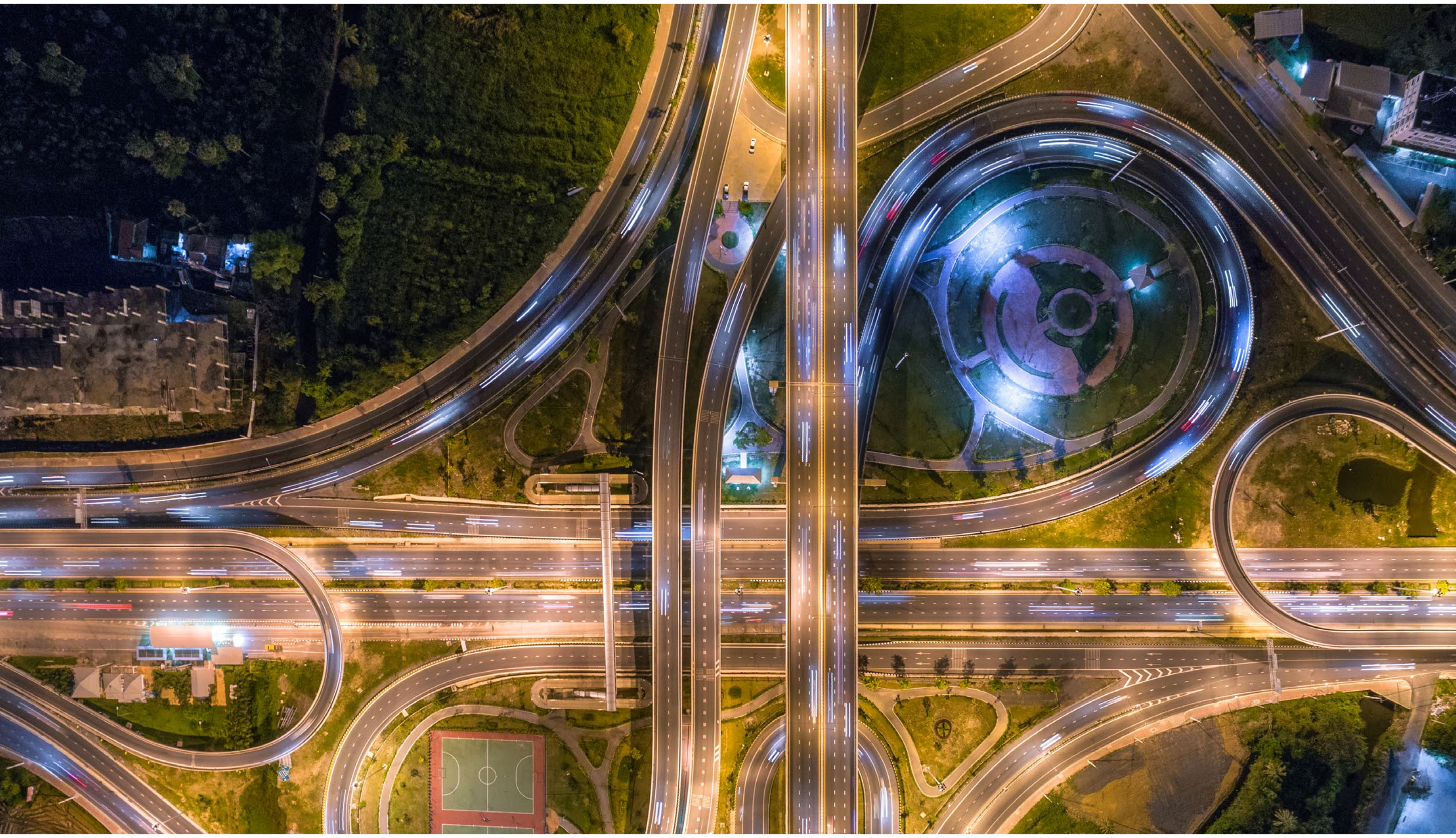
NOKIA



Dynamic Experience Management

User plane access network congestion detection and policy control

Application note

Abstract

The traffic load generated by users of wireless networks continues to grow to unprecedented levels, with highly variable loads across the network and over time. Internet of Things and machine-type communications (IoT/MTC) will further increase data demand and diversify consumption behaviors, with traffic management and quality of experience requirements that are different from consumer applications.

These trends place enormous demands of traffic growth on access networks for any access technology that uses wireless licensed or unlicensed spectrum, or for fixed networks where resource sharing and overbooking will likely remain a common practice. When network congestion occurs, users are impacted in a variety of ways that can disrupt the user experience of the applications they use, with no fairness between users on the shared congested resource.

To address these problems, Nokia Dynamic Experience Management (DEM) is a systemic approach that automatically detects congestion in the user plane and allows it to be mitigated immediately without management or control plane intervention. This document provides an overview of DEM, its potential applications and capabilities.

Contents

The need for automated access network congestion control

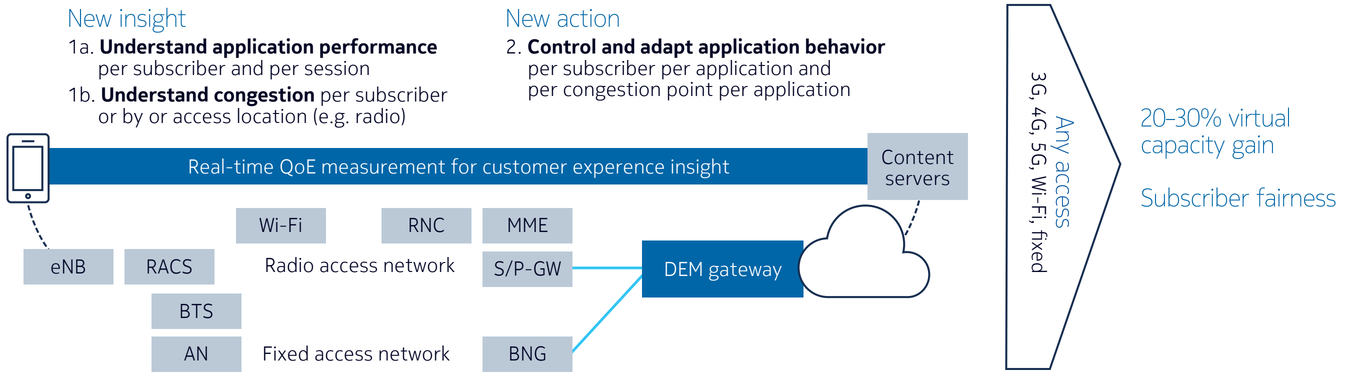
Video and multimedia continue to dominate growth in broadband traffic, with augmented reality (AR) and virtual reality (VR) experiences on the horizon. The traffic load generated by users of wireless networks continues to grow to unprecedented levels, with highly variable loads across the network and over time. Internet of Things and machine-type communications (IoT/MTC) are set to further increase data demand in the future. These trends place enormous demands of traffic growth on access networks for any access technology that uses wireless licensed or unlicensed spectrum, or for fixed networks where resource sharing and overbooking remain common practices.

Due to these circumstances, combined with dynamic and highly variable user behavior, it is a regular occurrence for access networks to congest, meaning more traffic needs to be delivered through the network than can be supported. In the absence of dynamic congestion control solutions, traffic is queued and eventually discarded with no awareness of the value of the traffic, which subscribers on that network are being impacted, or if the resulting use of resources is fair between subscribers. For instance, subscribers with traffic that existed before the congestion event could get much better performance than new users trying to connect over the congested network.

To address these problems, a systemic approach that automatically and simply detects congestion in the user plane and allows it to be managed immediately without management or control plane intervention is essential. The parameters for how to treat subscribers, services and traffic in a manner based on application awareness and quality of experience (QoE) should be proactively defined as a matter of network policy, such that when congestion occurs, it can be managed immediately, avoiding any delays caused by external systems' involvement in real-time policy decisions.

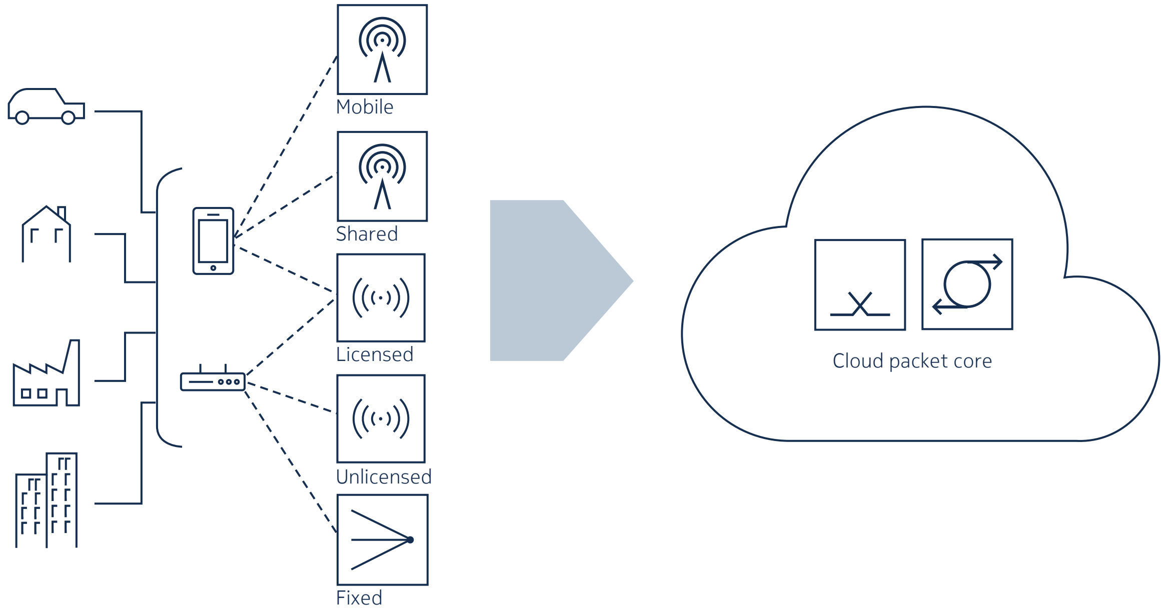
The Nokia solution to this problem is called Dynamic Experience Management (DEM). This is an optional capability available in network functions based on the Nokia Service Router Operating System (SR OS) used for fixed and mobile subscriber management and policy enforcement across a range of wired and wireless access network technologies. It is Dynamic in that is responds automatically in real time to access network congestion events; Experience Management means that the traffic mitigation polices invoked are based on monitoring the QoE being achieved by the actual user traffic on the network, rather than by synthetic OAM traffic parameters or the measured network performance by the operations support system (OSS). The traffic mitigation policies implemented may be service aware, application aware and subscriber aware, with any combination and weight of these as desired.

Figure 1. Dynamic Experience Management


New insight
1a. Understand application performance per subscriber and per session
1b. Understand congestion per subscriber or by or access location (e.g. radio)
New action
2. Control and adapt application behavior per subscriber per application and per congestion point per application
Real-time QoE measurement for customer experience insight
Content servers
Wi-Fi
RNC
MME
eNB
RACS
Radio access network
S/P-GW
DEM gateway
BTS
AN
Fixed access network
BNG
Any access
3G, 4G, 5G, Wi-Fi, fixed
20–30% virtual capacity gain
Subscriber fairness

With DEM, network congestion events become an opportunity to manage and improve the end user's satisfaction with the network services, rather than have them be subjected to network overload events with unintended, sub-optimal consequences. In doing so, it also maximizes the value of the access network during times of congestion, which may not be preventable, but which certainly can be controlled by policy.

Figure 2. Access networks for any applications and services



Intelligent congestion control

DEM technology enables intelligent network congestion control. The Nokia SR OS network function managing DEM is called a DEM Gateway (DEM-GW). If congestion is detected, DEM automatically cuts back traffic that is not delay-sensitive and gives priority to applications that are more delay-sensitive. Applications are managed by their respective resource needs to provide the best QoE. Over-the-top (OTT) applications and users are managed by their respective resource needs and configured preferences.

The DEM-GW builds on Application Assurance (AA) Layer 3 to Layer 7 deep packet inspection (DPI) capabilities to detect and manage applications per subscriber as well as per congestion point. This allows the DEM-GW to take subscriber and application aware actions when congestion occurs in the access network.

A DEM-GW is integrated directly into the Wireless LAN Gateway (WLAN-GW), Gateway GPRS Support Node (GGSN), Packet Data Network Gateway (PGW) or 5G User Plane Function (UPF) as an (optional) license-enabled software capability of AA.

In mobile networks, DEM runs on a subscriber basis, referred to as non-location-based DEM (NLB-DEM). This means that the detection and control of congestion is done on a per-subscriber basis, independent of what access network the subscriber is on, or what resource in that network (e.g. a cell) is congested. NLB-DEM is by far the simplest way to deploy mobile network DEM, since it avoids any integration between DEM and the access network technology vendors or OSS.

For fixed wireless networks, DEM may be non-location-based if the network is shared with mobile services. If there is a dedicated access network for fixed wireless, location-based DEM (LB-DEM) may be used. This is described in more detail later.

In the WLAN-GW, DEM models the congestion points as access network locations (ANLs), which it learns from the WLAN-GW subscriber attributes, and manages them accordingly to achieve the configured QoE/quality of service (QoS) target on a per-ANL basis. The access location managed for Wi-Fi networks is the Wi-Fi access point (AP).

The DEM-GW achieves congestion control by:

- Running DPI to classify flows (including encrypted traffic) into applications
- For location-based DEM, dynamically learning access network congestion points and estimating their maximum capacity:
  - Through real-time detection, sniffing, measurements and profiling
  - Continuous monitoring of user equipment locations and associating them to the right AP radio congestion points
- For NLB-DEM, dynamically learning per-subscriber congestion state through real-time detection, sniffing, measurements and profiling
- QoE enforcement using flexible Application QoS Policy (AQP) rules:
  - Efficient AP radio congestion detection, localization and management provided via configurable "adaptive policers"
  - Per-subscriber policy control using congestion-override policers that rate limit subscriber bandwidth during different stages of congestion: mild and severe congestion.
  - Congestion policy control rules that can factor in application, application group, day and time of day, as well as service tier(s) or APNs

Experience management benefits

An intelligent network-enabled congestion control algorithm can be operated at the evolved packet core level with the following benefits:

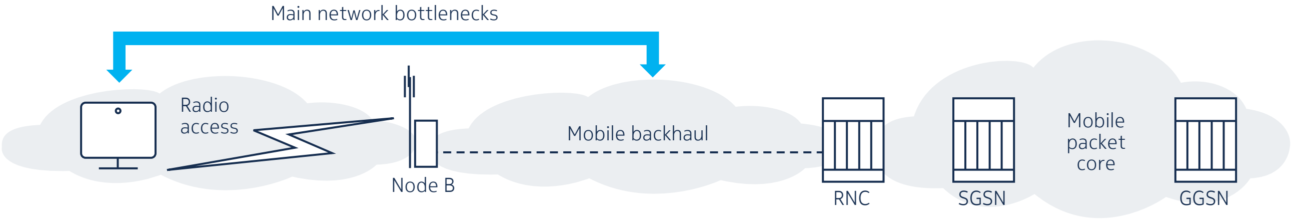
a. For operators

- During congestion, DEM will:
  - Control how subscriber services manage QoE for applications that matter
  - Ensure minimum rates that depend on the congestion level for lower priority traffic flows, including encrypted application traffic
- Allow higher bursts during times of no congestion
- Simple to deploy – no RAN probes needed

b. For subscribers

- Receive a good user experience (higher satisfaction rate)
- Receive the traffic priority and performance that was paid for (no unnecessary service disruption)
- Have more choices (for example, subscription and upgrade services in real time during times of network congestion)

Figure 3. 3G Mobile RAN congestion



For 3G, the use of high-capacity 3G/LTE smartphones and USB dongles in many cases leads to frequent user plane congestion in the access networks, which in turn causes:

- Service degradation for mobile subscribers attached to congested cell sites
- Challenges in implementing fair usage policies to manage network congestion in the core network and in the RAN
- Increased OPEX due to packet loss and retransmission.

Managing RAN access congestion points at the mobile core to provide QoS guarantees to different applications can be challenging as the cells are loaded differently at any point in time, both in quantity (bandwidth) and application type (video, web, mail, and so on).

DEM support in Nokia products

DEM is a feature supported in the Nokia Cloud Mobile Gateway (CMG) and the Nokia 7750 Service Router (SR), available on these Nokia products:

- CMG: GGSN, PGW and UPF as an integrated packet processing function
- CMG Subscriber Services Gateway (SSG): for operating as a standalone TDF policy control function for use with third-party network gateways
- 7750 SR and Virtualized Service Router (VSR): WLAN-GW, BNG and BNG for integrated congestion control.

DEM: a use case of Application Assurance

Nokia AA capabilities on SR OS deliver a range of high-touch packet processing functions. Nokia AA provides Layer 3 to Layer 7 packet inspection, classification, control and charging capabilities for advanced packet processing and service intelligence in addition to traffic optimization, dynamic congestion control, and Layer 7 stateful firewall. By enabling AA within the subscriber gateway, the Policy and Charging Enforcement Function (PCEF) including AA capabilities are offered from a single subscriber policy charging and control (PCC) point.

AA identifies bearer plane applications by implementing real-time, stateful traffic inspection and identification of packets and flows. After network traffic is identified and classified, enforcement and charging actions can be performed in accordance with statically or dynamically configured policies as part of the PCC rules and AA configuration of the gateway.

AA supports all applications and traffic types:

- Web, video, audio, voice, social networking, file transfer, software updates, mail, gaming, etc.
- Encrypted traffic support: HTTPs, HTTP2C, SPDY and QUIC
- IPv4, IPv6 and dual-stack.

Dynamic application policy control is the ability to dynamically apply prescribed actions to different subscriber traffic flows. The per-subscriber application control via the Diameter or Radius policy interface gives the service provider the freedom to use the different AA actions dynamically on a specific subscriber application.

AA subscriber control enables many other PCEF use cases that can be deployed at the same time in addition to DEM, including:

- Per-subscriber, per-application HTTP redirect
- Per-subscriber, per-application header enrichment
- Per-subscriber tethering detection and control
- Per-subscriber web filtering/parental control
- Per-subscriber, per application or application group throttle and control
- Network-based URL and domain blacklisting
- In-browser notification
- Layer 7 stateful firewall
- TCP optimization.

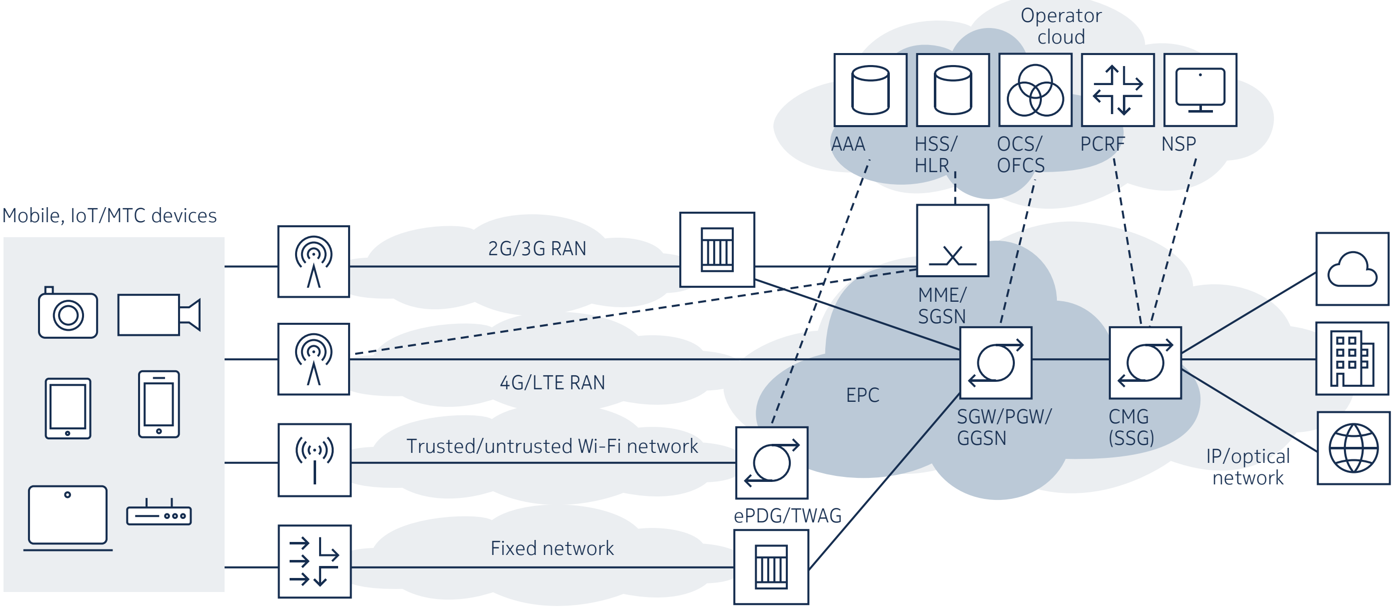
DEM integrated in PGW/GGSN/UPF

As a 3G GGSN, 4G PGW or combined SGW/PGW, the CMG provides connectivity for user equipment to external packet data networks and is the mobile network IP anchor point. The CMG performs policy enforcement, packet filtering, charging support, lawful interception and packet inspection for each traffic flow.

Similarly, as the anchor point for inter-RAT mobility and the interconnect to the data network, the 5G UPF provides packet forwarding and routing within the mobile network. The CMG supports standalone UPF on either virtualized or physical hardware platforms.

CMG also supports an SSG function (figure 4), which is located on the mobile operator SGi-LAN interface. The SSG enables CMG value-added use cases to be deployed as a standalone function for use when needed with external PGW/UPF functions.

Figure 4. SSG use for enhanced services on the SGi-LAN interface



DEM for access network types

Nokia offers multiple solutions for DEM congestion control that address various access network market requirements. The following implementations of the DEM-GW are available, depending on the network access deployment:

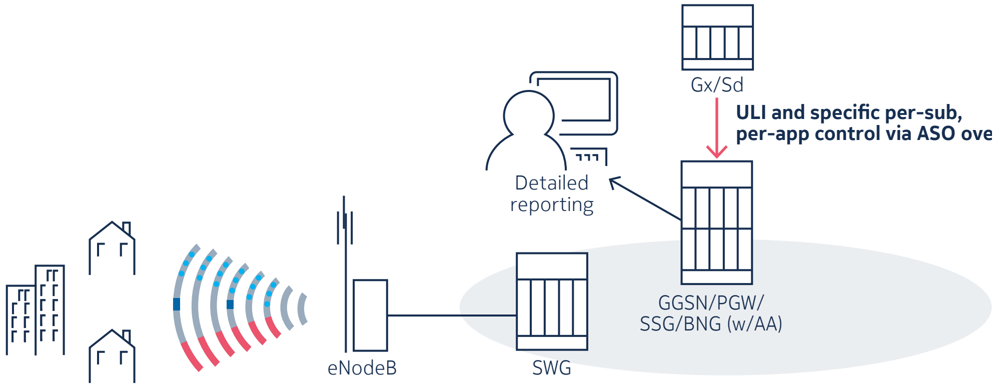
- NLB-DEM for any access network
- LB-DEM for fixed wireless access (FWA) for 3GPP fixed wireless services
- LB-DEM for WLAN-GW access networks.

“Location-based” means that the DEM system maintains tracking of the ANL of the subscribers, where the “location” is an access network resource such as a cell-ID or Wi-Fi AP. Non-location-based DEM manages subscribers independent of their location on an access network.

Fixed wireless access networks

Fixed wireless access DEM (FWA-DEM) is used when subscribers of a wireless network tend not to move, such as wireless broadband subscribers.

Figure 5. FWA-DEM architecture



For FWA on a 3GPP access network, the User Location Information (ULI) provided by the Policy and Charging Rules Function (PCRF) at bearer setup time does not change while the bearer is active. For FWA-DEM, the 3G, 4G or 5G cell is modeled as the congestion point.

Fixed residential networks may also be used for FWA on non-3GPP access networks. In these cases, the FWA service may not even propagate ULI to the IP edge network termination point, which is typically a BNG. In these cases the BNG should use NLB-DEM to manage congestion.

NLB-DEM for mobile networks

Dynamic Experience Management (DEM) provides the optimum user experience within the actual, overall network capabilities. In situations of high network load and congestion, the subscriber QoE degrades due to restricted resources across the network (such as in the radio and transport layers). In this congestion network condition, DEM enables bulk background traffic to be differentiated from real-time QoE-sensitive traffic efficiently and dynamically. The ability to limit background traffic to preserve available resources for real-time applications like video streaming is the key value of DEM.

NLB-DEM operates with subscriber policy scope. As such, no subscriber network location information is required. It operates with any access network, independent of the vendor deployed or technology in use.

Regardless of the user's location or where the congestion is taking place in the RAN, the NLB DEM-GW runs a congestion detection algorithm at the subscriber level. If congestion is detected, then different per-subscriber congestion bandwidth policers can be triggered for a mildly or/and severely congested subscriber's traffic, if configured. These are the same DEM-GW subscriber policers for FWA-DEM.

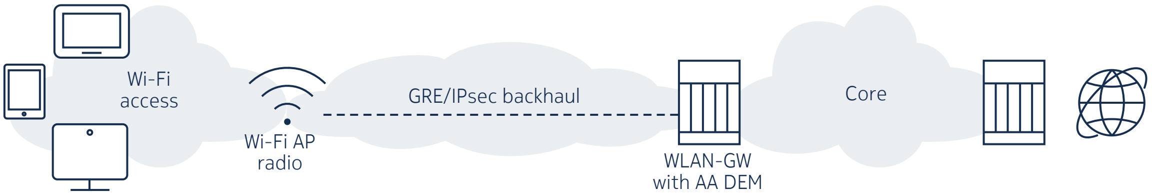
Unlike FWA-DEM, NLB-DEM does not offer reporting or actions at the ANL level (per ANL policers), such as a per-cell basis. However it still offers per-subscriber policing using multi-level-congestion override policers. Most importantly, NLB-DEM does not have any deployment requirements on the RAN, does not require enabling ULI updates into the packet core, nor any policy interface for passing location information. This makes it flexible and lightweight, and it can be deployed in any type of access network.

Wireless LAN access networks

DEM is offered as a Nokia WLAN-GW capability that monitors user plane traffic to build a network-wide view of congestion on the subscriber, application and AP radio levels.

In Wi-Fi networks, the network congestion points are typically located in the AP Wi-Fi radio.

Figure 6. Wi-Fi network congestion at the AP



Increased penetration of Wi-Fi-enabled devices (for example, mobile handsets, tablets, laptops, and TVs) and widespread use of streaming video result in frequent data plane congestion events in Wi-Fi networks. This congestion results in service degradation for Wi-Fi subscribers attached to congested APs and creates challenges in implementing fair usage policies to manage network congestion in the access network.

DEM provides the capability of managing Wi-Fi access congestion points at the WLAN-GW to provide some level of QoS guarantees to different applications, which otherwise poses challenges as the loading of the different APs at any point in time is different, both in quantity (bandwidth) and application types (for example, video, web or mail).

WLAN-GW DEM relies on location information relayed by subscriber management attributes for the Access Point MAC and VLAN. DEM uses the user location information to apply traffic management at specific impacted access sites, while not restricting users during times of non-congestion. This ensures different applications within an AP radio get their fair share of available resources, and controlling low-value traffic during times of congestion.

DEM configuration

DEM congestion detection

For each subscriber or ANL in a DEM-GW deployment, the system detects the subscriber's real-time congestion level toward the access network. The system continuously measures RTT between the DEM-GW and the subscriber's device. DEM only measures subscriber side delay; servers exhibiting long response delays or reached over very long WAN distances with high transmission delay will not trigger congestion detection. Congestion detection is debounced for many seconds to avoid triggering on spurious events. The algorithm includes hysteresis between the threshold for declaring congestion and the subsequent level of measured RTT performance needed to clear the congestion state.

There are two configurable aspects to the DEM congestion detection algorithm:

threshold — This parameter is used by the DEM-GW algorithm to determine ANL congestion or subscriber congestion in the case of NLB-DEM. It specifies the maximum acceptable RTT for connections under no congestion. Any measured RTT above the threshold is considered an indication of possible congestion. The DEM-GW auto-learns the normal RTT across the access to the network and uses it to derive the congestion's threshold. Alternatively, the operator can configure a static congestion threshold:

Values: integer 0 --500 ms – default 173 ms

If auto-threshold is enabled, the operator can configure a tolerance delay. The tolerance delay is added to the auto-learned RTT to derive a value for the congestion RTT threshold.

When auto-congestion threshold is selected, the DEM-GW continuously makes adjustments to the RTT norm threshold to ensure that it reflects changing environments.

tolerance — This parameter is used by the DEM-GW algorithm to determine ANL level or subscriber level congestion. It represents the ratio of RTTs above the configured threshold (rtt-threshold) over the total RTT measurements.

i.e. rtt-threshold-tolerance = #(RTTs > rtt-threshold)/(Total #RTTs)

If the rtt-threshold-tolerance ratio is exceeded, then all ANL subscribers (or just one subscriber in the case of NLB-DEM) are declared congested.

Values: integer 0 --100 % – default: 50%

DEM traffic control

ANL policers

The DEM-GW employs adaptive bandwidth policer variants of AA single rate leaky bucket bandwidth policers called ANL policers. These policers are used exclusively with location-based DEM congestion points. They are similar to existing AA single bucket policers, but differ in the following aspects:

- The policer rate is configured using a ratio (%) instead of absolute rates.
- The ratios are applied against the total estimated measured capacity of the congestion point to derive the actual policer's rate. For example, for a measured capacity at a congestion time of 1.5 Mb/s, and for a configured policer rate of 30%, the actual policer rate applied is calculated as follows: 1.5 × 30% = 0.5 Mb/s.
- ANL policers are applied only in the downstream traffic direction.
- ANL policers run only while the associated link or cell is in a congestion state. No action is taken when there is no congestion.

These policers are invoked using existing AQP mechanisms that match configured parameters, such as those configured with the application and app-group commands and execute the configured actions. ANL policers are used to throttle traffic going through backhaul links and radio cells during congestion states. Multiple ANL policers can be configured per congestion point types, where type is a link Iub or radio cell.

For example:

- ANL policer 1, rate = 20%, backhaul links called from AQP entry with “email” app-group match condition
- ANL policer 2, rate = 10%, backhaul links called from AQP entry with “OTT video” app-group match condition
- ANL policer 3, rate = 0%, backhaul links called from AQP entry with P2P app-group match condition. This effectively drops point-to-point traffic during congestion.

Per-subscriber congestion policers

Although ANL adaptive congestion policers apply to all traffic going through the ANL to maintain a positive customer experience and ensure priority traffic is not starved during congestion, they do not differentiate or provide any fairness among the same traffic class to different subscribers.

The per-subscriber congestion policers are enabled automatically when the congestion-override command is enabled and when the subscriber is in a congestion state. Typically, the policing rate is set as a common value for all users, meaning also that it affects the heavy users more than users sending less traffic (fairness means load-proportional control). The congestion override policers can be used for all DEM use cases, such as NLB-DEM and ANL-based DEM.

Similar to Time-of-Day (ToD) policers, per-subscriber congestion policers can be applied to all the traffic of a subscriber or to some specific applications or application groups as configured in the matching section of AQPs.

The DEM congestion state can, if configured, trigger a policing override to the per-subscriber's bandwidth policers. When a subscriber is declared to be in a congestion state, the per-subscriber congestion policer rates are triggered. This overrides any pre-existing per-subscriber policer rates, including ToD policer rates. These per-subscriber congestion policer rates are applied for the duration of time that the subscriber is in a congestion state. DEM allows the operator to optionally configure two sets of congestion policers to be applied under two congestion conditions: mild congestion and/or severe congestion conditions. When a congestion state transitions from "no congestion" to "congestion", the "mild" congestion policers are invoked, if configured. If the congestion state becomes severe, the second stage (severe) congestion policers are invoked, if configured, overriding any prior policers.

If stage 2 congestion policers are not configured, the "mild" congestion policers are applied under all congestion levels/states until the congestion is cleared. Under mild congestion state, the stage 2 policers are only applied if there are configured "mild" congestion policers. Once the subscriber's state is changed to uncongested, the per-subscriber congestion policer rates are no longer applied to the subscriber's traffic. The adjusted policing limits are applied immediately to any pre-existing or new flows of the subscriber.

The per-subscriber congestion override policers are only applicable to bandwidth policers, both single and dual leaky buckets. They are not applicable to per-subscriber flow count or flow rate policers.

To configure the per-subscriber bandwidth policer override rates, use the following commands:

- config>app-assure>group>policer>congestion-override
- config>app-assure>group>policer>congestion-override>cbs
- config>app-assure>group>policer>congestion-override>mbs
- config>app-assure>group>policer>congestion-override>pir
- config>app-assure>group>policer>congestion-override>cib

DEM performance impact

DEM is an optionally licensed software function of Application Assurance, and DEM has no measurable impact on the processing cost or throughput of AA. AA performance and scale vary depending on the host system on which AA is used. AA performance is described outside the scope of this application note on a per-product basis.

In networks that are already doing AA processing, such as many mobile core networks, this means that there is no incremental processing cost to using DEM congestion control, and the only costs to manage RAN congestion are the operational time to define and verify the AA DEM control policies, and the software licenses to enable use of DEM in the Nokia product.

Use cases and policy examples

Nokia DEM provides a flexible dynamic traffic management solution that is fully configurable by the network operator to implement any desired policy. Some examples are shown in this section for illustration purposes, but ultimately the network operator will define their own policies as suitable for their needs. The policy control is implemented using operator-configured rules in the AA AQP, defined as part of AA using match conditions and action rules. The AQP match conditions that are typically used for DEM policy include:

- app-group {eq | neq} <application-group-name>
- application {eq | neq} <application-name>
- characteristic <characteristic-name> eq <value-name>
- charging-group {eq | neq} <charging-group-name>
- traffic-direction {subscriber-to-network | network-to-subscriber | both}.

Note that most rules should include "app-groups" or "charging-group" instead of "application" match condition, since this allows the same treatment for similar sets of applications, without targeting policy to specific apps that change over time within an app-group, and without invoking treatment to specific application content providers (e.g., manage streaming video, rather than YouTube in particular).

AQP action control mechanisms typically include: bandwidth-policer <policer-name> (other options are possible for exception cases)

Once the AA DEM congestion detection algorithm determines congestion state, this instantly enables any configured AA-sub congestion policer or ANL policer that is referenced in the AQP rule actions. These policers can be used in AQP actions to implement application-aware control of the traffic including UDP/QUIC (i.e., policer values can differ based on AG/App/CG).

For more information on configuring AA AQPs, see the Multi-Service Integrated Services Adapter user guide.

Mobile RAN congestion control

Mobile networks generally implement DEM using NLB-DEM due to its simplicity of deployment, which avoids costly and complex integration with any RAN probes, OSS systems or mobile control plane elements such as the MME or PCRF.

The AQP rules possible for NLB-DEM on mobile networks can be illustrated with this example.

1. AQP rules to block or significantly rate limit all non-real time applications, since these generally do not affect the user experience if they are delayed, and thus can wait until the network congestion is alleviated. This would include:
    - "Software Update" App Group, key to managing large-scale congestion events triggered or aggravated by IOS and Android device updates, or anti-virus security updates
    - "Peer to Peer" App Group, since peer to peer apps generate background scavenger traffic that should use excess bandwidth, and not use priority bandwidth under congestion
    - "File Transfer" App Group, such as FTP bulk data file transfers.

An example for an AQP policy rule to block this traffic would be:
> entry # > match app-group eq “software update” action bandwidth-policer sub-rate-zero

Where that policer name is defined as a single-bucket subscriber policer with congestion-override PIR= 0 kb/s (effectively blocking traffic)

2. AQP rules to limit per-subscriber streaming video to a specific per-subscriber rate such as 500 Mb/s, which will result in average bit rate (ABR) video downshifting to stream at approximately 480p resolution. This rule is vital to enforce per-subscriber fairness for video rates within a congested network, since existing connections already streaming at 4k or 1080p may continue to send at those rates even while the network cannot accept any new sessions.

   - AQP > entry # > match app-group eq “Multimedia Streaming” action bandwidth-policer sub-500

     Where that policer name is defined as a single-bucket subscriber policer with congestion-override PIR= 500 kb/s

2.1 Option: If congestion persists and becomes severe, policy can be configured to block all video streaming traffic. This ensures that network resources are still available for mission-critical applications, such as email, maps and browsing.

To achieve this, the “sub-500” bandwidth policer is configured with a stage 2 congestion-override policer that has a PIR=0 kb/s (effectively blocking OTT video traffic).

3. AQP rules to limit per-subscriber total traffic (all apps including video) to a specific per-subscriber rate such as 700 Mb/s. This rule is vital to enforce per-subscriber overall fairness within a congested network independent of what application is being used.

   - AQP > entry # > match any action bandwidth-policer sub-aggregate-700

     Where that policer name is defined as a dual-bucket subscriber policer with congestion-override PIR= 1000 kb/s and CIR = 700 kb/s (or whatever values are desired…)

4. Option: Specific applications requiring preferential QoE treatment from the remaining best effort traffic could be separated out from the subscriber aggregate policer as illustrated in this example. Let’s say a service provider has decided that an application (or a set of applications defined in custom App-Group or Charging Group “Premium”) shall be exempt from DEM congestion rate limiting rules. This can be implemented by mapping all other remaining applications into one app or charging group, with the subscriber aggregate rate policer defined in the previous step conditioned to exclude this app-group.

   - AQP > entry # > match app-group neq “Premium” action bandwidth-policer sub-aggregate-700

     An example of an application or set of applications that could benefit a service provider by being exempt from congestion management are applications that measure network performance.

5. Option: Specific subscribers requiring preferential QoE treatment could be separated out and excluded from any policers as illustrated in this example. Let’s say a service provider has decided to offer a “Priority” service. Those subscribers who opt in to this package shall be exempt from DEM congestion rate limiting rules. This can be implemented by using application service option (ASO) characteristics match conditions against an ASO policy attribute assigned to these subscribers, such as Characteristic “Service” value “Priority”; then the policy rules:

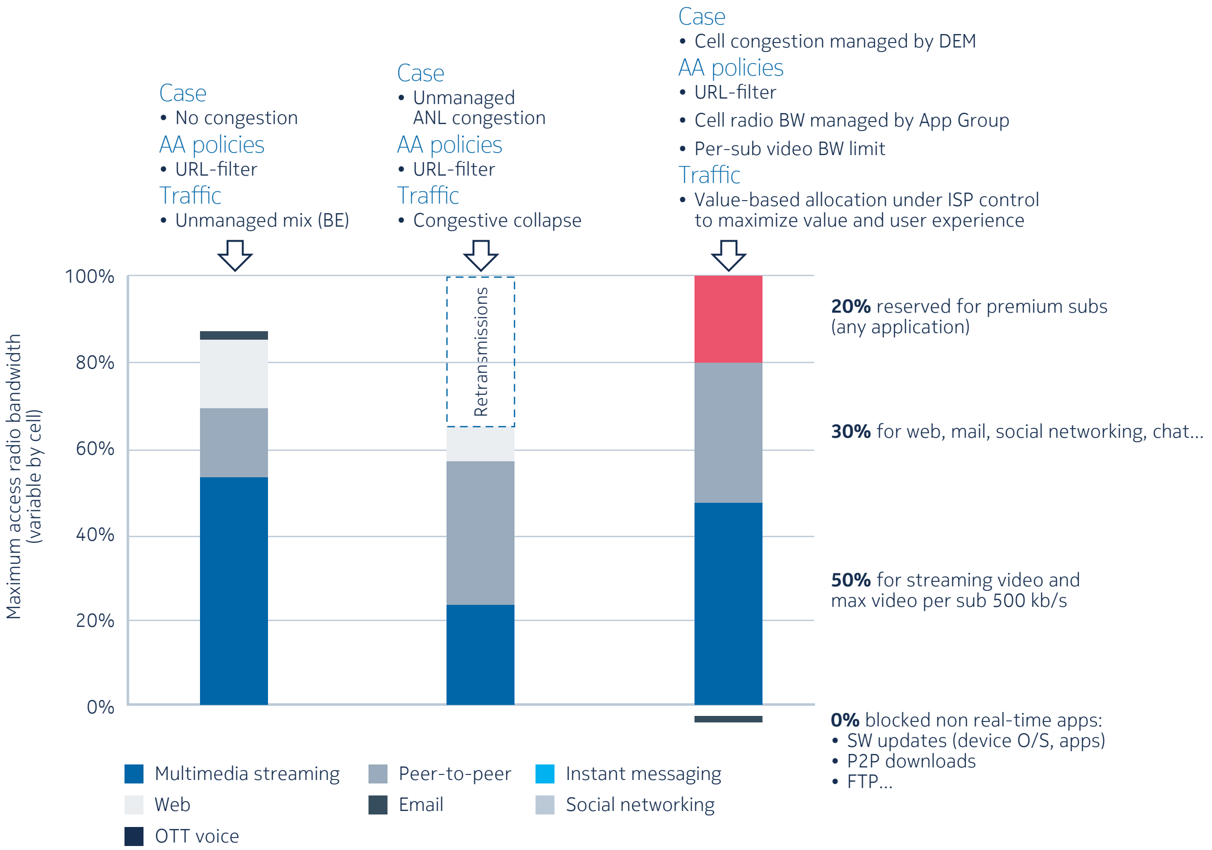
   - AQP > entry # > match characteristic neq “Priority” action bandwidth-policer sub-aggregate-700 (everyone without the Priority characteristic is policed)

Location-based congestion control

For fixed wireless or WLAN-GW networks, DEM can track all the ANLs that subscribers are using and build a real-time view of the maximum network capacity that is supported at each access location (independent of vendor, technology, etc.). As shown in Figure 7 in the stacked chart on the left, when there is no congestion there will typically be no network policy invoked, and DEM runs in the background, tracking network capacity and per-user real-time user plane traffic performance.

When congestion events occur, congestion management policy prevents congestive throughput collapse (caused by packet drops, TCP retransmissions, increased RAN latency, etc.). Congestion collapse of goodput is shown in the center case.

Figure 7. LB-DEM policy



Case
- No congestion

AA policies
- URL-filter

Traffic
- Unmanaged mix (BE)

Case
- Unmanaged ANL congestion

AA policies
- URL-filter

Traffic
- Congestive collapse

Case
- Cell congestion managed by DEM

AA policies
- URL-filter
- Cell radio BW managed by App Group
- Per-sub video BW limit

Traffic
- Value-based allocation under ISP control to maximize value and user experience

20% reserved for premium subs (any application)

30% for web, mail, social networking, chat…

50% for streaming video and max video per sub 500 kb/s

0% blocked non real-time apps:
- SW updates (device O/S, apps)
- P2P downloads
- FTP…

With DEM control enabled, DEM control policy can be enforced at the total aggregate cell level as expressed in terms of the percent of available bandwidth at the time of congestion, in addition to policy on a per-subscriber basis. For example, the DEM policy may include:

1. AQP rules to block or significantly rate limit all non-real-time applications, since these generally do not affect the user experience if they are delayed, and thus can wait until the network congestion is alleviated. This would include
    - “Software Update” App Group, key to managing large-scale congestion events triggered or aggravated by IOS and Android device updates, or antivirus security updates

- “Peer to Peer” App Group, since point-to-point apps are background scavenger apps that should use excess bandwidth, not priority bandwidth under congestion
- “File transfer” App Group, such as FTP bulk data file transfers.

  To achieve this, an example of AQP policy would be:
  > entry # > match app-group eq “software update” action bandwidth-policer ANL-rate-5

  Where that policer name is defined as granularity access-network-location with rate = 5%

2. AQP rules to limit streaming video application to 50% of the total bandwidth of each ANL (e.g., cell). This limits the total amount of video across all subscribers to not exceed 50% of network capacity.
   - AQP > entry # > match app-group eq “Multimedia Streaming” action bandwidth-policer anl-video-congestion-50

     Where that policer name is defined as granularity access-network-location with rate = 50%
3. AQP rules to limit per-subscriber streaming video to a specific per-subscriber rate such as 500 Mb/s, which will result in ABR video downshifting to stream at approximately 480p resolution. This rule is vital to enforce per-subscriber fairness within a congested network, since existing connections already streaming at 4k or 1080p may continue to send at those rates even while the network cannot accept any new sessions.
   - AQP > entry # > match app-group eq “Multimedia Streaming” action bandwidth-policer sub-500

     Where that policer name is defined as a single-bucket subscriber policer with congestion-override PIR= 500 kb/s

3.1 Option: If congestion persists and becomes severe, policy can be configured to block all video streaming traffic. This ensures that network resources are still available for mission-critical applications, such as email, maps and browsing.

To achieve this, the “sub-500” bandwidth policer is configured with a stage 2 congestion-override policer that has a PIR=0 kb/s (effectively blocking OTT video traffic).

4. Optional: rate limit consumer traffic vs higher priority emergency service or premium access point name (APN) traffic.

   To do this, define AQP rules to limit the total amount of consumer APN traffic for all remaining traffic (app groups) to 30% of the available ANL capacity. APN traffic is not directly an AQP match criteria; this would be implemented using an Application Service Option (ASO) characteristic name such as “APN” with a value such as “Consumer”, and configuring PCC PRB policy statically or dynamically to assign this ASO characteristic to any subscriber in the consumer APNs.
   - AQP > entry # > match characteristic “APN” eq “Consumer” action bandwidth-policer anl-congestion-30

     Where that policer name is defined as granularity access-network-location with rate = 30%

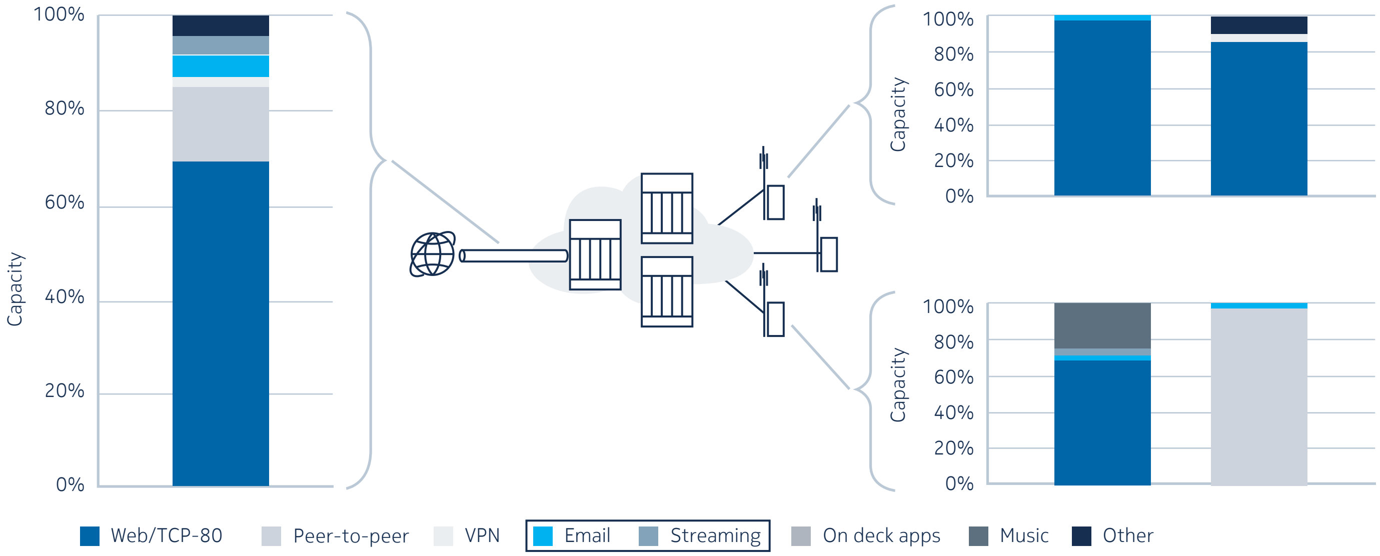
In doing this, the remaining 20% of network capacity is reserved for other APNs that would include emergency services or other premium APNs.

5. As with the video streaming case, for per-subscriber fairness within this traffic segment, the traffic within the APN 30% ANL rate limit control should also be complemented by a per-subscriber AQP rate limit.

DEM location-based analytics

Location-based analytics provides the operator with an accurate view of the subscriber's location (ANL) and application usage for a specified location in Wi-Fi networks for the purpose of data mining.

Figure 8. AP radio per application reporting



To provide an accurate reporting of the subscriber location via analytics tools such as the Network Services Platform, AA exports location information and congestion status in both volume and comprehensive cflowd reports. The off-line cflowd collector then allows per ANL (Access Point and AP radio) per application or application group statistics.

Net neutrality and congestion control

In some jurisdictions, the network regulator may have some form of net neutrality rules implemented.

Most net neutrality regulations exempt network congestion events from the neutrality rules, to continue to allow a variety of widespread techniques for mitigating and managing congestion. Under these scenarios, DEM may be fully and flexibly used in an application-aware manner to control QoE during congestion based on the real-time needs of each type of application.

In some more restrictive jurisdictions, any use of application awareness may be forbidden. This can easily be accommodated with DEM, since the congestion detection mechanism is not application dependent, and the congestion control polices are flexibly defined by the service provider to fit each regulatory case. DEM control rules may be implemented only on a subscriber basis, for the aggregate of all traffic within the subscriber context. With LB-DEM, the control policies for both the ANL and the subscriber context may be defined on the traffic rates alone, with no application awareness if that is preferred. If no other AA use case is using AA DPI traffic classification, the AA system can have DPI-based classification disabled using the shallow-flow-inspection command.

However, whenever possible under regulations, Nokia recommends using DEM application awareness to implement application-aware control policies, mainly due to the widely varying transmission requirements and characteristics between real-time traffic such as streaming video, and non-real time traffic such as software updates and other large file transfers.

Conclusion

The Nokia DEM solution is a tool enabling network operators to easily and effectively implement intelligent access network congestion detection and control from within the IP networking user plane.

DEM active intelligent congestion control can be provided as:

- Location-based: cell site or AP-specific
  - Only apply traffic management at the impacted sites
  - Don't restrict users during times of non-congestion
- Non-location-based
  - When integrated with AA enforcement specific applications can be targeted for control based on entitlements
  - Provide notification to subscriber regarding service adjustment and network conditions (also input into data offload decisions)

Using these capabilities, providers can flexibly mitigate network congestion:

- In real time without overbuilding the network
- Control/restrict relevant traffic by application during times of congestion
- Enforce subscriber fairness for access bandwidth use during congestion
- With location awareness, manage how different applications within a cell receive a fair share of available resources.

Abbreviations

3GPP	3rd Generation Partnership Project
AA	Application Assurance
ABR	available bit rate
ANL	access network location
AP	access point
AQP	Application QoS Policy
AR	augmented reality
ASO	application service option
BNG	Broadband Network Gateway
CMG	Cloud Mobile Gateway
CPC	Cloud Packet Core
DEM	Dynamic Experience Management
DEM-GW	DEM Gateway
DPI	deep packet inspection
GGSN	Gateway GPRS Support Node

IoT	Internet of Things
LB-DEM	location-based DEM
MTC	machine-type communications
NLB-DEM	non-location-based DEM
OTT	over the top
PCC	policy charging and control
PGW	Packet Data Network Gateway
QoE	quality of experience
QoS	quality of service
RAN	radio access network
RAT	radio access technology
RNC	radio network controller
SGSN	Service GPRS Support Node
SR OS	Service Router Operating System
SSG	Subscriber Services Gateway
TDF	Traffic Detection Function
ULI	User Location Information
UPF	User Plane Function
VR	virtual reality
WLAN-GW	Wireless LAN Gateway

About Nokia

We create the technology to connect the world. Powered by the research and innovation of Nokia Bell Labs, we serve communications service providers, governments, large enterprises and consumers, with the industry's most complete, end-to-end portfolio of products, services and licensing.

From the enabling infrastructure for 5G and the Internet of Things, to emerging applications in digital health, we are shaping the future of technology to transform the human experience. networks.nokia.com

Nokia operates a policy of ongoing development and has made all reasonable efforts to ensure that the content of this document is adequate and free of material errors and omissions. Nokia assumes no responsibility for any inaccuracies in this document and reserves the right to change, modify, transfer, or otherwise revise this publication without notice.

Nokia Oyj
Karaportti 3
FI-02610 Espoo, Finland
Tel. +358 (0) 10 44 88 000

Document code: SR1903033255EN (March) CID206161