# A Context-Aware User-Driven Framework for Network Selection in 5G Multi-RAT Environments

F. Bouali, K. Moessner
5G Innovation Centre (5GIC)
University of Surrey, UK
Email: {f.bouali, k.moessner}@surrey.ac.uk

M. Fitch
BT Research
Adastral Park, Ipswich IP5 3RE, UK
Email: michael.fitch@bt.com

***Abstract*—To improve interworking of future 5G systems with existing technologies, this paper proposes a novel context-aware user-driven framework for network selection in multi-RAT environments. It relies on fuzzy logic to cope with the lack of information usually associated with the terminal side and the intrinsic randomness of the radio environment. In particular, a fuzzy logic controller first estimates the out-of-context suitability of each RAT to support the QoS requirements of a set of heterogeneous applications. Then, a fuzzy multiple attribute decision making (MADM) methodology is developed to combine these estimates with the various components of the context (e.g., terminal capabilities, user preferences and operator policies) to derive the in-context suitability level of each RAT. Based on this novel metric, two spectrum selection (SS) and spectrum mobility (SM) functionalities are developed to select the best RAT in a given context. The proposed fuzzy MADM approach is validated in a dense small-cell environment to perform a context-aware offloading for a mixture of delay-sensitive and best-effort applications. The results reveal that the fuzzy logic component is able to efficiently track changes in the operating conditions of the different RATs, while the MADM component enables to implement an adjustable context-aware strategy. The proposed fuzzy MADM approach results in a significant improvement in achieving the target strategy, while maintaining an acceptable QoS level compared to a traditional offloading based on signal strength.**

## I. Context/Motivation

The fifth-generation (5G) of wireless networks is being developed to meet the strict requirements of future applications (e.g., two-way gaming and the Tactile Internet [1]). In this respect, an extensive amount of research has been devoted to develop various technologies and key-enablers (e.g., ultra-densification, design of new radio interface and use of higher frequencies) to boost the performance of future 5G radio access technologies (RATs) compared to what can be achieved using the current technologies [2]. However, less effort has been made to ensure interworking with the existing wireless systems and standards (e.g., WLANs and LTE).

From a practical standpoint, new 5G RATs need to interwork with the existing RATs for various reasons. First, from a commercial perspective, operators often prefer to continue to use their existing network infrastructures to serve the traditional applications (e.g., voice and web-browsing) as much as possible. Second, whenever user equipments (UEs) go out of 5G coverage, traditional RATs need be used to provide a seamless perception to the end-users. In such circumstances, multi-RAT operation could be considered as a complementary option to the new 5G RATs.

The multi-RAT selection problem has been extensively investigated in the context of 3GPP (e.g., WCDMA and LTE) and non-3GPP (e.g., WLANs) access networks [3–5]. Most of these works considered a common radio resource management (cRRM) based on a tight coupling architecture of the considered RATs. However, such a coordinated approach may not be valid to complement new 5G RATs. As a matter of fact, recent 5G architectures do not integrate most of legacy RATs (e.g., GSM, WCDMA and WLANs) or prefer to interconnect them only at the core network level because a full integration would be too costly in terms of multi-RAT measurements and interworking [6]. This means that a user-driven decision-making would be much more suitable particularly when the selection decision has to be made fast for specific applications (e.g., delay-sensitive) or under particular circumstances (e.g., fast degradation of radio conditions). However, the limited amount of information that is typically available to UEs through beacons and pilot channels (e.g., SS and SNR) is not enough to select the best RAT in a given context (e.g., QoS requirements, terminal capabilities and network constraints). This calls for a form of network assistance to inform UEs about all the relevant information that would be needed to perform an efficient decision.

To cope with the lack of information and uncertainty associated with UEs, most proposals have relied on fuzzy logic to infer the best RAT out of the available pieces of information. In this respect, some works developed a single fuzzy logic controller (FLC) that is fed with all the relevant attributes of the available RATs [4, 7]. The main limitation of these works is the lack of scalability. As a mater of fact, when more RATs and attributes are considered, the number of inference rules increases exponentially. Other works proposed a single FLC per RAT that combines the set of available radio parameters with the various components of the context [8, 9]. Such approach improves scalability, but is not flexible. First, it assumes that all data is fuzzy, while some attributes may be obtained precisely (e.g., QoS requirements). Second, using fuzzy logic rules to combine both QoS- and context-related attributes does not offer flexibility in adjusting the importance (i.e., weight) of each of attribute depending on the operating conditions.

To offer a higher degree of flexibility to fuzzy logic, few proposals have combined it with multiple attribute decision making (MADM) that is known for its ability to efficiently combine various heterogeneous attributes [10–12]. All these proposals have made no distinction between the radio parameters that are directly related to the achievable QoS (e.g., SS, bandwidth and SNR) and the various components of the context (e.g., speed, battery consumption and price) whose importance vary from case to case. It follows that the suitability level of each RAT to meet the set of QoS requirements cannot be explicitly assessed before considering the various contextual information, which means that, in practice, the selected RAT may not meet the target QoS level. Another key limitation of these schemes is that they all assumed that the decision-maker has full access to all required information (e.g., MADM attributes and weights). Therefore, they are not valid to tackle the unique issues and constraints associated with a user-driven mode of operation, e.g., which attributes can be in practice obtained by UEs or how the network may adjust the controlling parameters to achieve a target strategy. Clearly, all the architectural constraints should be taken into account early in the design of any feasible user-driven decision-making.

Therefore, the first main contribution of this paper is to construct a novel functional architecture to enable context-aware user-driven operation in multi-RAT environments. In particular, the proposed architecture relies on a UE connection manager (CM) that selects the best RAT according to a policy that is remotely adjusted by a network policy designer. Correspondingly, the second contribution is to develop a feasible fuzzy MADM implementation of the CM to select the best RAT for a set of heterogeneous applications. Fuzzy logic is used to first estimate the out-of-context suitability level of each RAT to support the various QoS requirements. Second, an MADM component combines these estimates with the various components of the context (e.g., user preferences and operator policies) to derive the in-context suitability levels of the considered RATs. Finally, the third contribution is to validate the proposed approach to perform a context-aware offloading in a dense small-cell environment to support a mixture of delay-sensitive and best-effort applications.

The remainder of this paper is organized as follows. A context-aware user-driven functional architecture is constructed in Section. II to support spectrum management in multi-RAT environments. Then, the CM is implemented in Section. III to select the best RAT based on a novel metric that assesses the in-context suitability levels of the various RATs. The proposed approach is instantiated in Section. IV to perform an intelligent offloading in a dense small-cell environment for a mixture of VoIP and FTP file transfer applications. The initial results are presented in Section. V, comparing several variants of the proposed approach to a traditional offloading approach. The conclusions and future directions are provided in Section. VI.

## II. The proposed context-aware user driven framework

A set of $K$ available RATs ($\{RAT_k\}_{1\leq k\leq K}$) are considered by UEs to establish a set of $L$ applications ($\{A_l\}_{1\leq l\leq L}$) that are characterised in term of various sets of QoS requirements ($\{Req_l\}_{1\leq l\leq L}$). At the time of establishing each of these applications, the various contextual information that may be available about the UE (e.g., velocity, remaining power and remaining balance) and the network (e.g., operator strategy and regulation rules) should be also taken into account as they may have a strong impact on the suitability of each of these RATs.

Therefore, the problem considered here is whenever an application $A_l$ needs to be established, how to:

- make the **UE** select the best RAT,
- to meet the set of **QoS requirements**,
- in the considered **context**?

To enable such context-aware user-driven mode of operation, a functional split between the UE and network domains should be clearly made to identify the logical entities and scope of each side. In this respect, the functional architecture described in Fig. 1 is proposed. Specifically, a connection manager (CM) is introduced at the UE to collect the relevant components of the context from both the terminal and network sides to implement a given decision-making policy (e.g., the proposed fuzzy MADM). The collected contextual information is combined with a radio characterisation of each available RAT in terms of a set of short-term attributes (e.g., SS, SNR and load) obtained e.g., through beacons and some medium- and long-term attributes (e.g., cost and regulation rules) stored in a policy repository together with all the policy-related parameters (e.g., algorithm, fuzzy logic membership functions and MADM weights). The content of the policy repository may be retrieved in practice from a local instance following a pull or push mode using e.g., the Open Mobile Alliance-Device Management (OMA-DM) protocol [13]. To offer a higher degree of flexibility to the network manager, a policy designer entity enables to build and update the policy repository content based on a set of measurement reports collected from UEs and the various network-level strategies and constraints (e.g., operator strategy and regulation rules). For instance, the policy designer may dynamically adjust some of the policy-related parameters (e.g., MADM weights) or the RAT attributes (e.g., cost) to optimise some network-level metrics (e.g., spectrum efficiency and energy efficiency) or implement a form of traffic steering (i.e., push UEs to use specific RATs during some periods of time). Finally, the CMs of different UEs may collaborate to further improve their individual performances.

## III. Fuzzy MADM Multi-RAT selection

This section proposes to select, for a given application $A_l$, the $RAT_k$ that would best meet each of the associated QoS requirements in a given context. To this end, a three-steps approach is considered:

1) Design a fuzzy logic calculator to estimate the "out-of-context" QoS suitability level ($s^{oc}_{k,l}$) based on the available radio parameters.
2) Develop a fuzzy MADM methodology to combine $s^{oc}_{k,l}$ with the various components of the context to derive the so-named "in-context" suitability level ($s^{ic}_{k,l}$).
3) Select the RAT that maximizes the in-context suitability level $s^{ic}_{k,l}$.

### A. Out-of-context suitability levels

Given the uncertainty and lack of information associated with UEs, this section develops a fuzzy logic calculator to estimate the suitability levels of each RAT to meet the QoS requirements of the various applications.

The key building block of fuzzy logic reasoning is the fuzzy logic controller (FLC) whose block diagram is described in Fig. 2 [14]. It is composed of three main modules, namely the fuzzifier, inference engine and defuzzifier. During fuzzification, crisp (i.e., real) input data are assigned a value between 0 and 1 corresponding to the degree of membership in a given fuzzy set. Then, the inference engine executes a set of *if-then* rules on the input fuzzy sets. These rules, referred to as inference rules, are maintained in a rule base that is typically built based on previous expert knowledge. Finally, the aggregated output fuzzy sets are converted into crisp outputs using a given defuzzification method.

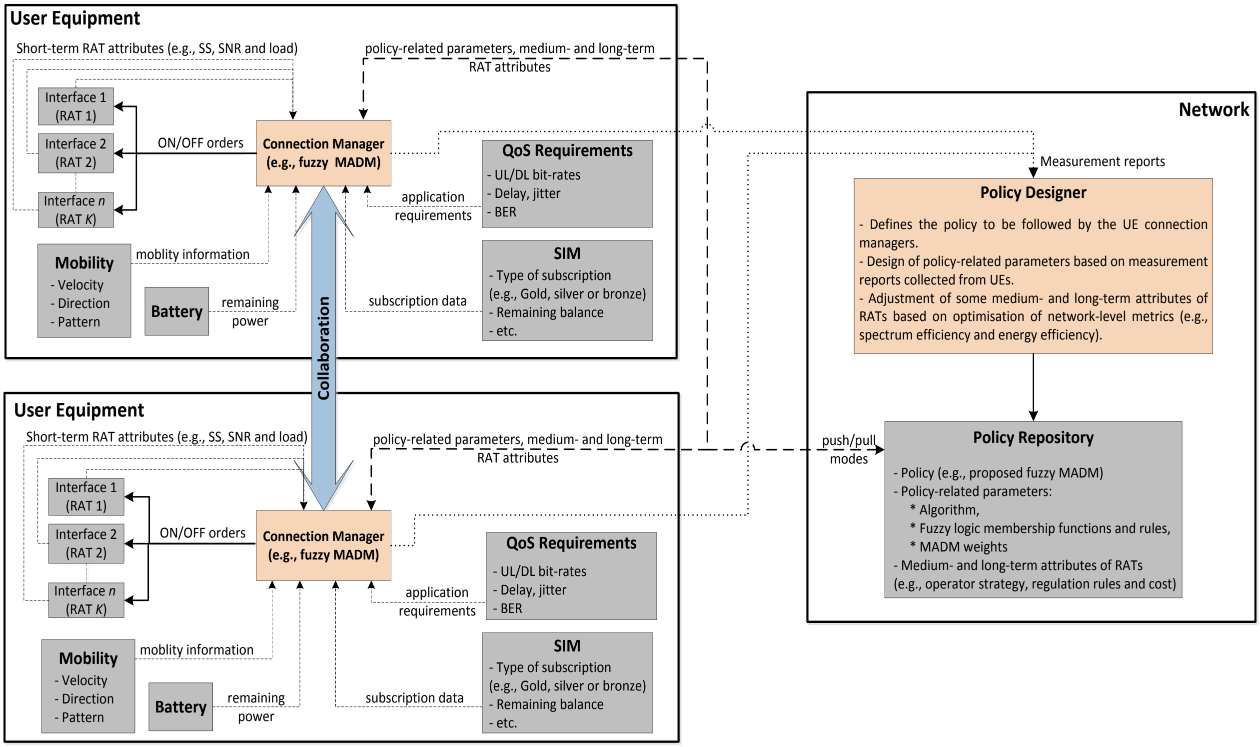



Fig. 1. Functional architecture of the proposed context-aware user driven framework

In our case, the main challenge is how to design a FLC that would reliably estimate the QoS suitability level ($s^{oc}_{k,l}$) in the most scalable way. To this end, the main design requirements can be highlighted as follows:

1) How to minimize the number of FLC inputs?
2) How to adjust all the relevant parameters (e.g., membership functions and inference rules)?

To tackle these issues, it is proposed to develop a single FLC for each pair of $RAT_k$ and application $A_l$. This enables to directly incorporate the set of associated QoS requirements (i.e., $Req_l$) when designing the set of rules. To further improve scalability, the minimum set of input radio parameters that are relevant for the considered application should be designed/considered. Finally, all membership functions and inference rules need to be adjusted by the policy designer of Fig. 1 based on the actual measurements reported by the various UEs to capture any dependency on the proprietary features and settings of the network (e.g., scheduler in LTE).

To better illustrate how these guidelines can be followed in practice, the FLCs designed for the use case considered in Section. IV will be presented in Section. IV-D.

### *B. In-context suitability levels*

In this section, the previously determined QoS suitability levels are combined with the various components of the context to derive the in-context suitability levels.

To particularly cope with the heterogeneity of the various components of the context, the following MADM methodology is considered:

- MADM formulation: The decision-maker is in our case a UE who wants to establish an application $A_l$ and has to select among a set of alternatives (i.e., RATs). For

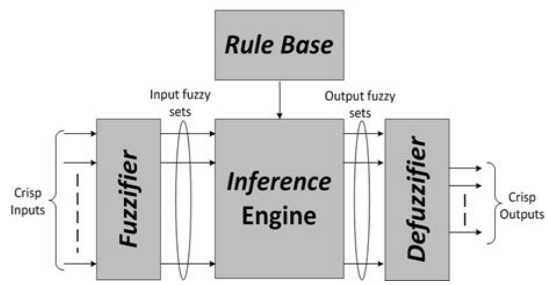



Fig. 2. Block diagram of fuzzy logic controller

each $k \in \{1,\ldots,K\}$, $RAT_k$ is characterized in terms of the following $M{=}4$ attributes:

  - $s^{oc}_{k,l}$: the out-of-context suitability to meet the set of QoS requirements. Recall that this is the output of the previous sub-section.
  - $cost_k$: the monetary cost of $RAT_k$.
  - $power_k$: the power consumption level when using $RAT_k$.
  - $range_k$: an assessment of the range to reflect the appropriateness from the UE velocity perspective.

Therefore, the RATs are fully characterized in terms of a *KxM* decision matrix $\mathbf{D_l}$ whose element $d^l_{k,m}$ denotes the performance of $RAT_k$ in terms of the *m*-th attribute:

$$\mathbf{D_l} = \begin{pmatrix} s^{oc}_{1,l} & cost_1 & power_1 & range_1 \\ \vdots & \vdots & \vdots & \vdots \\ s^{oc}_{k,l} & cost_k & power_k & range_k \\ \vdots & \vdots & \vdots & \vdots \\ s^{oc}_{K,l} & cost_K & power_K & range_K \end{pmatrix} \quad (1)$$

To reflect the relative importance of each attribute for the various applications, a vector $\mathbf{w_l}$ of $M$ weights ($\{w_{l,m}\}_{1\leq m\leq M}$) is introduced:

$$\mathbf{w_l} = \begin{pmatrix} w_{l,QoS} \\ w_{l,cost} \\ w_{l,power} \\ w_{l,range} \end{pmatrix} \tag{2}$$

When both performance metrics $d^l_{k,m}$ and weights $w_{l,m}$ are crisp numbers, the traditional MADM ranking methods, such as simple additive weighting (SAW) and technique for order preference by similarity to ideal solution (TOPSIS), can be efficiently used to first rank the various alternatives and then select the best one [15]. Without loss of generality, SAW will be considered in the remainder of this paper for its simplicity.

- Cope with fuzziness: In our case, most of the attributes (e.g., cost and power) and their corresponding weights are difficult to quantify and would be much better expressed in terms of linguistic terms (e.g., LOW, MEDIUM and HIGH). The considered problem becomes fuzzy MADM and classical MADM methods cannot be directly applied to solve it. To handle such imprecision, over a dozen fuzzy MADM methods have been developed [16]. However, these approaches are too cumbersome to be implemented in UEs because fuzzy data are operationally difficult. Therefore, it is proposed to solve the considered fuzzy MADM problem using the simplified approach proposed in [16]. It consists in two steps, namely first converting fuzzy data to crisp numbers, and second applying classical MADM methods to rank the alternatives. The reader is referred to [16] for more details about how to convert the linguistic terms to crisp numbers. In what follows, it is assumed that all fuzzy data in $\mathbf{D_l}$ and $\mathbf{w_l}$ have been converted to crisp numbers. Therefore, classical MADM ranking methods can be applied.
- MADM ranking: According to SAW ranking, the vector $\mathbf{s^{ic}_l}$ of in-context suitability levels ($\{s^{ic}_{k,l}\}_{k\in\{1,\ldots,K\}}$) can be obtained by combining the various MADM attributes and weights as follows:

$$\mathbf{s^{ic}_l} = \begin{pmatrix} s^{ic}_{1,l} \\ \vdots \\ s^{ic}_{k,l} \\ \vdots \\ s^{ic}_{K,l} \end{pmatrix} = \mathbf{D_l}\cdot\mathbf{w_l} \tag{3}$$

### C. Decision-making

Based on the previous section, the best RAT that maximises the in-context suitability level is selected for application $A_l$:

$$k^*(l) = \arg\max_{k\in\{1,\ldots,K\}}\left(s^{ic}_{k,l}\right) \tag{4}$$

To track the variability in the various attributes (e.g., radio conditions and contextual information), the CM implements the following functionalities based on the above criterion:

- Spectrum selection (SS): the best RAT is selected at the time of establishing each of the considered applications.
- Spectrum mobility (SM): a handover (HO) to the best RAT is performed during sessions. This may be triggered on an event-basis (e.g., emergency situation due to QoS degradation) or periodically (i.e., comfort HO).

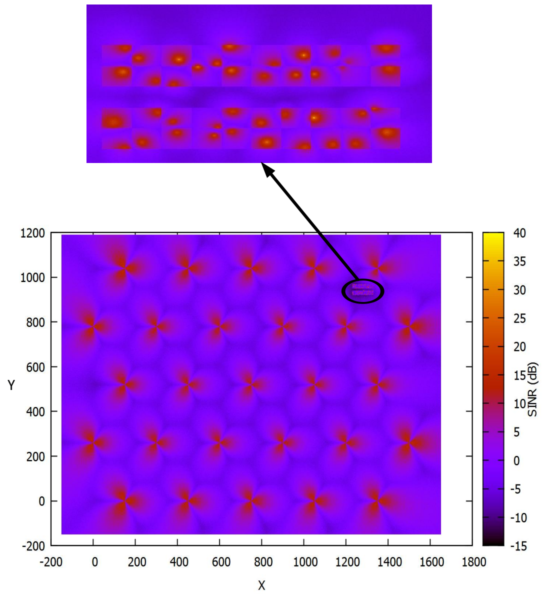



Fig. 3. Illustrative example of SINR map

To get insight into the relevance of each of these functionalities, different combinations will be compared in Section. V.

## IV. Use Case: Context-aware Offloading in Dense Small-cell Environments

To evaluate the effectiveness of the proposed framework in assisting in the spectrum management decision-making process, performing a context-aware offloading in dense small-cell environments is considered in this section as an initial use case.

### A. Considered environment

The considered environment is an hexagonal setting of LTE macro cells overlaid by a set of buildings where several WLAN/LTE small-cells are dropped randomly (i.e., $K$=2). According to the dual stripe model [17], each building is modelled as two stripes of rooms with a corridor in-between, which corresponds in practice to e.g., the set of stores inside a shopping mall. The various propagation losses (i.e., indoor-indoor, outdoor-outdoor, indoor-to-outdoor and vice versa) are modeled using the hybrid building model that combines several well known propagation loss models to consider the phenomenon of indoor/outdoor propagation in the presence of buildings [18]. As an illustrative example, Fig. 3 describes the signal-to-interference-and-noise-ratio (SINR) map obtained for a co-channel configuration where a building a 20 rooms is dropped on top of an hexagonal layout of 27 LTE macro cells, with one small-cell placed in each room.

### B. Traffic mixture

To get insight into the effectiveness of the proposed methodology in supporting various applications, the following $L$=2 applications are considered:

- VoIP: The VoIP traffic model is based on G.729A. It generates packets of 60Bytes (i.e., payload plus IP header overhead) at an inter-arrival time of 20ms, which corresponds to a data rate of $24Kbps$. The associated set of QoS requirements $Req_1$ is composed of a maximum end-to-end delay of $D_{max}$=$150ms$ and frame loss ratio of $L_{max}$=$5\%$. In this respect, the VoIP receiver accepts only in-sequence frames whose end-to-end delay does not exceed $D_{max}$. Any other frame is dropped with no subsequent retransmission.

- FTP file download: an ON/OFF model is used to model file download sessions (i.e., ON periods) and the inactivity intervals in-between (i.e., OFF periods). Both ON/OFF periods are exponentially distributed with rate $\lambda$. Whenever a file download session is established, it uses the whole capacity of the in-use radio link (i.e., WLAN or LTE) with loose QoS requirements ($Req_2$=∅).

### C. Target strategy

The most common scenario of a free broadband WLAN connection is considered, where indoor UEs prefer to use WLAN whenever possible and jump to the outdoor LTE macro otherwise (e.g., when UEs move outside or the WLAN backhaul connection gets lost). For our particular scenario, this means that FTP file download should be always established on WLAN, while VoIP needs to maximize WLAN usage as long as its QoS requirements are met.

To achieve the considered strategy, the generic fuzzy MADM approach proposed in Section. III will be instantiated in the following sub-sections.

### D. Design of fuzzy logic controllers

Following the guidelines given in Section. III-A, two separate FLCs are designed to estimate the suitability level of each RAT to meet the VoIP QoS requirements:

- WLAN: The considered FLC together with the corresponding membership functions are described in Fig. 4. The minimum set of input radio parameters is designed as follows:
  1) *SINR*: the signal-to-interference-and-noise-ratio of the AP beacon that reflects the radio and interference conditions.
  2) $Dwell\_T$: the average dwell time in the AP MAC queue assumed to be advertised by its beacon. It jointly captures the channel load (i.e., due to contention) and the traffic load of the various UEs served by the AP. No 802.11e QoS support is considered initially, which means that all traffic types share the same MAC queue.
- LTE: The considered FLC together with the corresponding membership functions are described in Fig. 5. In particular, the following radio parameters are considered:
  1) *RSRQ*: the reference symbol received quality that captures the radio and interference conditions.
  2) $T\_Sched$: the average time each packet waits before being scheduled. It reflects the load condition on the eNodeB and may be broadcasted in one of its system information blocks (SIBs). A non QoS-aware scheduler (e.g., proportional fair (PF)) is initially assumed, which means that all packets are treated equally.

For both FLCs, the $3^2$=9 required inferences rules have been designed based on a sensitivity analysis to the various combinations of the input parameters, which is omitted for the

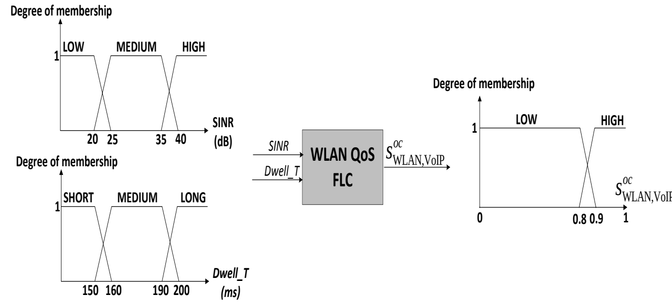



Fig. 4. WLAN/VoIP FLC

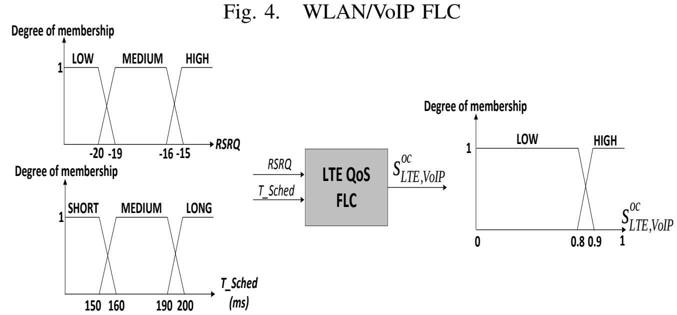



Fig. 5. LTE/VoIP FLC

sake of brevity. This mimics the adjustment performed by the policy designer of Fig. 1 based on the actual performance measurements collected from the various UEs. Finally, the deffuzification process is based on the commonly used centroid method for its accuracy [19].

Note that a possible extension to support QoS-aware bearer traffics (e.g., ViLTE and VoLTE) is to make the above traffic load metrics (i.e., $Dwell\_T$ and $T\_Sched$)) separate for each application type, which is left for future consideration.

Finally, it is worth pointing out that, for the other considered application (i.e., FTP file download), there is no need to develop sperate FLCs. Both WLAN and LTE are assumed to meet the associated loose QoS requirements as long as the corresponding UEs are associated/attached.

### E. Fuzzy MADM settings

This section instantiates the fuzzy MADM approach described in Section. III-B to achieve the target strategy defined in Section. IV-C. In what follows, $RAT_1$ and $RAT_2$ arbitrarily correspond to WLAN and LTE, respectively.

First, the decision matrix $\mathbf{D_l}$ is defined for each of the considered applications as follows:

$$\mathbf{D_{VoIP}} = \begin{pmatrix} s^{oc}_{WLAN,VoIP} & LOW & MEDIUM & SMALL \\ s^{oc}_{LTE,VoIP} & HIGH & HIGH & LARGE \end{pmatrix} \tag{5}$$

$$\mathbf{D_{FTP}} = \begin{pmatrix} HIGH & LOW & MEDIUM & SMALL \\ HIGH & HIGH & HIGH & LARGE \end{pmatrix} \tag{6}$$

Note that, compared to LTE, WLAN is qualified as cheaper, less power-consuming and smaller for both applications. Recall that, for both RATs, the first attribute associated with VoIP (i.e., $s^{oc}_{k,VoIP}$) is the output of each of the FLCs designed in the last sub-section and is therefore a crisp number.

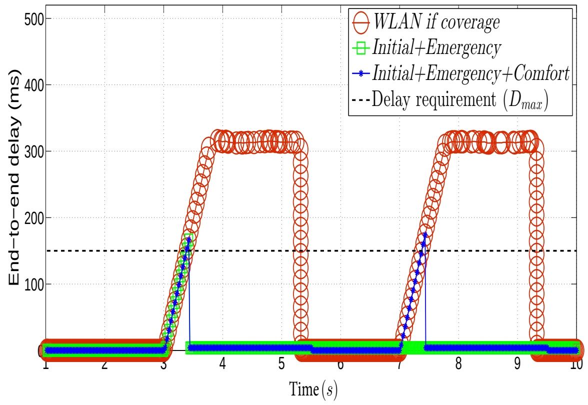



Fig. 6. Evolution of the end-to-end delay of VoIP

Finally, the corresponding MADM weights are adjusted as follows:

$$\mathbf{w_{VoIP}} = \mathbf{w_{FTP}} = \begin{pmatrix} HIGH \\ HIGH \\ LOW \\ LOW \end{pmatrix} \quad (7)$$

Note that both $w_{l,QoS}$ and $w_{l,cost}$ are set to HIGH to achieve the target strategy of Section. IV-C, while $w_{l,power}$ and $w_{l,range}$ are set to LOW for the sake of simplicity (i.e., the UE power consumption and velocity are not considered initially).

## V. Simulation results

To get insight into the relevance of the proposed framework in performing a context-aware offloading, a set of system-level simulations are performed using the NS-3 simulator [20].

### A. Initial assumptions

- The macro LTE is operating in a licensed FDD mode at $2.120Ghz$.
- A single room scenario of the dual stripe model described in Section. IV-A is considered. A WLAN 802.11n AP operating on channel 36 (i.e., $5.180GHz$) together with a single UE per application (i.e., VoIP and FTP file download) are randomly dropped inside the same room. Both applications are established in the downlink direction only, where a remote client sends traffic to UEs.
- The full list of WLAN/LTE physical parameters is that of the outdoor scenario considered in the Annex A of [21].

### B. Benchmarking

To assess the influence of the different components of the proposed framework, the following variants will be compared:

- *Initial+Emergency*: The fuzzy MADM selection scheme developed in Section. III-C is applied both initially and during sessions. Only emergency HOs are triggered (i.e., when QoS requirements are not met).
- *Initial+Emergency+Comfort*: In addition to the triggers of *Initial+Emergency*, periodic comfort HOs are triggered each $\Delta t$. If a better $RAT_{k^*}$ is identified (i.e., $s^{ic}_{k^*,l} > s^{ic}_{serving,l}$), the UE is moved to it. To avoid ping-pong effects, comfort HOs are blocked whenever any of the QoS requirements is not met.
- *WLAN if coverage*: a benchmarking scheme that always selects WLAN if the corresponding UE is associated.

### C. Performance evaluation

This section evaluates the performance of the proposed fuzzy MADM approach in achieving a context-aware offloading in the considered indoor environment.

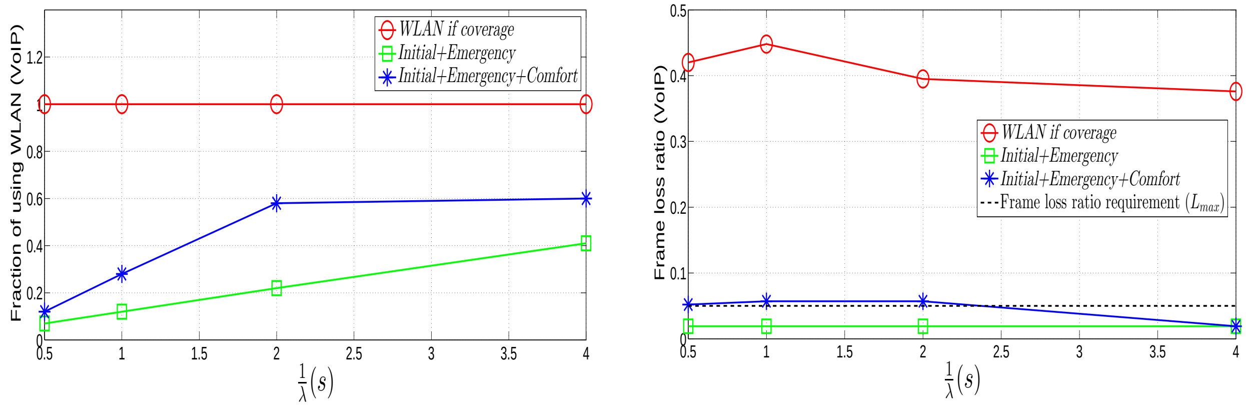
Fig. 6 shows the time evolution of the end-to-end delay of the VoIP application for all variants introduced in Section. V-B with the threshold $D_{max}$ shown in dashed lines. For better analysis of the obtained results, constant ON/OFF durations for FTP sessions are initially considered (i.e., $\frac{1}{\lambda}{=}2s$) with comfort HOs triggered each $\Delta t{=}0.5s$.

The results show that *Initial+Emergency* introduces significant improvement in terms of the end-to-end delay compared to *WLAN if coverage*. After VoIP is initially established, the WLAN link gets quickly saturated when the first FTP session starts (i.e., around $t{=}3s$), which strongly degrades the end-to-end delay for *WLAN if coverage*. In turn, when *Initial+Emergency* is used, an emergency HO to the macro LTE is triggered shortly after the VoIP receiver starts to receive some frames whose delay exceeds the threshold $D_{max}{=}150ms$. Given that all subsequent FTP sessions are established on WLAN, *Initial+Emergency* will not experience any further degradation and no subsequent HO will be triggered. On the contrary, shortly after the end of the FTP session, *Initial+Emergency+Comfort* triggers a comfort HO back to WLAN (e.g., around $t{=}5.4s$), which makes it subject to a future degradation when the next FTP session starts before a new emergency HO can be triggered (e.g., around $t{=}7.3s$).

To better illustrate the effectiveness in achieving the target strategy, Fig. 7(a) plots the WLAN usage fraction achieved by each of the considered schemes for different average ON/OFF durations. Fig. 7(b) shows the corresponding average frame loss ratio (FLR) with the threshold $L_{max}$ shown in dashed lines.

The results show that *Initial+Emergency* introduces significant improvement in terms of FLR compared to *WLAN if coverage*, which confirms the behavior previously observed in Fig. 6. When comparing *Initial+Emergency* and *Initial+Emergency+Comfort*, it can be seen that the additional comfort HOs significantly increase the fraction of using WLAN with gains up to 160% (Fig. 7(a)). When long ON/OFF durations are used (i.e., $\frac{1}{\lambda}{>}2s$), few comfort HOs are executed, which keeps the number of lost frames relatively low and maintains an acceptable FLR (Fig. 7(b)). As a result, comfort HOs are never blocked, which justifies the higher fraction of using WLAN achieved by *Initial+Emergency+Comfort* in Fig. 7(a). When shorter ON/OFF durations are used, more comfort HOs are initially executed with the associated increase in the number of lost frames. Interestingly, the FLR does not further degrade, but stabilises around the requirement $L_{max}$ as can be observed in Fig. 7(b). This is because as soon as FLR degrades, comfort HOs back to WLAN are blocked, which helps to increase the WLAN usage fraction only to the extent that does not hurt the FLR requirement.

In summary, *Initial+Emergency* avoids to select WLAN when QoS requirements are not met, which significantly outperforms the traditional offloading in terms of FLR. Executing comfort HOs on top of it further increases WLAN usage to the maximum extent that does not hurt any of the QoS requirements, which efficiently achieves the target strategy.




(a) WLAN usage fraction of VoIP

(b) Frame loss ratio of VoIP

Fig. 7. Analysis of the effectiveness in achieving the target strategy

## VI. Conclusions and Future Directions

This paper proposes a novel context-aware user-driven framework for supporting spectrum management in 5G multi-RAT environments. It relies on a UE connection manager (CM) that collects the relevant contextual information from both the terminal and network sides, and acts according to a policy that may be adjusted remotely by a network policy designer. Based on the proposed architecture, a fuzzy MADM methodology is developed to determine the best RAT in a given context subject to the lack of information usually associated with UEs. It first relies on fuzzy logic to estimate the out-of-context suitability of each RAT to support the QoS requirements of the various applications. Then, a fuzzy MADM approach is proposed to combine these estimates with the various components of the context to assess the in-context suitability level of each RAT. Based on this novel metric, a context-aware strategy combining SS and SM functionalities is proposed. As an initial use case, the proposed framework is applied to perform a context-aware offloading in a dense small-cell environment for a mixture of delay-sensitive and best-effort applications. The results reveal that the fuzzy logic component is able to efficiently track changes in the operating conditions of the different RATs, while the MADM component enables to implement an adjustable context-aware strategy. The proposed fuzzy MADM approach results in a significant improvement in achieving the target strategy while maintaining an acceptable QoS level compared to a traditional offloading based on signal strength.

As part of future work, it is intended to develop more advanced context-aware strategies and extend the proposed framework to support other RATs (e.g., LTE-U and mmWave-based) and applications (e.g., video streaming and IoT traffic).

## Acknowledgement

The authors would like to gratefully acknowledge the support of BT through the University of Surrey 5G Innovation Centre (http://www.surrey.ac.uk/5gic), and the EU funded H2020 5G-PPP project SPEED-5G under the grant agreement no. 671705.

## References

[1] G. P. Fettweis, "The Tactile Internet: Applications and Challenges," *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, pp. 64–70, March 2014.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What Will 5G Be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[3] V. Marques, R. L. Aguiar, C. Garcia, J. I. Moreno, C. Beaujean, E. Melin, and M. Liebsch, "An IP-based QoS architecture for 4G operator scenarios," *IEEE Wireless Communications*, vol. 10, no. 3, pp. 54–62, June 2003.

[4] R. Agustí, O. Sallent, J. Pérez-Romero, and L. Giupponi, "A fuzzy-neural based approach for joint radio resource management in a beyond 3G framework," in *QSHINE 2004*, Oct 2004, pp. 216–224.

[5] X. Gelabert, O. Sallent, J. Pérez-Romero, and R. Agustí, "Radio access congestion in multiaccess/multiservice wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 8, pp. 4462–4475, Oct 2009.

[6] "5G Whitepaper: The Flat Distributed Cloud (FDC) 5G Architecture Revolution," Institute for Communication Systems, 5G Innovation Centre, University of Surrey, Tech. Rep., January 2016.

[7] L. Giupponi, R. Agustí, J. Pérez-Romero, and O. Sallent, "A novel joint radio resource management approach with reinforcement learning mechanisms," in *PCCC 2005*, April 2005, pp. 621–626.

[8] B. Ma, X. Liao, and X. Xie, "Vertical handoff algorithm based on type-2 fuzzy logic in heterogeneous networks." *JSW*, vol. 8, no. 11, pp. 2936–2942, 2013.

[9] A. Kaloxylos, S. Barmpounakis, P. Spapis, and N. Alonistioti, "An efficient RAT selection mechanism for 5G cellular networks," in *2014 International Wireless Communications and Mobile Computing Conference (IWCMC)*, Aug 2014, pp. 942–947.

[10] W. Zhang, "Handover decision using fuzzy MADM in heterogeneous networks," in *Wireless Communications and Networking Conference, 2004. WCNC. 2004 IEEE*, vol. 2, March 2004, pp. 653–658 Vol.2.

[11] V. Ramkumar, A. Mihovska, N. Prasad, and R. Prasad, "Fuzzy-logic based call admission control for a heterogeneous radio envioroenment," in *WPMC 2009*, September 2009, "Invited paper.

[12] B. Z. Aymen, M. Ayadi, and S. Tabbane, "A fuzzy logic algorithm for RATs selection procedures," in *Networks, Computers and Communications, The 2014 International Symposium on*, June 2014, pp. 1–5.

[13] "OMA Device Management Protocol," Open Mobile Alliance, Tech. Rep. version 1.2.1, June 2008.

[14] L. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338 – 353, 1965.

[15] C.-L. Hwang and K. Yoon, *Multiple Attribute Decision Making: Methods and Applications A State-of-the-Art Survey*. Springer Berlin Heidelberg, 1981, ch. Methods for Multiple Attribute Decision Making, pp. 58–191.

[16] S.-J. J. Chen and C. L. Hwang, *Fuzzy Multiple Attribute Decision Making: Methods and Applications*, M. J. Beckmann and W. Krelle, Eds. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 1992.

[17] "Simulation assumptions and parameters for FDD HeNB RF requirements," 3GPP, Tech. Rep. 3GPP R4-092042, May 2009.

[18] "NS-3 Model Library," , Tech. Rep. Release NS-3.25, 2016. [Online]. Available: https://www.nsnam.org/docs/release/3.25/models/ns-3-model-library.pdf

[19] T. J. Ross, *Fuzzy logic with engineering applications*. Chichester, U.K. John Wiley, 2010.

[20] The network simulator-3 (NS-3). [Online]. Available: https://www.nsnam.org/

[21] "Study on licensed-assisted access to unlicensed spectrum, (Release 12)," 3GPP, Tech. Rep. 36.889-V13.0.0, June 2015.