Efficient Beam Sweeping Paging in Millimeter Wave 5G Networks

Chung-Wei Weng, Biswa P. S. Sahoo, Ching-Chun Chou, and Hung-Yu Wei
*Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan
Email: hywei@ntu.edu.tw

***Abstract*—In millimeter-wave communications, beamforming is a new enabling technology for the next generations (5G) new radio (NR) system which requires beam sweeping for transmission and receptions. Generally, users in the network are informed about the incoming services, new or updated system information by a procedure called Paging. This procedure is triggered by the network to request and establish signaling connection with the user equipment (UE). In 5G NR, the next generation NodeB (gNB) transmits paging messages on different beams to notify the target UE on discrete time slots. Since, the paging is transmitted using beam sweeping, the overhead of paging transmission will expand to multiple time slots limiting the resource in paging occasion. Additionally, on the other hand, the UE will also need wake up increased number of time to monitor the paging occasion, which directly influences the energy consumption of UE. In this paper, we propose an efficient beam sweeping paging mechanism taking beam sweeping overhead and UE energy consumption into consideration. Our mathematical analysis and simulation experiments show that the proposed scheme can enhance the energy saving efficiency while reducing the paging sweeping beam overhead.**

I. Introduction

Millimeter-wave (mmWave) has been recognized as a promising candidate for next generation wireless communication systems [1]. The use of large number of antennas in mmWave systems leads to an equally large number of RF chains. This results in unaffordable energy consumption, as the energy consumption of RF chain is high at mmWave frequencies [2]. Hence, the energy consumption for user equipment (UE) is a crucial problem in 5G mmWave communications. In order to reduce the power consumption and prolong the battery life of a UE, a discontinuous reception (DRX) mechanism is specified in the 3rd Generation Partnership Project (3GPP) standard [3]. If no data transmission is scheduled to an UE, the evolved NodeB (eNB) coordinates with the UE to enable DRX operation by radio resource control (RRC), thus UE leave CONNECTED state and go to sleep. Since the UE sleeps when it is not scheduled for data transmission, the energy consumption of UE can be reduced.

In long term evolution (LTE), when DRX is configured, the UE periodically wakes up in the paging occasion (PO) of paging frame (PF) to read the physical downlink control channel (PDCCH). Since the next generation NodeB (gNB) would be equipped with directional antennas to counteract the high free-space loss in mmWave [4], the 3GPP has specified that the PO can consist of multiple time slots and paging can be transmitted by beam sweeping [5]. Thus, in order to receive the paging transmitted using beam sweeping, UE has to monitor increased number of time slots. As a result, this will directly influence the UE power consumption. On the other hand, the gNB also need to transmit same paging in each direction with different beams in discrete time slots. Subsequently, this additional beam sweeping cause large overhead for the gNB. The paging mechanism of mmWave 5G using beam sweeping is illustrated in Fig. 1.

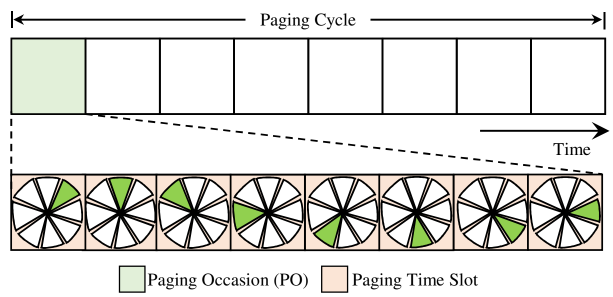

Fig. 1. Illustration of paging transmission using beam sweeping in PO. The total number of beams is eight.

In LTE cellular systems, UEs are configured with two RRC states, i) IDLE and ii) CONNECTED [6]. The two states of LTE shall be not comprehensive enough to satisfy the delay requirement of fast packet access and efficient power saving in 5G New Radio (NR) systems. Thus, 5G NR technology has introduced a new RRC state called INACTIVE to meet the requirement of 5G services. The UE paging in INACTIVE state can initiated by radio access network (RAN) or core network (CN). In contrast, the UE paging in IDLE state can only be initiated by the CN. [7]. Additionally, the UE context in INACTIVE state is not released while the UE context in IDLE state is released. Thus, the paging mechanism need to be redesigned for the UE in INACTIVE state for NR.

In this paper, we investigate the beam sweeping paging mechanism and its challenges for mmWave systems. To deal with beam sweeping overhead for paging based on beamforming technique, we propose an efficient beam sweeping paging mechanism by incorporating the total beam sweeping and selected beam sweeping for the UEs in INACTIVE state. This approach minimize the paging sweeping beams subject to paging delay as the constraint. We evaluate the performance of the proposed efficient beam sweeping paging by mathematical analysis and simulation of realistic channel model.

II. Related Works

Several works have been investigated the paging operation in UMTS evolution and LTE [8] [9]. However, in the literature, there has been very limited discussion on NR paging. The author in [7] proposed a hybrid paging scheme incorporating RAN and CN initiated paging for RRC INACTIVE state UE. Nonetheless, the paging scheme deals with only cell paging but not considers directional beam sweeping which is a enabling factor for mmWave systems. The author in [3] modeled the beam sweeping DRX mechanism for mmWave, but they assume simple total beam sweeping for paging. The standardization body for mobile networks, 3GPP started discussion on paging operation in NR during RAN2 Adhoc. The lesson learned during the discussion of RRC states in NR further progressed the RRC states design in NR through beam sweeping [10].

The focus in 3GPP discussion on paging operation in NR is mostly addressed the new RRC state i.e. INACTIVE state and required beam sweeping operation. The author in [11], discussed the NR paging mechanism in the high frequency scenario based on 3GPP RAN1's agreements. As the transmission and reception of paging information will be performed through beam sweeping, it leads to sweeping overhead which must be addressed. The author in [5], discussed issues, particularly, related to paging transmission and reception with beamformed paging operation. During paging operation, it is also important to determine how the paging occasions are to be calculated and assigned over time through subframes.

III. Proposed Efficient Beam Sweeping Paging

The 3GPP provides a framework for allocating paging transmission in different time slot on different beam. There are several ways to allocate paging transmissions. A naive way is to apply total beam sweeping, nevertheless, it is not efficient. Therefor, we proposed an *Efficient Beam Sweeping Paging* mechanism that effectively reduces the paging sweeping beams and required number of time slots based on UE assisted feedback. Our proposed mechanism incorporates two part as *Full Paging* and *Fast Paging*. In *Full Paging*, the paging is transmitted in all directions using beam sweeping. In contrast, in *Fast Paging*, the paging is transmitted only on some selected beams and directions.

A. Main Idea

In 5G, as there may be multiple device types e.g. general UE which performs normal data transmission and delay-sensitive UE which performs some mission critical operation; which require different paging delay. We assume, the different UE type can select appropriate DRX cycle length. In our mechanism, the delay-sensitive UE reports the receiving beam details to the gNB after it receives paging indication in *Full Paging*. Since the UE context will not be released for UE in INACTIVE state, the gNB stores the beam information for delay-sensitive UEs. In contrast, the general UEs do not required to report the receiving beam information to the gNB. For delay-sensitive UEs, the gNB use short paging cycle to perform *Fast Paging* by transmitting only on selected beams. In contrast, for the general UEs, the gNB use long paging cycle to perform *Full Paging* by sweeping all the beams.

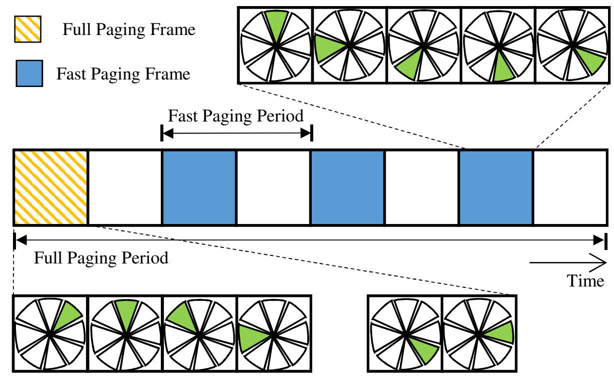

Fig. 2. Illustration of *Efficient Beam Sweeping Paging*, this example shows that the period of *Full Paging* cycle is four times of which of *Fast Paging* cycle and the number of beams to transmit paging in *Fast Paging* is five.

An example of the proposed mechanism is illustrated in Fig. 2. As illustrated in the figure, the *Full Paging* and *Fast Paging* compose *Efficient Beam Sweeping Paging*. The key idea is that one or many PFs of *Fast Paging* between the PFs of *Full Paging*, so as to reduce the beam sweeping overhead over *Full Paging*. The detailed explanation is presented in the following subsection.

1) Full Paging: In *Full Paging*, the gNB performs paging beam sweeping using all beams in all directions of the cell. This beam sweeping pages general UEs with long DRX cycle as well as delay-sensitive UEs with short DRX cycle. This paging triggers the delay-sensitive UEs for beam reporting and can be used to broadcast system information update by gNB. Note that, the gNB transmits paging indication on each beam, and then all the UEs including delay-sensitive UEs and general UEs will receive paging indication in PDCCH and check whether or not the UE identity is in the list of paging message on PDSCH. Also, delay-sensitive UEs monitor paging indication beam sweeping to check if any change of associate beam. If delay-sensitive UEs find any change of associate beam, then the UE reports the new received paging beam information to the gNB. This beam information is utilized in *Fast Paging*.

2) Fast Paging: In *Fast Paging*, gNB transmits paging only on recorded beams which will reach the delay-sensitive UEs. These recorded beam information is updated from UE beam reporting during the *Full Paging*. In our proposed mechanism, only delay-sensitive UEs wakes up to monitor the paging indication in *Fast Paging* PF. For instance, in the example of Fig. 2, the gNB spend eight time slots sweeping all eight beams to transmit paging in the PO of the first radio frame which is the paging frame. The delay-sensitive UEs monitor all eight time slots of beam sweeping in this PO and realize each of them is on which beam. Then, each of them reports received paging beams, beam 2, beam 4, beam 5, beam 6 and beam 7 to the gNB. The gNB then realizes only this five beams associate with delay-sensitive UEs and transmits only this five beams in each following three PF of *Fast Paging*. Hence, the gNB only spend five time slots in each PO of *Fast Paging*. By incorporating *Fast Paging* in the paging mechanism, it not only reduces paging sweeping beams but also saves the general UE power to monitor paging occasion (PO).

B. Analytic Modeling

We provide an analytic model to give insight into the performance of the proposed paging mechanism. First, we will develop the mathematical formulation of paging delay of delay-sensitive UE; second, we will formulate the optimization problem of minimizing paging sweeping beams subjecting to paging delay of delay-sensitive UE. Let t_r be beam residence time that represent the time in which UE stays in a beam's coverage. We assume that the beam residence time in each beam are independent and identically distributed (i.i.d) random variables, and then the number of times that UE changes beam is a renewal process.

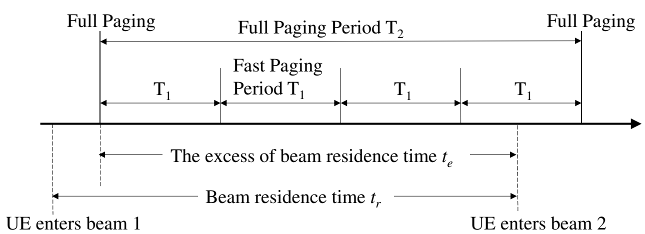

Fig. 3. The timeline of paging

We calculate the excess of beam residence time at time of *Full Paging*. The excess life of a renewal process at time of *Full Paging* is the time from time of *Full Paging* until the next renewal [12]. That is, the time t_e between time of *Full Paging* and UE changing beam. An example of paging timeline is illustrated in Fig. 3. In this example, the UE first enters beam 1 coverage and stays in beam 1 coverage for beam residence time t_r until it leaves beam 1 coverage and enters beam 2 coverage. The excess of beam residence time is the time during which UE stays in the same beam after *Full Paging*. The length of time between two *Fast Paging* is T_1, and the length of time between two *Full Paging* is T_2. In other word, *Fast Paging* is transmited periodically with period T_1 and *Full Paging* is transmited periodically with period T_2. In the example in Fig. 3, the period of *Full Paging* is four times of the period of *Fast Paging*. Let F_r be cumulative distribution function and ρ be mean of beam residence time. If F_e is cumulative distribution function of excess of beam residence time at time of *Full Paging*, derivation follows from the analysis of excess of renewal process in [12],

$$F_e(t) = \frac{1}{\rho}\int_0^t (1 - F_r(\tau))d\tau \tag{1}$$

Also, we have probability density function $f_e(t)$,

$$f_e(t) = \frac{1}{\rho}(1 - F_r(t)) \tag{2}$$

Let T_1 be the period of *Fast Paging* cycle and T_2 be the period of *Full Paging* cycle. The period of *Full Paging* cycle is a multiple of the period of *Fast Paging*, i.e., $T_2 = kT_1$, where $k \in \mathbb{N}$. The time between *Full Paging* and the last *Fast Paging* which is before next *Full Paging* is $T_2 - T_1$. The arrivals of packet in this time may be paged by *Fast Paging*. For the arrivals in the first *Fast Paging* cycle, let the probability that UE will be paged by first *Fast Paging* cycle be $P_{fast,1}$. The probability $P_{fast,1}$ is the sum of the probability that UE remains in the same beam as *Full Paging* and UE moves to beams in first *Fast Paging*. The probability that UE does not change beam between *Full Paging* and the first *Fast Paging* is $1-F_e(T_1)$. Let N_d be the number of beams in the cell and C_f be the number of beams transmitted in *Fast Paging*. We assume that if UE changes beam, it will change to any beams with equal probability, which is consistent with the uniform angle of departure measurement [13] and adopted in realistic beam tracking millimeter wave simulator [14]. If UE changes beam after *Full Paging*, the probability that UE moves to the beams in first *Fast Paging* is $\frac{C_f}{N_d}F_e(T_1)$, and then the probability that the first *Fast Paging* hits the UE is,

$$P_{fast,1} = 1 - F_e(T_1) + \frac{C_f}{N_d}F_e(T_1) \tag{3}$$

To the best of our knowledge, there is no mathematical distribution model for beam residence time. The assumption of exponential residence time have been used for many paging research and location management for simplicity [8] [9]. If beam residence time is exponential distribution with mean ρ, then we can derive (3) to be,

$$P_{fast,1} = e^{-\frac{T_1}{\rho}} + \frac{C_f}{N_d}(1 - e^{-\frac{T_1}{\rho}}) \tag{4}$$

If the first *Fast Paging* does not find the UE, then the subsequent *Fast Paging* may find the UE with probability $\frac{C_f}{N_d}F_e(T_1)$. Since, the length of time between two *Full Paging* cycle includes $k-1$ *Fast Paging* cycle, we can derive the probability that the arrival of packet in the first *Fast Paging* cycle trigger *Fast Paging* hitting the UEs with probability as the following,

$$P_{fh,1} = 1 - (1 - P_{fast,1})(1 - \frac{C_f}{N_d}F_e(T_1))^{k-2} \tag{5}$$

By the same way, we can derive the probability that the arrival of packet in the m^{th} *Fast Paging* cycle trigger *Fast Paging* hitting the UEs with probability as the following,

$$P_{fh,m} = 1 - (1 - P_{fast,m})(1 - \frac{C_f}{N_d}F_e(T_1))^{k-1-m} \tag{6}$$

where,

$$P_{fast,m} = 1 - F_e(mT_1) + \frac{C_f}{N_d}F_e(mT_1) \tag{7}$$

If beam residence time is exponential distribution with mean ρ, then we can derive (7) to be,

$$P_{fast,m} = e^{-\frac{mT_1}{\rho}} + \frac{C_f}{N_d}(1 - e^{-\frac{mT_1}{\rho}}) \tag{8}$$

Next, we assume the traffic arrival is Poisson process. By Poisson arrivals see time averages (PASTA) [15], we can obtain the probability that UE will be found in *Fast Paging* by the arrivals in $T_2 - T_1$.

$$P_{fast,hit} = \frac{1}{k-1} \sum_{m=1}^{k-1} P_{fh,m} \tag{9}$$

If the gNB does not find the UE by *Fast Paging* before *Full Paging*, the UE will be paged in *Full Paging*. This occurs with the probability $(1 - P_{fh,m})$ for packet arrival in m^{th} *Fast Paging* cycle. If the packet arrives in first $T_2 - T_1$ and the UE miss all *Fast Paging*, then the average delay of this situation will be,

$$d_{\overline{fast}} = \frac{\sum_{m=1}^{k-1}(1 - P_{fh,m})(k - \frac{2m-1}{2})T_1}{\sum_{m=1}^{k-1}(1 - P_{fh,m})} \tag{10}$$

The average paging delay for *Fast Paging* is d_{fast} which is derived from summation of multiplying probability that UE is found in m^{th} *Fast Paging* cycle and the delay in m^{th} *Fast Paging* cycle. Note that the arrivals in last T_1 before *Full Paging* will not trigger *Fast Paging*. Then, the average paging delay for a delay-sensitive UE is,

$$D = \frac{T_1}{2k} + \frac{k-1}{k}[P_{fast,hit}d_{fast} + (1 - P_{fast,hit})d_{\overline{fast}}] \tag{11}$$

If the number of UEs present in the cell is n and the ratio of delay-sensitive UEs is x, then the expected number of beams need to page in *Fast Paging* is $\mathbb{E}[C_f]$,

$$\mathbb{E}[C_f] = N_d(1 - (\frac{N_d - 1}{N_d})^{xn}) \tag{12}$$

the total number of paging sweeping beams in one $T_2 = kT_1$ period per radio frame is,

$$C = \frac{N_d + (k-1)C_f}{T_2} \tag{13}$$

Our goal is to minimize the paging sweeping beams per radio frame and meet the delay-sensitive UE paging delay requirement D_b. That is to say,

$$\min C = \frac{N_d + (k-1)C_f}{T_2} \tag{14}$$

$$\text{subject to } D \le D_b \tag{15}$$

$$T_2 = kT_1 \tag{16}$$

$$\text{where } k, T_1, T_2 \in \mathbb{N} \tag{17}$$

Because both T_1 and $T_2 = kT_1$ are numbers of radio frames, these numbers must be natural numbers. Note that although increasing T_1 can reduce paging sweeping beams per radio frame, it also affects the probability of paging hit and can also increase the paging delay. Proper choice of this parameters will be discussed in the next section.

C. Period of Paging Cycle

Since the probability $P_{fast,m}$, which determines the constraint (15), follows nonlinear property, the optimization problem (14) is still an integer nonlinear programming problem which is known as NP-hard. We solve this problem using branch-and-bound way as following Algorithm 1.

Algorithm 1 Find the period of paging cycle

Input: Delay constraint D_b, shortest DRX cycle $T_{1,min}$.
Output: The optimal period of fast paging cycle T_1^*, the optimal period of full paging cycle T_2^*.

1: **for** $k \leftarrow 2$ **to** k_{max} **do**
2: $D \leftarrow D(T_{1,min}, k \times T_{1,min})$ by Algorithm 2
3: **if** $D > D_b$ **then**
4: $T_{2,max} \leftarrow (k-1) \times T_{1,min}$
5: break
6: **end if**
7: **end for**
8: **for** $T_2 \leftarrow T_{2,max}$ **to** $2 \times T_{1,min}$ **do**
9: find T_{1List} as all nonnegative divisors except 1 and T_2 of an integer T_2
10: sort the elements of T_{1List} in descending order
11: **for** $j \leftarrow 1$ **to** length of T_{1List} **do**
12: $T_1 \leftarrow T_{1List,j}$
13: $D \leftarrow D(T_1, T_2)$ by Algorithm 2
14: **if** $D \le D_b$ **then**
15: $C \leftarrow \frac{N_d + (k-1)C_f}{T_2}$
16: **if** $C < C_{min}$ **then**
17: $C_{min} \leftarrow C$
18: $T_1^* \leftarrow T_1$
19: $T_2^* \leftarrow T_2$
20: **end if**
21: break
22: **end if**
23: **end for**
24: **end for**

Algorithm 1 is presented as two parts. First, for loop from line 1 to 7 calculates the maximum possible value of T_2. This works by the observation that the maximum feasible value of T_2 is a multiple of the minimum value of T_1. Any larger T_2 as a multiple of larger T_1 fails to meet the paging delay constraint. The minimum value $T_{1,min}$, which may be provided by system information, is shortest DRX cycle of UEs. Second, for loop from line 8 to 24 calculate the optimal solution T_1^*, T_2^* of the problem (14). To be more specific, any feasible solution will enter line 14 if to compare with the current minimize number of paging sweeping beams per frame. If new number of paging sweeping beams per frame is smaller than the record, the optimal value and optimal solution will update accordingly. Note that without explicitly

Algorithm 2 Calculate paging delay $D(T_1, T_2)$

Input: The period of paging cycle T_1, T_2.

Output: The paging delay D.

$k \leftarrow T_2/T_1$

$P_{fast,hit} \leftarrow 0$

for $m \leftarrow 1$ **to** $k-1$ **do**

  $P_{fast,m} \leftarrow 1 - F_e(mT_1) + \frac{C_f}{N_d} F_e(mT_1)$

  $P_{fh,m} \leftarrow 1 - (1 - P_{fast,m})(1 - \frac{C_f}{N_d} F_e(T_1))^{k-1-m}$

  $P_{fast,hit} \leftarrow P_{fast,hit} + P_{fh,m}$

end for

$P_{fast,hit} \leftarrow P_{fast,hit}/(k-1)$

calculate $d_{fast}, d_{\overline{fast}}$

$D \leftarrow \frac{T_1}{2k} + \frac{k-1}{k}[P_{fast,hit} d_{fast} + (1 - P_{fast,hit}) d_{\overline{fast}}]$

initializing C_{min} in Algorithm 1, C_{min} is initialized as largest number in implementation. Also, k_{max} can be set as the largest integer in system. The *break* in line 21 works from the observation that the objective function (13) increases with k when T_2 is fixed.

IV. Evaluation and Analysis

In this section, we present the results for paging delay of a UE in INACTIVE state with the FTP traffic based on 3GPP [16] which is Poisson distribution with the average packet arrival rate of 1 packet per 60 sec. The beam residence time can be estimated by exponential moving average or other estimation method. In Fig. 4, we validate the simulation result by event trigger simulator and the mathematical results. The error rate between event trigger simulation and mathematical analysis is below 1%. Fig. 4 shows that the shorter the beam residence time, the longer the paging delay because the short beam residence time results in low fast paging hit probability as formula (8) shown.

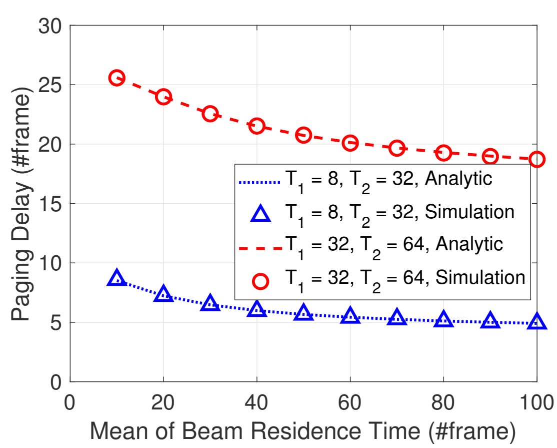

Fig. 4. Effects of beam residence time on paging delay on varying period.

For numerical results in Fig 5, we assume sixteen beams in a cell, and some delay-sensitive UEs randomly associates with six beams, and mean of beam residence time $\rho = 100$ radio frames. The figure is ploted as the number of paging sweeping beams per frame versus the constraint on paging delay of delay-sensitive UEs. The optimal full paging cycle and fast paging cycle is obtained from Algorithm 1. The proposed paging mechanism outperforms total paging beam sweeping which is similar the method in [5], especially in short delay constraint. Because our proposed mechanism reduces the paging sweeping beams in fast paging, gNB use fast paging between full paging to easily meet the delay constraint. In contrast, although total paging beam sweeping can guarantee that UE receives paging in each paging occasion, it also implies that it needs to sweep all paging beams once within paging delay constraint. Therefore, total paging beam sweeping cost more paging beams under the same paging delay constraint.

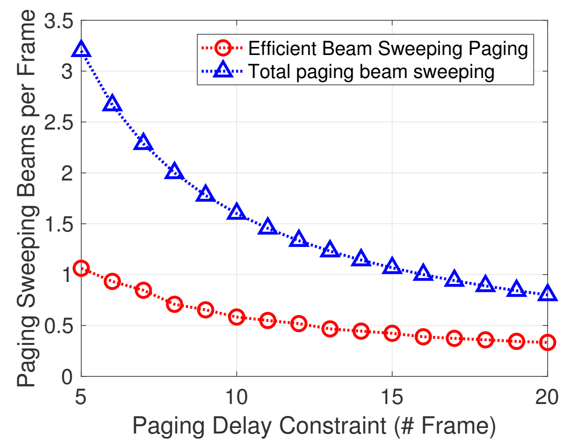

Fig. 5. Compare proposed paging mechanism with total paging beam sweeping under the same paging delay constraint for delay-sensitive UEs

Also, Fig. 6 demonstrates the paging DRX cycles with the same parameter setting in Fig. 5. Fig. 6 demonstrates the period of fast paging cycle is longer than total beam sweeping paging cycle. This implies UEs can wake up once with longer DRX cycle and save more power. Therefore, the proposed scheme also achieves better UE power-efficiency.

V. Simulation Experiments

In this section, we verify the performance of efficient beam sweeping paging under realistic mmWave channel model [14] and UE moving pattern is based on the random way-point model [17]. The stopping duration of random way-point model is uniformly distributed between 0 and 36 seconds. The UE speed is uniformly distributed between 1 and 10 meter per second. The simulation channel parameters are adopted from 3GPP technical report [18]. The gNB can beamform twenty-four directional beams. The beam residence time is estimated by exponential moving average. The delay constraint is 10 radio frames, i.e. 100 ms.

The average number of paging beams per frame for the proposed paging mechanism varies when UE moves and beam residence time changes in each iteration in Fig. 7. Since, UE may change moving velocity or stop randomly at points, this

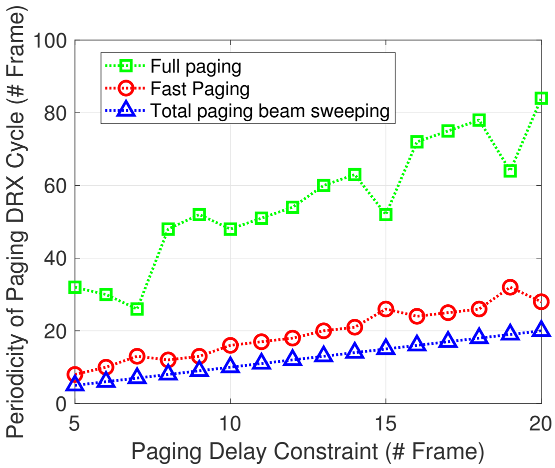

Fig. 6. Compare DRX cycle of proposed paging mechanism with total paging beam sweeping under the same paging delay constraint for delay-sensitive UEs

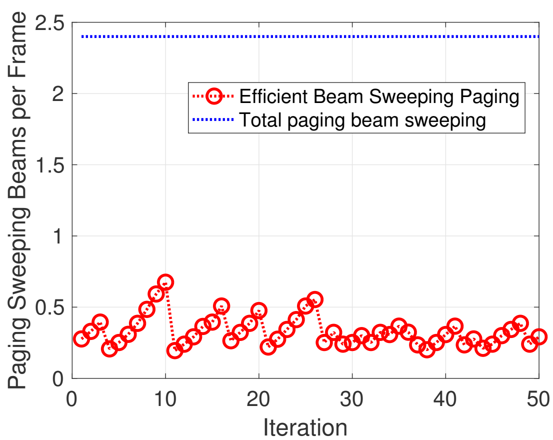

Fig. 7. Proposed paging mechanism vs total paging beam sweeping under the same paging delay constraint 100 ms for delay-sensitive UE in random way-point mobility model and realistic mmWave propagation simulation.

can make estimated value of beam residence time vary in each iteration. As the simulation is executed, the estimation value of beam residence time keep updating in each iteration. In contrast, the total paging beam sweeping can always page the UE in spite of UE mobility because gNB page the UE using all beams. Hence, the paging sweeping beams per frame is fixed as $\frac{N_d}{D_b}$ for total beam sweeping during simulation. The result demonstrates that in spite of complicated UE mobility pattern, the proposed efficient beam sweeping paging method can also reduce large overhead comparing with total beam sweeping.

VI. Conclusion

In this paper, we proposed an efficient beam sweeping paging mechanism to reduce paging sweeping beams. Also, the DRX cycle of UE is able to be longer to achieve better power-efficiency. We introduce *Fast Paging* between *Full Paging* to efficiently page the delay-sensitive UE in inactive state with RAN paging. We also design an algorithm to find the optimal period of paging cycle with minimum paging sweeping beams while meet the paging delay constraint. The simulation results show that the number of paging sweeping beams of proposed scheme is over 50% smaller than which of another simple scheme. The implementation of proposed scheme in real world testbed ramains as the future work.

References

[1] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, 2011.

[2] W. B. Abbas, F. Gomez-Cuba, and M. Zorzi, "Millimeter Wave Receiver Efficiency: A Comprehensive Comparison of Beamforming Schemes with Low Resolution ADCs," *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 8131–8146, 2017.

[3] C. Ho, A. Huang, P. Hsieh, and H. Wei, "Energy-Efficient Millimeter-Wave M2M 5G Systems with Beam-Aware DRX mechanism," in *IEEE 86th Vehicular Technology Conference (VTC2017-Fall)*, Toronto, Canada, 2017, pp. 1–5.

[4] Theodore S. Rappaport, R. W. Heath Jr., R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*, 1st ed. Pearson, 2014.

[5] 3GPP R2-1710097, "Paging in NR- Beamforming Aspects," 2017.

[6] 3rd Generation Partnership Project (3GPP), "Technical Specification Group Radio Access Network; Interlayer procedures in Connected Mode," 3GPP 25.303 V14.0.0, Tech. Rep., 2017. [Online]. Available: www.3gpp.org

[7] S. Hailu and M. Saily, "Hybrid paging and location tracking scheme for inactive 5G UEs," *2017 European Conference on Networks and Communications (EuCNC)*, pp. 1–6, 2017.

[8] T. Deng, X. Wang, P. Fan, and K. Li, "Modeling and Performance Analysis of a Tracking-Area-List-Based Location Management Scheme in LTE Networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6417–6431, 2016.

[9] K. Li, "Location distribution of a mobile terminal and its application to paging cost reduction and minimization," *Journal of Parallel and Distributed Computing*, vol. 99, pp. 68–89, 2017.

[10] 3GPP R2-163441, "Discussion of RRC States in NR," 2016.

[11] 3GPP R2-1709554, "Paging mechanism with beam sweeping," 2017.

[12] Sheldon M. Ross, *Introduction to Probability Models*, 11th ed. Academic Press, 2014.

[13] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, 2014.

[14] J. C. Guey, M. P. Chang, C. H. Yu, and C. C. Su, "Modeling and evaluation of beam tracking in mobile millimeter wave communication," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, vol. 2015-Decem, pp. 775–780, 2015.

[15] Ronald W. Wolff, "Poisson arrivals see time averages," *Operations Research*, vol. 30, no. 2, pp. 223–231, 1982.

[16] 3rd Generation Partnership Project (3GPP), "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects," 3GPP 36.814 V9.2.0, Tech. Rep., 2017. [Online]. Available: www.3gpp.org

[17] C. Bettstetter, H. Hartenstein, and X. Pérez-Costa, "Stochastic properties of the random waypoint mobility model," *Wirel. Netw.*, vol. 10, pp. 555–567, 2004.

[18] 3rd Generation Partnership Project (3GPP), "Technical Specification Group Radio Access Network; Study on channel model for frequency spectrum above 6 GHz," 3GPP 38.900 V14.2.0, Tech. Rep., 2016. [Online]. Available: www.3gpp.org