided services and information about current characteristics of computation nodes.

(4) Controller chooses the optimal route (OpenFlow Switch-to-OpenFlow Switch) and creates a VLAN after the resource for service provision is founded.

(5) Controller generates reply message with information required to access the service – computation resources allocation (IP address of computation nodes).

The structural scheme of SDN architecture and allocation of resources for service provision process are shown on Fig.1.

The proposed fragment of SDN (Fig. 1) is divided into tree LANs: LAN1, LAN2 and LAN3. Each LAN includes a tree logical groups of network nodes – group of end users' nodes, group of compute nodes combined the mix of physical (PM) and virtual computation resources (VM), group of servers (Web, Mail, FTP, etc.).

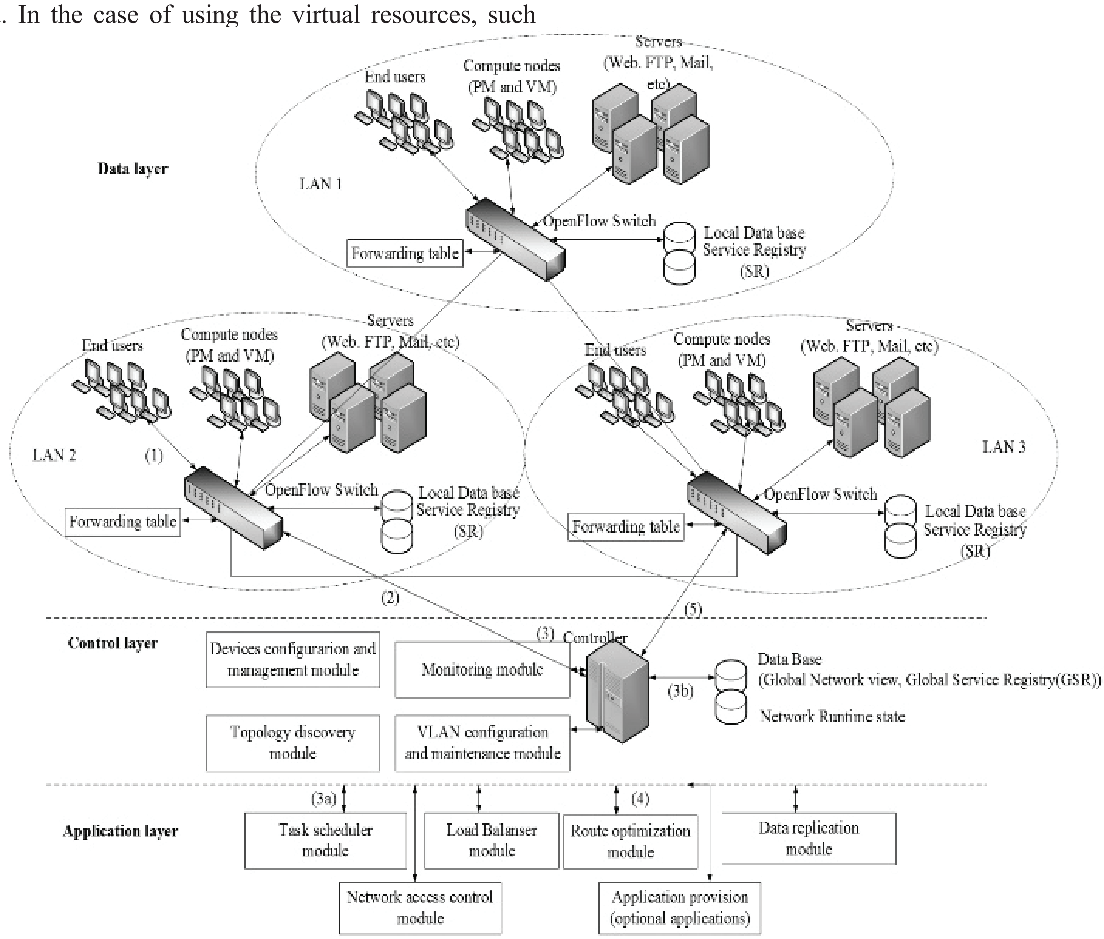

Fig. 1. Model of resource allocation process for SDN architecture

The information about services that can be delivered by the nodes stores in Service Registry (SR). In this case, each LAN has its own *SR*. The global information about network, in particular, the information about services that provided by different LANs stored in Global Service Registry (GSR) and placement on Control Layer [9].

Assume that a set of end users is defined as $U = \{ u_1, u_2, ..., u_m \}$, a set of services that can be provided in SDN defined as $TS = \{ TS_1, TS_2, ..., TS_n \}$. Amount of services and their QoS is predetermined by SLA [9].

The information about the services stored in local Data Base in SR and in global Data Base in GSR. Amount of compute nodes within LAN is defined as $N = \{ n_1, n_2, ..., n_m \}$ and includes a set of virtual machines N_{VM} and physical machines N_{PM}, $N \subset \{ N_{PM}, N_{VM} \}$.

The information about physical and virtual nodes is saved in local Data Base on Data Layer and the information about virtual machine is centralized and stored in global Data Base on Control Layer.

The policy regulated the service provision process based on chosen QoS metrics. In accordance with IETF recommendation for service provision, the next metrics are identified: cost efficient, performance utilization, response time, delay variation, reliability. In the case of using SDN architecture information about each service stored in service registry characterized by follows:

$$\begin{aligned} & SR_j[TS_i] = \{loc(TS_i), d(TS_i), c(TS_i), t(TS_i), p(TS_i), r(TS_i)\}; \\ & \forall tk_i, loc(TS_i) = \bigcup_k loc(ts_{ki}); \\ & d(TS_i) = \max_k d(ts_{ki}); \\ & c(TS_i) = \sum_{k=1}^{p} c(ts_{ki}); \\ & p(TS_i) = \min_k p(ts_{ki}); \\ & r(TS_i) = \min_k r(ts_{ki}); \end{aligned} \quad (1)$$

where SR_j is a registry where stored information about service; TS_i is complex services that include combination of monatomic $TS_i = \{ ts_{1i}, ts_{2i}, ..., ts_{ki} \}$, $loc(TS_i)$ is IP address of hardware resource that provided TS_i service; $c(TS_i)$ is a service provision cost efficiency; $t(TS_i)$ is average response time of service provision; $p(TS_i)$ is a performance of compute node required to the service execution; $d(TS_i)$ is a delay variation during service provision. Two factors affect to the value of delay variation – computation delay and transmission delay; $r(TS_i)$ is a reliability of service provision; ts_{ki} is a monatomic service.

The equation (1) shows the characteristics in the case of atomic services static placement of computing resources. In the case of virtual computing resources change dynamically, then compute node performance in the provision of services is calculated based on the following statements:

$$p(TS_i) = p(N_{PM}) - \sum_{j=1, j \neq i}^{k} p(TS_j), \quad (2)$$

where $p(N_{PM})$ is the nominal value of the compute node performance; $p(TS_p)$ is a compute node performance that are required for *p*-th complex service.

The complex service characterized by a set of metrics or criterions. The integral function is used for estimate the value of each metrics. The integral function represents a weighted sum of criterion multiplied by corresponding coefficients:

$$\begin{aligned} IF(TS_i) = & K_1 d(TS_i) + K_2 c(TS_i) + K_3 t(TS_i) + \\ & + K_4 p(TS_i) + K_5 r(TS_i), \end{aligned} \quad (3)$$

where $K_1, K_2, ..., K_5$ is value of each criterion. The decision about service composition is based on value of each criterion and end user requirements.

Each *SR* has its own responsibility of services provision that determined by the network service provider and can be changed by devices configuration and management module. The global information about services stored in GSR $GSR[TS]$ can be represented by follow:

$$\begin{aligned} & GSR[TS] = \{ SR_1[TS], SR_2[TS], ..., SR_q[TS] ..., R_n[TS] \} \\ & \forall SR[TS], SR_i[TS] = \{ TS_1, TS_2, ... TS_i \} \end{aligned} . \quad (4)$$

The model of SDN architecture where is taking into account the location of the service registry is shown in Fig. 2. The set of services that are available in LAN1 is stored in SR1[TS], LAN2 – SR2[TS], LAN3 – SR3[TS], global service registry is matched as GRS[TS].

TS_i service may include a set of monatomic services. To provide the services with guarantee quality the next conductions should be taking into account in resources allocation process:

III. Service Provision Method With Optimal Resource Allocation

The set of complex service realizations by the composition of monatomic services is a finite [13]. The QoS resulted complex service should corresponding SLA requirement:

$$\begin{aligned} & F_{QoS}(TS_i) = v_1 \cdot ts_{1i} + v_2 \cdot ts_{2i} + ... + v_k \cdot ts_{ki} = \sum_{k=1}^{n} v_k \cdot ts_{ki} \to \max; \\ & F_{QoS}(TS_i) \geq F_{SLA}(TS_i). \end{aligned} \quad (5)$$

$$\begin{aligned} F_c(TS_i) = & c_1 \cdot ts_{1i} + c_2 \cdot ts_{2i} + ... + c_k \cdot ts_{ki} = \\ & = \sum_{k=1}^{n} c_k \cdot ts_{ki} \to \min; \\ & F_c(TS_i) \leq F_{nom}(TS_i), \end{aligned} \quad (6)$$

where $F_{QoS}(TS_i)$ is resulted quality of complex service; $F_{SLA}(TS_i)$ is quality of provided services corresponded SLA requirements;

$v_1, v_2, ..., v_k$ is a value of each monatomic service, $c_1, c_2, ..., c_k$ is a cost of each monatomic service; $F_c(TS_i)$ is current meaning of complex service provision; $F_{nom}(TS_i)$ is the nominal cost corresponded SLA requirements.

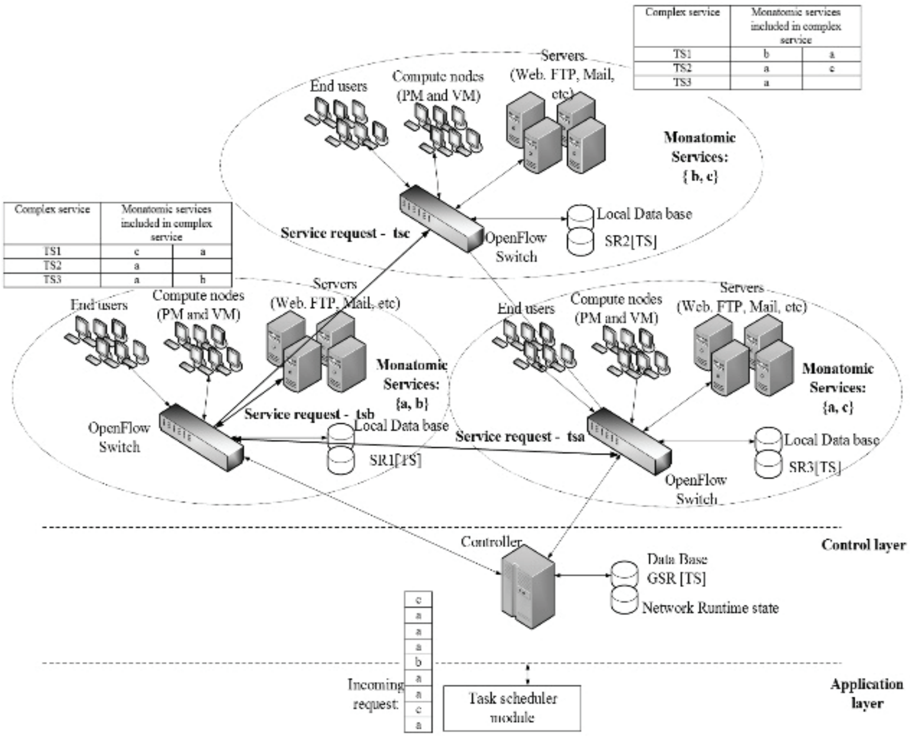

Fig. 2. Model of service orchestration process

In final decision about resource allocation, the controller selects the set of services with minimal cost and maximal QoS. The monatomic services in (2) can be replaced by other monatomic services with value that is more appropriate.

The flow-chart of proposed algorithm of complex service provision with optimal resource allocation is depicted in Fig.3.

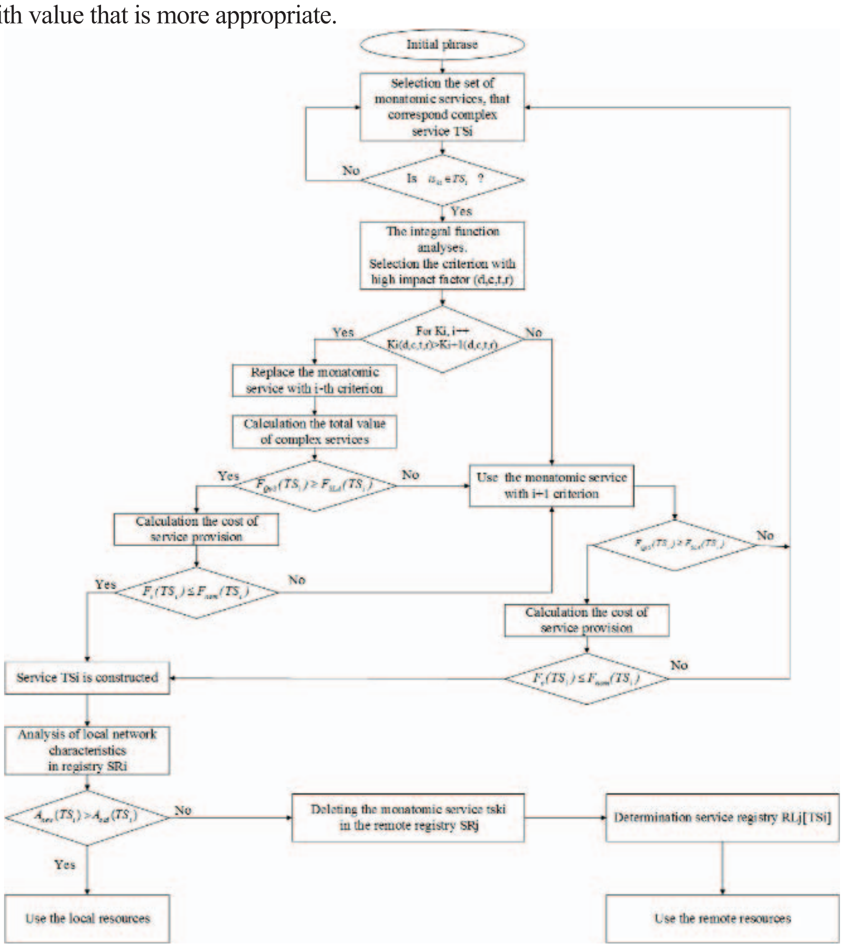

Fig. 3. Proposed algorithm of complex service provision

The proposed algorithm service provision with optimal resource allocation includes next steps:

1) Definition of the monatomic services that include required complex service.

2) Definition the optimal composition of monatomic services based on value integral function.

3) Reducing the set of obtained composition according to the value of criterion defined in (2) and (3) and local optimum of resources allocation definition.

IV. Conclusion

Application SDN concept in multiservice distributed networks can significantly improve the quality of end-user services. The improvement is achieved by forming a centralized management system and reduces the load on the data network by separating the control plane from the data plane.

However, along with the many benefits remain a number of unresolved problems. One of these tasks is to ensure the required level of service availability, the effective solution that is the use of complex mechanisms of services and resource allocation.

The proposed method of service provision with optimal resources allocation is the solution of the multicriteria task, where the take into account a contribution of each monatomic services in complex service. Application of the algorithm allows providing the required QoS of the requested service due to optimal resource allocation both regarded local service registry or remote service registry.

References

[1] F. Akyildiz, “A roadmap for traffic engineering in SDN-OpenFlow networks”. *Computer Networks*, vol. 71, pp. 1-30, 2014

[2] “Architecture SDN. Open Networking Foundation”, [On-line], Available: https://www.opennetworking.org/ 2014.

[3] T. Egawa. *SDN standardization Landscape from ITU-T Study Group*. ITU Workshop on SDN. 2013

[4] Qinghong Zhong, Ying Wang, Wenjing Li and Xuesong Qiu. "A min-cover based controller placement approach to build reliable control network in SDN", in *2016 IEEE/IFIP Network Operations and Management Symposium (NOMS)* , 2016, pp. 481-487. DOI: 10.1109/NOMS.2016.7502847

[5] Daphne Tuncer, Marinos Charalambides, Stuart Clayman and George Pavlou. "Adaptive Resource Management and Control in Software Defined Networks", in *IEEE Transactions on Network and Service Management*, vol. 12, issue 1, pp. 18-33, 2015. DOI: 10.1109/TNSM.2015.2402752

[6] S. Liu, Y. Wei, K. Tang, A. K. Qin and X. Yao, "QoS-aware long-term based service composition in cloud computing," *2015 IEEE Congress on Evolutionary Computation (CEC)*, Sendai, 2015, pp. 3362-3369.. DOI: 10.1109/CEC.2015.7257311

[7] Qin Shengjun, Chen Yan, Mu Xiangwei. “An Optimal Service Selection with Constraints Based on QoS.” *Physics Procedia*. vol. 25, pp. 2050–2057, 2012.

[8] S. Sezer, S. Scott-Hayward and P. Chouhan. “Are we ready for SDN? Implementation challenges for Software-Defined Networks”, in *IEEE Communications Magazine,* vol. 51, no. 7, pp. 36–43, 2013. DOI: 10.1109/MCOM.2013.6553676

[9] M. E Michael and P. Singh Munindar. “Agent-based trust model involving multiple qualities”. *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, 2005, pp. 519–526.

[10] N. B. Mabrouk, N. Georgantas and V. Issarny, "Set-Based Bi-level Optimisation for QoS-Aware Service Composition in Ubiquitous Environments," *2015 IEEE International Conference on Web Services*, New York, NY, 2015, pp. 25-32.

[11] R. Jain and S. Paul, "Network virtualization and software defined networking for cloud computing: a survey," in *IEEE Communications Magazine*, vol. 51, no. 11, pp. 24-31, 2013. DOI: 10.1109/MCOM.2013.6658648

[12] “OpenFlow Switch Specification. Version 1.3.0 (Wire Protocol 0x04)”. *Open Networking Foundation,* [On-line] Available: https://www.opennetworking.org/images/stories/downloads/sdn-resources/onf-specifications/openflow/openflow-spec-v1.3.0.pdfGerardo

[13] H. Abdalla and D. Ageyev, "Application of multi-layer graphs in the design of MPLS networks," *Proceedings of International Conference on Modern Problem of Radio Engineering, Telecommunications and Computer Science*, Lviv-Slavske, 2012, pp. 336-337.

[14] D. Ageyev and A. Ignatenko, "Describing and modeling of video-on-demand service with the usage of multi-layer graph," *Proceedings of International Conference on Modern Problem of Radio Engineering, Telecommunications and Computer Science*, Lviv-Slavske, 2012, pp. 340-341.

[15] D. V. Ageyev and F. Wehbe, "Parametric synthesis of enterprise infocommunication systems using a multi-layer graph model," *2013 23rd International Crimean Conference "Microwave & Telecommunication Technology"*, Sevastopol, 2013, pp. 507-508.

[16] A. A. Ignatenko and D. V. Ageyev, "Structural and parametric synthesis of telecommunication systems with the usage of the multi-layer graph model," *2013 23rd International Crimean Conference "Microwave & Telecommunication Technology"*, Sevastopol, 2013, pp. 498-499.

[17] D. Ageyev, A. Ignatenko and F. Wehbe, "Design of information and telecommunication systems with the usage of the multi-layer graph model," *2013 12th International Conference on the Experience of Designing and Application of CAD Systems in Microelectronics (CADSM)*, Polyana Svalyava, 2013, pp. 1-4.

[18] Q. Wu, F. Ishikawa, Q. Zhu and D. H. Shin, "QoS-Aware Multigranularity Service Composition: Modeling and Optimization," in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 11, pp. 1565-1577, Nov. 2016. DOI: 10.1109/TSMC.2015.2503384