

3GPP TSG RAN Rel-19 workshop

Taipei, June 15 - 16, 2023

Agenda: 5

Document for: Discussion

RWS-230240

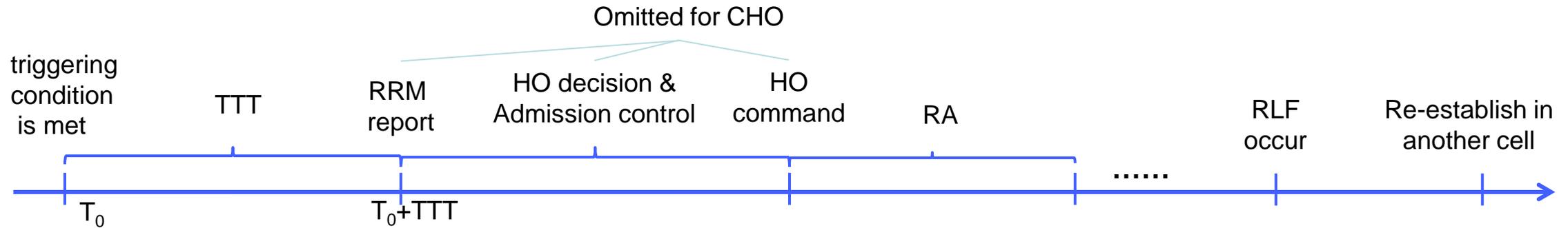


Motivations on AI/ML-based mobility enhancement

China Telecom, vivo

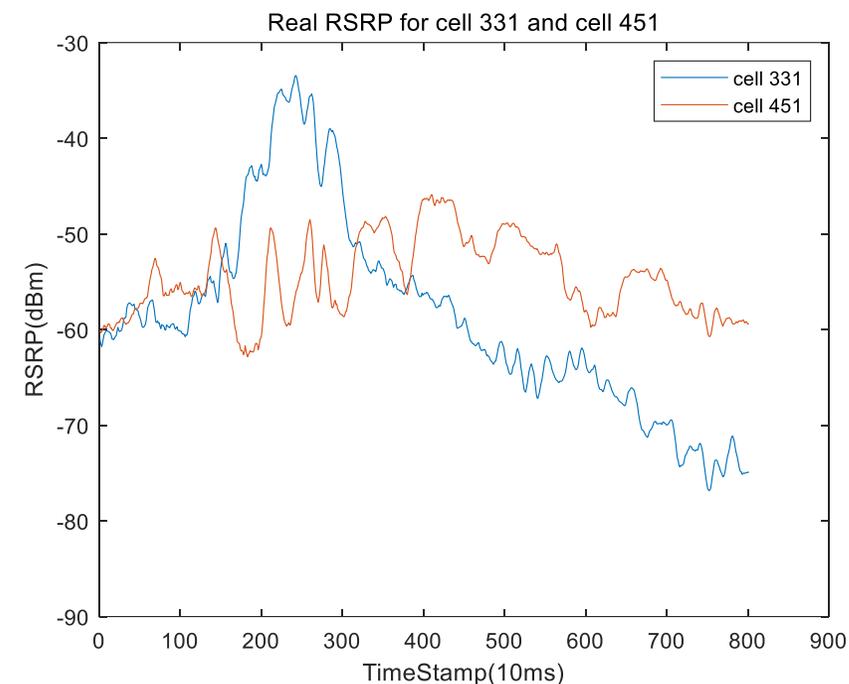
June 2023

- In Rel-18, the *NR_AIML_NGRAN-Core* WI specifies the AI/ML based mobility optimization:
 - The Model Inference functionality resides within the RAN node only supporting cell-level mobility prediction.
 - Some location-related information of UE (e.g., coordinates) was required as the input of AI/ML model, which may introduce UE privacy concerns. As a result, UE may choose not to provide these information.
 - Frequent exchange of the input between network and UE will introduce massive signaling overhead and latency.
- In Rel-18, the *NR_AIML_Air* SI would introduce the model life cycle management to enable the use cases at the air interface:
 - The framework can be reused to facilitate more use cases at the air interface.
 - The Model Inference functionality can reside on the UE side.
 - Local model inference at UE side may utilize more detailed location information and can reduce signaling overhead of input exchange.
- **Observation 1:** For mobility optimization, if the Model Inference functionality can be deployed on the UE side:
 - **More detailed local information from UE can be utilized as the input to improve prediction accuracy without privacy concerns.**
 - **Local model inference can reduce signalling overhead and inference latency.**



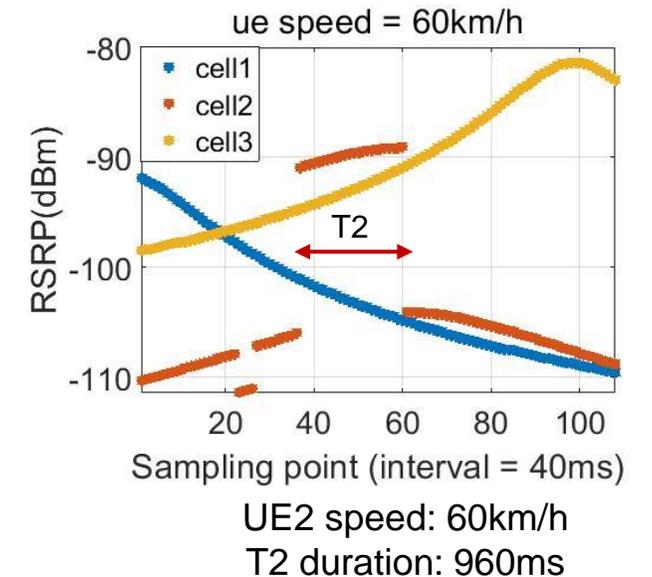
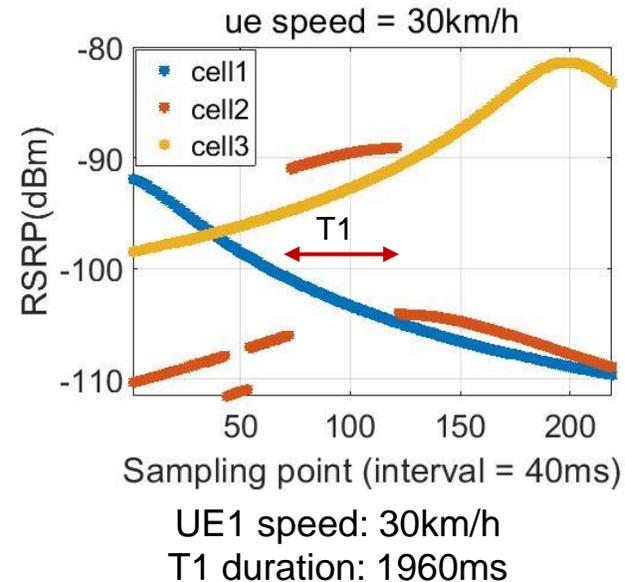
- For the legacy HO, the triggering condition for RRM reporting shall be met at T_0 and shall last for TTT (Time To Trigger) duration, which may lead to:
 - HO at a non-optimal time, poor user experience at source cell.
 - Failed to receive the HO command or failed to RA to the target cell, i.e., too-late HO.
- For CHO, the UE can RA to the target cell without receiving the HO command if the triggering condition is met during TTT. However, there is still a risk of RLF due to low SINR at the source cell during TTT and the UE cannot perform HO at a optimal time.
- The legacy solution to the above issue by reducing TTT duration may result in other unintended events, e.g., too-early HO, ping-pong HO, especially for the high-speed UEs.
- If an RLF occurs shortly after a successful HO, the UE may attempt to re-establish the radio link connection in a cell other than the source cell and the target cell, which is identified as HO to wrong cell.

Issues during field test - Highway



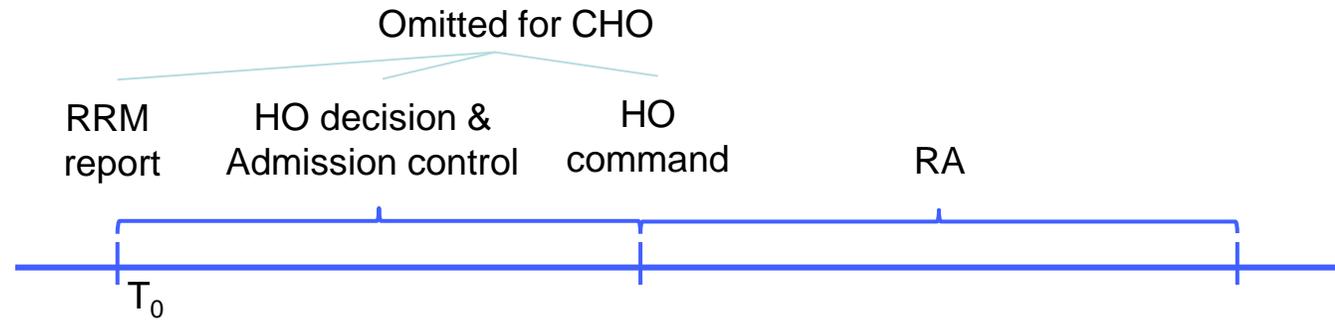
Carrier Frequency: FR1, 3.5GHz UE speed: 80~120km/h

- During UE high-speed movement, the RSRP of the neighbor may becomes better than the serving cell in a short period of time.
- UE will handover to the neighbor cell and handover back quickly to the last serving cell, i.e., ping-pong handover occurs.



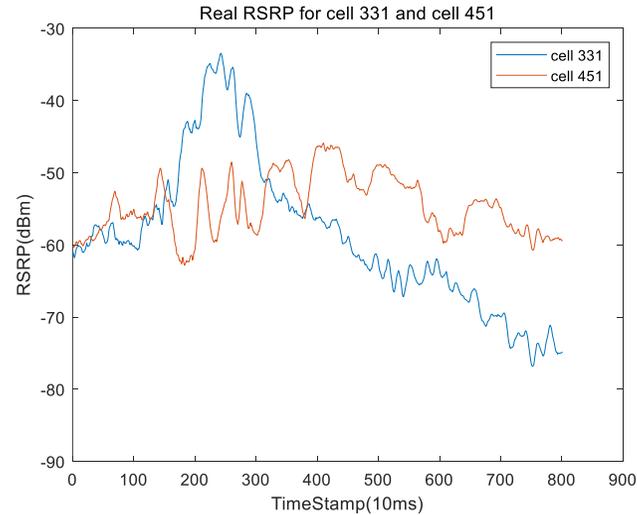
Simulation scenario and UE trajectory

- When the UE passes through the crossover, the RSRP of cell2 will change dramatically and UE handover to cell 2.
- For UE1 at a low speed, it will be served by cell 2 for a longer time (over 1 second).
- For UE2 at a high speed, it will only be served by cell 2 for less than 1 second.
- Both UEs may experience RLF when leaving the crossover and will reestablishment RRC connection on cell 3.

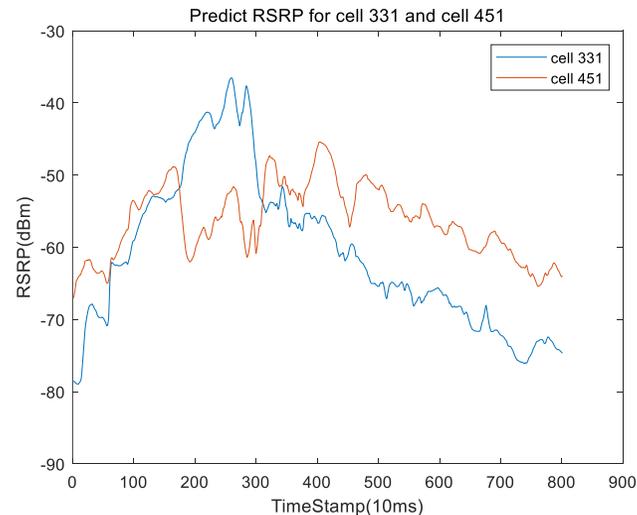


- UE may perform RRM prediction, e.g. RSRP prediction.
- For the legacy HO, with RSRP prediction, the UE can send the RRM report once the triggering condition is met at T_0 (i.e., no need to wait for TTT duration).
 - RSRP prediction within the $T_0 + \text{TTT}$ period shall meet the triggering condition,
 - The UE may send the RSRP prediction of neighbor cells during RA in the RRM report for HO decision,
 - Higher success rate for receiving HO command when the RRM report was sent at an optimal time.
- For CHO, with RSRP prediction, the UE can RA to the target cell once the triggering condition is met at T_0 .
 - RSRP prediction within the $T_0 + \text{TTT}$ period shall meet the triggering condition,
 - Lower risk of RLF at the source cell when the HO is performed at a optimal time.
- During the HO decision and target cell selection, the RSRP prediction can be used to reduce the unintended events, e.g., ping-pong HO, too early HO, HO to wrong cell.

Field test RSRP



Predict RSRP



UE speed: 80~120km/h

Dataset

- Training dataset: 15 UE trajectory
- Test dataset: 1 UE trajectory

AI/ML Model

- Fully Connected Neural Network

Input

- History RSRP of serving & neighbor cell (interval = 10ms)
- History UE location and speed (interval = 1s)
- Observation window = 2s

Output

- Predict RSRP of serving cell
- Predict RSRP of neighbor cell
- Prediction timing = 2s

Performance

- RMSE (root mean square error) = 0.8dB
- Field test results show that the predicted RSRP is basically consistent with the filed test RSRP

■ Accuracy of RRM measurement prediction

	Cell 1	Cell 2	Cell 3
Prediction 1	RMSE = 0.0044dB	RMSE = 1.08dB	RMSE = 0.26dB
Prediction 2	RMSE = 0.0062dB	RMSE = 1.11dB	RMSE = 0.26dB
Prediction 3	RMSE = 0.0844dB	RMSE = 1.23dB	RMSE = 0.28dB

Carrier Frequency: FR2, 30GHz

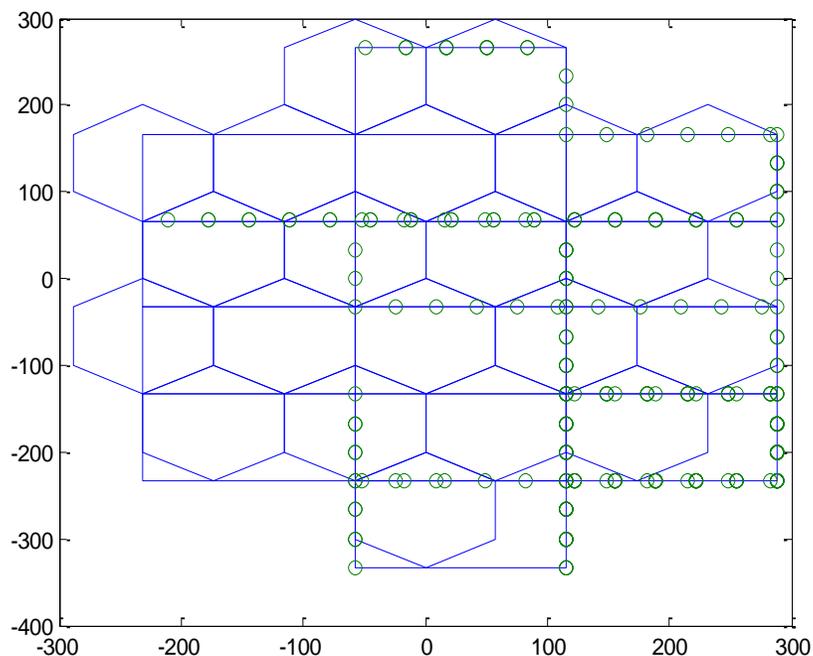
Prediction 1: RSRP of every 80ms in 320ms after T_0

Prediction 2: RSRP of 1s after T_0

Prediction 3: RSRP of 2s after T_0

Simulation results show that the RRM measurement predictions on different time scale have very low RMSE.

Simulation scenario



Simulation scenario and UE trajectory

Simulation assumption

Attributes	Values or assumptions
Carrier Frequency	FR1: 4GHz; FR2: 30GHz
TRP Number	7 sites, 3 sector per site
Channel Model	3D-Uma in TR 38.901, support Spatial consistency ISD = 200m
UE speed	120km/h
Mobility management	Event: A3; Hysteresis: 2dB; Offset: 1dB; TimeToTrigger: 320ms, 40ms Handover preparation time: 50ms; Handover execution time: 40ms
RLM	L1 measurement period: 20ms Qin sliding window length: 100ms Qout sliding window length: 200ms Qin threshold: -6dB; Qout threshold: -8dB N310: 1; N311: 1; T310: 1s
Handover model and corresponding metrics	As defined in TR 36.839 Short time of stay: served by the target cell for less than 1s after HO

■ Accuracy of RRM prediction

	Prediction 1	Prediction 2
FR1	RMSE = 0.38 dB	RMSE = 1.6 dB
FR2	RMSE = 1.3 dB	RMSE = 3.3 dB

Training dataset: Same large scale channel parameters for different drops

Prediction 1: RSRP of every 80ms in 320ms after T_0

Prediction 2: RSRP of 1s after T_0

Simulation results shows that the RRM measurement predictions on FR1 and FR2 have reasonable RMSE.

■ Usage of RRM prediction

- Legacy HO:
 - UE can decide whether to trigger the RRM reporting based on the predicted RSRP of every 80ms in 320ms after T_0 ,
 - Source cell can determine the target cell based on the predicted RSRP of 1s after T_0 to avoid too early HO or HO to wrong cell,
 - Source cell can forward the predicted RSRP to target cell for admission control.
- CHO:
 - UE can decide to trigger the target cell selection based on the predicted RSRP of every 80ms in 320ms after T_0 ,
 - UE can choose the target cell based on the predicted RSRP of 1s after T_0 to avoid too early HO or HO to wrong cell.

■ RRM prediction based HO

		Legacy HO, TTT = 320	Legacy HO, TTT = 40	AI/ML based HO	CHO, TTT = 320	CHO, TTT = 40	AI/ML based CHO
FR1	HOF rate	9.16%	2.2%	1.95%	0.28%	0.15%	0.32%
	Ping-pong HO rate	1.1%	3.6%	0.37%	1.0%	3.7%	0.37%
	Short Time of Stay (1s) rate	13.4%	18.9%	5.7%	13.6%	18.8%	5.67%
FR2	HOF rate	7.4%	2.5%	2.0%	0.42%	0.43%	0.44%
	Ping-pong HO rate	5.2%	10.3%	2.7%	5.2%	10.3%	2.7%
	Short Time of Stay (1s) rate	24.1%	36.7%	10.4%	24.4%	36.5%	10.8%

- **Observation 2:** With RRM prediction, the unintended events rate during HO and CHO can be significantly reduced, including HOF rate, ping-pong HO rate and short time of stay rate.

Study and finalize the sub use cases from the following for evaluation and specification impact analysis for AI/ML based mobility optimization: [RAN2]

- E.g., RRM measurement (e.g., RSRP, SINR) prediction, target Cell prediction, unintended events prediction
- UE sided model [network sided model and two sided model]

For the selected sub-use case for evaluation, evaluate performance benefits of AI/ML based algorithms [RAN2]

- Methodology based on statistical models (from [TR 38.901]), for system level simulations.
- KPIs: Determine the common KPIs and corresponding requirements for the AI/ML operations.

Assess potential specification impact, specifically for the selected sub-use cases, including [RAN2, RAN1, RAN4]

- Identify AI/ML framework applicable for the selected sub-use case
 - o Identify applicable levels of collaboration between UE and NW
 - o Characterize lifecycle management of AI/ML based mobility optimization: e.g., model training, model deployment, model inference, model monitoring, model updating
 - o Data collection aspects
 - o Note: Federated learning can be considered for model training.
- PHY layer aspects and protocol aspects of the identified framework
 - o E.g, identify impacts of different collaboration levels, the input and output for model training and model inference purpose, the performance metrics for model monitoring purpose
- Interoperability and testability aspects

- **Observation 1:** For mobility optimization, if the Model Inference functionality can be deployed on the UE side:
 - » More detailed local information from UE can be utilized as the input to improve prediction accuracy without privacy concerns.
 - » Local model inference can reduce signalling overhead and inference latency.
- **Observation 2:** With RRM prediction, the unintended events rate during HO and CHO can be significantly reduced, including HOF rate, ping-pong HO rate and short time of stay rate.
- **Proposal 1:** Agree on a RAN2 leading AI/ML-based mobility enhancement study item in Rel-19.

Thanks!
