Intent Classification

draft-li-nmrg-intent-classification-01 updating to -02

Chen Li, China Telecom
Ying Cheng, China Unicom
John Strassner, **Olga Havel, Shucheng Liu (Will)**, Huawei Technologies
Pedro Martinez-Julia, NICT
Jeferson Campos Nobre, Federal University of Rio Grande do Sul
Diego R. Lopez, Telefonica I+D

Oct 2019

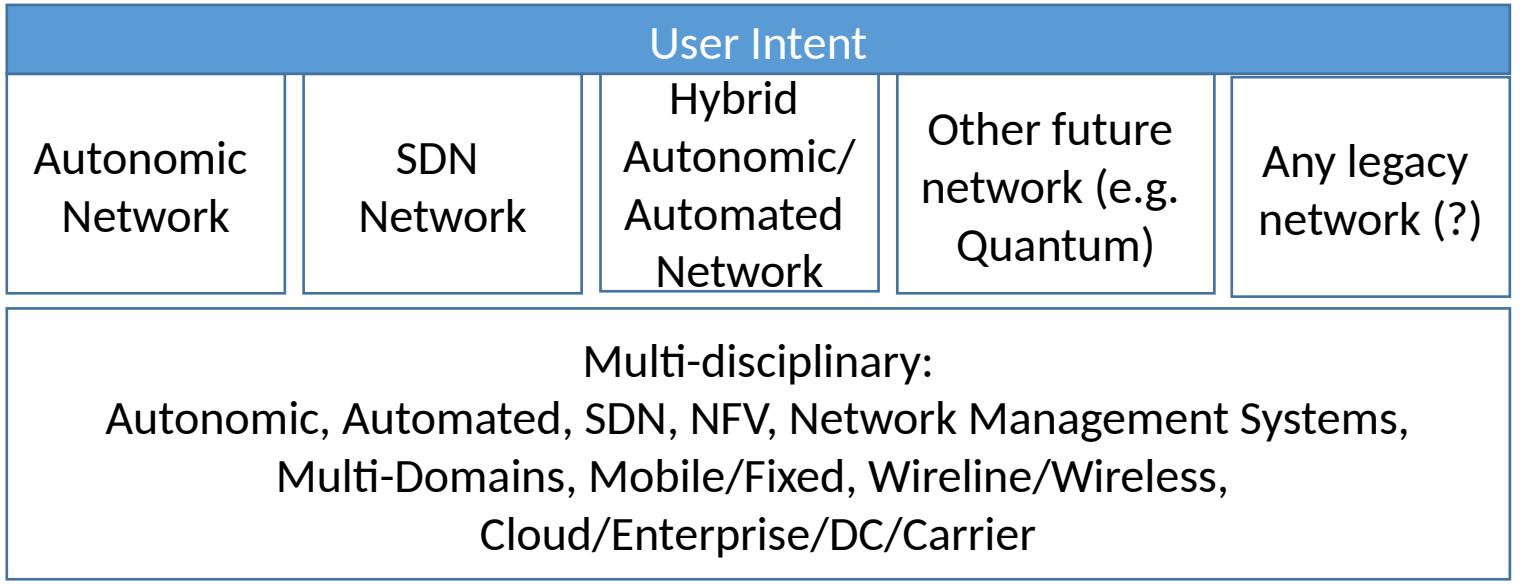
Brief Intro

- Goal: achieve agreed “Intent” related terminology and classification for NMRG, as well as guide how the term is used by other groups in IRTF/IETF, even other SDOs.

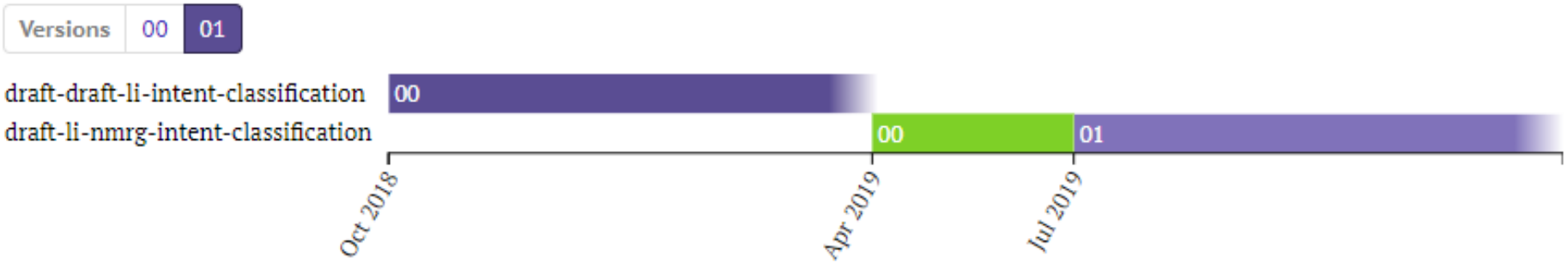
- Focus: user intents, intents definition and classification
- Scope:
  - relevant for any system or node that expects interaction with human user in the intent driven network
  - Intent driven approach is applicable to both autonomic and traditional networking, including controllers, network management systems, autonomic systems and autonomic nodes.
- Proposed classification based on:
  - Solutions, Users and their Purpose
  - When to Activate
  - Lifecycle Management Requirements
  - Granularity
- Examples were listed for each class above

Intent Classification

Intents need to be technology independent & easily transferrable; a robust system of classification will make it easier to transfer Intents, as well as easier to catalogue, search & retrieve suitable Intents



History & Update



Comments received:

- Adding solid use cases. - Brian // addressed by adding use cases in section 3.2
- Provide a framework for further intent evolution. - Diego //addressed by new draft: draft-sun-nmrg-intent-framework-00
- Suggest to use parameter to do classification – Laurent
- Classify based on granularity is not accurate, may change granularity to abstraction of service description – Ericsson
- Terminology is important for discussing intent, as it's new area. - Benoit
- Mapping un-professional language to progessional, such as define good network for 720 video, may mean no jitter, smooth, etc.
- User to specify what is the priority, e.g. , some might want instant, some want quality - Alex

- Recent updates:

- Section 3.2: Added detailed use cases
- Section 4.2: Removed Granularity part and changed to feedback
- Section 6 : Added explanation about AI related content

-01 version ToC

Table of Contents

Main Updates

Main Updates (1)

- Added use cases after the table in section 3.2:
  - For carrier networks scenario, for example, if the end users wants to watch high-definition video, then the user intent for video must be converted to 1080p configuration.
  - For DC networks scenario, administrators have their own clear network intent such as load balancing. For all traffic flows that need NFV service chaining, restrict the maximum load of any VNF node/container below 50% and the maximum load of any network link below 70%.
  - For Enterprise Networks scenario, enterprise administrators express their intent from an external client (application service provider). For example, when hosting a video conference, multiple remote access is required. The intent expressed to the network operator: For any user of this application, the arrival time of hologram objects of all the remote tele-presenters should be synchronised within 50ms to reach the destination viewer for each conversation session.

Main Updates (2)

Change Section 4.2 to:

- 4.2 Feedback
  - Intent can be classified by whether it is necessary to feedback the network information to the intended proponent after the intent is executed.
  - For ordinary users, they don't care how the intent is executed and they do not care about the details of the network. As a result, they don't need to know the configuration information of the underlying network. They only focus on whether the intent execution result achieves the goal, and the execution effect such as the quality of completion and the length of execution.
  - For administrators, such as network administrators, they perform intents, such as allocating network resources, selecting transmission paths, handling network failures, etc. They require multiple feedback indicators for network resource conditions, congestion conditions, fault conditions, etc. after execution.

Main Updates (3)-1

Added explanation text in section 6:

- AI technology has played an important role in the different stages of the intent network implementation.
  - Help identify and prevent security threats: Classification algorithms can attempt to identify malware or other undesirable web content or usage;
  - Intentional translation: use AI algorithm to assist the translation module, split translation into the requirements contained in the semantics of the intention; automatic delivery and execution strategy; Automate tasks and appropriate network changes based on the existing network infrastructure configuration according to the policy model;
  - Adaptive adjustment: perceive the quality of the user experience and perform predictive analysis to proactively optimize performance, such as excessive access time;

Main Updates (3)-2

Added explanation text in section 6:

- For instance, in the intent classification, the machine learning algorithm can be utilized to extract the intent feature values and classify the intent according to the intent feature distribution. For example, using artificial intelligence clustering algorithm, a large number of intents proposed by different users are used as training data to extract multiple feature dimensions, such as vocabulary information intended to be used, related feature parameters, context proposed by the intent, and the like. Cluster analysis is performed in the same form as the coordinate system, and multiple categories are classified according to the characteristics of the sample point distribution. For the input intent later, the category of the intent is judged based on the similarity with all categories.
    - For specific classification intents, such as safety or fault information, conditions can be preset in advance, and once a common error message occurs, it will automatically alarm.
    - For the network resource information, set the corresponding threshold information. When there is a certain number of link users or the network traffic is too large, the adjustment intention is started.
    - For users with higher priority, the resources can be configured preferentially.

Conclusion and Next Steps

- The current version addressed most of comments received online and offline.

- Next steps:
  - Update to -02 version before Singapore meeting, to address those comments un-addressed.
  - Add a section briefly describing Intent work status in other SDOs, such as 3GPP, ETSI, MEF…

- Ask for volunteers to review and contribute.
- (And ask for adoption as an RG draft if the group think it’s ready?)

Thank You

Intent Classification based on Solutions, Users and their Purpose

Intents may be classified based on solutions and users:

- Different Intent ***Solution Types*** e.g. Enterprise, Data Center
- ***Intent Users*** e.g. administrator/operator/end-user/customer/app developer/etc.

Intents may be classified based on its purpose:

- Customer network service intents *'I want to stream 4K Video to Sites A & B'*
- Network resource management intents *'Ensure Hosts in Eng don't exceed 40% avgCPU'*
- Cloud & cloud resource management intents *'I want a Safe-DNS & Firewall service with up-to-date white/blacklists'*
- Network Policy intents *'Use MPLS for Video and Internet for e-mail'*
- Task based intents *'Create new repo & give access to all leads'*
- System policies intents *'Use Host A for video & Host B for gaming'*
- etc.

Intent Classification based on When to Activate

Intent can be used to operate:

- ***Immediately*** on the target
  - E.g. *'Add firewall protection around RnD-Net'*
- ***Whenever required***
  - When some event happens: *'If an intrusion is detected, isolate all systems'*
  - Specific time in the future: *'Migrate hosts during maintenance window'*
  - Periodically at specific time: *'Scale back all servers over the weekends'*
  - When some condition occurs: *'If video quality degrades, switch to MPLS'*

Intent Classification based on Lifecycle Management Requirements

Intents can be classified into transient/persistent intents.

- If intent is ***transient***, it has no lifecycle management. As soon as the specified operation is successfully carried out, the intent is finished, and can no longer affect the target object.
  - E.g. *'Decommission host A and relocate its services'*
- If the intent is ***persistent***, it has lifecycle management (activate, monitor, correct, optimize). Once the intent is successfully activated and deployed, the system will keep all relevant intents active until they are deactivated or removed.
  - E.g. *'Don't allow hosts in Eng-NET to talk to those in Finance-NET'*

Intent Classification based on Granularity

Intents can have different granularities: high granularity, low granularity and anything in between.

- ***High granularity intents*** are more complex to design but are the most valuable. Intent translation, intent conflict resolution and intent verification are very complex and require advanced algorithms. Examples: e2e network service, like customer network service over physical & virtual network, over access, metro, dc and wan with all related QoS, security and application policies.
  - E.g. *'ensure the service quality for 720p video transmission to user A'*
- ***Low granularity intents***, like some path checks (can A talk to B) or individual network service/network/application/user policies, are the least complex. Their intent translation, intent conflict resolution and intent verification are much simpler than for high granularity intents.
  - E.g. *'ensure packet loss rate between device B to C is no higher than x%'*

Policy Continuum, Abstracted Intent Operation, Policy Targets and Policy Scope

- Policy Continuum for defining different types of Actors and their characteristics
- Intent Context / Capabilities / Constraints:
  - Context selects policies based on applicability
  - Capabilities describe the functionality provided by the policy
  - Constraints restrict the capabilities offered and/or the behaviour of the policy
- Policy Target is a set of managed objects which may be affected in the policy enforcement.
- Policy Scope (solutions, users)