#### Making (Virtualized) Service Interactions More Flexible Within and Across 5G Private Networks

Dirk Trossen Huawei, Advanced Network Technology Lab Munich, Germany



### SBA (Service-Based Architecture) in 5G

Move to Common HTTP-based Function Invocation: A Dramatic Change in 5G



- Moves from point-to-point protocols to service-based model with HTTP-based invocation across all network functions
  - > HTTP (web services) used for invocation
  - > HTTP service routing is realized as service communication proxy (SCP) within Release 16 (TS23.501)
- Major change in Rel15 and 16 of 3GPP with intention to allow for vertical-specific control planes
- Limited to control plane network functions so far with no support for user/data plane functions



## SBA Key Technologies

Realize Service Routing between Virtualized Network Functions



- Network functions for vertical use cases
  - > Provided by SMEs for specific use cases identified in proposal

- Lifecycle management for virtualized network functions
  - > Service orchestration & Virtualization management
- Service routing between those (dynamic) network functions
  - > Service-specific and across networks, if needed



#### Key Aspects: Within and Across 5G Private Networks

- Realized in Limited Domains (see RFC8799)
  - > Optimized for deployment within single, e.g., location, L2 technology, administrative domain
- Interconnected
- > Enabling the interconnection between different sites
- > Utilize the existing internet, too!
- Not purely relying on tunnels
- > Particularly a problem when utilizing Internet services
- > L2 tunneling (over L3) may become an issue in terms of overhead and complexity





#### Key Aspects: Flexible Routing

- Service instance may exist in many locations
  - > Workloads may fluctuate across those
  - -> Shortest path is simply not enough
- Policy to steer traffic may be highly servicespecific
  - > Workload: give me least loaded instance
  - > *Net latency*: no more than 5ms net latency
  - > *E2E latency*: no more than 10ms E2E latency
  - > Endpoint capability: give me instance with GPU
  - -> Enable service-specific scheduling
- Adherence to app semantic is key

5

- > State may be created, making traffic sticky
- -> instance affinity is crucial to adhere to





#### SBA on Steroids: Constraint-Based Service Routing

- Introduce semantic routing on top of IPv6
  - Service names instead of locators for service requests
- > IP addresses for affinity requests
- Flexible constraint model allows for service-specific traffic steering
  - e.g., pull video chunks from flexibly from CDN service instances at request level
- Interconnect using existing
   Internet
  - > Use other private as well as existing public services with no client awareness





#### **Compare Possible Solutions**

Solution	Support for Service Routing	Support for Service Constraints	Affinity Support	Transport Network Integration
DNS + IP (standard HTTP-TCP-IP solution)	Change in DNS with DNS aggressive DNS client cache flush ()	In DNS-based load balancer (+)	Long affinity due to DNS re- registration latency ()	IP transport (++)
Service Mesh (part of TS23.501 Appendix G)	Centralized service broker for discovery and delivery ()	Done via service broker (+)	Platform-specific, generally down to single requests, albeit with path stretch due to service broker (-)	IP transport (++)
Name-based path- based forwarding (part of TS23.501 Appendix G)	Centralized (reactive) discovery with distributed forwarding (+)	Shortest path only ()	Explicit instance notification with shortest path re- computation (-)	SDN-based L2, BIER overlay transport, or L2 tunnel (-)
Constraint-based Service Routing	Semantic routing with no centralized discovery (++)	At routing level (++)	Affinity level chosen by application, supporting down to single requests with no path stretch (++)	IP transport (++)



#### Conclusions

- SBA is just beginning its path in 5G
  - > Next stop is inter-connection of private networks (over public Internet)
  - > Most exciting will be introduction at user plane -> true micro-services on highly distributed devices
- Flexibility and interconnection is key
  - L2 was a good start but moving to L3 is key
     -> service routing is a, well, routing problem
  - > Using private network services is a start but you easily end up wanting to utilize what the Internet provides
    - -> interconnection is a must
  - > App-level solutions are simply not enough if we want to achieve request-level utilization of compute resources
    - -> we are coming back to this being a routing problem!

#### Constraint-based service routing tackles problem at right level with desired results



# Thank you.

Bring digital to every person, home and organization for a fully connected, intelligent world.

Copyright©2018 Huawei Technologies Co., Ltd. All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

