FUSION: Future Service Oriented Networks (www.fusion-project.eu)

Problem space and the uniqueness of the FUSION solution

FUSION will enable highly-demanding and personalised services to be flexibly deployed across the Internet; services that depend upon real-time processing of high-bandwidth streams with very low-latency to large numbers of geographically distributed users. Data centres and cloud-computing infrastructures have not been designed with such decentralised, bandwidth/processing-intensive, real-time applications in mind. In-network processing nodes need to be strategically positioned closer to the users of the services they deliver to provide faster application responsiveness and to reduce traffic within and between ISPs. A new networking paradigm is required to break down the barriers between data centres/server farms and the wide-area networks that interconnect them. We envision a ***fusion*** of service deployment and execution technologies with native service-centric routing capabilities throughout the network to provide a service-oriented network ecosystem for delivering a wide range of novel data- and processing-intensive services that have so far been impossible to deploy at large scale over the Internet.

FUSION foresees a situation where large numbers of service nodes are distributed throughout the Internet: in access points close to the users; collocated with routers within an ISP's network; in local data-centres owned and operated by ISPs; and in traditional data-centres and service farms operated by cloud and service providers. Given this rich set of resources, FUSION will enable services to be flexibly deployed over this distributed service-execution platform and will optimise the location of individual service component instances according to the performance requirements of the application, the location of its users and according to the experienced demand. Replicas of service components may be provisioned according to predicted load levels and furthermore they can be instantiated on-the-fly to deal with demand elasticity.

To meet performance targets and to support resilience in case of service node failure or network or service-level congestion there will be many replicas of the same service component instance running throughout the Internet and the users, the service providers or the network itself must be able to select an appropriate one. FUSION adopts a service-centric networking approach and will deploy a service-anycast capability in the network so that service instance selection can be optimised on the grounds of proximity and network load, maximising their availability. Furthermore, FUSION will develop lightweight protocols to also allow server load to be taken into account so that load balancing algorithms running in service-centric routers can discover the best service instances to route user request towards.

To make this vision a reality, new service-routing protocols are required at the network layer. FUSION will develop anycast routing and forwarding paradigms based on service names to achieve this goal. A major research challenge is how to achieve service routing and the exchange of routing information on the state of individual service-component instances in a scalable manner. Novel algorithms are also required at the service layer to decompose data- and processor-intensive applications into a set of service components and to optimise server selection and component placement across such a distributed environment.

Cooperation between the service and network layers is key to the success of FUSION and the project intends to innovate in this area too. Network-driven instantiation and replication of service components to reduce network congestion and adapt to highly dynamic fluctuations in usage patterns will be developed. In addition, FUSION will research and develop a service-node execution platform leveraging the state-of-the art in novel, container-based virtualisation technologies that simplify network-driven session instantiation, replication and migration.

FUSION's key performance indicators:

- Reduce the start-up time for remotely executed service component instances to within the order of seconds compared to today's equivalent operation of instantiating a virtual machine in 10s to 100s of seconds.
- Reduce total network traffic footprint (bits/s x number of network links traversed) by 50% for applications remotely processing large bandwidth streams by optimising the placement of service processing nodes.
- Reduce the network component of application latency by 50% for remotely processed services such as personalised video and networked games by optimising the placement of service processing nodes.
- The service resolution, selection and routing mechanisms should select service instances within no more than 200% of the optimal, in the worst cases, according to a combined metric that includes parameters such as RTT, throughput and service load.
- The exchange of routing information between service-centric routers will not exceed 5% of the capacity of the interconnecting links.