March 2022

REDUCTION OF HANDOVER DURATION FOR LATENCY SENSITIVE APPLICATIONS IN 5G MOBILE NETWORKS

Krushang Shah

Rajaneesh Shetty

Soundarya Pendyala

REDUCTION OF HANDOVER DURATION FOR LATENCY SENSITIVE APPLICATIONS IN 5G MOBILE NETWORKS

AUTHORS:
Krushang Shah
Rajaneesh Shetty
Soundarya Pendyala

ABSTRACT

Handovers in a Third Generation Partnership Project (3GPP) 5G mobile network occur quite often. During handovers, user packets are buffered either at a user equipment (UE) level or at a network level. In 5G network scenarios, primarily those involving ultra-reliable low latency communication (URLLC) use cases, buffering user packets, as well as handover delays, are highly undesirable. Higher delays can be observed between inter-Radio Access Technology (RAT) handovers or between 3GPP and non-3GPP networks, due to longer handover procedures. Presented herein are techniques that may reduce the 5G handover latency for time sensitive applications through the use of a fast handover procedure.

DETAILED DESCRIPTION

The introduction of 5G mobile networks promises guaranteed Quality of Service (QoS) along with reliable communication methods, as well as the introduction of new technology areas, such as Augmented Reality (AR) and Virtual Reality (VR), that can be designed and developed in a manner that maintain the key deliverables of a 5G core (5GC) network.

However, handovers in any mobile network are quite inevitable and one of the primary procedures that induces latency in user path communications. Currently, 3GPP standards provide a network function (NF) within the 5GC, referred to as the Network Data Analytic Function (NWDAF), which can be leveraged to predict current UE movements and the probably path of a UE, patterns of probably UE locations based on time of day information, probably duration of stay in a current location, and/or probable duration of stay in a current Radio/Non-Radio Access Network ((R)AN).

While the procedures prescribed by current 3GPP Technical Specification (TS) 38.913, 3GPP TS 33.521, and 3GPP TS 29.520 provide guidance around the use of data analytics, as well as possible uses of the NWDAF, current standards do not describe utilizing the NWDAF to improve handover efficiencies. Further, current handover technologies provide guidance regarding algorithms and methods that can be used to collect data and generate some level of actionable insights, does not provide details that indicate how exactly to collect such data, the NF(s) to which to provide the raw data, or how to practically leverage the generated insights.

Presented herein is a fast handover (H/O) procedure through which 5G handover latency can be reduced for time-sensitive applications. In particular, as shown in Figure 1, below, subscribers for which the fast handover procedure may be enabled can be identified within the Access and Mobility Management Function (AMF).

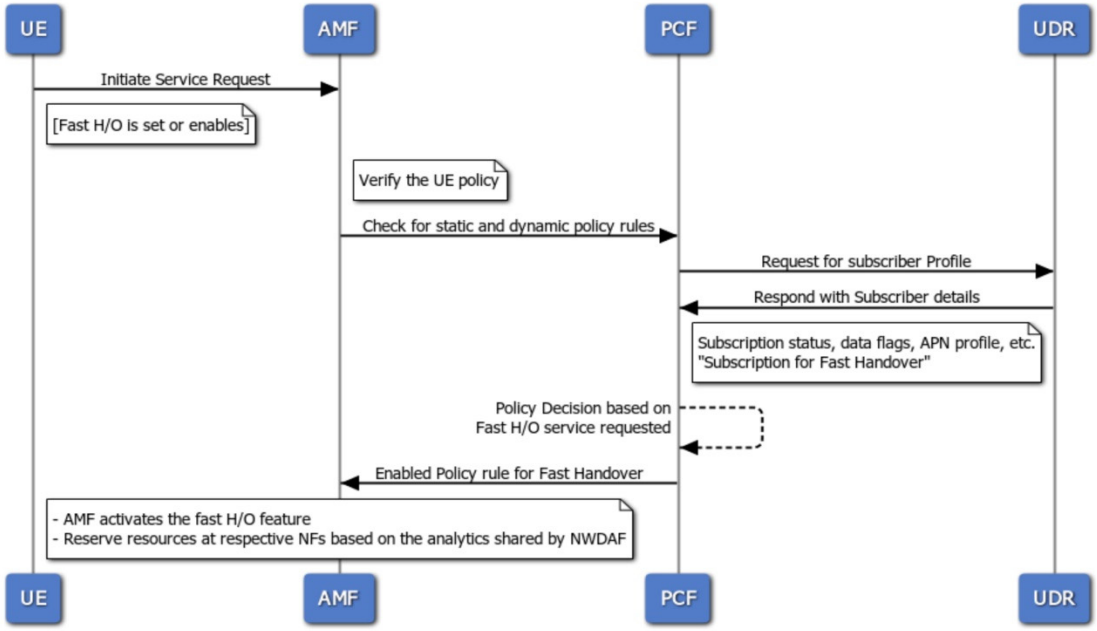

Figure 1: Identifying a UE for which Fast Handover is Enabled

New parameters can be defined on the N8 interface between the AMF and a Unified Data Management (UDM)/Unified Data Repository (UDR) in order to communicate UE subscription information indicating that fast handover is enabled for a given UE. Following identification of a UE for which fast H/O is enabled, the AMF activates a fast H/O feature, through which an algorithm can be executed on the AMF to reserve resources

at one or more NFs based on analytics shared by the NWDAF, as shown in further detail in Figure 2, below, prior to receiving a handover indication for the UE.

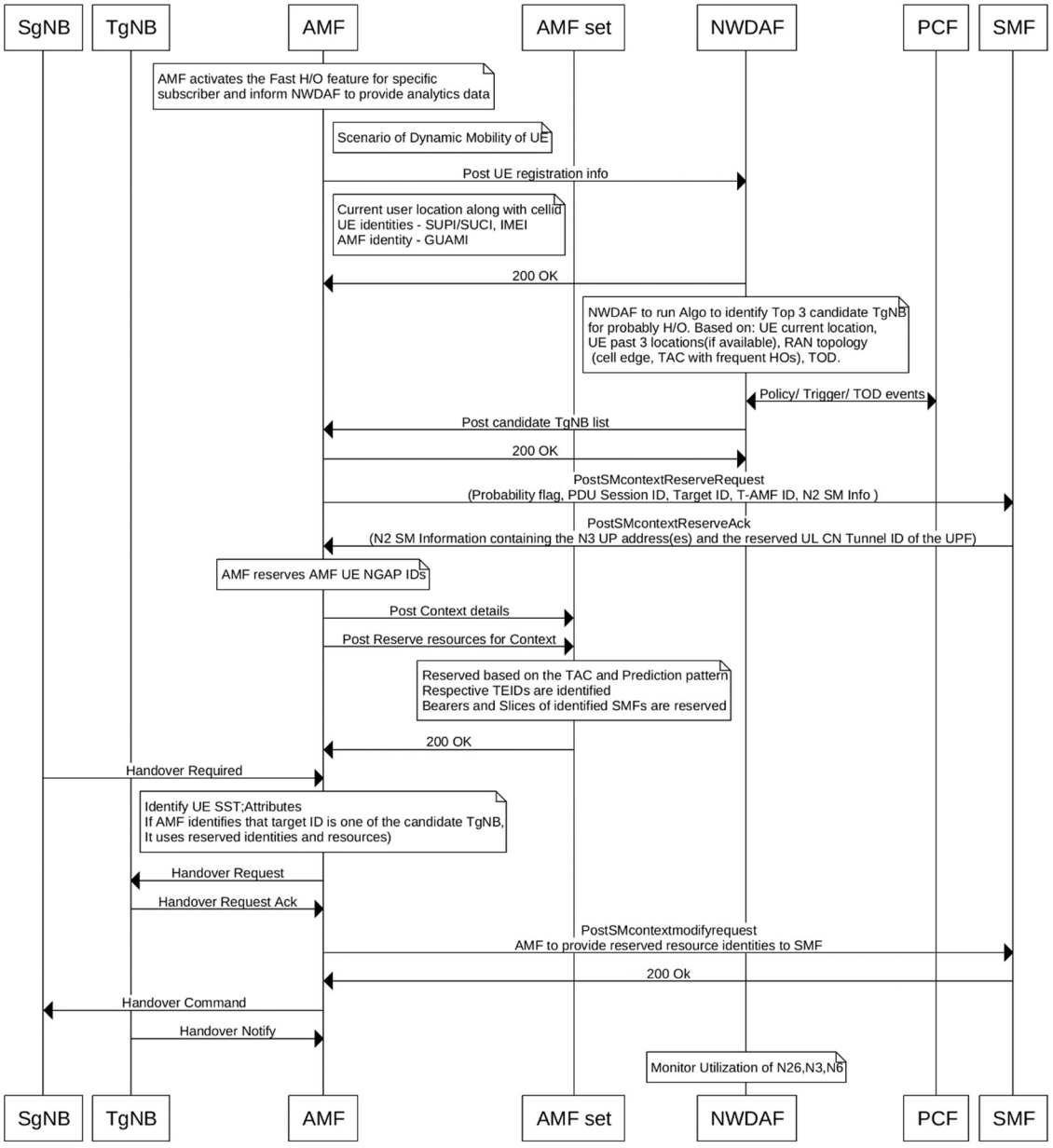

Figure 2: AMF and SMF Resource Readiness Algorithm

With reference to the features illustrated in Figure 2, logic for the NWDAF can identify candidate handover cell(s) for a given UE. New parameters can be introduced between the AMF and the NWDAF in order to provide exchanging feature subscription information, as well as a candidate handover cell indication, between the AMF and the NWDAF.

Following identification of the candidate handover cell(s), the AMF performs a temporary reservation of resources on the Session Management Function(s) (SMF(s)) serving the cell(s) for potential handover of the UE, as well as a temporary reservation of User Plane Function (UPF) resources for potential handover of the UE. Following handover to a particular cell, temporary resources of non-utilized SMF(s)/UPF(s) can be cleared. The technique proposed herein may be useful for Session and Service Continuity (SSC) modes 2 and 3.

As compared to existing standards-based solutions to address handover latencies, such as Dual Active Protocol Stack (DAPS) handovers, DAPS is directed to overcoming handover delays and bearer handling issues at the radio side of a UE. In contrast, the techniques proposed herein provide a soliton to identify and reserve handover resources via various core network functions (e.g., AMF, SMF, UPF, and NWDAF), and do not involve radio side enhancements, thereby avoiding any UE and/or gNodeB impacts. Further, the techniques herein may be implemented as a subscription-based feature, which may have no impact on UEs and/or the air interface between UEs and a given access network.

Advantageously, techniques herein provide for the ability to address potential handover delays by facilitating handover resource allocations on the core network side, in anticipation of potential UE handovers. This can be achieved by executing a predictive algorithm and reserving resources for possible handover scenarios on the core network nodes so that when a measurement report is received from a given UE, there is no extra delay for reservation/allocation on core network nodes, thereby facilitating almost instantaneous handover.

Although a signaling reduction may not be realized through the techniques presented herein, reservation latency and resource allocation latencies can be reduced at various core network nodes. Further, in one instance while executing the predictive algorithm, the current utilization of the AMF/SMF can be considered by the NWDAF. Thus, while reserving resources, if a particular node/function is over-utilized, the node/function can be excluded from a potential candidate cell list, which may facilitate a slight reduction of signaling at over-utilized nodes.