**SA WG2 Meeting #162 S2-2404033**

**15 – 19 April 2024, Changsha, Chin (merge of S2-2403923, S2-2403964, S2-2404752)**

**Source: NTT DOCOMO, Vivo, Lenovo ?, Nokia**

**Title: KI#4: New Sol: NWDAF analytics for signalling storm prediction**

**Document for: Approval**

**Agenda Item: 19.15 (FS\_AIML\_CN)**

**Work Item / Release: FS\_AIML\_CN / Rel-19**

*Abstract of the contribution: This contribution proposes a solution for KI#4 where NWDAF provides a new analytics “Abnormal Control Plane Signalling Pattern” to trigger signalling storm prevention/mitigation mechanisms in other network entities.*

# 1 Discussion

The following aspects are identified in the KI#4 of TR 23.700-84 to be studied in the context of FS\_AIML\_CN:

*- Whether and how existing analytics or new analytics can be used to assist detection and prediction of signalling storm, including aspects of input /output data that needs to be collected/provided by the NWDAF.*

*- What NF(s) will be consumer of such analytics and whether and how they can use them.*

*- Whether and how signalling storm can be prevented or mitigated based on the inputs provided by NWDAF.*

To address these aspects, in this paper, given that signalling storm scenarios depends on the network deployment, it is suggested to design a flexible analytics procedure that is designed to support different signalling storm scenarios in multiple network deployments. It is proposed that NWDAF provides a new analytics, named "Control Plane Signalling Pattern Abnormality", to predict/detects abnormalities such as SBI signalling storm in the control plane signalling between network entities when a certain "network event" (for example, massive mobile registrations received at an AMF) occurs. A network event identifier can be defined to allow the NWDAF to determine the signalling storm use case. This analytics can be used by different entities including the control plane service producer and/or consumer network functions to trigger an appropriate protection and/or mitigation mechanisms.

In the request to the NWDAF, the consumer specifies the scenario (e.g. UE mobility scenario) that it is interested to be monitored, involved NF, and other parameters like reporting threshold. The NWDAF derives the nominal signalling traffic pattern and monitors the signalling when the NWDAF detects abnormalities it notifies the consumer (e.g., if the NWDAF is configured to monitor mobile registrations at AMF, it notifies the consumer when there is massive signalling load. A consumer of such analytics may be an SCP. The SCP uses this information to determine how to handle the extra signalling storm generated.

An overloaded NF may fail to react in a timely manner to incoming signalling, leading to repetitions of incoming requests that further increase the signalling. It is thus desirable to direct signalling to other NFs when an NF becomes, or is expected to become, overloaded. For Indirect Communication, an SCP is inserted between NF service consumer and NF service producer. The SCP is also expected to monitor the availability of NFs and information about NF load obtained and use that information when selecting or re-selecting NFs at target of request that the SCP forwards. However, the SCP is currently only reacting to mitigate impacts when NF failure or overload is observed. In this paper, as a use case of NWDAF analytics, it is suggested that the SCP subscribes at the NWDAF analytics (including NF load, signalling abnormality, etc.) for proactively selecting or re-selecting NF instances.

# 2 Proposal

It is proposed to update TR 23.700-84 according to the following text.

\*\*\*\*\*\*\*\*\*\* First Change \*\*\*\*\*\*\*\*\*\*

## 6.0 Mapping of Solutions to Key Issues

Table 6.0-1: Mapping of Solutions to Key Issues and Use Cases

|  |  |  |
| --- | --- | --- |
|  | Key Issues | Use cases (optional) |
| Solutions | <Key Issue #1> | <Key Issue #2> | <Key Issue #3> | <Key Issue #4> | <use case #3> | <use case #6> |
| #1 |  |  |  |  |  |  |
| #2 |  |  |  |  |  |  |
| #X |  |  |  | X | X | X |

\*\*\*\*\*\*\*\*\*\* Next Change (All new text) \*\*\*\*\*\*\*\*\*\*

6.x Solution #X: NWDAF Analytics based Signalling Strom Prediction and Protection

### 6.X.1 Description

The 5G core network, with its exponential growth in connected devices (e.g., IoT devices), complex control plane traffic (considering the number of network functions and corresponding services) and unexpected misconfiguration or failure of NFs, is susceptible to "signalling storms," which overwhelm the network function with excessive control messages, leading to dysfunctionality of NFs and consequently service outages. Predicting and preventing these storms is crucial to minimize and ideally to avoid the outrage period. AIML offers powerful tools like anomaly detection and traffic forecasting to anticipate storm risks, while automated state-of-the-art mitigation strategies like resource scaling and service throttling can prevent them from crippling the network. This proactive approach helps to safeguard the smooth operation of 5G core.

Communications between entities, such as the communication between RAN and AMF or inter-NFs or between AF and NEF, to get services from each other in the corresponding reference points via the corresponding service operations can be represented as a control plane signalling pattern. Each pattern can be characterized by a number of (time varying and/or location dependent) features such as max/mean/min number of request in a time window, max/mean/min number of responses in a time window, the number of the timed-out requests, max/mean/min response time, etc.

Signaling storms can erupt due to different reasons and in various scenarios in the network such as:

- A large number of Internet of Things (IoT) devices registering simultaneously during device rollout or a huge number of UEs registering simultaneously after a disaster recovery may lead to signalling storm in N1 reference point between RAN (on behalf of UEs) and AMF.

- An event in an area could trigger many IoT devices to report the event, and for that, they need to establish PDU session. Consequently, there will be an overwhelming number of PDU session establishment requests on N11 reference point between AMF and SMF.

- So frequent requests from multiple AFs to NEF may lead to signalling storm on N33 reference point between AF and NEF.

- The numerous session updates by SMF caused by a faulty policy update can also trigger the storm on N4 between SMF and UPF.

However, regardless of the actual root-cause of the signalling storm in the previous examples, the control plane signalling pattern between the mentioned entities is abnormal, e.g., too many requests than the normal case or unusual too large response time, or large number of rejected requests, etc.

Putting these observations together, one possible solution to predict and/or detect signalling storm in the network (and accordingly prevent and/or mitigate via taking an appropriate action) is to train a ML model(s) to predict/detect abnormalities in the control plane signalling pattern in a specific reference point between network entities and possibly for a particular set of the associated service operations. By training a ML model to recognize typical SBA signalling patterns, the NWDAF can predict the occurrence of SBA signalling storm when there is a trend of abnormal SBA signalling.

This is the idea of the proposed solution where NWDAF collects metrics/KPIs of service operations (e.g., the number of the requests messages in a time window, the number of responses messages in a time window, etc.) between two particular network entities and train a model for the analytics of detecting the control plane abnormal signalling patterns. Other network entities such as the service producer entity or OAM can subscribe for this analytics ID and get notified if a control plane signalling abnormality is predicted/detected by NWDAF and then takes the predefined prevention/mitigation actions. The framework of the proposed solution is shown in Figure 6.x.1-1 and details are explained as follows.



**Figure 6.x.1-1 – The framework of the proposed solution**

Currently, NFs can collect and report various metrics/KPIs of their operations; for example, the following KPIs are defined in TS 23.552:

- Number of successful initial registrations,

- Number of mobility registration update requests,

- Mean time of Registration procedure,

- Total number of attempted service requests,

- Number of PDU session creation requests,

By extending this capability of NFs, it is supposed that they can provide more information about their service operations such as number of requests per time window, number of responses per time windows, timing information of the request/responses, number of timed-out requests, number of rejected requests, etc.

NWDAF functionality is extended to provide the "Control Plane Signalling Abnormality" analytics ID. For this purpose, NWDAF collects the metrics/KPIs of the operations for a reference point (e.g., N11) and (optionally) for a particular service (e.g., PDU session create context) exchanged between the corresponding service producer and consumer from the corresponding network entities. In addition to these service operation KPIs, NWDAF may also collect other input data related to signalling storms such as the load information of the NFs. Using the collected input data, NWDAF trains ML model(s) to detect/predict the abnormality in the control plane signalling in a particular reference point and (optionally) for a particular service. Other network entities such as the service producer and consumer NFs can subscribe to this analytics ID. To provide the analytics, NWDAF continuously collects the input data from the corresponding source and use them as inputs to the ML model which can identifies various abnormalities/exceptions in the control plane signalling such as too many (parallel / on-going) requests, too many requests reject/time-out, too late responses, etc. Table 6.x.1-1 and Table 6.x.1-2 show the input data and output prediction/detection of the "Control Plane Signalling Abnormality" analytics.

Table 6.x.1-1: Input data from 5GC related to Control Plane Signalling Abnormality

|  |  |  |
| --- | --- | --- |
| Information | Source | Description |
| NFs ID | Service Producer and Consumer entities | The NF instance ID of the service producer and consumer entities. |
| Service Operation information | Service Producer and Consumer entities | The information of the reference point and corresponding service operation |
| Service Operation Request Counters | Service Producer entity | The counters related to the requests received by the service producer entity |
| > Number of requests |  | The number of request received in a time window |
| > Rate of requests |  | The rate of the service requests |
| > Unhandled requests |  | The number of request received but not responded  |
| > Timing of request |  | Max/min/variance of inter-request periods |
| > Rejected requests |  | The number of rejected requests |
| > Redundant requests |  | The number of redundant request |
| Service Operation Response Counters | Service Consumer entity | The counters related to the requests sent by the service consumer entity |
| > Number of requests |  | The number of request sent in a time window |
| > Rate of requests |  | The rate of the service requests |
| > Unhandled requests |  | The number of request sent but not replied  |
| > Timing of request |  | Max/min/variance of inter-request periods |
| > Rejected requests |  | The number of timeout requests |
| > Redundant requests |  | The number of redundant request |
| NF Context data | Service Producer and Consumer entities | The information about the current UE context, or network context and NF’s load |
| > Number of UEs | NF | Total number of UEs that is served by the NF. |
| > Number of network context | NF | Total number of network context (e.g. interface context) that is maintained by the NF. |
| > NF load | NRF | The load of specific NF instance(s) in their NF profile |
| > NF status & resource usage | OAM | The usage of assigned virtual resources currently in use for specific NF instance(s) (mean usage of virtual CPU, memory, disk) as defined in clause 5.7 of TS 28.552  |
| Unexpected operational  | Service Producer and Consumer entities |  |
| > Load  |  | NF load crossed a threshold |
| > Message buffer |  | Number of buffered messages crossed threshold |
| > Rate of memory |  | Rate of memory allocation crossed threshold |
| > Process/Thread  |  | Number of process/thread crossed threshold |
| > I/O latency |  | I/O operations latency crossed threshold |
| > Sockets |  | Number of allocated protocol stack sockets crossed threshold |

Table 6.x.1-2: Outputs of Control Plane Signalling Abnormality

|  |  |
| --- | --- |
| Information | Description |
| Source NF(s) ID | The NF ID of the source network entities |
| Source NF ID | The NF ID of the destination network entity |
| Exceptions ID | The ID of detected abnormality including:- Too many requests - Too many parallel requests (without response)- Too less responses - Too late responses - Too inconsistent responses time- Too many request reject- Too many request timeout - Unexpected messages pattern |

Currently various mechanisms in the implementation (e.g., Request throttling, Back-pressure, Connection Pooling) and standardized solution (e.g., load-balancing, N2 overload control, NAS congestion control as defined clause 5.19 of TS 23.501) can be used in the case of signalling storm. In the proposed solution, the following entities can subscribe to the "Control Plane Signalling Abnormality" analytics ID and once reception of the notification of abnormality, may take the actions mentioned as examples:

- The service producer entity: Request throttling, Back-pressure, Connection Pooling, etc.

- The service consumer entity: Request throttling, NF reselection, etc.

- OAM: Horizontal/Vertical scaling, Proper configuration of the network entities, etc.

- SCP (in the case of using SCP for service communication between NFs): Request throttling, Back-pressure, Connection Pooling, etc.

- NRF: Local policy to isolate the impacted NFs (e.g., avoid selecting the overwhelmed service producer for a time period).

6.x.2 Procedures

Figure 6.x.2-1 depicts the procedure for the proposed solution where NWDAF collects input data including the metrics/KPIs of the service operations, and produce the notification of the abnormal control plane signalling pattern, which is a trigger to take a prevention/mitigation action in other entities.



**Figure 6.x.2-1: NWDAF Control Plane Signalling Abnormality Analytics ID for Signalling Strom**

0. NWDAF is configured for detecting/predicting specific signalling storm scenarios either by instruction from OAM, or a request from other NF that uses the NWDAF analytics for signalling storm mitigation the NF may be the service provider NF, service consumer NF, NRF, SCP, or any other NF in the network. There are two options to specifies the signalling storm scenarios: 1) specifying the "Scenario ID" of a set of predefined scenarios (e.g., UE registration, PDU establishment, Network Exposure, etc.) or 2) specifying service operation in a reference point, i.e., services and message types to filter messages to be observed (e.g., Nsmf\_PDUSession\_CreateSMContex in N11).

1. NWDAF subscribes to other network entities for the input data needed for analytics. The metrics/KPIs of a particular service operations in a specified reference point are collected from the service consumer and service producer entities in steps 1a and 1b respectively. If SCP is deployed in the network and the service consumer and producer entities are communicating through the SCP, the service operation metrics/analytics can also be collected from the SCP in step 1c. In step 1d, NWDAF collects NF load information from NRF. Finally, in step 1e, NWDAF subscribes to the NF load information from OAM.

2. The NWDAF trains ML model(s) to predict/detect abnormalities in the control plane signalling in a specific reference point and specific service operations.

3. The analytics consumer, i.e., the network entities may be notified in the case of signalling storm of a service operations of a reference point, subscribes to the NWDAF indicating the reference point and service operations. More specifically, the service consumer entity in step 3a, the service producer entity in step 3b, SCP in step 3c, NRF in step 3d, and OAM in step 3e may subscribe to the analytics. Any other NF may also subscribe to the analytics in step 3f.

4. The NWDAF collects the input data needed for inference, i.e., the metrics/KPIs of service operations and NF loads.

5. The collected input data are used to detect/predict control plane signalling abnormalities.

6. The analytics consumers are notified by NWDAF for detected/predicted abnormalities indicating the exception(s), the affected service operations of a reference point and the NF IDs of the service consumer and producer. In step 6a, NWDAF sends notification to SCP, in step 6b, it notifies the service consumer entity, in step 6c, the notification is sent to service producer entity, notification is sent to NRF in step 6d, and finally in step 6e, NWDAF notifies OAM regarding the prediction/detection of signalling abnormalities. If any other NF subscribed to the analytics, it is notified by NWDAF in step 6f.

7. The analytics notification by NWDAF triggers prevention/mitigation mechanism already available in the network entities, e.g., back-pressure in the service consumer. In step 7a, SCP may activate its internal overload control mechanism. In step 7b, the service consumer entity may decide to reduce the number of the requests. In step 7c, the service producer entity may activate overload control such as back-pressure, in step 7d, NRF may decide to isolate the impacted NFs based on this notification, and finally in step 7e, OAM takes this notifications into account and enable prevention/mitigation mechanisms such as scaling up/out of the entities or proper configuration of other entities. In a particular signalling storm, if any other NF is notified by NWDAF regarding the abnormality, it may take appropriate mitigation/protection action.

8. The OAM may also provide extra configurations for signalling storm prevention/mitigation.

6.x.2.1 UE registration signalling storm prevention/mitigation using Control Plane Signalling Abnormality Analytics

The proposed procedure can be used to prevent/mitigate various signalling storm scenarios. Figure 6.x.2-2 depicts how it is applied for the UE registration signalling storm scenario.



**Figure 6.x.2-2: UE registration signalling storm prevention/mitigation using Control Plane Signalling Abnormality Analytics**

1a. NWDAF subscribes to AMF to collect metrics/KPIs of the UE registration process in the N1 reference point.

1b. NWDAF subscribes to OAM to collect AMF load information.

2. The NWDAF trains ML model(s) to predict/detect abnormalities in the control plane signalling in a specific reference point and specific service operations.

3. AMF and OAM respectively in steps 3a and 3b subscribe to NWDAF for the Control Plane Signalling Abnormality Analytics of UE registration process in the N1 reference point.

4. There is a large number of UE registration requests that are going to overwhelm AMF.

5. NWDAF predict/detect abnormalities (e.g., too many requests and/or too late responses) in the UE registration process in reference point N1.

6. NWDAF notifies the OAM in step 6a and the AMF in step 6b respectively regarding the exception(s) predicted/detected in the signalling pattern of the N1 reference point.

7a. To mitigate the signalling storm, the OAM, based on the operator’s policy, may decide to scale up/out the AMF VNF.

7b. Based on the notification from the NWDAF, the AMF may attempt to reduce the request, e.g., NAS level congestion control mechanism defined in clause 5.19.7 of TS 23.501.

8. The AMF may also use N2 overload mechanisms defined in clause 5.19.5.2 of TS 23.501 to ask RAN to reduce the volume of the request.

9. Based on the request from the AMF, RAN may also activate congestion control mechanisms defined in TS 38.300 and 38.413 such as RRC reject and/or Access barring.

Editor’s Note: The mitigation and prevention methods above may affect the network access of normal UEs. To reduce the impact on normal UEs, more precise and efficient mitigation and prevention methods targeting abnormal UEs are ffs.

6.x.2.2 Inter-NF Signalling storm prevention/mitigation by SCP using NWDAF Analytics



1. The NF service consumer intends to communicate with an NF service producer. The NF service consumer sends the service request to an SCP. The request may include discovery and selection parameters necessary to discover and select a NF service producer instance. For a subsequent request to the same resource, the request may contain binding information.

2. The SCP may perform discovery upon the request either by interacting with an NRF using Nnrf\_NFDiscovery service NRF or may use information collected during the previous interactions with an NRF (by the Nnrf\_NFDiscovery service or Nnrf\_NFManagement\_NFStatusNotify service operation

3. The NRF returns NF profiles of candidate NFs matching the discovery parameters

4. If the SCP did not yet subscribe to NF load analytics for NF candidate, it now subscribes at the NWDAF for load predictions based on the NF load analytics. Then, the SCP obtains Notifications about NF load predictions for the candidate NFs from the NWDAF.

5. If the SCP did not yet subscribe to CP Signalling Abnormality Analytics ID for NF candidate, it now subscribes to the analytics ID. Then, the SCP obtains Notifications about abnormal signalling pattern for the candidate NFs from the NWDAF.

6. The SCP selects (or reselects) an NF service producer from the candidate NFs. It prefers NFs with lower load predictions and avoids sending requests to NFs which are predicted to become overloaded/abnormal/malfunction.

7. If the NF service consumer is authorized to communicate with the NF service producer, the SCP forwards the request to the selected NF service producer.

6.x.3 Impacts on existing services, entities and interfaces

The solution has the following impacts:

NWDAF:

- Extending functionalities to produce Control Plane Signalling Abnormality Analytics ID.

 - Collects training data for signalling storm detection and prediction.

- Trains ML model for signalling storm detection.

5GC NFs (e.g., AMF/SMF/NEF/SCP):

- Extending service to report metrics/KPIs of service operations.

NRF:

- Set local policy to isolate the source NF based on NWDAF signalling abnormality analytics.

SCP

- Subscribe at NWDAF to predictions of NF load or Abnormal CP signalling related to NFs the SCP communicates with.

- If an NF is predicted to become overloaded or abnormality in the CP signalling, direct incoming service requests to other NFs

OAM, AMF, SCP, NEF:

- Subscription to NWDAF analytics to be notified regarding signalling storm

- Taking signalling storm mitigation action based on the notification from NWDAF

Editor's note: Impacts on other NFs are FFs

\*\*\*\*\*\*\*\*\*\* End of Changes \*\*\*\*\*\*\*\*\*\*