Can Online Learning Increase the Reliability of Extreme Mobility Management?

Yuanjie Li
Tsinghua University

Esha Datta
UC Davis

Jiaxin Ding
Shanghai Jiaotong University

Ness B. Shroff
The Ohio State University

Xin Liu
UC Davis

***Abstract*—Seamless Internet access under extreme user mobility is highly demanded on high-speed trains and vehicles. However, existing mobile networks (e.g., 4G LTE and 5G NR) cannot reliably satisfy this demand, with a 5.5%–12.6% handover failure ratio at 200–350 km/h. A root cause is that, the 4G/5G handovers have to balance the *exploration* of more measurements for satisfactory handover and the *exploitation* for timely handover before the fast-moving user leaves the coverage.**

We design `BaTT`, an online learning solution for reliable handovers in extreme mobility. `BaTT` decomposes the exploration-exploitation tradeoff into two multi-armed bandit problems. It uses ϵ-binary-search to optimize the threshold of a serving cell's signal strength to initiate the handover with $\mathcal{O}(\log J \log T)$ regrets. It further adopts opportunistic Thompson sampling to optimize the sequence of target cells measured for reliable handovers. `BaTT` can be implemented using the recent Open Radio Access Network (O-RAN) framework in operational 4G LTE and 5G NR. Our evaluations over a dataset from operational LTE networks on the Chinese high-speed rails show a 29.1% handover failure reduction at the speed of 200-350 km/h.

***Index Terms*—Mobile network, 5G and beyond, extreme mobility, reliability, online learning, multi-armed bandit**

I. Introduction

The wide adoption of high-speed rail has made extreme mobility a reality. Today, passengers on the high-speed train want always-on Internet access at up to 350km/h. A common solution is mobile networks, such as 4G LTE and 5G NR. However, 4G/5G struggles to retain reliable services in extreme mobility. Empirical studies from real high-speed rail show the handover failure ratio ranges from 5.5% to 12.6%, which is 2× compared to low-mobility scenarios (§III-A).

A challenge for reliable handovers in extreme mobility is the *exploration-exploitation tradeoff*. To decide the next cell a user should migrate to, the serving cell asks the user to measure available candidate cells (Figure 1). But in extreme mobility, the fast-moving user may leave the serving cell's coverage before initiating a handover, thus losing network services. It has to balance the need to take more measurements (exploration) for a satisfactory decision, and the demand to make a timely, successful handover (exploitation) before the fast-moving user leaves its coverage. The optimal tradeoff depends on diverse factors, such as the train's speed, wireless qualities, multi-path fading, Doppler effect, and the external environment change. A static, manually-crafted, or offline handover decision policy will fall short in responsively optimizing handovers in such a dynamic environment.

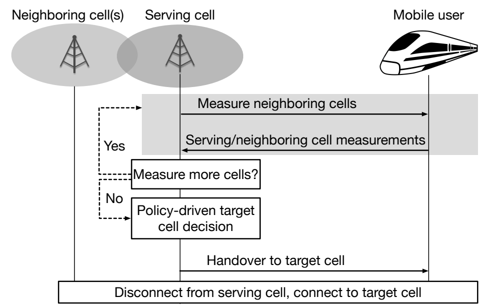


Fig. 1: Wide-area mobility management in 4G/5G today.

This paper studies if online learning can help automate the optimization of reliable handovers in a dynamic environment. The exploration-exploitation dilemma has been extensively studied in online learning. It is provably responsive, adaptive, and robust to the dramatic environmental changes in extreme mobility. We show online learning can help enhance handover reliability but be customized for extreme mobility to become a practical solution. In §III, we formulate this exploration-exploitation tradeoff as two distinct problems:

1) *When to initiate the exploration:* For each user, the serving cell should first identify an optimal threshold to trigger the handover and measurement procedure.
2) *What target cell sequence to explore:* With the threshold, the serving cell determines when and in what sequence to take a measurement, and when to execute a handover.

This decomposition is aligned with the readily-available mechanisms in 4G/5G, thus facilitating implementations in reality.

We devise `BaTT`, an online learning-based solution to both problems (§IV). To determine *when* to start the measurements, we formulate the problem as a J-armed stochastic bandit problem over T rounds, and solve it with ϵ-Binary-Search-First with $\mathcal{O}(\log J \log T)$ regret. To solve *what sequence* of target cells to measure, `BaTT` formulates it as an opportunistic bandit with side observations. We adopt opportunistic Thompson sampling to solve this problem with $\mathcal{O}(\log T)$ regret. Both algorithms can be realized under the recent O-RAN framework [1] (§V). To our best knowledge, `BaTT` is the first use case in O-RAN for reliability rather than other performance QoS metrics, thus complementing existing use cases [2], [3]. Our experiments in §VI with an operational LTE dataset on the Chinese high-speed trains show `BaTT` reduces 29.1% handover failures compared to the 4G/5G handover policies and has lower regret than traditional multi-armed bandit policies.

978-0-7381-3207-5/21/$31.00 ©2021 IEEE

User speed (km/h)	200	300	350
Total handover failures	5.5% (100%)	12.1% (100%)	12.6% (100%)
• **Due to the serving cell**	4.9% (90.0%)	9.3% (77.1%)	11.0% (87.3%)
• **Due to the target cell**	0.6% (10.0%)	2.8% (22.9%)	1.6% (12.7%)

TABLE I: Handover failures in extreme mobility

II. Wide-Area Mobility Management in 4G/5G

The 4G LTE and 5G NR mobile networks offer wide-area mobility support for ubiquitous network access. They deploy base stations to cover geographical areas. Each base station may run multiple cells under various frequency bands (using separate antennas) with different coverage and performance. Each user is primarily served by a single cell, and migrated to another (*handover*) if it leaves the current cell's coverage.

Figure 1 shows the handover in 4G/5G based on the radio resource control (RRC) protocol [4], [5]. When a user connects to a serving cell, it receives a list of neighboring cells. The user can measure these cells' radio quality sequentially. If any neighboring cell satisfies the standard triggering criteria (e.g., its signal strength is offset better than the serving cell's), the user will report this measurement to the serving cell. The serving cell will then run its local handover policy to decide if more neighboring cells should be measured, whether handover should begin, and which target cell the user should hand over to. If the serving cell chooses to take new measurements, it will provide the user with a new cell list. If it chooses to hand over, the serving cell will send the handover command with the target cell's identifier to the user. The user will then disconnect from the serving cell and connect to the target cell.

III. Problem and Formulation

A. Is 4G/5G Reliable in Extreme Mobility?

The current 4G/5G handover is primarily designed for static and low-mobility scenarios. Recent studies [6], [7] have shown that fast-moving users suffer from non-negligible handover failures, thus frequently losing Internet access. Table I shows the LTE handover failure ratios from a Chinese high-speed train from Beijing to Shanghai based on the dataset from [7]. On average, 5.5%, 12.1% and 12.6% handovers fail at the train speed of 200km/h, 300km/h and 350km/h, respectively.

By analyzing the LTE signaling messages, we find the *late handover* causes 77.1%–90.0% of these failures, i.e., by the user not receiving the handover command from the serving cell by the time it leaves the serving cell?s radio coverage. The remaining failures occur when the user receives the handover command from the serving cell but fails to connect to the new target cell, i.e., the selected target cell is unreliable.

B. New Challenge: Exploration-Exploitation Trade-off

Frequent handover failures in extreme mobility arise from the dilemma between *exploration* (more measurements for satisfactory target cell selection) and *exploitation* (fast measurements for timely handover). As explained in §II, the serving cell relies on the user to measure and report the cells' signal strengths for the handover decision. The moving user must deliver these measurements *before* it leaves the serving cell?s radio coverage to retain Internet access. But finding a satisfactory target cell may require scanning and measuring *all* available cells sequentially. As shown in Figure 2a, on average, a user on a Chinese high-speed train measures 16 different neighboring cells before making a handover decision. If the user is moving fast, it may not be able to deliver all its measurements and trigger a handover before leaving its serving cell's radio coverage (resulting in a late handover failure). Reducing the number of cells to measure can mitigate late handovers, but risks missing better cells and therefore committing a handover to an unreliable target cell (thus failures) [1]. This exploration-exploitation dilemma only becomes apparent with the recent emergence of extreme mobility.

C. Problem Formulation

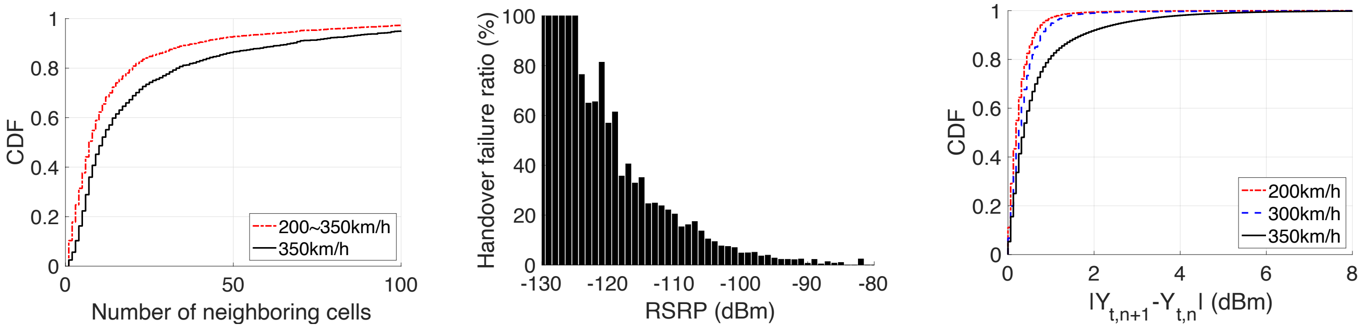
As shown in §III-B, a fast-moving user has a short but critical time window to conduct measurements for handover. It should use this period effectively by measuring the right target cells *before* losing the service. For reliable handover, we must answer two questions: 1) When should measurement start? 2) what is the right sequence of target cells to measure?

To answer both questions, we formulate the reliable handover problem as follows. Consider a serving cell with K neighbor cells. Given a set of mobile users $t = 1, \ldots, T$, our goal is to minimize the handover failure ratio for all T users.

When should the measurements for handover begin? As a user leaves the serving cell's coverage, its signal strength weakens. So the user's critical time starts when the user-perceived signal strength of the serving cell is below a certain threshold. In 4G/5G, this threshold has been defined (A2 in [4], [5]) and configurable for each cell. Manually tuning this threshold is a hard task. On one hand, it should be high enough so that 1) handover failure will not often occur due to weak serving cell; and 2) the user has sufficient time to measure neighboring cells to obtain a good target cell for handover. On the other hand, this threshold should also be low enough to avoid "ping-pong" loops (where a user oscillates between two cells due to signal fluctuations) and a false start (when a desirable target cell is too far away to measure appropriately).

To automatically learn this threshold, we formulate this "when" question as the closest sufficient threshold identification problem. We consider *discrete* signal strengths standardized in 4G/5G [4], [5]. Given a serving cell, let $\{Z_j\} = Z_1, \ldots Z_j, \ldots, Z_J$ be the sequence of J serving cell signal strength observed by a mobile user. Let $[J]$ denote the list $\{1, 2, \ldots, J\}$. Let the random variable $f(Z_j)$ indicate a handover failure event due to the serving cell's signal strength Z_j. Note $f(Z_j) \in \{0, 1\}$, where 0 indicates handover failure and 1 indicates success. The probability of a handover failure due to signal strength Z_j is $r_j = \mathbb{P}[f(Z_j) = 0]$. Let R be

[1] Although high-speed trains operate on the fixed route at a predictable speed, the available candidate cells at each time are *not* always fixed or predictive. This is because each cell's quality varies over time [6], [7] and depends on various unknown factors like operators' cell breath (for energy saving), multi-path fading, external environmental changes, to name a few. Without runtime measurements, simply caching or predicting the target cells will risk missing better cells and causing handover failures.



(a) Number of neighboring cells (b) Failures w.r.t. signal strength (350km/h) (c) Dynamics of adjacent measurements

Fig. 2: Characteristics of LTE handovers over Chinese high-speed train.

the given serving cell handover failure tolerance level. We assume that as the serving cell's signal strength Z_j increases, the handover failure probability r_j monotonically decreases. This is coherent with the physical laws and empirical results from the high-speed rail dataset in Fig. 2b. Our goal is to find a threshold $M \in \{Z_j\}$ that is the smallest Z_j such that $r_j \leq R$. That is, M is the lowest signal strength at which the probability of handover failure is no larger than R.

What target cells to measure? Given threshold M, the next issue is to decide the sequence of target cells to measure and the time to stop measurement for handover. Consider a user t. Once the measurement procedure is triggered, the serving cell starts a sequence of measurements of neighboring cells indexed by n. Whether to take more measurements or to execute a handover is the central exploitation-exploration dilemma faced by the serving cell. This tradeoff may vary from user to user, depending on their movement speed.

At the n-th measurement, let $I_{t,n}$ be the index of the target cell to measure. In 4G/5G, the user can observe the serving (target) cell's signal strength $Y_{t,n}$ ($X_{I_{t,n}}$). Let X_{best} be the strongest target cell observed thus far. After n measurements, if the handover is decided, the user will migrate to the best target cell with signal strength X_{best}. We define $g(X)$ as the handover failure caused by the target cell with signal strength X, where $g(\cdot)$ and $f(\cdot)$ may be distinct functions. The handover failure probability of user t is $\mathbb{E}[f(Y_{t,n})g(X_{best})]$ when the handover happens after n measurements and X_{best} is the best target cell. In general, $Y_{t,n}$ decreases with n as the mobile user is moving away from the serving cell. Therefore, the tradeoff is whether to make more measurements, which improves X_{best}, but at the risk of decreasing $Y_{t,n}$. Today, the typical practice is to measure target cells following a fixed sequence and trigger actual handover when X_{best} is greater than or equal to $Y_{t,n}$ plus an offset quantity determined by the network provider [4], [5]. The objective of the "what sequence" question is to decide the best order of target cell measurement and when to stop measurement for handover.

Online or offline learning? Both online learning and offline learning may solve the "when" and "what" problems. We note that online and offline algorithms have fundamentally different assumptions concerning whether there exists sufficient prior data to learn an optimal policy. In online learning, we do not assume such prior information. The algorithm adapts to learned information and adapts to the optimal base station selection (quickly). For comparison, offline learning requires sufficient information to learn a good policy for a *fixed* environment. The cellular networks experience frequent dynamics and configuration changes. In this situation, online learning can much better capture such dynamics.

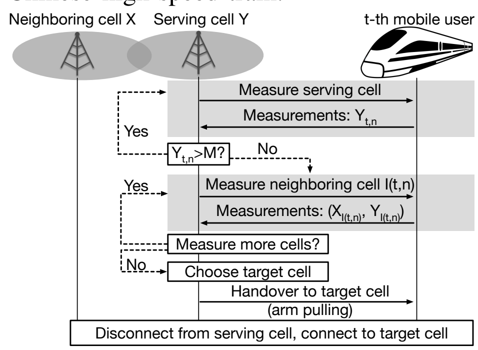


Fig. 3: Workflow of BaTT in extreme mobility.

IV. BaTT: Reliable Handover via Bandits

We show how online learning helps solve both problems for reliable handovers in extreme mobility. Our solution, BaTT, explores *multi-armed bandit* algorithms. Compared to other learning algorithms, bandit algorithms are lightweight and responsive for fast-moving users. Moreover, bandit algorithms are highly adaptive to environmental dynamics, network configuration changes, and user movement variations.

A. When: ϵ-Binary-Search-First

Recall that our objective is to find the handover threshold M, i.e., the lowest signal strength at which the probability of serving cell handover success is no smaller than R. Clearly, exploring each value of the J signal strengths is expensive. Instead, we should leverage the monotonicity property between the signal strength and handover failure rate.

To this end, we propose the ϵ-Binary-Search-First based on multi-armed bandit algorithms. We provide an exploration parameter $0 \leq \epsilon \leq 1$. Each arm $j \in [J]$ is associated with a random variable $f(Z_j)$ where $\mathbb{E}[f(Z_1)] \leq \mathbb{E}[f(Z_2)] \leq \cdots \leq \mathbb{E}[f(Z_J)]$. The goal is to identify the optimal threshold:

$$M = \arg\min_{Z_j} \{\mathbb{E}[f(Z_j)] \geq R\} \quad (1)$$

Algorithm 1 shows how ϵ-Binary-Search-First works. In exploration, it pulls the arms in a binary search manner (Binary-Arm-Search subroutine). Exploration lasts $\leq \epsilon T$ rounds. In exploitation, Algorithm 1 identifies the estimated best arm and pulls it for the remainder of the game.

Algorithm 1 ϵ-Binary-Search-First

Input: J, T, R, $0 \leq \epsilon \leq 1$
Explore: Binary-Arm-Search(J,$\lfloor \frac{\epsilon T}{\log J} \rfloor$,1,$J$,R)
Select arm j such that $\hat{r}_j \geq R$ and $j \in \arg\min_{i \in [J]} |\hat{r}_i - R|$
for remaining rounds $n \leq T$ **do**
  Play arm j
end for
$M = \arg\min_{Z_j} \{\mathbb{E}[f(Z_j)] \geq R\}$

Binary-Arm-Search(J, P, R, Start, End)

if End $\geq$ Start **then**
  Play arm $j = \lceil Start + \frac{End-Start}{2} \rceil$ for a total of P times. Denote the empirical mean reward $\hat{r}_j$.
  if $\hat{r}_j \geq R$ **then**
    Return Binary-Arm-Search(J, P, Start, $j-1$, R)
  else
    Return Binary-Arm-Search(J, P, $j+1$, End, R)
  end if
end if

Algorithm 2 Opportunistic Thompson Sampling (TS)

Input: $t, K, \hat{M}$, current TS posterior
1: $n = 0, X_{best} = 0, Y_{t,n} = \infty, B = \emptyset$
2: **if** $X_{best} < \hat{M}$ **then**
3:   **if** $Y_{t,n} > X_{best}$ **then**
4:     Measure target cell $I_{t,n}$ using TS, where $I_{t,n} \notin B$.
5:     Receive ($X_{I_{t,n}}$, $Y_{t,n}$); Update()
6:   **else**
7:     Handover to X_{best}
8:   **end if**
9: **else if** $Y_{t,n} \geq \hat{M} + c$ **then** % "free" observation
10:   Measure target cell $I_{t,n}$ using TS where $I_{t,n} \notin B$.
11:   Receive ($X_{I_{t,n}}$, $Y_{t,n}$); Update()
12: **else**
13:   Handover to X_{best}
14: **end if**
def Update()
15: **if** $X_{I_{t,n}} > X_{best}$ **then**
16:   $X_{best} \leftarrow X_{I_{t,n}}$
17: **end if**
18: $n \leftarrow n+1$, $B = B \cup I_{t,n}$. Update TS posterior distribution of arm $I_{t,n}$

B. What Sequence: Opportunistic-TS

Consider a mobile user t with a given threshold $\hat{M}$ decided by the ϵ-Binary-Search-First algorithm. Once the handover measurement is triggered by $Y_{t,0} < \hat{M}$, our goal is to determine the optimal sequence of target cells to measure.

We propose an opportunistic Thompson sampling (TS) for extreme mobility. As empirically validated in Fig.2c, changes in signal strength over consecutive measurement are bounded. We assume there exists some positive constant c such that $|Y_{t,n} - Y_{t,n+1}| < c$. This ensures the serving cell signal strength does not change quickly between consecutive measurements. Under this assumption, we gain "free" measurements when the best target cell so far is good enough ($X_{best} \geq \hat{M}$) and the serving cell is still strong enough ($Y_{t,n} \geq M + c$). So the next measurement is risk-free. Therefore, we can find the best target cell and then use the "free" observations to satisfy the need for exploration.

Algorithm 2 shows BaTT's opportunistic TS. It requires the number of available neighboring cells K, the user index t, and the threshold $\hat{M}$ from ϵ-Binary-Search-First. If the best target cell X_{best} is not satisfactory ($X_{best} < \hat{M}$), then Algorithm 2 compares the serving cell to X_{best} (Line 2). If $Y_{t,n} > X_{best}$, then the algorithm continues to measure the best unmeasured target selected using TS. If $Y_{t,n} < X_{best}$, the user t handovers to X_{best}. Otherwise, if X_{best} is satisfactory and $Y_{t,n} \geq \hat{M} + c$, then the algorithm can make "free" measurements. Then Algorithm 2 explores an unmeasured target selected using TS.

C. Regret Analysis

When: ϵ-Binary-Search-First. To analyze Algorithm 1, we first define its regret. Let $N_T(j)$ be the number of times a threshold setting j is pulled under a given policy Γ. We define the regret over T rounds as

$$R_\Gamma(T) = T - \mathbb{E}[N_T(a^*)] = \sum_{a \neq a^*} \mathbb{E}[N_T(a)]$$

Define $\Delta = r_M - R$, $D = \min_j |r_M - r_j|$, and $d = \min_j |r_j - R|$, and $\delta = \min(\Delta, D/2)$, where r_M is the probability associated with the signal strength M. Then the regret bound is as follows (with proofs available in [8]):

Theorem 1. *Algorithm 1 achieves regret bounded by*

$$R(T) \leq \log J \left(\frac{\log 6\delta^2 T J}{2\delta^2} - \frac{\log\log J}{2\delta^2} + \frac{1}{2\delta^2} + 1 \right)$$

when $d < \sqrt{\frac{\log(T \log J)}{2P}}$, *where* d *is the minimum absolute distance between a searched arm and* R, *and* $\delta = \min(\Delta, D/2)$.

What Sequence: Opportunistic-TS. We note the exact regret of the general "what sequence" is difficult to evaluate. The reason is $Y_{t,n}$ is an unknown and non-stationary process over n. A simpler case is where the user is only allowed to measure one target cell and then handover to it. In this case, Opportunistic-TS reduces to the classic Thompson Sampling with $O(\log T)$ regret. We note the ability to select among multiple target cells should yield better performance than classic TS in general. This is empirically validated in §VI.

V. Software-defined BaTT in 4G/5G with O-RAN

Traditionally, most 4G/5G handover policies are hard-coded in base stations' proprietary firmware or dedicated hardware, making it hard to adopt new policies like BaTT. Fortunately, the recent advances in O-RAN framework [1] makes it possible for software-defined, learning-driven handover customizations. We next sketch an O-RAN-based BaTT implementation.

An O-RAN primer: Initiated in 2018 by global telecom industry, O-RAN offers open standards for infrastructure vendors and operators to quickly customize their mobile network functions. Figure 4 shows its framework (adopted from the standards [9], [10]). O-RAN provides built-in support for machine learning modules to empower the network intelligence. It introduces the software-defined near-real-time RAN intelligent controller on top of the control and user planes. The intelligent controller reads the radio information base, runs operator-customized AI algorithms, and adapts the control and user-plane operations. It can host diverse applications such

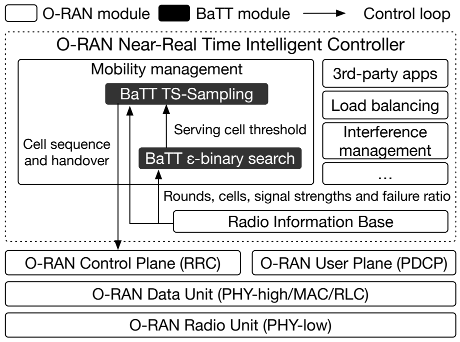
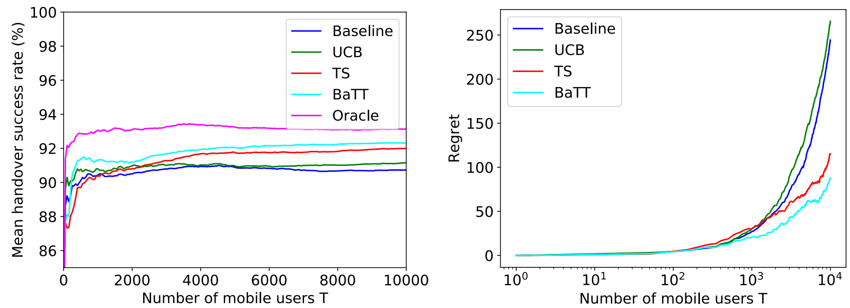


Fig. 4: `BaTT` implementation via O-RAN.



(a) Reliability with M=-120 dBm (b) Regret with M=-120 dBm

Fig. 5: `BaTT` and other algorithms with c=4.

as mobility management, load balancing, radio interference management, and 3rd-party applications from cloud and edge.

`BaTT` implementation with O-RAN: `BaTT` can be realized as an extension of 4G/5G mobility management in O-RAN's near-real-time intelligent controller. As shown in Figure 4, `BaTT` adds two modules in existing mobility management: the ϵ-binary-search in §IV-A, and the opportunistic-TS in §IV-B. Both algorithms' inputs are available from the 4G/5G RRC signaling messages, i.e., RRC measurement report and RRC connectivity reconfiguration. At runtime, O-RAN collects these messages from its control-plane central point and uploads them to the radio information base. Given these inputs, `BaTT` runs Algorithm 1 to output the serving cell's threshold (i.e., A2 in 4G/5G RRC [4], [5]). This threshold is passed to the opportunistic-TS (Algorithm 2). If Algorithm 2 decides to explore more cells, it notifies O-RAN to issue an RRC connectivity reconfiguration message to the user device with the new cells to measure. Otherwise, Algorithm 2 notifies O-RAN to issue a handover command to the end device.

Interoperability with other network functions: As shown above, `BaTT` reuses the standard RRC procedures for measurements (exploration) and handover (exploitation). In practice, these procedures are shared by other functions such as the radio interference management and load balancing. Additional measurements by other functions can prolong `BaTT`'s exploration phase and cause late handovers and failures. To avoid so, we recommend the operators prioritize `BaTT`'s actions since reliability is the prerequisite of other network functions. Other functions can reuse `BaTT`'s measurements via shared radio information base, and refine the handover decisions given multiple equally-reliable target cells from `BaTT`.

Impact on the devices and infrastructure: As a network-side solution, `BaTT` does not modify end devices. Moreover, `BaTT`-empowered base stations can co-exist with legacy ones. It will not negatively impact the legacy base stations and users.

VI. Evaluation

We evaluate `BaTT` using operational LTE dataset in extreme mobility and compare it with existing handover policies.

A. Experimental Setup

Dataset: We use a large-scale 4G LTE dataset on Chinese high-speed trains from [7]. This dataset was collected on the rails between Beijing and Shanghai over 135,719 km of trips. In these tests, a phone using China Mobile or China Telecom 4G LTE runs continuous iperf data transfer on the train at 200–350 km/h. Meanwhile, the phone runs MobileInsight [11] to collect LTE signaling messages from the hardware modem. These messages include 38,646 runtime configurations of neighboring cell lists and thresholds, 81,575 measurement reports of serving/neighboring cell's signal strengths, and 23,779 handover commands as exemplified in Figure 2.

Benchmarks: We conduct a two-step evaluation of `BaTT`. We first evaluate `BaTT` with a given threshold M, and compare its handover success rate with four algorithms: **(1) Oracle**: This is the theoretically optimal solution. It assumes the average handover failure rates of the target cells are known, and select the target cell with the lowest failure ratio. **(2) Baseline**: This is the state-of-the-art 4G/5G handover [4], [5], [11]. It compares the serving cell and target cell's signal strengths and selects the first neighboring cell with $X_{best} > Y_{t,n}$ as the target cell. The user's device measures the target cells randomly. **(3) UCB**: The serving cell maintains UCB estimates for the target cells and instructs a user to measure target cells based these estimates. **(4) TS**: Similar to UCB, except using Thompson sampling. Note in all these algorithms the handover happens when $X_{best} > Y_{t,n}$, which is same as existing 4G/5G handovers. The algorithms differ in deciding the measurement order. The regret is defined as the handover failure rate difference between a the Oracle and a given algorithm. Recall that a failure happens when $f(Y_{t_n}) = 0$ or $g(X_{best}) = 0$, where we draw $f(\cdot)$ and $g(\cdot)$ from real traces as shown in Figure 2b. Also note the evaluation results are driven by real traces and not limited by the assumptions made for the analysis. Next, we assess `BaTT`'s efficiency by comparing its `BaTT`'s ϵ-Binary-Search-First with a uniform threshold search.

B. Results

We test a cell with $K = 10$ neighboring cells. We draw each cell's expected reward (handover success rate) by mapping their signal strength distribution to the handover success rate based on $f(\cdot)$ and $g(\cdot)$ from real traces in Figure 2b. This results in the reward vector [0.76, 0.88, 0.90, 0.91, 0.92, 0.93, 0.94, 0.95, 0.97]. We then replay all sequences of serving cell's measurements before each handover in the dataset. We generate each neighboring cell's measurements based on its empirical distribution of signal strengths from the dataset. Then we run each algorithm to decide the handover target.

Comparison with state-of-the-art Figure 5a shows that `BaTT` improves mean handover success rate from 89.7% to 92.7% compared to the baseline in 4G/5G today. It prevents 29.1% handover failures in 4G/5G today. Note `BaTT` approximates the Oracle, which is the optimal performance we can expect in reliable mobility today. Compared to the baseline, `BaTT` reorders the cells to measure when the serving cell's quality is decreasing, thus mitigating late handover failures.

Comparison with other bandit algorithms Figure 5b shows `BaTT` outperforms UCB and TS in terms of regret (i.e., failed handovers). This is because `BaTT` balances the exploration and exploitation based on the runtime serving cell quality, while UCB and Thompson sampling do not. `BaTT` can accelerate the exploration when the serving cell quality is good and mitigate late handover failures when serving cell quality is not. This is crucial, since late handovers due to the serving cells dominate the handover failures in reality (Table I).

Effectiveness of ϵ-Binary-Search-First Following [4], [5], we consider a serving cell with $J = 81$ signal strength threshold values available to search from -140dBm to -60dBm. We draw each corresponding serving cell failure rate by mapping their signal strength distribution to the handover success rate based on $f(\cdot)$ from the real traces as shown in Figure 2b. We run ϵ-Binary-Search-First over T = 25000 rounds, with results averaged over 10 trials. We compare ϵ-Binary-Search-First with Uniform-Search-First algorithm, which takes ϵT rounds to uniformly sample all available arms and then picks the one with sample mean closest to optimal. Our experiments shows that, compared to Uniform-Search-First, `BaTT` significantly reduces the number of users (from 470 to 189) who select a serving cell with less desirable serving cell signal strength.

VII. Related Work

Reliable mobility management has been actively studied recently, such as its sub-optimal coverage [12], policy conflicts [13], [14], late handovers [15], to name a few. Our work studies a different aspect of handover failures in extreme mobility. In this scenario, [16], [7] report the non-negligible handover failures in reality and [6] unveils the exploration-exploitation tradeoff in handovers and alleviates it by refining wireless communications. In contrast, our work moves further to explicitly address the exploration-exploitation trade-off using online learning. There are efforts to refine the performance of handover with machine learning techniques like XGBoost [15], fuzzy logic [17], neural networks [18], and SVM [19]. Our work differs from them because we focus on the handover reliability. `BaTT` is motivated by recent advances in multi-armed bandits. Its problem formulation is inspired by the cascading bandit [20], [21]. But our problem differs from them since our problem does not assume a known cost.

VIII. Conclusion

This work strives for reliable 4G/5G handover in extreme mobility using online learning. We formulate and decompose the exploration-exploitation dilemma in extreme mobility into two online learning problems. Then we showcase a multi-armed bandit-based strategy to search for the optimal threshold of signal strength to address this dilemma *and* opportunistically balance the exploration and exploitation of target cells based on the runtime serving cell's signal strength. Our analysis shows $\mathcal{O}(\log J \log T)$ overall regret of handover failures. `BaTT` can be incrementally deployed in 4G LTE and 5G NR under the recent O-RAN framework. Experiments with operational LTE datasets from the Chinese high-speed trains demonstrate the viability of handover failure reduction.

Acknowledgements

We thank our anonymous reviewers for valuable comments. This work was partly funded by NSF CNS-1901218, 1901057, 2007231, 1618520, 1409336, and an Office of Naval Research under N00014-17-1-241. Jiaxin Ding acknowledges support from Shanghai Sailing Program 20YF1421300.

References

[1] "O-RAN Alliance: Operator Defined Open and Intelligent Radio Access Networks," https://https://www.o-ran.org/, 2021.

[2] "O-RAN Use Cases Analysis Report (v3.0)," 2020.

[3] "O-RAN Use Cases and Deployment Scenarios," 2020.

[4] 3GPP, "TS36.331: Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC)," Jun. 2020.

[5] ——, "TS38.331: 5G NR: Radio Resource Control (RRC)," Jun. 2020.

[6] Y. Li, Q. Li, Z. Zhang, G. Baig, L. Qiu, and S. Lu, "Beyond 5G: Reliable Extreme Mobility Management," in *ACM SIGCOMM*, 2020.

[7] J. Wang, Y. Zheng, Y. Ni, C. Xu, F. Qian, W. Li, W. Jiang, Y. Cheng, Z. Cheng, Y. Li, X. Xie, Y. Sun, and Z. Wang, "An active-passive measurement study of tcp performance over lte on high-speed rails," in *ACM MobiCom*, 2019.

[8] Y. Li, E. Datta, J. Ding, N. Shroff, and X. Liu, "Proof of Theorem 5.1," https://www.dropbox.com/s/midckn7ia6iam78/iwqos21-proof.pdf?dl=0, 2020.

[9] "O-RAN Architecture Description (v2.0)," 2020.

[10] "O-RAN Operations and Maintenance Interface Specification," 2020.

[11] Y. Li, C. Peng, Z. Yuan, J. Li, H. Deng, and T. Wang, "Mobileinsight: Extracting and analyzing cellular network information on smartphones," in *ACM MobiCom*, 2016.

[12] E. Amaldi, A. Capone, and F. Malucelli, "Radio planning and coverage optimization of 3g networks," *Wireless Networks*, 2008.

[13] Y. Li, H. Deng, J. Li, C. Peng, and S. Lu, "Instability in distributed mobility management: Revisiting configuration management in 3g/4g mobile networks," in *SIGMETRICS*. ACM, 2016.

[14] Z. Yuan, Q. Li, Y. Li, S. Lu, C. Peng, and G. Varghese, "Resolving policy conflicts in multi-carrier cellular access," in *ACM MobiCom*, 2018.

[15] F. B. Mismar and B. L. Evans, "Partially blind handovers for mmwave new radio aided by sub-6 ghz lte signaling," in *ICC Workshops*, 2018.

[16] L. Li, K. Xu, T. Li, K. Zheng, C. Peng, D. Wang, X. Wang, M. Shen, and R. Mijumbi, "A measurement study on multi-path tcp with multiple cellular carriers on high speed rails," in *ACM SIGCOMM*, 2018.

[17] S. Horrich, S. B. Jamaa, and P. Godlewski, "Adaptive vertical mobility decision in heterogeneous networks," in *IEEE ICWMC*, 2007.

[18] S. Horrich, S. B. Jemaa, and P. Godlewski, "Neural networks for adaptive vertical handover decision," in *IEEE WiOpt*, 2007.

[19] M.-S. Ong, F. Magrabi, and E. Coiera, "Automated categorisation of clinical incident reports using statistical text classification," *Quality and Safety in Health Care*, vol. 19, no. 6, pp. e55–e55, 2010.

[20] B. Kveton, C. Szepesvari, Z. Wen, and A. Ashkan, "Cascading bandits: Learning to rank in the cascade model," in *International Conference on Machine Learning*, 2015, pp. 767–776.

[21] R. Zhou, C. Gan, J. Yan, and C. Shen, "Cost-aware cascading bandits," *arXiv preprint arXiv:1805.08638*, 2018.