Intent Profiling and Translation Through Emergent Communication

Salwa Mostafa*, Mohammed S. Elbamby^,
Mohamed K. Abdel-Aziz^, and Mehdi Bennis*

*Centre for Wireless Communications,University of Oulu, Finland

^Nokia Bell Labs, Espoo, Finland

Agenda

1. Introduction
2. Motivation
3. System Model
4. Problem Formulation
5. Proposed Framework
6. Simulation Model and Results
7. Conclusion

The work is funded by the project SCENE (G.A no. 00164501.0)

Introduction

- 5G and beyond networks support various services and applications.
- Traditional manual configuration and management cannot support the stringent and diverse demands of services and applications.
- Network service providers are driven to move toward automated network and service management.
- IBN introduces a simple and efficient autonomic and autonomous way to configure and manage networks.
  - Understanding what network users/applications want and what network operators can offer to optimize the alignment of the network operations with service/application needs.
- Network capabilities are offered through network slices, which allows the creation of multiple end-to-end logical networks on a shared physical and virtual infrastructure.

Augmented/Virtual Reality (AR/VR)

Cloud Gaming

Vehicle-to-everything (V2X)

Smart Industry

IBN: Intent-based networking

  The work is funded by the Nokia project SCENE (G.A no. 00164501.0)

Introduction

Intent is defined as a high-level and abstract description of the network services.

- IBN provides a complete life cycle to the intent, which takes place over five main steps to form a CLA.
  1. <u>Intent profiling</u>: the user interacts with the network and they collaborate towards expressing a meaningful intent for the network. (i.e., what the user expects as an outcome from the network or service).
  2. <u>Intent translation:</u> the expressed intent is converted into network policy and low-level configuration to the network functions and devices.
  3. <u>Intent resolution</u>: solves the potential conflict between independently submitted intents.
  4. <u>Intent activation:</u> activates the network functions and services to provide the intended customized service.
  5. <u>Intent assurance</u>: indicates the success of the deployed intent in the network throughout its dynamic life cycle.

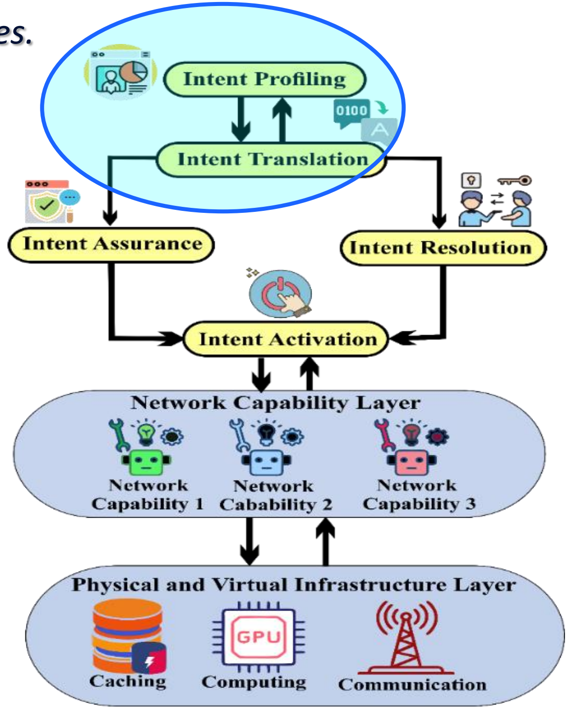

Fig.1 Interaction of the main IBN components.

CLA: closed-loop automation.

The work is funded by the Nokia project SCENE (G.A no. 00164501.0)

NOKIA BELL LABS

Introduction

Intent profiling

- IETF classified the intent profiling based on the intent type, the intent scope, and the network scope.
- *State of the art:*
  - The studies in [1–3] proposed human-friendly interfaces such as GUI and templates:
    - ✓ Network users choose, from a drop-down menu and template filling choices, what they want in their requested services.
    - ✓ Do not allow network users to express demands not supported as a possible option.
  - The studies in [4,5] proposed NLP-based intents:
    - ✓ Network users write down/orally express what they expect from a network service in a human language, and then the network interprets and translates it to a network configuration.
    - ✓ NLP requires a specific grammar format so that important information can be detected and a correct intent is expressed.
  - The studies in [6-8] proposed a domain-specific language such as Nile language and NEtworking MOdeling (NEMO)
    - ✓ Express the intent in readable and abstracted technical details and declare intent with information about network services and resources.
    - ✓ The language requires technical users (i.e., network operators/ administrators), which makes it not general enough for all types of applications.

IETF: Internet Engineering Task Force. GUI: graphical user interface. NLP: natural language processing.

The work is funded by the Nokia project SCENE (G.A no. 00164501.0)

Introduction

Intent translation

- *State of the art:*
  - The studies in [1], [6] proposed a template/blueprint-based translation:
    - ✓ Relies on pre-defined configuration files that contain the main configuration set up to network devices and functions.
    - ✓ Low efficient utilization to network resources and handling multiple intents.
  - The study in [9] proposed NSD–based translation:
    - ✓ contains deployment templates that are directly used by an orchestrator to manage, configure, and deploy a network service.
  - The studies in [10,11] proposed keyword-based translation:
    - ✓ Identifies the associated keywords with each intent and maps them to specific rules or template policies.
  - Traditional mapping and translation mechanisms mentioned above do not guarantee an accurate translation.

NSD: network service descriptors.

The work is funded by the Nokia project SCENE (G.A no. 00164501.0)

Motivation

- Most of the aforementioned work focused on human-to-machine scenarios and no attention has been paid to machine-to-machine scenarios.
- H2M interaction humans can articulate their needs through verbal, GUI, and drop-down menus.
- M2M interactions require learning a common language (i.e., non-verbal) for applications to express their intents.
- Existing literature assumes intents are easily expressed as network QoS requirements by applications.
  - ✓ Applications have to be able to express their needs in their domain language.
  - ✓ This condition is complex since it requires networks to learn how to interpret these domain languages which is not practical nor scalable.
- Confining intents to pre-defined lists of generic intents that the network can understand restricts the potential of future networks' ability to support the variety of applications, each having its own needs/goals.

- <u>Proposed</u>: a framework based on emergent communication for intent profiling:
  - ✓ Applications express their abstract QoE intents to the network through emergent communication messages.
  - ✓ The network learns how to interpret these communication messages and map them to network capabilities (i.e., slices) to guarantee the requested QoS.

H2M: human-to-machine. **M2M:** machine-to-machine

The work is funded by the Nokia project SCENE (G.A no. 00164501.0)

NOKIA BELL LABS

Motivation

Deep Emergent Communication

- Human intelligence relies on the ability to communicate and use language to interact in complex social situations.

- Deep neural agents learn their own language for cooperation/coordination and facilitate knowledge transfer.

- **Definition:** The field of research wherein agents with no pre-existing communication protocols engage in interactions with the goal of training agents that can bootstrap a communication protocol that acts as a shared language.

The work is funded by the Nokia project SCENE (G.A no. 00164501.0)
NOKIA BELL LABS

System Model

- We consider a time domain network system consisting of :
  - **N IIoT-MDs:** Each runs a different application and generates an intent instance $I_{n,t}$ requesting a certain QoE levels .
  - **A network** with **M** network slices with different capabilities (i.e., communication, computing).
- The requested QoE (i.e., intent) is then mapped to a QoS expressed in a communication $t_{n,req}^{up}$ and computation $t_{n,req}^{comp}$ deadlines.
- The network assigns a network slice m with capabilities $(R_m; f_m)$ to each IIoT-MD, where R_m: uplink rate in bps, f_m: CPU cycles/second.
- Based on the allocated network slice **m**, for IIoT MD *n*
  - Uplink time: $T_n^{up}(m) = \frac{A_n}{R_m}$
  - Computation time: $T_n^{comp}(m) = \frac{A_n \times C_n}{f_m}$

A_n: the generated application instance task size of the IIoT-MD *n*.

C_n: no. of required CPU cycles per bit of the IIoT MD-generated application task.

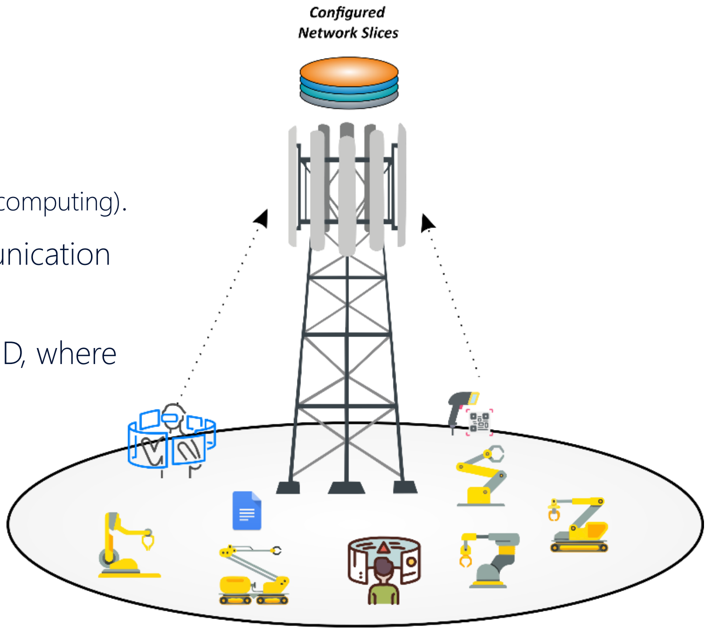

Fig.2 System Model.

IIoT : Industrial Internet of Things. MDs : mobile devices. QoE : quality-of-experience. QoS : quality-of-service.

The work is funded by the Nokia project SCENE (G.A no. 00164501.0)

NOKIA BELL LABS

Problem Formulation

We divide the intent profiling and translation problem into two subproblems.

1- <u>For the intent profiling subproblem:</u>

- Each IIoT MD communicates its intent to the network through a communication message chosen from a vocabulary set U.
- The IIoT MDs aim to be associated with network slices that can satisfy their intents.
- Each IIoT MD n optimizes a mapping function expressed as

$$f_n(I_{n,t}) : I_{n,t} \rightarrow u_{n,t} \forall\, n \in N, t \in T, u \in U \qquad (3)$$

that maps each intent instance $I_{n,t}$ to a communication message $u_{n,t} \in U$ to *maximize the successful number of intents to communication messages association.*

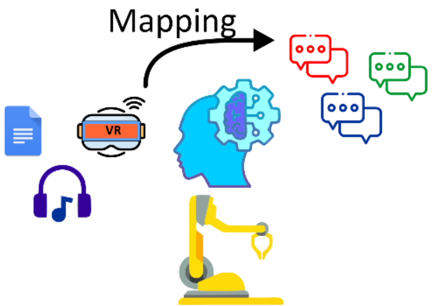

Fig.3 Intent to message mapping.

The work is funded by the Nokia project SCENE (G.A no. 00164501.0)
NOKIA BELL LABS

Problem Formulation

2- For the intent translation subproblem:

- The network slices are deployed based on predefined network slice templates that can guarantee a certain QoS.
- We assume that the expressed intents are within the capabilities of the deployed network to guarantee QoS satisfaction.
- The network receives the communication messages from all IIoT MDs and optimizes a mapping function expressed as

$$g(u_t):\ u_t \rightarrow\ c_t \qquad (4)$$

maps the communication messages to network slices to *maximize the number of successful communication messages to network slices association.*

- $u_t \triangleq [u_1, u_2, \cdots, u_N]$ the received communication messages from all IIoT MDs at time slot t.
- $c_t = \ [c_1, c_2, \cdots, c_N]$: slices allocated to each IIoT MD at time slot t, where $c_n \in M$.

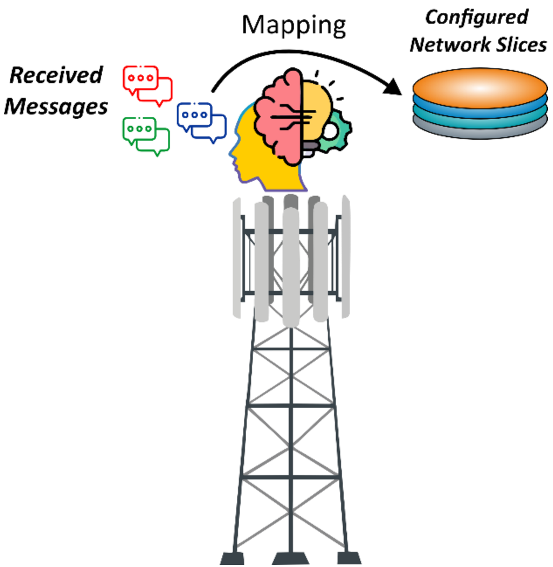

Fig.4 Messages to network slices mapping.

The work is funded by the Nokia project SCENE (G.A no. 00164501.0)
NOKIA BELL LABS

Problem Formulation

- *The objective :*

  maximize the number of successful associations of intents to network slices.

- The optimization problem can be stated as follows

$$\max_{\mathbf{Y},\boldsymbol{u}} \sum_{t\in\mathcal{T}} \sum_{n\in\mathcal{N}} x_{t,n}$$

subject to

$$C1: \sum_{m\in\mathcal{M}} y^t_{n,m} = 1, \;\; \forall\, n \in \mathcal{N},\; t \in \mathcal{T}$$

$$C2: x_{t,n} \in \{0,1\}.$$

$$x_{t,n} = \begin{cases} 1 & \text{if } \; t_n^{\text{comp}} \le t_{n,req}^{\text{comp}} \text{ and } t_n^{\text{up}} \le t_{n,req}^{\text{up}}, \\ 0 & \text{otherwise.} \end{cases}$$

$Y^t \triangleq [y^t_{n,m}]$: the $N \times M$ binary association matrix of network slices to IIoT MDs at time slot t.

- We consider the mapping is *successful* if the allocated network slice m characteristics can *satisfy the requested QoE*.

The work is funded by the Nokia project SCENE (G.A no. 00164501.0)

Proposed Framework

To solve the intent profiling and translation problem

- The main objective can be *maximized through maximizing the objective functions Eq. (3) and (4)* that maximize the successful mapping between intents to communication messages and communication messages to network slices.
- Unfortunately, both mapping functions Eq. (3) and (4) are unknown and hard to model explicitly mathematically.
- We propose an AI framework to approximate and solve the problem.
- The proposed AI framework utilizes ***cooperative MARL***, where the network and IIoT MDs are modeled as reinforcement learning agents.
- We adopt ***emergent communication technology*** to learn a communication protocol that provides a common ground between the intent expression and the network capabilities.

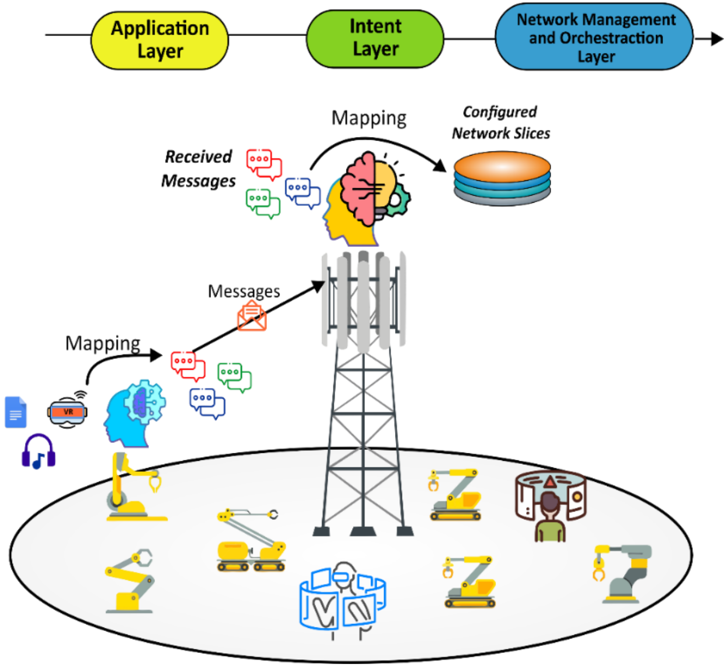

Fig.5 Proposed Framework.

MARL: multi-agent reinforcement learning.

The work is funded by the Nokia project SCENE (G.A no. 00164501.0)

Proposed Framework

- The MARL is described with a Dec-POMDP $P = \langle S, O, A, T, R, \gamma \rangle$:
  - S is the state space, O is the observation space, A is the action space.
  - $\mathrm{T}: S \times A \rightarrow \Delta(S)$ a non-deterministic transition function maps the state and action space to a probability distributions $\Delta(S)$ over S.
  - $R: S \times A \rightarrow R$ is a reward function, which maps the states and actions to a set of real numbers and γ is the discount factor.
- The agents communicate over an episode of maximum length T time instances.
- At time step t, the agent receives an observation O_t depending on its current state s_t and the previous action a_{t-1} then takes an action.

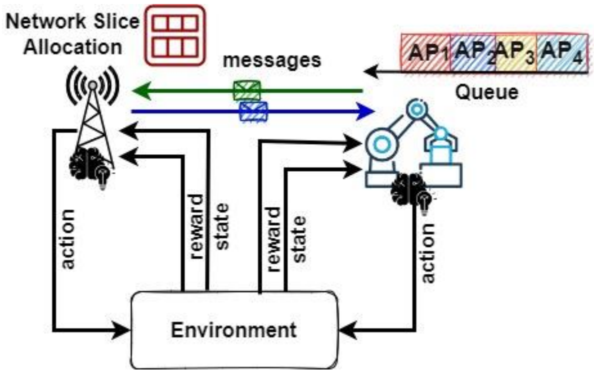

Fig.6 IIoT MDs and network are cooperative MARL.

MARL: multi-agent reinforcement learning. Dec-POMDP: decentralized partially observable Markov decision process.

The work is funded by the Nokia project SCENE (G.A no. 00164501.0).

Proposed Framework

- State Space :
  - For network: recent l uplink and downlink communication messages and the environment action

$$S^b = [U_t^n, \cdots, U_{t-1-l}^n, D_{t-1}^b, \cdots, D_{t-1-l}^b, a_{t-1}^b, \cdots, a_{t-1-l}^b]$$

  - For IIoT MD: recent l uplink and downlink communication messages and the generated intent instance

$$S^n = [U_{t-1}^n, \cdots, U_{t-1-l}^n, D_{t-1}^b, \cdots, D_{t-1-l}^b, I_{n,t}, \cdots, I_{n,t-1-l}]$$

- Action Space
  - Environment action: a_e represents the network slice allocation to each IIoT MD.
  - Communication action: a_c includes the uplink and downlink communication messages.
    - The uplink messages $U \in \{1,2,\cdots,M\}$.
    - The downlink messages $D \in \{0,1\}$ indicating the success or failure of the network slice allocation in satisfying the requested QoE.
- Reward Function:

$$R_n(t) = \begin{cases} +\rho & \text{if the intent is satisfied} \\ -\rho & \text{otherwise} \end{cases}$$

  - The team reward : $R(t) = \sum_{n\in N} R_n(t)$
- Note that the uplink and downlink communication messages are not pre-defined and the meaning associated with each message emerges through communication.
- To solve the above-formulated Dec-POMDP problem, we adopt the multi-agent proximal policy optimization (MAPPO) algorithm.

The work is funded by the Nokia project SCENE (G.A no. 00164501.0)

Simulation Model

- We consider a warehousing logistic area with a network and 5 IIoT MDs.
  - The network has ten network slices with different capabilities.
  - The applications supported in the system are either URLLC or eMBB.
  - The uplink messages set has a cardinality equal to the number of network slices supported on the system.

- The proposed framework is compared with the following baselines:
  - Perfect Knowledge: The network checks all possible network slices available that can guarantee the requested QoE to IIoT MD (i.e., intent). Then, it chooses one of them at random.
  - Random Assignment: The network allocates network slices to the IIoT MDs (i.e., intents) in a random way.
  - Self-Learning Slice Selection: The IIoT MDs learn to access the network slices without any prior assignment or communication with the network through interaction via RL.

URLLC: Ultra-reliable low latency communication. eMBB: enhanced mobile broadband

The work is funded by the Nokia project SCENE (G.A no. 00164501.0)

Simulation Model

Simulation Parameters

Parameters	Values
No. of IIoT MDs	5
No. of network slices	10
Tasks Size	100 - 500 bits
Tasks Computation Requirement	$1 \times 10^{2} - 5 \times 10^{4}$
Tasks Storage Requirement	200 - 600 bits
IIoT MDs Reliability Requirement	$1 \times 10^{-2} - 5 \times 10^{-5}$
Tasks Offloading Tolerance	$1 \times 10^{-2} - 5 \times 10^{-2}$ second
Tasks Computation Tolerance	$1 \times 10^{-2} - 5 \times 10^{-2}$ second
Probability of Task Arrival	1
Duration of episode	15

MAPPO Hyperparameters

Parameters	Values	Parameters	Values
Number of episodes	6000	Learning rate	10^{-3}
Minibatch size	64	Discount factor (γ)	0.99
GAE parameter ()	0.95	Clipping parameter	0.2
VF coeff. (c1)	0.2	Entropy coeff.	0.2
Optimizer	Adam	Optimizer epsilon	10^{-5}

The work is funded by the Nokia project SCENE (G.A no. 00164501.0)

NOKIA BELL LABS

Simulation Results

Fig.1 Normalized successful QoS translations versus the number of episodes.

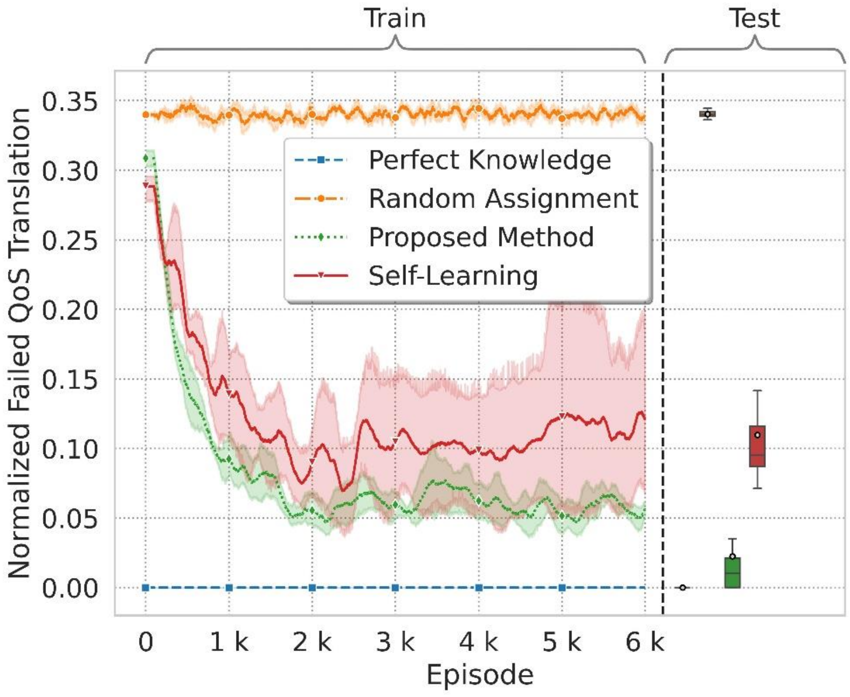

Fig.2 Normalized failed QoS translations versus number of episodes.

The work is funded by the Nokia project SCENE (G.A no. 00164501.0)

Simulation Results

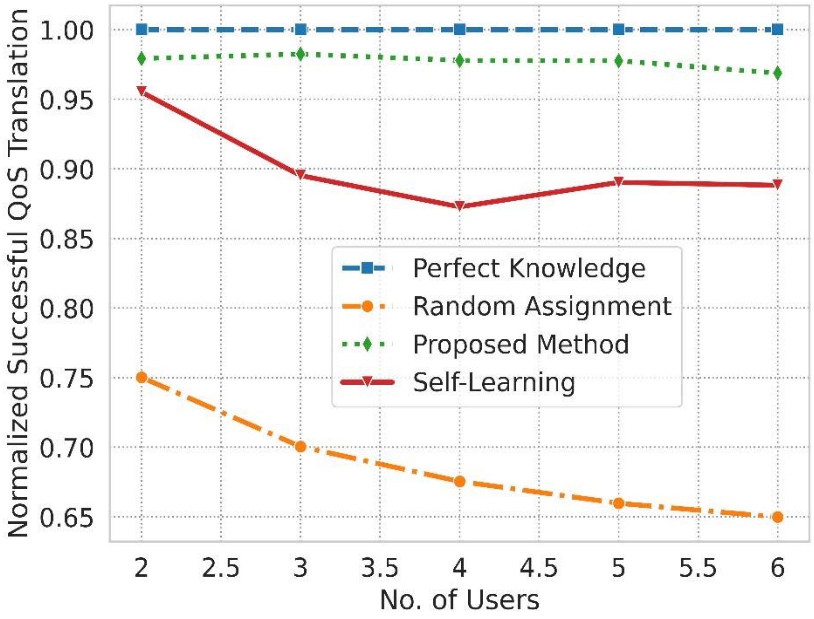

Fig.3 Normalized successful QoS translations versus the number of users.

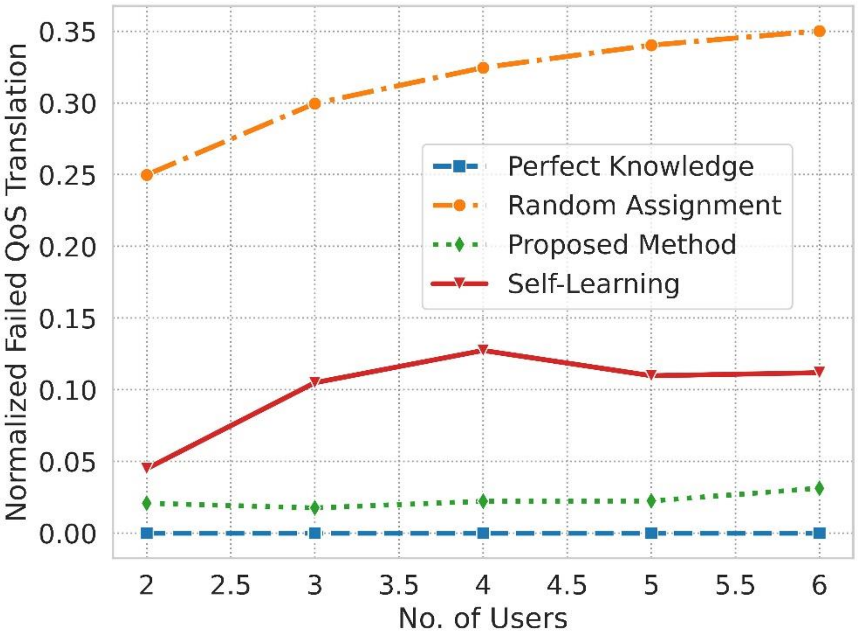

Fig.4 Normalized failed QoS translations versus the number of users.

The work is funded by the Nokia project SCENE (G.A no. 00164501.0)

Conclusion

- The problem of intent profiling and translation is investigated to provide a simple, efficient, and automated way to manage and operate intent-based networks.
- The proposed scheme leverages machine learning and emergent communication
  - IIoT MDs (i.e., applications) learn a policy to map their intents to communication messages to the network.
  - The network learns a policy to translate these messages to network resources (i.e., network slice allocation).
- The proposed scheme outperformed the random assignment and self-learning selection strategies during training and testing.
- The proposed scheme gives a very close performance to the perfect knowledge benchmark scheme.

  The work is funded by the Nokia project SCENE (G.A no. 00164501.0)

NOKIA BELL LABS

Q&A

Patents

1. A Method for Mapping Differentiated Application Requirements to Network Slices , Salwa Mostafa, Mohammed S. Elbamby, Mohamed K. Abdel-Aziz, Mehdi Bennis.

Conference Papers

1. Intent Profiling and Translation Through Emergent Communication , Salwa Mostafa, Mohammed S. Elbamby, Mohamed K. Abdel-Aziz, Mehdi Bennis, 2024 IEEE International Conference on Communications (ICC 2024),Denver, CO, USA.

The work is funded by the Nokia project SCENE (G.A no. 00164501.0)

Reference

[1] A. Leivadeas and M. Falkner, "VNF placement problem: A multi-tenant intent-based networking approach," in IEEE 24th Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN), pp. 143–150,2021.

[2] J. Kim, E. Kim, J. Yang, J. Jeong, H. Kim, S. Hyun, H. Yang, J. Oh,Y. Kim, S. Hares, et al., "Ibcs: Intent-based cloud services for security applications," IEEE Communications Magazine, vol. 58, no. 4, pp. 45–51, 2020.

[3] A. Rafiq, M. Afaq, and W.-C. Song, "Intent-based networking with proactive load distribution in data center using IBN manager and smart path manager," Journal of Ambient Intelligence and Humanized Computing, vol. 11, pp. 4855–4872, 2020.

[4] M. Kiran, E. Pouyoul, A. Mercian, B. Tierney, C. Guok, and I. Monga,"Enabling intent to configure scientific networks for high performance demands," Future Generation Computer Systems, vol. 79, pp. 205–214,2018.

[5] E. J. Scheid, P. Widmer, B. B. Rodrigues, M. F. Franco, and B. Stiller, "A controlled natural language to support intent-based blockchain selection," in IEEE International Conference on Blockchain and Cryptocurrency (ICBC), pp. 1–9, 2020.

[6] M. Riftadi and F. Kuipers, "P4i/o: Intent-based networking with p4," in IEEE Conference on Network Softwarization (NetSoft), pp. 438–443, 2019.

[7] A. S. Jacobs, R. J. Pfitscher, R. A. Ferreira, and L. Z. Granville, "Refining network intents for self-driving networks," in Proceedings of the Afternoon Workshop on Self-Driving Networks, pp. 15–21, 2018.

[8] Y. Xia, S. Jiang, T. Zhou, S. Hares, and Y. Zhang, "NEMO (Network MOdeling) Language," Internet-Draft draft-xia-sdnrg-nemo-language-04, Internet Engineering Task Force, Apr. 2016. Work in Progress.

[9] D. Borsatti, W. Cerroni, G. Davoli, and F. Callegati, "Intent-based service function chaining on ETSI NFV platforms," in IEEE 10th International Conference on Networks of the Future (NoF), pp. 144–146, 2019.

[10] A. Alsudais and E. Keller, "Hey network, can you understand me?," in 2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 193–198, IEEE, 2017.

[11] M. Toy, "Intent-based networking for connectivity and cloud services," Adv. Netw, vol. 9, no. 1, pp. 19–22, 2021.

  The work is funded by the project SCENE (G.A no. 00164501.0)