SUPERFLUIDITY

a super-fluid, cloud-native, converged edge system

Research and Innovation Action GA 671566

Deliverable D1.2:

Project Vision and Roadmap, v1

Deliverable Type:	Report
Dissemination Level:	Public
Contractual Date of Delivery to the EU:	01/07/2016
Actual Date of Delivery to the EU:	15/07/2016 (as agreed with the PO)
Workpackage Contributing to the Deliverable:	WP1

Editor(s):	Nicola Blefari Melazzi (CNIT)
Author(s):	Nicola Blefari Melazzi, Giuseppe Bianchi, Luca Chiaraviglio, Stefano Salsano (CNIT), Philip Eardley (BT), George Tsolis (Citrix), Walner Saint-Aubin (EBlink), Michael J. McGrath (Intel), Bessem Sayadi (NOKIA-FR), John Thomson (OnApp), Pedro A. Aranda (Telefónica, I+D)
Internal Reviewer(s)	George Tsolis (CITRIX)

Abstract: This deliverable reports the first version of the Project Vision and Roadmap of the project; a second version will be delivered at the end of the project. The document starts from the Description of Work document, and takes into account results generated within the project, other relevant scientific, technological or market developments,

	and the long term strategies of the EU and project partners. Thus, it includes parts of the original proposal, updated when needed, concise description of work performed, status of the project roadmap as of now, and new/updated concepts. Its aim is to concisely describe the project as a whole, and publicly.
Keyword List:	SUPERFLUIDITY, Project Vision, Roadmap

Version Control Table			
Version N.	Purpose/Changes	Author	Date
1	First release	Nicola Blefari Melazzi	30/06/2016
1.1	Revision and contributions from partners	Several contributors	12/07/2016
1.2	Final Revision	Nicola Blefari Melazzi	15/07/2016

Executive Summary

This deliverable reports the first version of the Project Vision and Roadmap of the project; a second version will be delivered at the end of the project. The document starts from the Description of Work document, and takes into account results generated within the project, other relevant scientific, technological or market developments, and the long term strategies of the EU and project partners. Thus, it includes parts of the original proposal, updated when needed, concise description of work performed, the current status of the project roadmap, and new/updated concepts.

In some cases this document duplicates information provided in other deliverables, but its aim is to more concisely describe the project as a whole, and publicly. More details can be found in project deliverables (http://superfluidity.eu/results/deliverables/), which are quoted in this document by their official abbreviation, Dx.y for final deliverables and Ix.y for internal, intermediate deliverables. All deliverables are available for reviewers and advisory board members, while only those marked as “P(ublic)” are publicly available.

As for the organisation of the document, after a brief introduction, in Section 2 we report the vision of the project, detailing major features and expected benefits, potential market, and comparison with other solutions. In Section 3, we describe how we plan to reach our vision, by: i) listing specific objectives; ii) specifying scientific/technical, market and societal aspects of the overall vision; iii) providing updated roadmap, KPIs and risks. Finally in Section 4, we give an account of the work performed so far, including both design and demonstration activities and related expected impact.

INDEX

1 INTRODUCTION 9

2 VISION STATEMENT 11

2.1 Major Features and Key Benefits 12
2.1.1 Towards a Different Architectural Model 13
2.1.2 A Realistic Open Programming Paradigm 15
2.1.3 Security by design 15
2.1.4 Sample Use Cases Enabled by the SUPERFLUIDITY Architecture 16
2.1.5 Expected Benefits 17

2.2 Potential Market 18

2.3 Relationships to current solutions and initiatives 19
2.3.1 Related Projects 19
2.3.2 Progress beyond the SoA 21

3 CREATING THE VISION 22

3.1 Project Objectives 26

3.2 Scientific and Technological Vision 27

3.3 Market Vision 29

3.4 Societal Vision 29

3.5 Roadmap 29

3.6 Metrics / KPIs 31
3.6.1 Performance KPI 32
3.6.2 Societal KPIs 33
3.6.3 Business-related KPIs 35
3.6.4 Contribution to high-relevance KPIs 35

3.7 Dependencies and risks 41

4 WORK PERFORMED SO FAR 47

4.1 Current achievements: R&D 47
4.1.1 Work Done in a Nutshell 47
4.1.2 WP2 Use Cases, System Requirements, and Functional Analysis 48
4.1.3 WP3 Architecture and Programming Interfaces Specification 49
4.1.4 WP4 Heterogeneous Infrastructures and Abstractions 53
4.1.5 WP5 Virtualisation Platform Implementation and Network Dynamics 54
4.1.6 WP6 System Orchestration and Management Tools 55
4.1.7 WP7 System Integration and Validation 58
4.1.8 WP8 Communication, Dissemination, Standardisation and Exploitation 58

4.2 Current achievements: demonstrations 59
4.2.1 Demo 1: Orchestration 61
4.2.2 Demo 2: Software Defined Superfluid Wireless Network Demo 62
4.2.3 Demo 3: Verification of OpenStack 63
4.2.4 Demo 4: Mobile Edge Computing (MEC) 63

4.3 Scientific and Technological Impact 65

4.4 Market Impact 65

4.5 Societal Impact 66

5 SUMMARY 66

6 REFERENCES 66

List of Figures

Figure 1: Threefold Convergence 13
Figure 2: A conceptual view of the SUPERFLUIDITY 5G network 23
Figure 3: SUPERFLUIDITY function decomposition and matching process 25
Figure 4: Project roadmap 30
Figure 5: Architectural framework for SUPERFLUIDITY with an example mapping into the physical Data Centres 49
Figure 6: SUPERFLUIDITY conceptual architecture 52
Figure 7: SUPERFLUIDITY Architecture APIs 53
Figure 8: Initial setting for the SUPERFLUIDITY testbed #1 60
Figure 9: Initial setting for the SUPERFLUIDITY testbed #2 61

List of Tables

Table 1: Milestones 30
Table 2: Performance KPI 32
Table 3: Societal KPI 33
Table 4: Business-related KPIs 35
Table 5: Considered options for the control framework 56

List of Abbreviations

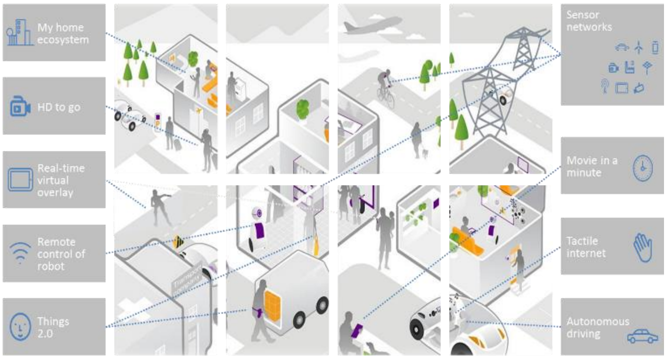
API	Application Program Interface
CDN	Content Distribution Network
DPI	Deep Packet Inspection
DRM	Digital Rights Management
ETSI ISG	European Telecommunications Standards Institute Industry Specification Group
FTTX	Fiber To The x
GGSN / P-GW	Gateway GPRS Support Node / Packet Gateway
IoT	Internet of Things
JBOA	Just a Bunch Of Accesses
KPI	Key Performance Indicator
LTE	Long Term Evolution
LTM	Late TransMuxing
M2M	Machine to Machine
MEC	Mobile Edge Cloud
MIMO	Multiple-Input Multiple-Output
NFV	Network Function Virtualization
NFVI	NFV Infrastructure
NIC	Network Interface Card
NS	Network Service
ONF	Open Networking Foundation
OSS	Operational Support Systems
OTT	Over-the-top
PNF	Physical Network Function
PoC	Proof of Concept
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RFB	Reusable Functional Block, for a definition see

	section 3.6.4 and I3.1
RRH	Remote Radio Head
RRM	Radio Resource Management
S/Gi	Reference point defined by 3GPP between the mobile packet core and PDN (Gi is between GGSN and PDN; SGi is between P-GW and PDN)
SDN	Software Defined Networking
SDO	Standard Defining Organisation
SFC	Service Function Chaining
SLA	Service Level Agreement
SON	Self-Organising Network
UBB	Ultra-Broadband
UE	User Equipment
vCS	virtual Convergent Services
vHGW	virtual Home GateWay
VM	Virtual Machine
VNF	Virtual Network Function
VNFC	Virtual Network Function Component
VNFM	Virtual Network Function Manager

1 Introduction

The vision of the SUPERFLUIDITY project emerges from powerful drivers that are shaping our society.

The first driver is the increase of population and the still growing globalisation and physical and virtual mobility: more people (2 billion and a half in 1950, almost 7.5 billion today, half of them living in cities), and more interconnections among them.

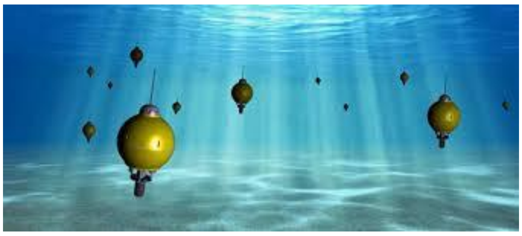


Source: Nokia

The second driver is the proliferation of new or improved applications and services that need network connectivity: social networks, video (high definition), IoT (metering, smart home, connected cars), industry 4.0 (or the fourth industrial revolution, the current trend of automation and data exchange in manufacturing technologies), low latency services (games, virtual reality, autonomous vehicles), advanced services (face recognition and speech translation, cognitive and expert systems, big data exploitation). The white paper on "View on 5G Architecture" published on July, 13, 2016 by the 5G PPP Architecture Working Group (https://5g-ppp.eu/white-papers/) identifies three groups of use cases: i) Massive broadband (xMBB) that delivers gigabytes of bandwidth on demand; ii) Massive machine-type communication (mMTC) that connects billions of sensors and machines; iii) Critical machine-type communication (uMTC) that allows immediate feedback with high reliability and enables for example remote control over robots and autonomous driving. The "5G Manifesto for timely deployment of 5G in Europe", presented by "key players in the telecoms industry and sectors which will use 5G technologies and services in the future" on July, 7, 2016 (http://ec.europa.eu/newsroom/dae/document.cfm?action=display&doc_id=16579), adopts another classification, grouping use cases according to so-called verticals: i) connected automotive scenarios; ii) connected eHealth scenarios; iii) connected planes, railway and high-speed transportation; public safety; iv) smart grids; v) smart City; vi) media and entertainment.

Whatever the classification, the network connectivity requirements of such applications and services are more and more demanding in terms of: i) data rates; ii) latency; iii) ubiquitous coverage.

The Digital Agenda Europe requires that all European citizens should access the Internet at bit rates greater than 30 Mbit/s by 2020, but in reality Ultra-Broadband (UBB) access performance is even higher and growing (FTTx, VDSL2-vectoring, G.fast, 4G, 4G+, 5G). The target latency is in the order of very few milliseconds. Network access is required everywhere.






Simply put, we have more connected devices, with each requiring higher data rates, lower latency, and ubiquitous coverage, with very high densities of users possible.

In addition, the importance of network connectivity and networked applications in our society and economy has the consequence of requiring significant improvements also in terms of: i) faster deployment of applications and services, so reducing their time to market and easing their evolution; ii) lower energy consumption; iii) enhanced security and privacy; iv) better reliability and dependability.

Furthermore, processing needs will be exacerbated in high capacity, dense networks. Current cloud computing solutions are not suitable for dynamic, real-time, high-bandwidth, low-latency applications because of issues such as granularity, localisation and configurability; service processing nodes should be distributed and located close to users or to routers or in local data-centres and not only in traditional data-centres.

As a matter of fact, the upcoming 5G network has two important characteristics, sometimes overlooked: the increasing role of cloud computing, and the fact that this generation of networks is concerned with the whole network and not only with its cellular part. The first generation of cellular networks was about portability; 2G marked the advent of digital technologies; 3G introduced widespread data services, in addition to voice services; 4G focused on Internet integration. 5G will be the whole network, including fixed sections and core, and will continue to disrupt the original Internet architecture, fully adopting cloud computing in its paradigm. The Internet will become a network of data centres, providing more often than not a 1-hop access to cloud, with a thin access section targeting infinite bandwidth/zero latency performance. Smartphones will increasingly provide access to artificial intelligence/cognitive services, with a computer-to-cloud-to-computer communication model, rather than person-to-person.

However, in spite of, or maybe because of, the great achievements that we witnessed in ICT, revenue growth for telecom operators is expected to halve from now to 2020. This means that demand cannot be satisfied by simply increasing network capacity, especially in networks that are becoming always more diverse, dense, mobile and changing unpredictably.

The answer to the challenging and sometimes contradicting necessities summarised above, consists in reducing capital and operating costs, by using low cost technologies, reducing energy consumption, sharing and optimising resources utilisation by dynamically allocating them in time and space, and in general resort to virtualisation techniques as much as possible. Benefits of a full virtualisation of network devices, at all layers, include [JAI13]: i) sharing: resources divided into multiple virtual pieces used by different users; ii) isolation: sharing of a resource does not endanger security and privacy of users; iii) aggregation: if resources are not big enough to accomplish a task, they can be aggregated; iv) dynamics: reallocation of resources in space and time on demand; v) ease of management: software-based devices are easier to manage and update.

In addition, it is necessary that the network be programmable, as a function of the needs of the services that it provides. An example of the capabilities of a virtualised and programmable network is the concept of a network slice; a virtual, end-to-end network, deployed in software, which runs in parallel to other slices on a common hardware infrastructure. A network slice also allows the isolation and support of different classes of services/customers.

The overall vision is thus the one of a software network with an application/service-centric network control able to dynamically share and allocate virtualised resources, allowing to: i) reduce costs, simplify network management, increase flexibility and ease evolution, and dynamically deploy network services.

5G will be, then, a fully "softwarised" network providing fixed and mobile UBB access to a distributed cloud infrastructure.

The SUPERFLUIDITY project contributes to the vision of a "superfluid" network, which will have the ability to instantiate services on-the-fly, run them anywhere in the network (core, aggregation, edge) and shift them transparently to different locations. Such capabilities are a key part of the converged cloud-based 5G future - they will enable innovative use cases in the mobile edge, empower new business models and allow almost instant roll-out of new services, and reduce investment and operational costs.

As for the organisation of this document, we first present the specific vision of the project, then how we intend to realise this vision and finally where we are now, and what is the work performed so far.

2 Vision Statement

The SUPERFLUIDITY project tackles crucial shortcomings in today's networks: long provisioning times, with wasteful over-provisioning used to meet variable demand; reliance on rigid and cost-ineffective hardware devices; daunting complexity emerging from three forms of heterogeneity: heterogeneous traffic and sources; heterogeneous services and needs; and heterogeneous access technologies, with multi-vendor network components.

The SUPERFLUIDITY solution is based on: a decomposition of network components and services into elementary and reusable primitives; a native, converged cloud-based architecture; the virtualisation of radio and network processing tasks; platform-independent abstractions, which permit the reuse of network functions across heterogeneous hardware platforms, while catering to the vendors' need for closed platforms/implementations; and high performance software optimisations along with leveraging of hardware accelerators.

As a result, the 5G network will benefit from: i) location-independence: network services deployable in heterogeneous networks; ii) time-independence: near instantaneous deployment and migration of services; iii) scale-independence: transparent service scalability; and iv) hardware-independence: development and deployment of services with high performance irrespective of the underlying hardware.

2.1 Major Features and Key Benefits

The ever increasing traffic volume is expected to push the limits of 5G access infrastructures, with some experts expecting up to 1,000 times the volume of traffic than what current 3G/4G deployments carry. This increased volume is to be accompanied with stringent low-delay requirements, all the while trying to reduce investment and operational costs in an increasingly competitive network operator market. Technological enhancements in the wireless domain are of course pivotal to facing such performance and scalability challenges, and mandate for the ability to fully harness capillary distributed antennas (for massive MIMO, beamforming, interference cancellation) via cooperative multi-point processing and control primitives. The ability to virtualise signal processing tasks, and to execute them inside centralised computational environments directly deployed within the network infrastructure itself (i.e., by trading network capacity with properly localised CPU resources, placed not only in the edge cloud but also remotely placed near the antenna) will enable such technologies to profit from the rate of innovation of typical software-based solutions. Furthermore, well-designed basic processing primitives and application programming interfaces would allow deployment of advanced algorithms which are today blocked by the proprietary nature of the involved components.

5G networks will face issues beyond performance and scalability. SUPERFLUIDITY plans to offer a converged solution to counter the complexity emerging from three forms of challenging heterogeneity:

1. Heterogeneous data traffic and end-points make proper planning and prediction of loads incredibly hard. The ability to effectively cope with, and dynamically adapt to, different, suddenly emerging, and ever mutating conditions requires a substantial leap in the level of flexibility and agility with respect to today's 3/4G networks; we argue that a much more fluid network architecture is called for.
2. Heterogeneity in services and processing needs: operators have largely recognised the need to transform the wireless access network from a bit pipe to a "smart" pipe. A network that is able to instantiate platform-agnostic software-based processing when needed, where needed, would open up seemingly endless possibilities. Paying businesses could run blazingly fast location-aware processing for their clients in the Radio Access Network. Virtual CDN operators could deploy virtualised content caches at edge networks, growing their infrastructure as their business grows. Customers could pay for on-demand ad removal in order to improve their browsing experience without draining their battery power, or could use aggressive traffic compression at the edge when the cellular load is high.
3. Heterogeneity in access technologies and their scale. 5G networks should become access-agnostic: specific wireless or wired technology should be treated as "just a bunch of accesses" (JBOAs), and seamlessly exploited (switched) so as to offer an "always best served" model down to a per-application level of granularity (rather than users). This

process should be autonomously driven by service level agreements and congestion conditions and should be scale-agnostic, i.e., handled in the very same way irrespective of the size and scale of the network nodes handling traffic/application flows.

All of these challenges point to the need for a new convergence model (see Figure 1) well beyond the "everything-on-IP" convergence that paved the way in past research efforts and that has already been implemented, to a large practical extent, in current 4G network deployments.

Concretely, we will accomplish this through a "decoupled" approach which identifies and abstracts re-usable network primitives, so as to hide complexity inside: i) their (vendor-specific) adaptation to the specific technologies and hardware, and inside ii) specifically devised cloud-like technologies providing the relevant (distributed and edge-based) dynamic programming, scaling, execution and provisioning environment.

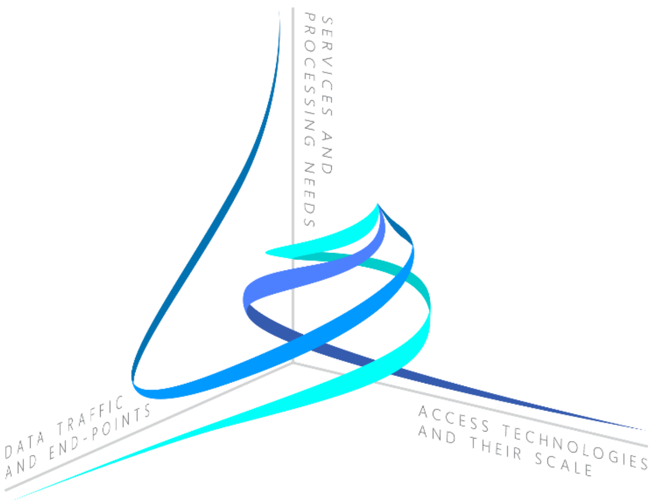


Figure 1: Threefold Convergence

What we target is the identification of elementary and reusable network and processing "primitives", as well as provisioning "elements", and their high performance software-based implementation for maximum portability, so as to permit us to rapidly deploy widely different network services where they are needed. As this concept is vital to understanding SUPERFLUIDITY's position, it is elaborated in the next section.

2.1.1 Towards a Different Architectural Model

In the past, and still in most of today's deployments, physical diversification of hardware-based network equipment has been the mainstream approach to providing an increased range of network functionalities, including developing different equipment to cater to the specific scale of target deployments. The crucial drawback resides in the proprietary nature of these highly heterogeneous devices, which brings about obvious provisioning and upgrading expense issues, as well as major management complexity in large scale multi-vendor networks.

More recently, the Network Functions Virtualisation (NFV) trend of implementing network services as (virtualised) software running on commodity hardware has gotten a lot of traction. NFV is an important part of a convergence solution for 5G networks but SUPERFLUIDITY goes beyond it by not limiting itself to exploiting NFV in an otherwise ordinary components/interfaces architecture, but tries to foster programmability as a main architectural feature of future 5G networks.

We believe NFV alone is insufficient to solve the convergence problems of 5G networks. NFV comes with the mantra of running everything as software on commodity hardware, which has two key implications: a) the hardware is homogeneous (x86) and b) all in-network processing can be efficiently implemented in software and run on x86. Both these implications do not hold in the real world: custom ASICs are still the way to implement fast network processing, and heterogeneous components such as TCAMs, FPGAs, GPUs can be added to x86 machines to dramatically increase the performance of certain workloads; any solution aiming for convergence should accommodate heterogeneous hardware. Secondly, there are cases where some functions are much more efficiently run in hardware (e.g., packet switching, or wireless protocol implementations with strict packet timings); in such cases, the best option is to create an API that abstracts this complexity and allows external control and innovation while still benefiting from the speed afforded by the available hardware.

Summing up these considerations, our concept to accommodate and leverage this hardware heterogeneity is a "divide and conquer" architectural paradigm for 5G networks. We want to depart from the usual 2/3/4G architectural approach based on monolithic network components/entities and their interfaces, and promote an approach where components are "constructed" via the programmatic composition of elementary "building blocks". In our view a specific 5G network deployment should comprise the combination of:

1. Elementary radio, packet, and flow processing primitives and events, formally specified and described independently of the specific underlying hardware, but implemented and automatically selected/instantiated so as to match the underlying hardware facilities while taking advantage of the relevant accelerators (e.g., GPUs or FPGAs) if/when available.
2. Platform-agnostic node-level and network-level "programs" describing how these primitives interact, communicate, and connect to each other so as to give rise to specific (macroscopic, and formerly monolithic) node components, network functions and services.
3. A computational and execution environment, supporting the execution (and deployment) of such "programs" and the relevant coordination of the signal/radio/flow/network processing primitives.

With this, operators will formally describe a desired network service as a composition of platform-agnostic, abstract elementary processing blocks; vendors will be in charge of providing efficient software implementations of such elementary blocks, possibly casting them to underlying hardware accelerated facilities; and the (cloud based) infrastructure will provide the brokerage services to match the design with the actual underlying hardware, to control and orchestrate the execution of the elementary blocks comprising the designed service, and to permit dynamic and elastic provisioning of supplementary dedicated computational, storage, and bandwidth resources to the processing components loaded with peak traffic. In short, we believe that one of the project's major contributions will be applying this architectural approach to 5G networks to create the "glue" that can unify heterogeneous equipment and processing into one dynamically optimised, superfluid, network.

2.1.2 A Realistic Open Programming Paradigm

We aim to dedicate a large part of our effort towards enabling software processing on heterogeneous hardware platforms. To this extent, we will design and prototype selected building blocks as virtualised, software-based network processing tasks, which can be run on heterogeneous commodity hardware anywhere in the network: in the data centre, in edge networks including mobile ones [MEC], in micro data-centres at Points-of-Presence [MICRODC] all the way out to access points, customer premises equipment, or even end user devices. Such processing will need to be instantiated on-the-fly and will be also available to third parties: entities who do not necessarily own the underlying infrastructure such as mobile application developers or businesses.

While the project will make its architecture, APIs and output as open as possible (several partners in the consortium have an excellent track record of releasing major software packages as open source), a pragmatic analysis shows that in reality we need to also accommodate the need for vendors to keep certain parts of their platforms closed: device manufacturers have spent time and money to develop their products internally, and ignoring this or their need to stay profitable would doom a solution to non-relevance.

To address this issue and emerge with a viable and pragmatic approach, our programming abstraction relies on the identification of a set of open reusable elementary primitives and processing tasks. These primitives can be openly available software run on commodity hardware, but this is not compulsory. As long as proprietary building blocks perform processing that is well defined by the API, they can be used to create higher-level, more complex solutions.

Finding the right mix of elementary primitives (so that they can be reused in several different more comprehensive network functions), which are sufficiently complex to permit vendors to differentiate from competitors, and sufficiently complete to permit operators to deploy desired services, is certainly not trivial. To our support, we note that in trying to accomplish our goal we will not start from scratch, but will be able to leverage, recent successful experiences in both the software defined community at large and in pioneering initiatives directly run by project partners in the wireless access and the wired network core domains.

Finally, although we target reuse of functionalities at all levels of the 5G hierarchy, e.g., from radio primitives in small cells and access points, SUPERFLUIDITY's practical approach will be to differentiate radio and signal processing tasks in the cloud-RAN from packet/flow processing network tasks in the rest of the network. Further, the project will target the convergence of control and management primitives needed to ensure the interoperability and integration of multiple tasks (e.g., triggering of per-flow differentiated radio behaviour) in the same network component, as well as similar programmatic composition and orchestration of elementary and reusable building blocks.

2.1.3 Security by design

Last but not least, SUPERFLUIDITY will pay great attention to security and robustness. Anecdotal reports from network operators show that, in many cases, admins do not really know exactly what processing a deployed monolithic network appliance is performing, and are outright hesitant to make any major changes to its configuration or even disable it, for fear of breaking their networks. In other words, network processing today is not only difficult to scale but it actually ossifies the network structure. In future 5G converged networks, flexible in-network processing that is instantiated on the fly has the potential to significantly disrupt network operations unless care is taken to understand the effects changes will have before they are applied. This is very important for the network operators' processing, and crucial for processing that is run on behalf of third parties, where security concerns are paramount.

SUPERFLUIDITY pioneers an approach where the processing blocks will be described in a way that is amenable to automatic offline verification. Given such a high-level description of processing, it is possible to use a technique from compilers called symbolic execution to reason about the effects network processing will have on traffic beyond simple packet reachability. We will contribute with a tool that can be used to check converged 5G networks, together with theoretical models of traditional building blocks that are proprietary (e.g. NATs, firewalls, DPI boxes). Additionally, we will propose mechanisms to help derive and in some cases automatically infer the model from source code given the public availability of such code.

2.1.4 Sample Use Cases Enabled by the SUPERFLUIDITY Architecture

To give a brief flavour of the possibilities of SUPERFLUIDITY, here are the few use cases that it could enable, as envisioned in the proposal:

- Minimum-Delay Cloud storage: Cloud storage has the potential for finally allowing people to throw out all of the clumsy hard drives, memory cards and USB sticks cluttering homes and travel bags. Unfortunately, the high data volumes and relatively low throughputs and high delays to the core/data centre mean that there is still a difference in the experience between local and cloud storage; deploying cloud storage services at the network edges would finally close this gap.
- RAN As A Service: individual functions constituting a Cloud RAN would be readily deployed, following a dynamic life cycle (creation, attachment to core network and antenna site / RRH, hot upgrade, etc.) involving optimised placement decisions about CPU, NIC, memory, hardware-acceleration capabilities; Moreover, a RAN could be flexibly build and adapted to the context using different types of schedulers, different physical layer blocks pointing to various waveforms, etc.
- Localised services: Many services that require some sort of mixing server (e.g., video conferencing, online gaming, to name a few) often end up using a distant one in terms of delay. Instead, such a virtualised server could be deployed on-the-fly at the edge, drastically reducing delays and improving user experience.
- Pooling: User-specific functions attached to various cells (and even baseband computation units [Wer13]) can be pooled in a same host so as to maximise the host load / minimise the required number of hosts; intra-cluster live migration of functions would optimise system KPIs (pooling gain, total radio capacity, energy efficiency, etc.), and would comply with RRM handover requirements (e.g. the current intra LTE handover < 50ms would be readily attained by our technology);
- Edge offloading: One of the drawbacks of mobile devices is their short battery life. Many services (e.g., firewalling, anti-virus software, ad blockers) could be offloaded to the edge to reduce battery consumption.

- Portable signal processing: platform independence would permit portability of signal processing tasks between the edge cluster and the antenna site so as to minimise front-hauling requirements and maximise radio capacity (as front-hauling requirements increase for larger radio bandwidth using carrier aggregation, more massive MIMO, more network MIMO…).
- On-the-fly Monitoring: The owner of the infrastructure could deploy a monitoring service in order to track usage of its tenants' services, or to for instance instantiate a DPI service on a particular suspicious flow.
- Virtualised CDN operators: It is well known that the performance of CDNs improves the closer that content is from users. This, however, is an expensive proposition, and so restricts all but the biggest players from the market. Instead, newcomers could deploy (virtualised) content caches at network edges, effectively renting out infrastructure and growing it as their business grows.
- And many others: For instance, context-aware services that take advantage of location information, low-delay services such as augmented reality (e.g., Google glass) or SIRI, edge-based video analytics, and application-aware performance optimisations, as described in an ETSI white paper on Mobile-Edge Computing [MEC].

However, during its first year of work, the project elaborated a more comprehensive set of 23 use cases, covering themes such as wireless access, mobile edge computing and on-the-fly monitoring. Each use case is described, along with its business and technical requirements in D2.1; the main outputs of D2.1 are also reported in this document in section 4.1.

2.1.5 Expected Benefits

As anticipated in the vision statement, SUPERFLUIDITY aims to achieve four key characteristics that together will enable dynamic processing in 5G networks: (1) location-independence, so that the processing can be placed in a number of different places and networks along the end-to-end path, as deemed optimal by the beneficiary of that processing. (2) time-independence, whereby processing can be deployed or moved near instantaneously, without end-users noticing or traffic being affected; (3) scale independence to achieve seamless scaling by decoupling network services from their scaling, and (4) hardware-independence, such that the processing can run efficiently irrespective of the different kinds of underlying, possibly heterogeneous, both commodity and proprietary hardware.

The resulting superfluid network is a network where multi-tenant, virtualised software-based network services co-exist with proprietary network functions, allowing network operators and third parties to quickly stitch together complex functionality that achieves high performance. Software processing will run on common, shared heterogeneous commodity hardware infrastructure deployed throughout the network. Operators will have the ability to instantiate such services on-the-fly, whenever needed (i.e., in milliseconds), and run potentially thousands of them on a single inexpensive platform (thus supporting a large number of concurrent users), migrating them near-instantaneously (in milliseconds, allowing for transparent adaptation to changing requirements and network conditions) and deploying them across a wide range of hardware and locations, ranging from base station and multi-cell aggregation sites all the way to data centres in the core network.

2.2 Potential Market

The advances in this project can influence and create new potential markets, such as in Internet of Things, and Big Data computing and Entertainment (gaming, multimedia).

For instance, a particular market that could benefit from the advances in SUPERFLUIDITY is the media industry that is increasingly moving to over the top (OTT) video delivery. In this case of OTT video content, owners can stream video directly over the current Internet infrastructure. According to a recent report the current over the top video streaming market amounts to around 26 billion USD in 2015, expected to nearly double to 51.1 billion in 2020 [HOLLY]. Famous OTT delivery includes companies like NETFLIX, but many operators and content owners are deploying OTT delivery. Large scale OTT video on demand streaming systems, often demand massive and duplicate storage of video content. The media industry typically encrypts its content for different users, and uses different delivery protocols for different devices. The location independence paradigm offered by the SUPERFLUIDITY platform enables video streaming functions to operate closer to the users. This can enable media operations on the edge such as Trans Muxing (protocol specific format generation), Digital Rights Management (encryption of video content) and/or ad insertion. Centralized execution of these highly personalized functions increases required backend bandwidth and storage. This inefficiency has made over the top video streaming (OTT) an unprofitable business for many content owners and distributors (due to high CDN bill, network bandwidth and storage). The Location independency introduced by the SUPERFLUIDITY architecture enables optimization in location of video processing operations (transcode, storage, DRM, personalization) leading to hugely reduced costs. By executing functions on the edge of the network, more efficient caching and reduced backhaul traffic can be achieved.

Further, SUPERFLUIDITY Orchestration and service Management and virtualization enables network programmability. This creates opportunities for telecom operators to increase their revenue by providing virtualized network capacities to OTT services. Further, the time independence (fast instantiation of services) and scale independence enable much better provisioning and adaptation of services that demand large amounts of bandwidth and computational resources. We expect that the SUPERFLUIDITY architecture can improve revenue and efficiency in the media industry, for content owner, distributor and telecom provider. Beyond this, users will benefit from increased bandwidth, reduced latency and so forth. In addition telecom operators can benefit by opening up API's to their network enabling charging for network and compute infrastructure, receiving a more fair share in OTT revenues than currently possible.

In addition, as mentioned above, a key differentiator between 4G and 5G is the idea of convergence of the whole network including the Cloud. Cloud solution providers can benefit from the converged vision of 5G. Cloud incorporates a lot of the core concepts that are of interest for 5G, including managing virtualised resources and varying workloads across a heterogeneous hardware environment. Other than the existing customer base, which will benefit from technology improvements that enable better or more efficient use of resources, the improvements that will allow greater convergence with the Telecommunications space will mean that there are more potential customers and in particular the larger, enterprise-focused businesses.

2.3 Relationships to current solutions and initiatives

SUPERFLUIDITY is one of the 19 projects belonging to Phase 1 of the 5G Infrastructure Public Private Partnership (http://5g-ppp.eu/), a 1.4 Billion Euro joint initiative between the European ICT sector and the European Commission to rethink the infrastructure and to create the next generation of communication networks and service. Such projects are described at https://5g-ppp.eu/5g-ppp-phase-1-projects/. SUPERFLUIDITY is working together with the other projects to reach the aims of the 5G PPP. Important collaborative achievements of SUPERFLUIDITY include: i) contribution to the white paper on 5G Architecture, already available (https://5g-ppp.eu/white-papers/), ii) contribution to another paper on Software Networks, currently being written; iii) work for achieving commonly agreed KPIs (see Section 3.6); iv) demonstration of exemplary functionality through testbeds (see Section 4.2) presented in joint events, such as the 5G Global event (https://5g-ppp.eu/2nd-global-5g-event/).

SUPERFLUIDITY is also linked to important international open source activities where several SUPERFLUIDITY partners are strong stakeholders: i) OpenStack, the world's fastest-growing open cloud platform and developer community, receives strong contributions from REDHAT and INTEL (part of the platinum members and technical committee), NEC (part of the gold members) and NOKIA and CITRIX (part of the corporate sponsors); ii) OpenDaylight, the open platform for network programmability, receives strong contributions from INTEL and REDHAT (part of the platinum members), CITRIX and NEC (part of the gold members), NOKIA (part of the silver members) and TELEFONICA (Dr. Pedro A. Aranda is a member of the User Advisory Board); iii) Xen is a collaborative projects of the Linux Foundation. INTEL and CITRIX are members of the XEN project and have representatives in the XEN advisory board, ONAPP has attended the latest XEN Hackathon held at ARM; iv) OPNFV, the new open source project focused on accelerating the evolution of Network Functions Virtualisation (NFV) in terms of Virtualised Infrastructure Management (VIM) and application programmable interfaces (APIs) receives strong contributions and steering from NOKIA, INTEL, NEC and REDHAT (part of the current board of directors).

In the following subsection we report the projects that are specifically relevant to SUPERFLUIDITY and the general advances with respect to the state of the art as planned in the proposal. The actual current advances obtained in SUPERFLUIDITY in its first year are reported in Section 4.

2.3.1 Related Projects

In the past years, a number of projects (EU and otherwise) have been carrying out work in the area of software-based networking and cloud infrastructure management. SUPERFLUIDITY aims to: (1) identify a set of functional building blocks for 5G function virtualisation; (2) define the most suitable programming abstraction; (3) develop a high-performance, secure and virtualised data plane platform for providing 5G network function virtualisation; (4) design novel control tools to provide transparent service migration to the edge cloud.

To the best of our knowledge SUPERFLUIDITY is the only project that aims for convergence at all levels: data and control plane convergence across heterogeneous commodity hardware and at different places in the network: radio/access, aggregation and core. Most related projects target the control plane, management frameworks or applications.

The EU FP7 CHANGE project [CHANGE] was aimed at re-enabling innovation in the Internet by designing and implementing an incremental Internet architecture around the concept of software-based network processing platforms. The Trilogy2 project [T2] aims to create "liquidity" in the Internet by developing resources (e.g., computing, storage, energy) that can be easily and quickly traded against each other in order to optimise a set of different scenarios (e.g., offloading computing to the cloud to save energy on a mobile device). Both CHANGE and T2 focused on running the NFV data plane on commodity hardware and core networks. SUPERFLUIDITY goes beyond their efforts aiming to cater for radio/access, aggregation and core networks at the same time, embracing heterogeneous hardware and allowing legacy hardware blocks to coexist with software functions via a new functional decomposition approach.

Work on Platform-as-a-Service systems (e.g. projects Coherentpaas [CP] and Cumulonimbo [CN]) aims to identify common and coherent programming APIs for data-oriented functionalities such as transactions and data stores. Programming paradigms and APIs in IaaS and NaaS cloud systems have been so far mainly devised to facilitate the development and administration of cloud applications, and the integration of heterogeneous resources; among others, see for instance mOSAIC [MOS] (negotiation and brokerage of cloud services), HARNESS [HAR] (integration of heterogeneous hardware and networking resources), CELAR [CELAR] (automatic scaling of resources), CONTRAIL [CONTR] (federation and sharing of resources), and so on. While sharing many of the goals, SUPERFLUIDITY is complementary to these efforts: our focus is on applying delay-sensitive processing to network traffic anywhere in the network, whereas the abovementioned cloud programming frameworks have much longer response times (e.g. a single map-reduce job runs for seconds, at the very least).

Open source projects such as OpenStack and CloudStack [OSTK,CSTK] (for orchestration), and Open daylight and Open Dataplane [ODLT, ODPL] (for API definition and programmability of data planes) are complementary to the objectives of SUPERFLUIDITY.

The EU Mobile Cloud Networking (MCN) project [JAM2013] is aimed at the control plane and management layers, and targets mobile deployments. MCN takes advantage of the OpenStack cloud platform as foundation for the infrastructural resources delivered as fully virtualised end-to-end MCN services. The MCN platform focuses on business needs, namely SLA management, authentication, authorisation, accounting, etc. EU T-NOVA [TNOVA] concentrates on the management layer, proposing to develop a management/orchestration platform for the automated provision, configuration, monitoring and optimisation of Network Functions-as-a-Service (NFaaS) over virtualised Network/IT infrastructures. The EU Unify project [CSA2013] also focuses on management and control, targeting the development of enablers and of a dynamic service creation platform based upon a service creation language and an orchestrator.

The FLAVIA project [FLA] targeted wireless platforms, trying to abstract technologies such as 802.11 and 802.16 to provide open programmable interfaces at different abstraction levels. The FP7 ALIEN project [ALIEN] aims to provide hardware-level forwarding platforms with an extensible interface between software and hardware for non-OpenFlow hardware platforms through the development of a Hardware Abstraction Layer (HAL). This abstraction mechanism aims to hide hardware complexity as well as technology and vendor-specific features from the OpenFlow control framework. HAL targets platforms that perform L2 switching, circuit switch devices (WDM, TDM), and any device that can be programmed to perform L2-L4 packet processing (NetFPGA). Further, ALIEN aims to build a network operating system (NOS) on top of these mechanisms. Its focus is on control and management plane issues, as well as L2-L4 (e.g., Openflow) packet processing. SUPERFLUIDITY concentrates on high performance network processing especially at higher layers, as well as ensuring the safety of running third-party provided processing, and will integrate with existing management frameworks (e.g., OpenStack, see Task 5.3) for control purposes.

The FP7 iJOIN project focuses on on-demand RAN function decomposition leveraging the benefits of virtualisation and software defined networking to enable a cloud base-station to cope with a non-ideal backhaul. iJOIN studies the RAN function placement considering the speed and technology characteristics of the backhaul and performs a joint RAN and backhaul optimisation in creating algorithms and mechanisms. These innovations may be considered as an input for the SUPERFLUIDITY architecture definition.

The CROWD project [CROWD] considers highly dense and heterogeneous wireless networks and also relies on SDN as an effective solution for MAC layer reconfiguration, dynamic backhaul reconfiguration, and connectivity management. Obviously, these works that target D-RAN (e.g., SoftRAN [GUD13]), CROWD [CROWD] would also benefit operators in the deployment of C-RAN. Dense Cooperative Wireless Cloud Network (DIWINE) [DIWINE] considers wireless communication in dense relay/node scenarios where Wireless Network Coding messages are flooded via dense massively air-interacting nodes in the self-contained cloud while the PHY air-interface between the terminals (sources/destinations) and the cloud is simple and uniform.

SUPERFLUIDITY's functional decomposition architecture provides a generic platform that allows great flexibility in creating new services and network functionalities regardless of the network hardware, topology and scale. For instance, the SUPERFLUIDITY architecture can support dense networks as well as sparse networks; it can support services which aim at increasing the network capacity, reducing the power consumption, providing different layers of security, etc. The SUPERFLUIDITY network can, in fact, be used as an optimised infrastructure to run many of the above proposals, e.g. the cloud computing frameworks.

2.3.2 Progress beyond the SoA

In this section we outline the project's progress beyond the state of the art in a set of research areas that we believe are more of importance to SUPERFLUIDITY. This progress is the one expected from the whole duration of the project. More detailed and current information about the project's progress beyond the state of the art in the first year of work, as well as the project's innovation potential, can be found in deliverable D8.3 (Innovation and Exploitation Plan), which reports the advancements reached as of today.

Cloud Networking: SUPERFLUIDITY will aim to meet the stringent requirements imposed by future 5G networks by designing and implementing a superfluid, converged network architecture that is location, hardware and time-independent. The work will push the boundaries of what is currently possible with virtualised, software-based packet processing (10-40Gb/s and higher, extremely fast service instantiation and migration in milliseconds, massive numbers of concurrent virtualised services on a single platform, significant power reductions, etc.). The goal is to bring the advantages of cloud and software-based networking to 5G networks so that services can be deployed whenever and wherever they are needed, and to leverage the availability of inexpensive, off-the-shelf hardware in the process.

Network Services Decomposition and Programmability: SUPERFLUIDITY will devise programming abstractions specifically targeted to 5G functions. The API design work will address three programming levels: service, function, and processing levels, and will attempt to maximise viability by reusing existing standard (when applicable) or research community best practices. Work will on one side target the definition of 5G specific actions and events, and on the other side will address the specification of the constructs needed to combine and orchestrate a desired execution of such actions (conditioned on the arrival of events). Particularly promising and forward-looking is SUPERFLUIDITY's approach of combining block-based composition abstractions (such as those exploited in Click routers, in some software defined radio architectures, or emerging in the ETSI NFV work on service chaining) with event-driven programming paradigms such as basic match/action based approaches or more powerful stateful abstractions based on extended finite state machines.

RAN Cloud and Mobile Edge Computing: Beyond the current vision of a static RAN function fully located in one "edge computing" place, SUPERFLUIDITY will support the ability to modularly "hot" replace eNB functions (such as scheduling) and to permit migration of such functions between edge clouds and the antenna subsystem, so as to balance algorithmic complexity with front-haul capacity. SUPERFLUIDITY will also transcend current Mobile Edge Computing vision where non-RAN functions (local caching, CDN, etc.) are envisaged to be co-located only at the eNB by enabling their migration between the RRH and the edge cloud, to maximise their performance.

Automated Security and Correctness: SUPERFLUIDITY will provide a two-pronged, complementary approach to security. First, it will go beyond the state of the art, providing a pre-deployment checking system that will ensure that virtualised network services do not negatively affect the network nor other tenants; unlike approaches in the literature, the system will be both scalable and stateful, able to model most types of services. Second, SUPERFLUIDITY will implement a post-deployment system that will learn the behaviour of traffic and detect any anomalies, thus providing a further security mechanism in cases where the checking system does not have information about the processing performed by a network function, or when static analysis is inaccurate.

3 Creating the Vision

We believe that the time is ripe to tackle the next stage in the network's evolution, towards what we termed the superfluid network. A number of different trends have recently started to come together to render such architecture an incrementally deployable possibility. First, hypervisor-based virtualisation technologies such as Xen or KVM provide the necessary isolation needed to be able to share common infrastructure across different, possibly competing, tenants. In addition, other virtualization technologies, such as containers, offer alternative solutions that may fit different scenarios.

Second, the availability of an increasing range of commodity hardware at affordable prices means that network processing can be carried out with high performance on a number of different heterogeneous hardware (e.g., GPUs, TCAMs, SSDs, FPGAs, etc.) and platforms ranging from high performance blade servers and stand-alone x86 servers to small-size, low-energy footprint microservers (e.g., CubieTrucks, Raspberry Pis, fit-PC, etc.). The micro-servers open up the possibility of pushing software and IP-based network services to the very edge of access networks, in places where energy consumption or space constraints might render the deployment of traditional servers impossible.

Finally, the accelerating pace of deployment of such platforms not only in the core of the network, but also at the edge, at network PoPs [MICRODC] or even next to base stations with different access technologies [MEC, SMART], brings the concept of running network processing "anywhere" in the network one step closer to reality.

One of the aims is to open-up the proprietary and difficult-to-innovate equipment and protocols that are deployed in heterogeneous access networks by relying on SUPERFLUIDITY's "basic blocks" concept to converge towards a common architecture and base protocol (IP) just like the rest of the network has done. While the RAN industry is yet to move towards open APIs, the success of Openflow has created large waves, and their ripple effect is likely to spread in the near future.

Figure 2 shows a conceptual view of the SUPERFLUIDITY 5G network. At the top of the picture we show a physical view with a set of SUPERFLUIDITY platforms (in red boxes) running on different types of hardware (microservers, small racks, larger x86 deployments). These platforms are set up at different points in the network: next to base stations and aggregation sites in access networks, at micro data centres at Point-of-Presence (PoP) sites in aggregation networks, and at full-fledged data centres in the core network. Each of these platforms is a multi-tenant and shared infrastructure, and network processing can be instantiated by third parties on-the-fly, when and where it is needed.

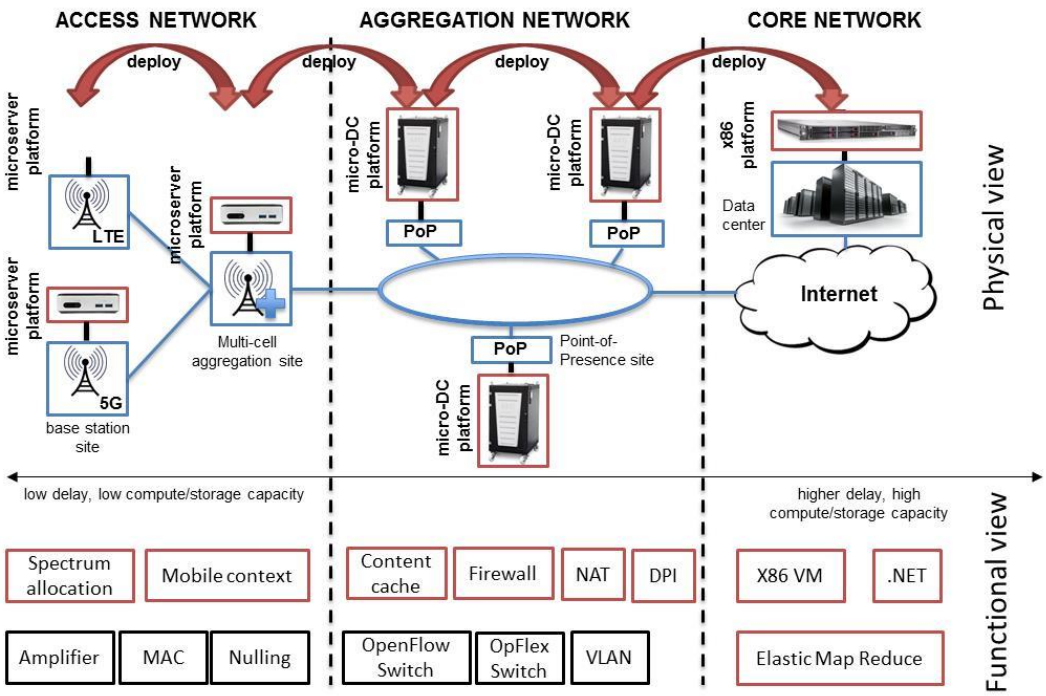


Figure 2: A conceptual view of the SUPERFLUIDITY 5G network

An orchestrator (not shown in the picture for the sake of clarity) is in charge of providing an API to platform users and of deploying the necessary network processing in a safe, high performance fashion. End-users, application developers and any SUPERFLUIDITY tenant can decide the trade-off between low-delay access near the edge (left-hand side of the figure) and high compute/storage capacity near the core (right-hand side).

At the bottom of the figure, we provide a functional view, based on decomposing processing into a series of functional blocks, within different heterogeneous environments. These functional blocks (that will be referred to as “Reusable Functional Blocks” – RFB, for a detailed definition see below in section 3.6.4 and in I3.1) may range from very basic and “small” functionality to complex and aggregated functions. As shown in the Figure, both hardware (black boxes) and software (red boxes) functions can be represented in the functional view. These functional blocks are then used by the operator and its customers to deploy more complex processing that relies on combinations of existing blocks. It is possible for the operator to apply static checking techniques to understand the impact of deploying certain processing.

To achieve this high-level architecture and vision, SUPERFLUIDITY will organise its work around five central pillars, presented here in a bottom-up form:

Pillar 1: Superfluid, Converged Architecture Design

The work will be driven by real-world requirements derived from the project’s operator partners, but also from other EU operators, as well as from the definition of novel use cases such as the ones listed in D2.1. The aim is to provide a converged, uniform and superfluid network that allows virtualised, software and IP-based network services to be deployed as-needed and throughout the end-to-end path. For access networks in particular, the idea is to push IP and software processing all the way to the edge, leaving base stations to provide basic heterogeneous connectivity and for the rest of the network to operate in a converged, IP-based fashion; Deutsche Telekom’s Terastream architecture, for instance, is already starting to push IP all the way to the edge in order to simplify management, reduce equipment costs, improve service deployment and remove ugly traffic breakout kludges [TERA]. The high-level goal is to provide an architecture that is location, time, and platform independent while providing high performance, ease-of-use, and security mechanisms to prevent the deployed services from harming each other or the network as a whole.

Pillar 2: Platform-Independent Block and Function Abstractions

SUPERFLUIDITY will seek to decompose (1) heterogeneous hardware and system primitives into block abstractions and (2) network services into basic functional abstractions. It will then implement a provisioning framework that will match function abstractions to the available hardware/block abstractions in order to automatically derive high performance and meet end-user SLA requirements, i.e., without forcing developers to have to understand low-level system details. Figure 3 depicts the process, where network processing is decomposed into function abstractions, and these are then matched to the underlying hardware. For instance, a simple firewall could be decomposed into functional abstractions to do rule matching, packet dropping or packet forwarding. These functions would then be matched to the available underlying hardware, for instance assigning the matching function to a GPU in order to parallelise and thus speed up this look-up.

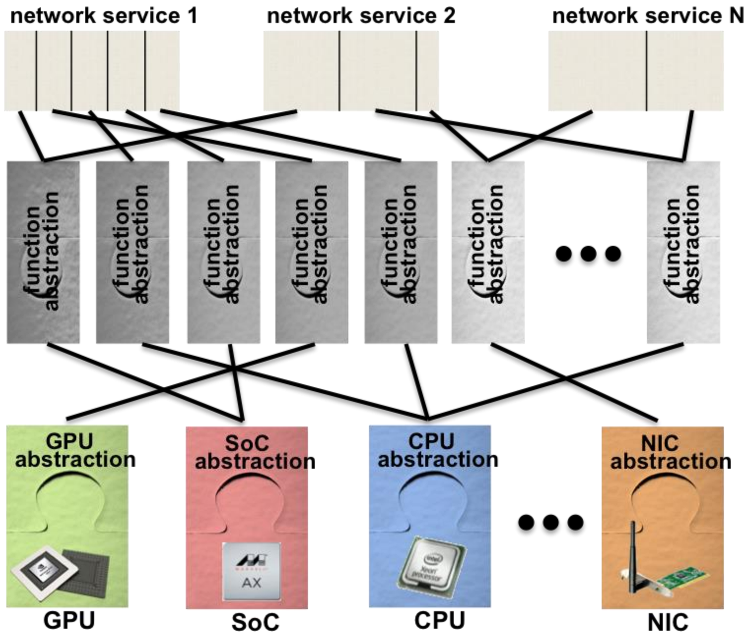


Figure 3: SUPERFLUIDITY function decomposition and matching process

Pillar 3: High Performance Block Abstractions Implementation

This pillar consists of deriving block abstractions for each of the underlying hardware components, and to optimise their performance. For instance, an abstraction could be defined for a System-on-Chip (SoC), describing its ability to perform 40Gb/s rule matching on IP's 5-tuple. A further abstraction could wrap the capabilities of an SSD drive: low-storage but high access speeds. Yet another one could abstract a GPU with high parallelisation potential but poor performance when using complex data structures (e.g., radix-trees, commonly used in IP forwarding). Beyond the abstractions, the work will involve an investigation and measurement study of the available commodity hardware, virtualisation schemes and packet I/O frameworks (Intel DPDK or netmap, to name a couple). This pillar will also look at revisiting commodity operating systems' concepts. For instance, processes and threads (which may not be congruent with the pipeline approach of packet processing) and the typical user/kernel split (which may result in unnecessary overheads given that the software already runs within an isolated virtual machine) may prove redundant.

Pillar 4: Superfluid Platform

This pillar will be in charge of putting together all of the work on abstractions into a single, coherent, SUPERFLUIDITY platform able to run virtualised network services for different tenants on shared infrastructure. The platform will implement an API for deploying services, and an algorithm for deriving the best way to match the necessary services to the underlying block abstractions such that SLAs can be met. Further, this pillar will look into optimising the number of network services that can be concurrently run, setting a target of as many as 10,000 virtual machines and/or containers on a single, shared, multi-tenant server; such high-capacity on an inexpensive server would make it easier to achieve the concept of running services anytime, anywhere. Finally, the work will include the ability to almost instantaneously instantiate services (as low as a few milliseconds is the target), as well as to migrate them quickly enough that the process is transparent to end-users and their connections (e.g., in hundreds of milliseconds or less). Clearly, some of these figures will depend on the underlying hardware, but the goal is to make the network I/O as efficient as possible so that even low-cost hardware (e.g., micro-servers running ARM processors) can act as SUPERFLUIDITY platforms.

Pillar 5: Orchestration Framework

The last pillar is in charge of gluing the different SUPERFLUIDITY platforms deployed throughout the access, aggregation and core networks into a coherent architecture. First, it will provide an API through which users can request the deployment of network services along with SLAs to which the platform(s) should conform; where possible, we will aim to integrate SUPERFLUIDITY's orchestration framework into existing management frameworks such as OpenStack. Second, this pillar will implement provisioning algorithms that will ensure that the required network processing is deployed at platforms and places in the various networks such that the SLAs are met (and should the network conditions change to migrate the processing in order to still comply).

Finally, in terms of security, the great flexibility derived from software-based network processing opens the door to introducing new problems and threats into networks. To address them, SUPERFLUIDITY will take a two-pronged, complementary approach. First, for pre-deployment security it will embrace static analysis to check network processing for safety before it is instantiated. Static analysis tools have shown great promise in checking programs for faults; SUPERFLUIDITY postulates that the same tools can be adapted to great effect for network processing. Static checking will remove the need for firewalls and other sandboxes, greatly reducing the overheads. For instance, pass-through virtualisation would be feasible to use for untrusted tenants, as long as their processing is deemed "safe" by static analysis. Second, as a backup to static analysis, this pillar will implement post-deployment security through anomaly detection algorithms that will ensure the correct functioning of the network.

3.1 Project Objectives

The objectives of the project as reported in the Description of Work are the following:

Objective 1	**Novel 5G data plane processing architecture** - Design a flexible, open and programmable 5G data plane processing architecture and relevant APIs for network functions' convergence.

Objective 2	**Converged 5G platform** - Design, implementation, and evaluation of a unified and high performance distributed cloud platform technology for radio and network functions support and migration.

Objective 3	**New Algorithms and functions** - Design, development and evaluation of algorithmic and design improvements for radio processing tasks, flow processing primitives, and service optimisation.

Objective 4	**Ultra-fast and efficient virtualisation** – Design, implementation and evaluation of beyond the state of the art quickly instantiable, low memory footprint, and high performance virtualisation technology

Objective 5	**Hardware adaptation and abstraction** – design and development of technologies and interfaces to exploit and integrate customised hardware.

Objective 6	**Control and provisioning framework** – Extensions of existing and widespread frameworks for platform's management, control, and elastic provisioning.

Objective 7	**Security framework** – Taking into account security aspects to be integrated in the platform

Objective 8	**Contribution to standardisation** – Feed SUPERFLUIDITY results into the relevant standards bodies and communities working on de-facto standard tools

3.2 Scientific and Technological Vision

The explosive growth in mobile data traffic, together with the proliferation of 'smart' connected devices such as smartphone, personal fitness trackers etc. is forcing service providers to transform their networks from one which is static and is inflexible and is based on monolithic fixed appliances to one which is dynamic, flexible and fluid in nature. At the centre of this network transformation is pervasive deployment of NFV and SDN technologies. While NFV is being adopted and rolled out by many service provides today, its role in the context of 5G will be vastly expanded and will become a foundational technology.

The characteristics of 5G networks will be significantly different to the current 3/4G networks. 5G is expected to provide an exponential increase in network capacities, support a much wider variety of use cases, improved operational characteristics such as lower latency and jitter, ubiquitous connectivity through increase connection densities as well and increase reliability and resilience. 5G networks will be highly virtualised and support high levels of flexibility through a software driven programmable network environment. 5G networks will fully embrace open technologies (both open source and derivative commercial offerings) from a diverse ecosystem built using open standards. This new ecosystem will provide support for increased and more rapid innovation, new value creation and will allow the development of new business models and revenue opportunities.

The advent of 5G will herald a new connectivity paradigm. The emergence of the smartphone has significantly changed the way in which we use our phone moving away from the consumption and generation of voice traffic to data. In Cisco's recent Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015–2020 White Paper it is predicted that mobile data traffic is expected to grow from 2.5 Exabytes of traffic per month to more than 23.4 Exabytes per month by 2019. Mobile users are increasing using applications that have high-bandwidth requirements e.g. on-demand video streaming, thus generating increasing levels of data traffic over the network. In the

5G paradigm everything will be connected, mobility is pervasive, usage and connectivity varying significantly by location and time. 5G is set to realise a vision of a hyper connected world in which the network is highly heterogeneous with convergence across multiple access technologies (wired and wireless) across a significantly expanded range of licence and unlicensed spectrums making the access network much more adaptable to user contexts and needs. The access networks will be much more spectrally efficient allowing them to support much higher densities of device connections, higher data rates and with more predictable characteristics such as latency, jitter etc.

5G will drive the proliferation of the IOT devices. Cisco have predicted that there be 50 billion by 2020. This means there will be almost 7 IOT devices for every human on the planet. 5G will need to have appropriate scalability which need be achieved through technologies such as virtualisation to enable scaling in orders of magnitude and with minimal cost implications. 5G will have to deliver low latency which will an order of magnitude lower than in 4G networks. In some specialist 5G uses cases such as tactical feedback for robotic surgery a latency of 1ms is required. Due to range and scale of expected use cases in 5G, control becomes a significant challenge and requires a much larger control plane.

To meet the diversity of anticipated use cases/services, heterogeneous access mechanism, and quality of service requirements will require the ability to provide network slices enabled through SDN capabilities in order to ensure required capacity, throughput, security, latency etc. This requires a significant movement away from proprietary, vertical integrated and specialised hardware to open platforms that offer significantly more flexibility, support from programmability with a high level of automation and autonomy. The foundational adoption of NFV and SDN is fundamental to realising these capabilities. An important differentiator of these technologies is their utilisation of an open source approach. The approach leverage significant investment from industry players such as Cisco, RedHat, Intel etc., solution providers, and community developers to provide a unified approach to implementation. Projects such as OpenStack, OpenDaylight, OPNFV, OpenFlow, Open vSwitch, DPDK etc. will be instrumental in providing the technology building blocks for 5G. These software-based building blocks will enable service providers to better monetise their network assets.

The advent of Cloud RAN will be another key architectural change for 5G. The approach allows aggregated traffic from radios to be processed by virtualised applications running on standard high volume servers. Functions which have flexible timing constraints such as mobility, subscriber management etc. are suitable for centralisation and virtualisation in a C-RAN environment.

The adoption of open and non-propriety technologies in 5G will afford new opportunities for many companies. For example, network infrastructures can be realised on standard X86 high volume services which offer various capabilities to support virtualised environments. For example, Intel is providing its Intel Open Networking Platform (ONP) as a reference architecture that combines leading open source technologies and Intel technologies such as QuickAssist Acceleration as building blocks that can be used by the ecosystem to create optimised solutions for SDN and NFV workloads to meet the needs of 5G.

3.3 Market Vision

The market vision for 5G is a technology ecosystem that is built around open standards, open source technologies and standard high volume hardware platforms. This move towards a more open and standards-based environment will encourage new entrants into market both from the infrastructure technologies and service provisioning perspectives. Increased diversity in the ecosystem drives increased and more rapid innovation, new value creation by reducing barriers to entry and adoption, and will enable the realisation of new business models and use cases.

The initial target market for 5G is likely to be fixed wireless broadband. Verizon and AT&T have indicated they are targeting initial commercial deployments in the 2017 timeframe. KT Corp who are the telecom sponsor of the 2018 Winter Olympics located at Pyeong Chang in South Korea have plans to demonstrate 5G technologies supporting broadcasting and data services. This is expected to act as a catalyst for the launch of 5G service by Korean operators by 2020 provided the standards are in place by then. In a similar timeframe NTT DoCoMo plan to launch 5G services in time for the 2020 Olympics [JAN15].

From a standardisation perspective, the 3GPP started the standardisation process in September 2015. It is expected that the Release 15 should be available in the 2018 timeframe. The Release 15 work items should form the basis for the first phase of 5G in the 2020 timeframe.

3.4 Societal Vision

5G is envisioned to deliver an end to end ecosystem that supports a diverse range of users and use cases enabling a fully mobile and connected society. The vision of 5G is support to a diverse range of use cases which affect the daily lives of most citizens. These include smart grids, smart utilities and better water quality monitoring and domotics which will impact and improve people's domestic lives. In transport domain we will have smart mobility with autonomous cars and intelligent traffic management. From a personal perspective smart wearables will enable people to live fitter and healthier lives. E-healthcare will also improve people health outcomes and improve the democratisation of healthcare by giving people access to healthcare specialisations such as remote surgery, specialists consultations etc. which are independent of location.

3.5 Roadmap

The following figure shows the project's roadmap, reporting step-by-step the main results obtained so far and official project milestones.

The results reported in the figure are the following:

1. August 2015 – Management structures and technical infrastructure needed to run the project fully operational
2. September 2015 – First definition of technical and business requirements
3. December 2015 - Definition of the Communication and Dissemination Plan
4. January 2016 - First draft of SUPERFLUIDITY's architecture
5. March 2016 - Final definition of use cases, technical and business requirements

6. May 2016 - Functional analysis and decomposition of functions; hardware selection, modelling and profiling; Function allocation algorithms; Initial design for control framework
7. June 2016 - Decomposition of existing monolithic network functionality into reusable components

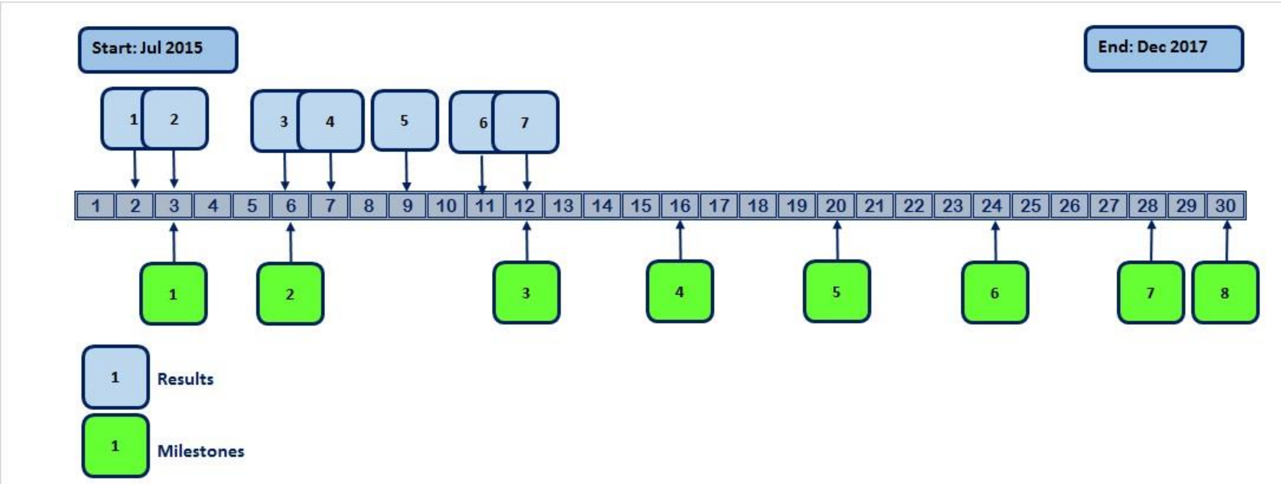


Figure 4: Project roadmap

The official project milestones are reported in the following table.

Table 1: Milestones

Milestone number	Milestone name (and short description)	Related WP(s)	Estimated date	Means of verification
1	**Project fully operational.** Management structures and procedures, including standard formats and forms for project documentation ready. Composition of boards and teams fully defined. Technological infrastructure to support cooperative work fully operational (web server, document server, version control system for sources files, mailing lists, management & report tools, etc.). First version of the use cases. Planning of Communication, Dissemination, Standardisation and Exploitation activities.	1, 2, 8	Month 3	D1.1, Project technological platform office fully operational (verified by all partners), I2.1, I8.1
2	**Intermediate Requirements and Functional Analysis.** WP2 producing intermediate Technical and Business Requirements and Functional Analysis and Decomposition for the Reuse&Sharing of Resources. Early results on System architecture	1,2,3	Month 6	D8.1, I2.1, I2.2

3	**End of phase 1**. First project review; first edition of the Project vision and roadmap; first annual review report; final results from WP2. Stable system architecture; intermediate results from core WP 4, 5, and 6; first report on communication, dissemination and open source contributions and standardisation and innovation and exploitation.	All	Month 12	D1.2 D1.3 D1.4, D1.2, I4.1, I5.1, I5.2, I6.1, I6.2, I6.3, D8.2, D8.3 and first project review
4	**System architecture complete**	3	Month 16	D3.1
5	**Public documentation of major project accomplishments;** in 2nd report on communication, dissemination and open source contributions and standardisation and innovation and exploitation	All	Month 20	D8.4, D8.5
6	**End of phase 2 and completion of core technical WPs (4,5,6)**. Second project review: second edition of the *Project vision and roadmap*; second annual review report; Platform components complete, initial platform available and demoed; Communication and Dissemination reaching wider audience.	1,4,5,6, 7,8	Month 24	D1.2 D1.3 D1.4, D4.1, D4.2, D4.3, D5.1, D5.2, D5.3, D6.1, I7.3; 2nd project review
7	**System integrated**. Platform integration and prototype available.	1,7,8	Month 28	Trial platform integrated and running (D7.2)
8	**End of phase 3 and of the project**. Third project review: third edition of the *Project vision and roadmap*; third annual review report; final release of platform and use case code; third report on dissemination and open source contributions and standardisation .	1,7,8	Month 30	D7.3, D8.6, D8.7, D8.8 and final project review

3.6 Metrics / KPIs

In this section we report the key performance indicators defined by the 5G Infrastructure PPP; we also detail first what is the envisaged contribution of SUPERFLUIDITY to each KPI (see third column) and then we give more details on SUPERFLUIDITY's contribution to four KPIs, which are believed to be the most relevant for our project. The first three are those labelled as P1, S3, B3 in the following table; the fourth one is “End-to-End latency of < 1ms”, and is not reported in the official 5G PPP table below, defined in 2013, but it was added afterwards to those of interest to 5GPPP (see https://5g-ppp.eu/kpis/).

3.6.1 Performance KPI

Table 2: Performance KPI

KPI		Relevance (High/Medium/ Low / N.A.)	Details on planned project contribution towards achieving the KPI
P1	Providing 1000 times higher wireless area capacity and more varied service capabilities compared to 2010.	High	Solutions to lower deployment and maintenance costs (e.g. BBU aggregation on the edge) through the use of standard high volume servers. Increased service diversity will be achieved by means of the slicing concept that will allow diverse services to coexist on the same infrastructure and allocate resources to different tenants. At the same time, our RFB concept and service definition tools provide a means to compose new services based on virtual network functions (VNFs) that goes beyond what can currently be achieved based on available standards.
P2	Reducing the average service creation time cycle from 90 hours to 90 minutes.	High	Automatic verification of service deployments before instantiation via static analysis to reduce the need for manual checks. Symbolic execution is a key enabler for this. Adoption of light-weight virtualisation approaches such as unikernels or container-based approaches to significantly reduce service instantiation times. Simplicity in network function deployment thanks to platform agnostic configuration/programming interfaces.
P3	Facilitating very dense deployments of wireless communication links to connect over 7 trillion wireless devices serving over 7 billion people.	Medium	Cloud RAN concept particularly effective for dense environments Synchronisation and collaboration between cells: Cloud-RAN, HW selection and acceleration tackle this issue with a

			centralised BBU. Work on mobility at layer 4, based on Multipath TCP, to allow seamless handovers across a range of communication technologies. Connect to multiple technologies at once, favouring slow handover while connected via many links rather than fast handover of a single link, as done today.
P4	Creating a secure, reliable and dependable Internet with a "zero perceived" downtime for services provision.	Medium	Almost instantaneous service instantiation. Use of symbolic execution to ensure robustness of the network, to check configurations and services before they are applied, and to ensure security by design, without sandboxing.

3.6.2 Societal KPIs

Table 3: Societal KPI

KPI		**Relevance** (High / Medium / Low / N.A.)	**Details on planned project contribution towards achieving the KPI**
S1	Enabling advanced User controlled privacy;	N.A.	
S2	Reduction of energy consumption per service up to 90% (as compared to 2010);	Medium	Reducing the Energy consumption in NFV data centres by improving infrastructure utilisation through more intelligent placement of workloads and continuous optimisation of placement decisions. Fast Service migration and shift or processing from one hardware to the other enables energy efficient scheduling (and having most of the platform in low energy consumption mode). Use of offloading to dedicated and more energy efficient hardware. For example, offloading OVS functionalities to NICs reduces the CPU utilization dramatically, and thus the energy consumption.

			More in general, cloud networking helps to adapt resources allocation, increasing efficiency by matching the workload characterisation to optimised allocation of resources through resource aware intelligent orchestration.
S3	European availability of a competitive industrial offer for 5G systems and technologies;	High	Consortium partners aiming at strategically placing themselves as a driving force in the area of converged 5G service and network architectures by becoming early adopters of SUPERFLUIDITY's system. Interface definition and contribution to standards. Working with the OpenStack community to improve its functionalities as an NFV infrastructure manager and bridging the gaps with the ETSI NFV requirements. Contributions to industry initiatives such as OPNFV to accelerate the adoption and deployment of open platform solutions to drive the adoption of NFV. Creation of high fidelity prototypes to demonstrate feasibility of the approach developed by SUPERFLUIDITY.
S4	Stimulation of new economically-viable services of high societal value like U-HDTV and M2M applications;	Medium	Edge cloud concept facilitates delay-sensitive services and crowd-targeted video services (e.g. multi-camera video streaming apps of events). Easing the development of M2M markets via virtualised infrastructure.
S5	Establishment and availability of 5G skills development curricula (in partnership with the EIT).	N.A.	

3.6.3 Business-related KPIs

Table 4: Business-related KPIs

KPI		Relevance (High / Medium / Low / N.A.)	Details on planned project contribution towards achieving the KPI
B1	Leverage effect of EU research and innovation funding in terms of private investment in R&D for 5G systems in the order of 5 to 10 times;	Medium	Exploitation of SUPERFLUIDITY results and prototypes by industrial partners into internal R&D programs. Contribution to the open source community.
B2	Target SME participation under this initiative commensurate with an allocation of 20% of the total public funding;	Medium	Budget allocated to SMEs=€ 1.206.375, corresponding to 15,28% of the total public funding.
B3	Reach a global market share for 5G equipment & services delivered by European headquartered ICT companies at, or above, the reported 2011 level of 43% global market share in communication infrastructure.	High	Use of SUPERFLUIDITY collateral with customers to demonstrate feasibility of 5G use cases and solution paths to major 5G technical challenges. Investments by industrial partners.

3.6.4 Contribution to high-relevance KPIs

In this section, we give more details on SUPERFLUIDITY's contribution to four KPIs, which are believed to be most relevant for our project:

KPI P1: Providing 1000 times higher wireless area capacity and more varied service capabilities compared to 2010.

The current set of services that is achievable with Network Function Virtualisation (NFV) is limited by the composition mechanisms offered to construct VNF forwarding graphs (VNF-FGs). SUPERFLUIDITY is developing the Reusable Functional Block (RFB) concept as well as defining means to define service graphs recursively so as to compose services based on virtual network functions (VNFs) that goes beyond what can currently be achieved based on available standards. A RFB is a logical entity that performs a set of functionality and has a set of logical input/output ports. In general, a RFB can hold state information, so that the processing of information coming in its logical ports can depend on such state information. RFBs can be composed in graphs to provide services or to form other RFBs (therefore a Reusable Functional Block can be composed of other RFBs). RFBs need to be characterised and described in a formal manner. Additionally, we need platform-agnostic node-level and network-level "programs" describing how the RFBs interact, communicate, and connect to each other so as to give rise to specific (macroscopic, and formerly monolithic) node components, network functions and services. A language that supports the description and the interaction of RFBs is referred to as *RFB Description and Composition Language* (*RDCL*). The heterogeneous computational and execution environments, supporting the execution (and deployment) of the RDCL scripts and the relevant coordination of the signal/radio/packet/flow/network processing primitives are referred to as *RFB Execution Environments* (*REE*). The RFB decomposition concept is applied to different heterogeneous environments. An RFB may be analogous to a traditional VNF or VNFC, implemented as a fully-fledged VM running on a hypervisor or in an OS container. An RFB can correspond to a small footprint Unikernel VM running in a specialised hypervisor. In the latter case, the execution environment is the hypervisor specialised in supporting lightweight VMs. RFBs can also be modules or components of special purpose execution environments, like extended finite state machines based on OpenFlow for packet processing [BIA14], software routers [Koh00], or radio signal processing chains [BAN12] (for more details on the RFB concept see also Section 4.1 and I3.1).

In the following we report specific contributions to this KPI from partners.

Telefónica is pushing the RFB concept in the IRTF NFVRG as an initial sanity check, to be later integrated into the ETSI NFV work. Telefonica believes that it will allow to define and implement novel services based on previously defined and tested services in an agile way, very much in line with current DevOps approaches.

BT believes that in the next 2-3 years the worlds of cloud computing and virtualised network functions are coming together into a completely new model for telecommunications combining cloud services, network function virtualisation, and software defined networks. Entire enterprises will be able to run on virtualised servers, virtualised network functions, and virtualised network control services. The BT Vision is described by Howard Watson, our CEO of BT Technology, Service & Operations, at https://player.vimeo.com/video/166341115 (starts at 14 minutes). As part of this convergence, NFV has a critical role. BT has already launched our first commercial service that uses NFV (Cloud Connect Intelligence), with more in the pipeline. The fast and flexible deployment and modification of services is what is most attractive about NFV, including the promised DevOps approach – as this simplifies and automates the process from design of the new network service to its deployment. BT is feeding the results into our internal OSS/BSS developments.

NEC has worked on virtualisation execution environment minimalisation as reported as part of I5.2 (see also Section 4.1). By reducing the size of VMs to the minimal set of resources, it is then possible to have more workloads running on a node. For this, NEC has evaluated mapping workloads to containers and minimal Linux execution environments called TinyX. This work has also looked at profiling of the workload startup time and then reducing the elements that take most of the time. In Xen this has led to improvements in the control stack tools to reduce the startup time of the workloads.

NOKIA IL is working on optimizing the cloud utilization through optimizing and automating the infrastructure configurations, improving the placement algorithms based on rigorous hardware benchmarking, and task offloading to dedicated hardware (e.g., OVS offload to NICs). To allow those optimization, NOKIA IL rely on and extend OpenStack services and in particular Mistral work flow engine. Here, Mystral is a key component that facilitates the applications deployment and life cycle management. For example, orchestrating application scaling to obtain adequate application performance or alternatively better resource utilization.

OnApp has been working on hypervisor-level improvements and virtualisation optimisations to allow its MicroVisor platform to handle the scale of network functions expected for 5G. Fundamental to performance in a distributed platform, which is likely the way that hardware architectures will go, at least in the core network, is improving the network latency and throughput between guests. The network latency between guests on the same physical node and also to nodes that are hosted on different physical machines is very important for the overall performance of network functions and for service function chaining of RFBs that is envisaged in SUPERFLUIDITY. To this effect OnApp have been working on improving the network performance for 10-40GbE network interfaces and reducing the overhead in the virtualisation layers.

Nokia Bell-Labs France has been working on Cloud RAN to address capacity and coverage issues, while supporting mobile xHaul (Fronthaul and Backhaul) solutions as well as network self-optimisation and self-adaptation with software control and management through SDN and NFV. Centralising the processing in the Edge Cloud, one of the main feature in Cloud-RAN, open the door to implement a very powerful optimisation technique like the coordinated multipoint (COMP), Interference Cancellation scheme, improving the throughput of the system. The testbed #1 (described in Section 4.1) hosted by Nokia FR supports a first Proof of concept of Cloud RAN demonstrating the fronthaul network and the deployment of the different RFBs composing the Core and the RAN. These RFBs are executed using containers technology. On the testbed, we tested the flexibility in the service deployment in the sense that each container could be deployed in a virtual machine or on bare metal, as it is the case in the compute Node 2 of Testbed #1, following the service requirement.

EBlink will contribute towards a wireless Fronthaul architecture to ease network deployment with more flexibility while increasing network capacity in dense area by developing µWave/mm-Wave wireless fronthaul to support CRAN network. EBlink will work with Nokia to make more efficient some techniques such as UL CoMP, ICIC, CSPC, CA etc. by enabling BBU pooling/virtualisation to reduce X2 delay and enable Gbps wireless links with a very small footprint and few impact on the environment. The project will bring a new technological step in terms of spectral efficiency. Thanks to the use of the high capacity and high spectral efficiency wireless Fronthaul link combined with different baseband functional split options, it will be possible to overcome the last distance gap where optical fibre connection is not always available and where the cost of installing it would be too high and the administrative process a lengthy one. By developing a new flexible fronthaul interface (XHAUL) that can support both CPRI and packet-based protocol (NGFI), the available resource can be used more efficiently and ensure a smooth transition from legacy network to 5G network. For a deployment based on distributed base stations architecture/Cloud-RAN, laying down optical fibre (Gbps links) from the network access point all the way up to the RRHs is often a challenge and consequently imposes significant limitations to bring capacity in dense areas. The wireless fronthaul will contribute to build local cloud RAN with high capacity centralised processing and low-cost distributed RHHs. Mobile operators can easily and cost-effectively install small-RRH for indoor coverage (iDAS) for outdoor coverage and capacity in dense urban areas that are hooked to baseband and backhaul resources. In summary it can be said that the EBlink solution will facilitate very dense deployments of wireless communication links via the local cloud RAN and at the same time will help to reduce network latency and cost since Local Cloud RAN will facilitate MEC at the edge of the mobile network, within the Radio Access Network (RAN) and in close proximity to mobile subscribers.

Unified Streaming will unleash the power of the SUPERFLUIDITY architecture to tailor it towards the needs of its clients in the Over the Top Video Streaming Industry. In doing this it will exploits its expertise in protocols and media standards such as those defined in MPEG and MPEG DASH industry Forum of which it is both a contributor and member. In particular it will provide its core multimedia functions such as trans multiplexing (serving both Apple and Samsung devices), Digital Rights Management (content protections), and personalization (ad insertion) at the edge of the network in virtualized superfluid compute instances. This is expected to highly increase the efficiency of over the top video streaming in access networks, benefiting from Radio Network information services. The SUPERFLUIDITY project provides a use case to demonstrate that using the Unified Streaming products Remix, Origin in the edge will help achieve this target KPI.

KPI S3: European availability of a competitive industrial offer for 5G systems and technologies

The SUPERFLUIDITY consortium includes key actors at different standards defining organisations (SDOs) (i.e. ETSI NFV ISG, ETSI MEC ISG) and initiatives (i.e. OSM) that are currently shaping the 5G network.

BT is one of the companies that created the original work at ETSI and created the concept of network virtualisation, and we continue to play a leading role. Where the SUPERFLUIDITY results have implications for standards, BT will discuss this within the project, as a form of socialisation, and then help bring the results to the relevant standards bodies and support their progress. BT believes that standardisation is key in the telecoms field, and certainly in the NFV space, in order to ensure interoperability and allow us to source components from different / multiple vendors.

Telefónica and BT have leadership roles at ETSI NFV ISG; BT is Rapporteur for the End-to-end Processes Work Item and plays a key role in the ISG’s initiative to bring together the Information Models of many SDOs and industry groups. In addition, BT is involved in other bodies such as: Next Generation Mobile Networks Alliance, which is a service provider alliance that puts in joint requirements into SDOs; OSM (Open Source MANO – founding member, vice chair, and chair of the End User Advisory Group); and the TMF’s ZOOM project (Zero-touch Orchestration, Operations and Management).

Telefonica is chairing the TSC of the ETSI NFV ISG. With regards to OSM, Telefónica is chairing the board. In addition, Telefónica is a member of the User Advisory Board of OpenDaylight. Currently there are a large number of open source projects relevant to NFV, some of which are competitors – at the moment we believe this is reasonable, as it is encouraging innovation and filling gaps in proprietary implementations.

CloudBand of NOKIA IL is developing an NFV platform composed of Infrastructure Software, Application Management, and Cloud Network Director. Enhancing CloudBand platform with SUPERFLUIDITY concepts and particularly with support for RFBs is part of CloudBand path to 5G. Extending Cloudband management and orchestration products to the MEC domain is yet another goal in CloudBand roadmap. This will allow CloudBand to bring a leading NFV platform ready for 5G application demand. In addition to Cloudband product goals, NOKIA IL is an active contributor to OpenStack, leveraging it as a leading virtual infrastructure manager for NFV deployments, by that fostering the whole NFV industry. In OpenStack we are working with the community to bridge the gaps between OpenStack and the ETSI NFV community. For example, we contributed to Heat to allow better support for NFV deployment. Furthermore, CloudBand initiated a new OpenStack project, named Vitrage, which was recently approved under the OpenStack tent, developing a platform for monitoring correlation and more specifically, root cause analysis. Vitrage is expected to provides triggers and inputs for fast scale and migration operation that are facilitated by the SUPERFLUIDITY architecture. In addition to OpenStack, and OPNFV, NOKIA IL also operates within TOSCA to define the VNF and NS descriptors.

ONAPP is an SME that provides an IaaS solution to 3,000+ customers. As the leading public cloud provider in terms of number of licensed customers (1 in 3), ONAPP offers the possibility of exposing new 5G services and solutions to a large customer base. Particularly relevant from SUPERFLUIDITY are the innovations into managing large numbers of small, virtualised work environments that can be tied together to form high performance services. Being able to provide fluid resources across the datacentre and cloud access points through its ONAPP Cloud and ONAPP CDN platforms are the core business areas of ONAPP and as such any improvements in this space, as proposed by SUPERFLUIDITY will be advantageous for ONAPP's customers and the respective end-users. ONAPP has an exploitation plan that has been detailed, but aside from those specific exploitation activities, ONAPP will benefit from providing the latest 5G systems and technologies as a market leader with continued innovation. Many of ONAPP's customers that are based in Europe have multiple European data centres and so will benefit from the technologies that are being promoted by SUPERFLUIDITY.

Unified Streaming is an SME software company that provides OTT (over the top) streaming software to over 200+ customers including large companies in the media industry such as BBC iplayer, RTL, etc. Beyond that, its software is also used outside of Europe such as HBO (USA) and Foxtel (Australia), Globo.com (Australia). The Unified Origin is a flagship product that provides DRM, Late Trans-muxing, Streaming manifest generation that is particularly well suited towards offering in a 5G environment. Unified streaming will be an exemplary super user of the SUPERFLUIDITY deploying its software in edge cloud, virtualized environment, as a key offering an exemplary 5G environment that will bear many of the concepts from SUPERFLUIDITY.

KPI B3: Reach a global market share for 5G equipment & services delivered by European headquartered ICT companies at, or above, the reported 2011 level of 43% global market share in communication infrastructure.

The SUPERFLUIDITY consortium includes industrial partners that make use of the collateral with their customers to demonstrate the feasibility of 5G use cases and solution paths to major technical challenges.

NOKIA is a one of the leading telecom vendor with expected huge impact on the 5G market. CloudBand products, developed by NOKIA IL, are in the heart of NOKIA portfolio, where all NOKIA's VNFs are designed and optimize to work with. SUPERFLUIDITY concepts are well within CloudBand path to 5G, and as such will have impact on all NOKIA's product.

ONAPP, as an SME, is fully adopting the approaches recommended in SUPERFLUIDITY for its next generation Cloud platforms. Work at the Emerging Technology Department has been focused on getting an optimal platform for RFBs and Virtual Network Functions. At the time of writing, the work is still at an early stage for being production ready but it is fully envisaged that OnApp will be offering some products by the end of the SUPERFLUIDITY project that offer RFB / VNF support.

Unified Streaming as an SME will adopt the approaches recommended in SUPERFLUIDITY for deploying its much used streaming software for its clients. In doing so it hopes to unleash the 5G enabled video streaming components that will run at the edge network near the RAN. Further, the scalability offered by cloud computing and virtualized network instances will benefit video streaming in virtualized infrastructure, which Unified Streaming expects to be a key paradigm for the future of video streaming. Unified Streaming hopes to be a key player in delivering this 5G service, working with its clients worldwide.

To increase the global market share, Telcaria plans to reduce costs by combining SDN and NFV into their products and services for software-based telecom operators and ICT solutions integration at request, and therefore increasing added value by service differentiation. One of the main areas in which Telcaria foresees exploitation potential of SUPERFLUIDITY is the advances in technologies for integrating heterogeneous wireless networks (up to RAN level) and in architectures to optimise the reuse of functionality across heterogeneous access technologies for 5G.

KPI: End-to-End latency of < 1ms

Considering that the laws of physics cannot be overtaken, the only way to continue reducing latency to magnitudes on the 1ms order (end-to-end) is getting the services closer to the end users.

Recently, ETSI has created the MEC (Mobile Edge Computing) ISG in order to cover this issue. At this moment, there is a basic architecture defined and some guidelines about how edge computing should work. SUPERFLUIDITY (via Altice Labs) is actively contributing to this group. The SUPERFLUIDITY project is implementing that architecture from scratch, and expects by the end of the project to be able to fully manage and orchestrate the whole solution. The MEC solution to be implemented has some key properties:

- Compatibility with existing mobile networks - there is no need for network changes; the network is MEC-unaware.
- Compatibility with existing applications - applications do not need to change to provide services, unless they want to take advantage of MEC Services (APIs).

- Transparency to end users - the user is not aware about whether the service is provided from the core or the edge (they just feel different latencies).
- Integration with NFV infrastructure - the operator can take advantage of NFV and MEC synergies, reducing overall costs.

3.7 Dependencies and risks

In this section we report the risk envisaged in the Description of Work and we add a column reporting related issues, if any.

Description of risk and related probability and impact	WP(s) involved	Proposed risk-mitigation measures	Risk State of Play
Drop-out by a partner Probability: Low Impact: Medium	All	Partners in the project include major public institutions or large companies and two "large" SMEs, very unlikely to fail. Smaller and newer SMEs might run greater risks, but their relative weight in the consortium is also smaller. Drop out is highly unlikely for academic partners. The risk may be marginally higher for large companies, where parent companies or central management may decide to discontinue commitments for reasons not under our control. To limit the impact of such withdrawals, all WPs include at least a second partner with sufficient expertise to take the lead of the activity. In addition, it is true that partners' planned work complement each other without significant overlaps, but we allowed for a limited redundancy and sharing of tasks, which mitigates this risk. Given that no partner has more than 10% of the total effort, drop out of partners with no leading roles will be easily handled by either redistributing the work between other partners, recruiting a new partner, or revising the Technical Annex to deal with the withdrawal.	Risk did not materialise
Delays in the design and development of key project's elements Probab.: Medium Impact: Medium	All	Academic partners of SUPERFLUIDITY have extensive experience not only in research work but also in the deployment of demonstrators and test-bed, as proven by a very successful track record. The consortium includes companies certainly capable of delivering the technological components and software foreseen in the work plan. The Work Plan contains intermediate steps	Risk did not materialise

		and partial results. The task schedule is such that there is sufficient overlap between the various phases of the project (design; platform development; application development; deployment) so that moderate delays can be accommodated and recovered without much trouble. Also, results from the various stages of the project are not monolithic entities so partial unavailability of some elements will not entirely block the next stage of the activity. As an example, system development can start with partial results from WP4-6, and same goes for the integration activity in WP7 as preliminary implementations of the components in WP4-6 are available.	
Technology is superseded by competing solutions and/or patents are filed by competitors Probab.: Low Impact High	All	Partners have connections with other leaders in the industry and academia so will be aware of other projects that are ongoing and where there may be overlaps. The partners are also involved in standardisation groups so will see motions by other groups in similar areas. If competing solutions are created, then collaborative efforts will be sought in the first instance, and then failing this, efforts will be made into revising the scope and target areas of SUPERFLUIDITY.	Risk did not materialise
Lack of exploitation of project results Probability: Low Impact: High	All	Commercial and institutional partners of SUPERFLUIDITY intend to exploit and push SUPERFLUIDITY solutions. The project's dissemination and communication activities will allow the project to reach a very large base of possible customers, and to test and advertise SUPERFLUIDITY in real markets, a further assurance of real exploitation.	Risk did not materialise
Inadequate coordination Probability: Low Impact: High	WP1	CNIT has extensive experience in the coordination of EU projects. The management structures and procedures outlined in this proposal ensure that project management can closely supervise the delivery of the expected results (internal intermediate results as well as official deliverables). Meetings of the Management Board will identify potential problems and react early, e.g., by reducing the functionality of a prototype or by organisational changes. The Consortium inherits some management procedures and working relationships among some of the Consortium members, tested during previous EU projects and have been shown to work very	Risk did not materialise

		effectively and which have been very favourably judged by reviewers during technical reviews. The "management risk" attached to the work is therefore minimal.	
Conflicts among the partners Probability: Low Impact: Medium	WP1	The work plan has been designed so that tasks and responsibilities are clearly assigned so that conflicts are unlikely to arise. However, in previous projects, the SUPERFLUIDITY management team has already successfully handled conflicts between partners. The strategy applied mixes strong leadership by the coordinator with consensual discussion and decision-making within the General Board. Should unresolvable conflicts arise, their outcome will be handled as detailed in the partner dropout and "delays" risk.	Risk did not materialise
No consensus on the new architectural vision Probab.: Medium Impact: Low	WP3	No consensus on the new architectural vision, due to the different ideas requirements, and viewpoint of the partners, each coming from a different domain (e.g., hardware, virtualisation, orchestration and wireless). Mitigating this risk, the architecture work will closely follow and comply with the work of WP2 on the system requirements and use cases. Additionally, the architecture specification will be organised into two phases, in order to correct and extend the initial design on the basis of emerging limitations. Since the goal of the project is to create a generic architecture we argue that different point of views promoted by each partner can be an extra value (rather than a limitation only) for the success of the project. It will leverage from the broad view of the partners that is covering all related domains.	We had several discussions on architecture and after initial disagreements we reached a consensus
Rapid changes in Cloud and NFV concepts Probability: High Impact: Low	WP3	NFV and Cloud RAN are in a mist of rapid changes from theory to practice. Specifically, hardware capacities are constantly revolutionising (e.g., from servers to micro servers) while the wireless signal processing are becoming more and more demanding. In accordance, cloud concept are constantly varying, e.g., from full abstraction with large virtual machines to small containers that rapidly deployed and scaled. Therefore, it is expected that the state of the art of the cloud technologies will change dramatically	Risk did not materialise

		in the course of the project and may influence the devised architecture. Indeed, our architecture should be designed to meet with future cloud concepts. We will verify its advantages by seeing if it adapts and comply with future changes. Having the cloud industry leaders and the strongest influential companies on these technologies, allow us to foresee future generations and design in advance. That said, we also included a second phase for the architecture design to get updated.	
Energy efficiency of platform is not improved Probability: Low Impact: Medium	WP4	By using thin HVs, the likelihood is that more HVs can be provisioned for a system with a given capability. Inefficient scheduling may cause more contention issues and reduce the energy efficiency of the system. Members of the consortium are involved with other efforts to reduce the energy overhead at various levels of the Cloud and so this will be taken into consideration. The profiling carried out within WP4 will include deep instrumentation to inspect the effective use of infrastructure specialisations which will be correlated against application/service level performance KPIs. With the goal being optimal service performance relative to Cap Ex and Op Ex, energy efficiency where appropriate can be weighted as a factor.	Risk did not materialise
Abstraction of infrastructure resources is sub-optimal. Probability: Low Impact: High	WP4	Achieving the appropriate level of abstraction whereby meaningful differentiations are exposed to the consuming layer but not at the cost of excessive operational maintenance of a data model that allows for future platform innovations to be readily consumed. Project partners such as Intel have in-depth experience of correlating time-series data across the entire stack and applying statistical processing to correlate the key metrics which represent performance both of the system and of the infrastructure resources. Intel also has expertise in model inference, identification of high-order application specific metrics and in data models to expose actionable parametric representations of resource utility.	Risk did not materialise
Performance of	WP5	Project partners such as NEC have good	Risk did not

network packet forwarding less than other solutions Probability: Medium/High Impact: Low		experience in this area having worked on ClickOS and working with other leading contributors in this area. CITRIX is the main supporter of the Xen hypervisor and has a group dedicated to performance optimisations. Further, ONAPP bases a number of its high performance solutions on the Xen hypervisor.	materialise
Sub-par performance of massive consolidation server Probability: Low Impact: Medium	WP5	Running large number of virtual machines (in the order of 10K) on a single x86 server has never been attempted before, as far as we know. However, very early prototypes developed by project partners show that this is possible with the currently available hardware.	Risk did not materialise
Microserver hardware does not perform as expected Probab.: Medium Impact: Medium	WP5	Microserver hardware has only recently hit the market, and the chipsets and hardware they are based on can vary significantly. However, early experience by project partners shows that type-1 hypervisors such as Xen and type-2 ones like KVM can run on ARM-based microservers.	Risk did not materialise
Control framework does not work with all VNFs Probab.: Medium Impact: Low	WP6	Indeed each VNF possess unique requirements for the control framework. For example, each VNF has unique deployment requirements or scaling scenarios that might not be covered by our SLA based descriptors. In SUPERFLUIDITY we will aim for the broader application and design our framework accordingly. However, it will tested and validated with the unique and most demanding application of C-RAN. In the phase of identifying such gaps, the framework would be updated. Furthermore, we will allow bypasses to support such unique cases.	Risk did not materialise
Coded traffic preventing meaningful monitoring of tenants' behaviour Probability: High Impact: Low	WP6	To address such problems, we plan to offer a variety of features which can be monitored even under coded traffic. Such features, for example, timing data or sizes, can still give valuable knowledge on the way tenants behave, and, especially, change significantly if a tenant's behaviour changes significantly. Moreover, such features can be monitored without compromising tenants' privacy.	Risk did not materialise
Measurable features do not	WP6	To address this problem, we will employ data fusion on the results of multiple anomaly	Risk did not materialise

give significant results for efficient monitoring Probability: Low Impact: Low		detection algorithms, in order to achieve one significant measure of tenants' behaviour out of several, possibly very noisy, features.	
Operators' policies prevent coding at specific locations Probability: Medium Impact: Medium	WP6	Legacy software/hardware or restrictions due to operators' policies prevent coding at specific locations. To cope with such problems, we plan to employ coding techniques which minimise the number of coding nodes, and, moreover, are able to adjust, under some limitations, to nodes which must adhere to stricter constraints. Moreover, a possible solution in such systems is to allow for ``coding servers" - services which perform the required coding and can be re-located based on the system current requirements.	Risk did not materialise
Complexity of managing low resource HVs leads to too much complexity in orchestration Probab.: Medium Impact: Medium	WP5/ WP6	It is expected that reducing the complexity of the management plane in the hypervisor will lead to an increased need for orchestration at a higher level. This is an expected trade-off. If the orchestration becomes too complex then control will be passed back to the control domain at the expense of performance.	Risk did not materialise
Unable to integrate OnApp PoC into existing solutions Probab.: Medium Impact: Low	WP7	OnApp have managed to get results of previous European projects into their products but given that the contributions will go through other groups, external to SUPERFLUIDITY there is a chance that the contributed versions will not make it into the project. The mitigation action will be to promote the PoC work as a prototype rather than utilising a full integration.	Risk did not materialise
Standards bodies not accepting SUPERFLUIDITY’s work Probab.: Medium Impact: Medium	WP8	Several partners have extensive experience in standardisation work and are already involved on topics related to the project’s ones. In addition, standardisation bodies operate on long timescales, often exceeding the lifetime of this project, thus contribution to a standardisation body also means providing useful input to the process without necessarily having a standard formally accepted.	Risk did not materialise

4 Work performed so far

4.1 Current achievements: R&D

In this section, we report the work performed as regards R&D activities. As already mentioned above, today's networks suffer from a variety of shortcomings, including a lack of service agility, a lack of implementation agility and increasing complexity. The lack of service agility prevents us from creating new services in a rapid, flexible and tailored fashion. The lack of implementation agility means that we have to rely on rigid, cost-ineffective hardware devices with long provisioning times. Increasing complexity arises from the continuous growth and heterogeneity of network traffic, services and hardware technologies.

Several emerging trends, mainly in the context of 5G networks are likely to exacerbate these issues: the forthcoming explosion of the Internet of Things (IOT) new radio techniques such as massive Multiple Input Multiple Output (MIMO) and beam-forming, and the desire for more flexible business models.

The architecture being developed by the SUPERFLUIDITY project is focused on alleviating these problems. It has the following features: i) Flexibility, via an architectural decomposition of network components and network services into elementary, reusable primitives, defined in SUPERFLUIDITY as Reusable Functional Blocks (RFBs); ii) Agility, via the rapid 'chaining' of these RFBs to form exactly the service required; iii) Simplicity, via virtualisation of radio and network processing tasks, network functions and services (a fully cloud-based architecture); iv) Portability, via platform-independent abstractions, permitting the reuse of network functions across multiple heterogeneous types of hardware; v) High performance, via software acceleration, specialisation and adaptation to hardware accelerators.

The rest of this Section is organised as follows. First, we briefly summarise our first year achievements, and then we detail the work done in each Work Package (WP).

4.1.1 Work Done in a Nutshell

The work performed by SUPERFLUIDITY during the first year of the project has covered several aspects, including:

1) definition of use cases, system requirements, and functional analysis, performed in WP2;
2) definition of the SUPERFLUIDITY architecture, in WP3, based on the on the concept of *Reusable Functional Blocks* (RFBs), which is applied to different heterogeneous *RFB Execution Environments* (REE);
3) heterogeneous infrastructure and abstractions, in WP4, by: i) modelling, profiling and selecting the hardware used by the project, and ii) investigating the performance, scalability and portability aspects of different hardware platforms;
4) virtualisation platform implementation and network dynamics, in WP5, by considering: i) the implementation and allocation of function allocation algorithms, and ii) the survey and the modelling/measurement of the available virtualisation techniques;

5) system orchestration and management tools, in WP6, including the following activities: i) control framework design, ii) modelling and design for symbolic execution and monitoring tools;
6) the System Integration and Validation WP (WP7) is not formally started, yet; however, two testbeds have been established (ahead of schedule). These will allow various aspects developed in the other WPs to be explored practically (see Section 4.2).
7) Communication, dissemination, standardisation and exploitation activities, in WP8.

We report in the following a more detailed description of the work performed in each WP.

4.1.2 WP2 Use Cases, System Requirements, and Functional Analysis

A "superfluid" network will have the ability to instantiate services on-the-fly, run them anywhere in the network (core, aggregation, edge) and shift them transparently to different locations. Such capabilities are a key part of the converged cloud-based 5G future - they will enable innovative use cases in the mobile edge, empower new business models and allow almost instant roll out of new services, and reduce investment and operational costs. Specifically, there are a large number of potential use cases for a superfluid network, covering themes such as wireless access, mobile edge computing and on-the-fly monitoring. By studying these use cases we can better understand both business requirements, such as service agility and cost savings, and technical requirements, for example quality of experience and scalability.

Whilst use cases are interesting in themselves, they also help guide the technical work of the project. We have not attempted a strict 'top down approach', where the architecture would be derived in a formal process from the requirements of the various use cases. Instead we are using an iterative approach where the requirements analysis takes into account our on-going architecture work. We include business requirements, such as service agility and cost savings, and technical requirements, for example quality of experience and scalability, as well as architectural concepts like reusable functional blocks.

To instantiate services on-the-fly, run them anywhere in the network and shift them transparently to different locations, SUPERFLUIDITY introduced the concept of Reusable Functional Block (RFB). This capability is a key part of the converged cloud-based 5G future. Specifically, during the first year of the project we have started to understand the benefits of the concept and identify its most important applications. More in depth, SUPERFLUIDITY proposes to decompose the architecture into elementary radio and network processing primitives and events, which can then be exploited as basic modules of more comprehensive (and traditionally monolithic) network functions and services. The decomposition of monolithic functions into RFBs permits a flexible placement, as well as the incorporation of adequate virtualisation techniques like Containers (e.g. Docker), UniKernels, and full Virtual Machines, by considering real-time constraints.

SUPERFLUIDITY has focused on a set of different contexts, such as: Cloud Radio Access Networks, Mobile Edge Computing platform, generic NFV environments, fixed networking equipment and packet processing state machines. In each of this context a preliminary functional analysis and decomposition into RFBs has been performed.

4.1.3 WP3 Architecture and Programming Interfaces Specification

The vision of the SUPERFLUIDITY project is to move from the current architectural approaches based on monolithic network components/entities and their interfaces, to an approach where network functions can be programmatically composed using "building blocks" and the deployment of these "building blocks" over the underlying infrastructure is highly dynamical, allowing a continuous real-time optimisation. The SUPERFLUIDITY architecture is based on the following main pillars: (i) standardisation convergence, (ii) Reusable Functional Blocks (RFBs), (iii) programmable and portable interfaces, (iv) a security framework.

Standardisation Convergence

The SUPERFLUIDITY project aims to design a unified, high performance and distributed cloud platform for radio and network functions support, as well as their migration. In our vision, CRAN, MEC and cloud technologies are integrated, by adopting an architectural paradigm able to create the glue that can unify heterogeneous equipment and processing into one dynamically optimised, superfluid, network. Figure 5 depicts a high-level view of the overall architectural framework that has been investigated during the first year of the project. The top layer of the Figure includes the different components involved (CRAN, MEC, virtual core and Data Centres (DC)), while in the bottom layer the different types of physical DCs are shown (namely Cell-site, Local, Regional and Central). This classification is somehow arbitrary and the infrastructure of different operators can be structured in different ways. The next layer down is a traditional Operational Support System (OSS), whose main goal is to deal with all the components in order to create services for end-users.

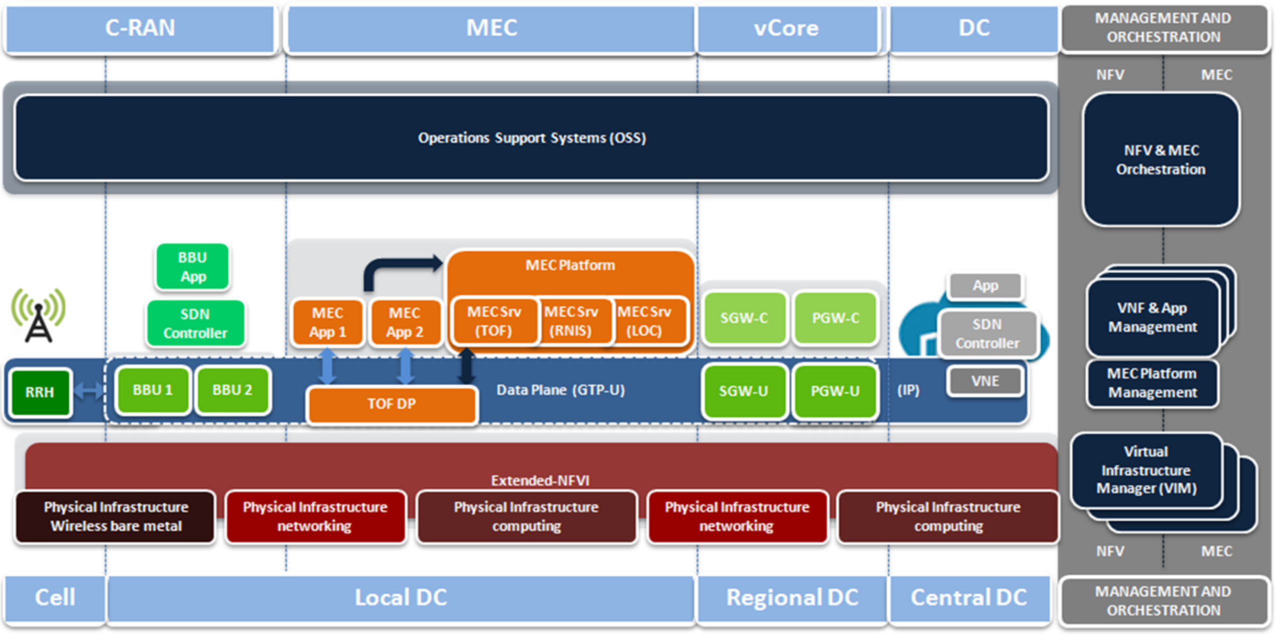


Figure 5: Architectural framework for SUPERFLUIDITY with an example mapping into the physical Data Centres

The underlying Extended-NFVI, located at the bottom of the proposed architecture, represents an evolution of the ETSI NFVI concept. The current NFVI focuses on supporting VMs or containers to run VNFs; the E-NFVI also considers heterogeneous execution environments (see I3.1 for more details). This extended-NFVI is common to all components, simplifies resource management and allows agile (superfluid) orchestration of services. The mapping of the components into the different physical DCs (Cell, Local, Regional) shown in Figure 5 and described hereafter is the one considered for the SUPERFLUIDITY testbed implementation. However, we stress that the dynamicity of the architecture, and the concept of an Extended-NFVI, allows support for different solutions, derived by considering the various trade-offs between performances and efficient utilisation of resources.

Starting from the left, the CRAN component is split into two blocks, corresponding to the Remote Radio Head (RRH) and the Base Band Unit (BBU) components. The RRH is placed in the cell-site, while the BBU is located in a specialised local DC, adapted for telecommunications. It is assumed that the local DCs will control a small number of cell-sites and that they are geographically distributed, but in general proximity to the cell sites. The BBU functionalities will be virtualised and thus they will benefit from the centralisation to scale in/out, according to the load and to be fully re-programmable under a holistic NFV/SDN controller. Figure 5 also shows the MEC components. In our architecture the MEC component is deployed in the same location (local DC) as the CRAN, given the fact that it utilises the exact same infrastructure with the latter or other VNFs. This endogenous characteristic of MEC significantly increases efficiency and ease of use. The next component shown in Figure 5 is the virtual Core (vCore), comprising the central nodes of a cutting-edge mobile network. The vCore runs on the common E-NFVI, usually located in regional DCs. In this way, both agility and fluidity of the overall architecture are improved, especially when live nodes need to be migrated and/or scaled. The Figure shows the expected evolution of mobile core networks towards the SDN model, where the control plane and data plane components are completely separated, with the former fully controlling the latter on data processing tasks. In particular, the components shown are the ones resulting from splitting the 4G/LTE Core elements. The DC component corresponds to the traditional datacentre segment, where a large number of services are deployed. These services are located at central points and deal with significant compute/storage/network resources. Beyond traditional services, central DCs also implement NFV and SDN technologies, in order to control the network components inside the DC. For this reason, a generic VNE (Virtual Network Element), a SDN controller and a generic App element are included in the architecture.

The last part of the architecture is the common management and orchestration vertical layer (right part of the Figure), which is responsible for providing the system intelligence. It includes a NFV-like set of functions, which supports an integrated view of the overall architecture, including networks, services and DCs. In this way, it is possible to take advantage of a common extended-NFVI and achieve an end-user centric view of the ecosystem. This layer is responsible for resource management over the different DCs illustrated at the bottom of Figure 5, thus building a federated environment. Moreover, it orchestrates VNFs to create complex services, taking the best decisions for services to be deployed, while considering customer needs.

Reusable Functional Blocks

The decomposition of high-level monolithic functions into reusable components is based on the concept of *Reusable Functional Blocks* (RFBs), which has been introduced during the first year of the project. An RFB is a logical entity that performs a set of functionality and has a set of logical input/output ports. In general, a Reusable Functional Block can hold state information, so that the processing of information coming in its logical ports can depend on such state information. RFBs can be composed in graphs to provide services or to form other RFBs (therefore a Reusable

Functional Block can be composed of other RFBs). The RFB decomposition concept is applied to different heterogeneous environments.

In order to achieve the vision of superfluid composition, allocation and deployment of “building blocks” (RFBs), the SUPERFLUIDITY project focuses on the following pillars:

- *Decomposition of monolithic functions into Reusable Functional Blocks in different heterogeneous environments* - SUPERFLUIDITY targets to decompose: network services into basic functional abstractions; node-level functions into more granular packet processing primitives; heterogeneous hardware and system primitives into block abstractions.
- *Measurements based RFB operations* - This pillar consists in the definition and realisation of tools for real time monitoring of RFB performances and of the status of the environments (e.g. resource utilisation). The information provided by such tools can be used to drive the RFB operations (e.g. allocation of resources to RFBs).
- *Semantic specification of Block Abstractions* – A crucial aspect in composition-based approaches is to provide the “programmer” with a clear and unambiguous understanding of what each block is expected to do, and how it processes input data and produces an output.
- *High Performance Block Abstractions Implementation* - This pillar consists of deriving block abstractions for each of the underlying hardware components, and to optimise their performance.

Programmable and Portable Interfaces

RFBs needs to be characterised and described, and we need platform-agnostic node-level and network-level “programs” describing how the RFBs interact, communicate, and connect to each other so as to give rise to specific (macroscopic, and formerly monolithic) node components, network functions and services. A language that supports the description and the interaction of RFBs is referred to as RFB Description and Composition Language (RDCL). The heterogeneous computational and execution environments, supporting the execution (and deployment) of the RDCL scripts and the relevant coordination of the signal/radio/packet/flow/network processing primitives are referred to as RFB Execution Environments (REE).

The SUPERFLUIDITY architecture shown in Figure 6 describes the relation among the RFB, RDCL and REE concepts. The figure highlights that the model (with due technical differences) recursively apply at different levels.

Considering the architecture proposed in Figure 6, we identify the logical entities that are involved in the realisation of services. We consider that the architecture can recursively be applied to a number of levels, therefore we define an approach that can be applied at all levels. At each level, we identify a REE Manager and a REE User. The REE User requests the realisation/deployment/execution of a service/service component described using a RDCL script to the REE manager. The REE Manager is in charge of deploying/executing the RCDL script using the resources in its REE. Within a REE, the REE Manager interacts with REE Resources Entities that are needed to support the realisation of the RCDL script.

Hence, we can identify two main APIs. The first one is the API used by the REE User that wants to deploy a service or a component into an RFB Execution Environment. We refer to it as the User-

Manager API (UM API). The second one is the API used by the REE Manager to interact with the resources in its REE. We refer to it as the Manager-Resource API (MR API).

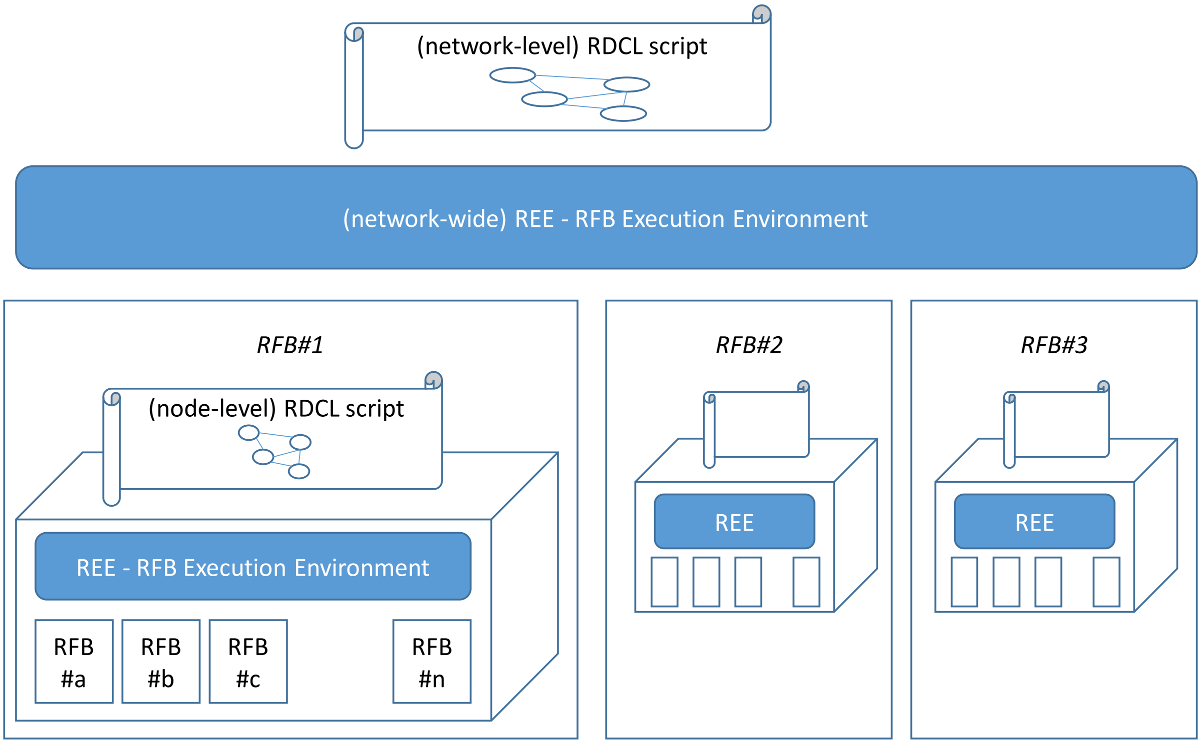


Figure 6: SUPERFLUIDITY conceptual architecture

Figure 7 illustrates these concepts, showing how the APIs can be mapped to different levels. Note that we refer to the APIs considering information exchange in both directions for each APIs.

The APIs for the different levels may use different modelling approaches, languages and tools, considering that SUPERFLUIDITY will not build them from scratch but will rely on existing work. Wherever possible, the project will try to relate the APIs at different level and proceed towards a harmonised vision.

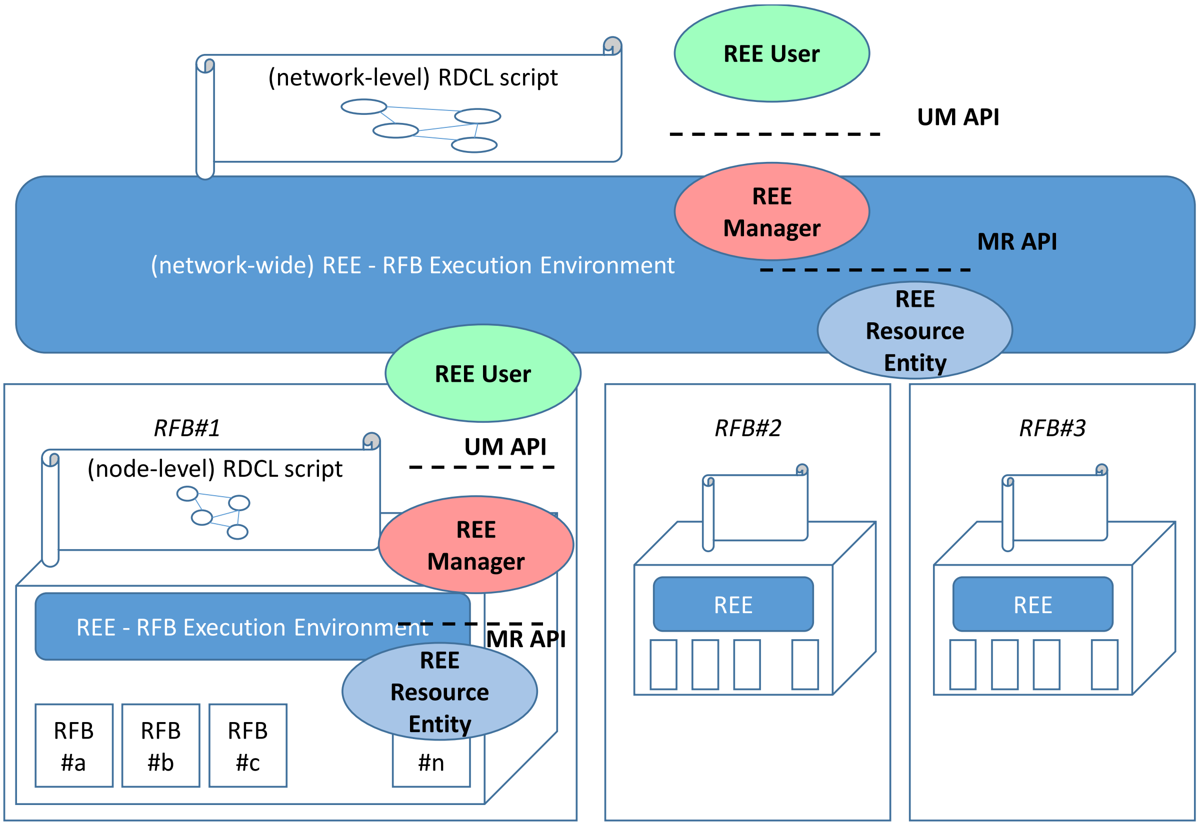


Figure 7: SUPERFLUIDITY Architecture APIs

Security Framework

Any architecture that relies on a real-time cloud infrastructure is susceptible to security hazards. This applies more to those relying on virtualisation of network functions. Accordingly, any such architecture is facing great challenges identifying and taking into account these security hazards. During the first year of the project we have started to study the security challenges and hazards that the SUPERFLUIDITY architecture need to address. Specifically, we have studied the issues concerning general cloud computing and virtualisation. Then, as a second step, we combine these two aspects and formulate the potential security challenges. Our preliminary results show that new security challenges and hence new security threats arise. These issues arise from to the migration of previously conventionally secluded application families to the virtual public environment. Moreover, pure security services can be deployed as isolated VNFs, providing services to other VNFs. Nonetheless, in some cases, the functional architecture should take into account and should be implemented in a symbiosis with accompanied security functions. Furthermore, we show that in different cases some security functionality should be integrated within the VNF design.

4.1.4 WP4 Heterogeneous Infrastructures and Abstractions

The scope of WP4 is twofold. On one hand, it aims to model, profile and select the HW used by the project. On the other hand, this WP targets the investigation of performance, scalability and portability aspects of multiple hardware platforms. Specifically, the first year of the project has been focused on the development of a framework to automatically map service level Key Performance Indicators (KPIs) to the platform level parameters in the host compute environment. The framework includes an analytics pipeline based on the use of eight ranking algorithms with a reliability scoring mechanism. The eight algorithms implemented to date are a mixture of clustering and machine learning classification approaches. The framework has been applied to a video commercial streaming platform from Unified Streaming called Unified Origin. The video transmuxing service is being evaluated under a structured experimental process to develop an understanding of how the service behaves under different usage scenarios. Initial results indicate that memory and associated parameters have the strongest influence on service performance. To investigate this relationship further, deployments of various VM flavours hosting the Unified Origin streaming platform are planned. This will enable the development of models, which relate service level KPIs to specific allocations of resources and user loads at deployment time. Benchmarking of software switches was also carried out using a Cloudband environment from NOKIA with SR-IOV accelerated and native versions of Open Virtual Switch (OVS). CPU pinning and CPU isolation were found to have major effect on increasing the performance and stability of the performance.

As second activity, WP4 has investigated the adoption of various models and techniques for presenting information collated from a set of input sources. The analysis has focused around investigating existing models and frameworks to determine which efforts have the most traction. This is an important input for the next steps of the work, which are to select and improve the models that will be used. This will require gap analysis to determine the limitations of the current models for representing concepts and data needed for performing the operations required by the scenarios captured and described in WP2. The models and API also have to be sufficient for addressing the integrated demonstrations that will be implemented by WP7.

Finally, the work in WP4 has been also concentrated on the development of innovative concepts for implementation of radio and networking functions. The proposed concept employs an abstraction methodology allowing seamless function portability and migration between network computing nodes. In addition, key functional blocks of LTE baseband system have been identified. These blocks have been then associated in a functional model. Moreover, an automated test and simulation framework has been developed. Preliminary results demonstrate the feasibility of proposed concept and methodology.

4.1.5 WP5 Virtualisation Platform Implementation and Network Dynamics

The goal of WP5 is to develop solutions for platform-agnostic programmability and configuration, taking into account the dynamics of virtual network functions. As first task, WP5 has started focusing on the implementation and the evaluation of function allocation algorithms. In the long term, this task will leverage the block and functional abstractions developed in WP4 to match a set of network services (decomposed into functional abstractions) to the underlying, available block hardware abstractions. Moreover, it will be possible to identify SLA (Service Level Agreement) metrics (e.g. throughput, delay, power consumption, etc.) and manage their mapping onto platform sources, etc.). However, solving the allocation problem that maps an SLA on a virtualised execution platform is a high-level endeavour. Before this problem can be tackled appropriately, the issues related to performance impact of traffic processing module placement within hardware components must be studied and understood. In order to face the aforementioned issues, we have first sought to characterise the performance of network traffic processing pipelines within standalone hardware processing unit. This characterisation leads to a low level component allocation framework called FastClick, which has been developed in the context of SUPERFLUIDITY. We have also studied FastClick's fitness for purpose through the development of a novel network cloud function called SplitBox, a privacy preserving filtering function for outsourced firewalls. Moreover, we have investigated the characterisation of a virtualised NFV infrastructure under different workloads. Finally, we have discussed how the proposed performance improvements on a bare-metal machine could translate into a virtualised setting.

As second task, WP5 has started to focus on the survey and the measurement of available virtualisation technologies, such as: type-1 hypervisors like Xen, type-2 hypervisors like KVM or lightweight virtualisation mechanisms like containers. Specifically, we aim to highlight the trade-offs between these solutions, and to select the most suitable ones to achieve the SUPERFLUIDITY goals of quick instantiation, migration and massive consolidation, to name a few. Moreover, this task, which will continue during the second year of the project, will further evaluate such mechanisms on different types of commodity hardware, from full-fledged x86 servers down to microservers. When the performance of light-weight containers and specialised Unikernels increases, in terms of reducing the instantiation and boot time up to the order of 10s of ms. or even less, the performance of Virtual Infrastructure Managers (VIMs), that control them may become a critical bottleneck. Therefore, we have started considering the VIM performance, by defining models for their characterisation and by identifying solutions for their improvements.

4.1.6 WP6 System Orchestration and Management Tools

This WP is devoted to the provisioning and control framework, including automated security verification of network processing code. We then report the main achievements in the following tasks: i) control framework design, ii) tools and algorithms for SLA based deployment, iii) modelling and design for symbolic execution and monitoring tools.

Design of Control Framework

In order to tackle the challenge of designing a control framework, our approach was split into several steps. As a first step we started by analysing the use cases from WP2 as our input. The objective was the identification of shared attributes and the identification of common requirements that the use cases shared. Such requirements come from NFV, MEC, C-RAN technologies. After doing so, we had the next step ready – investigation of the aforementioned requirements' support in existing orchestration solutions. As a last step, we need to identify the gaps between the requirements and each solution capabilities. The following table reports the options considered for each scope, as well as the preferred ones. Finally, part of the activities considered in this task has been devoted to the investigation of management software tools.

Table 5: Considered options for the control framework

Scope	Options	Comment
Cloud Infrastructure	One NFVI per Service	Inefficient and Complex
	Common NFVI for all Services eventually locations	Preferred
Cloud Infrastructure Management	One local VIM per NFVI	Acceptable
	Single centralised VIM for all NFVIs	Acceptable
	Hybrid solution between the previous two	Preferred
Cloud Management and Orchestration	One Orchestrator for all services and locations	Not realistic
	One Orchestrator per service	Preferred
Orchestration Layer	Northbound and Southbound Interfaces	Acceptable
	Eastbound and Westbound Interfaces	Difficult
	Hybrid solution between the previous two	Preferred

Initial Design for SLA based deployment (I6.2)

We have first presented NEMO – a language to be used as SLA descriptor, for an application owner to declare the QoS requirements of her application. NEMO is a human-readable command language used for Network Modelling. NEMO provides basic network commands (Node, Link, Flow, Policy) to describe the infrastructure and controller communication commands to interact with the controller. We are proposing to extend the Node definition command to import TOSCA or OSM based descriptors as Node definitions. Since Node models can make use of previously defined node models, the resulting language would be recursive and therefore support our notion of (recursive) reusable function blocks.

Then we considered resource allocation and placement problems in different contexts: i) service chains in NFV deployments, ii) MEC applications, iii) generic formulation of a service scaling problem in NFV environment.

For service chains, we study two possible placement strategies – one that gathers all components of each chain into the same host, and one that distributes them between different hosts. This is performed both in the presence of hardware acceleration (DPDK), and without it. Following the comprehensive evaluation of the two placement strategies of service chains, we also model the cost of network switching. Given an arbitrary number of service chains, our model accurately predicts the CPU cost of the network switching they require.

In the context of MEC applications, we discuss the management and orchestrator and the resource allocation problem. We also examine service migration for MEC applications.

As for the generic service scaling problem, we focus on the scale in/out operations in VNF allocation. In particular, we model the decision whether to increase the resources allocated to a certain VNF type (scale out operation) or to release some of these resources (scale in operation) based on the current and expected demand from the network service and the required QoS as a Markov Decision Process (MDP). Our formulation also incorporates the load balancing challenge steering the traffic flows to the different VMs and balancing the load between them.

Modelling and Design for Symbolic Execution and Monitoring Tools

As reported in the WP3 description, SUPERFLUIDITY proposes the concept of RFBs as a modelling tool that supports APIs for 5G networking. RFBs have a specification described in a higher level language that allows their users to compose them correctly, and can be implemented in multiple ways as long as they obey their specification: software (as a monolithic block), hardware (ASICs), or a composition of other RFBs. How can we secure the resulting 5G networks? Enforcing network security has two major phases: the network operator specifies higher-level policies and then implements them using (low-level) networking functionality typically provided by third-party vendors. High-level policies could include access control lists (who can talk to whom), firewall rules, routing protocol configurations, and so forth. Networking functionality includes switches, routers, middleboxes, etc. In traditional networks, high-level policies are fairly static and thus easy to manually check and deploy infrequently. Traditional networking hardware (switches, routers, simple firewalls) processes packets on custom-made ASICs that are thoroughly verified and seldom updated; such implementations give a fairly strong low-level security guarantee. In 5G networks, network functions will be instantiated dynamically, and the network will run services configured not only by the operator, but also by third parties. Running third-party processing will be a major revenue source for operators and is thus very attractive; however it can expose operator networks to many security risks that must be addressed. The great benefit of software network functions is that it can be easily upgraded and it avoids vendor lock-in. On the downside, diverse software with increasingly complex functionality and developed by many vendors is much more susceptible to low level exploits such as buffer overflows.

To enforce security in 5G networks SUPERFLUIDITY takes the following main directions:

1. Describe operator policies in a high-level specification language.
2. Describe RFBs in a way that is amenable to static analysis.
3. Perform static analysis of RFB configurations to ensure policy is obeyed before deployment.
4. Ensure that the implementation conforms to the specification at deployment time.

Specifically, there is a gap between the abstract model of the network that we can statically analyse (steps above) and the actual implementation. This gap appears because it is impossible to verify large C implementations of networking code in useful time.

Even if we assume the implementations of networking functionality are correct, similar gap appears when multiple abstractions are applied in a network, for instance network virtualisation.

To bridge this gap, we have relied on multiple techniques, some of which are known and tested but have limited coverage, and some which are novel and are still subject of on-going work. We list them here:

- Active testing. Intuitively, once the static analysis results are known, packets are generated for each path resulting from static analysis and injected in the real network; the outcome is then checked to see if it matches that predicted by static analysis. As all other active techniques, this is lightweight and useful, but it is not sufficient on its own because it has poor coverage and cannot give any type of strong guarantees.
- Monitoring and anomaly detection. This is another runtime technique that applies machine learning to understand the standard behaviour of the network and detect attacks when the behaviour deviates from standard.
- Static analysis at lower layers. Symbolic execution can be run on the model, and on the lower level implementation of the model; if the resulting outputs are equivalent, then the implementation is correct. Defining equivalence is not easy, and different definitions capture different parts of the problem we want to solve.
- Automatic generation of C implementations from SEFL code, which is memory-safe and easy to symbolically execute, thus allowing to prove it satisfies the high level properties of the operator.

4.1.7 WP7 System Integration and Validation

The System Integration and Validation WP (WP7) is not formally started, yet; however, two testbeds have been established (ahead of schedule). These will allow various aspects developed in the other WPs to be explored practically (see section 4.2).

4.1.8 WP8 Communication, Dissemination, Standardisation and Exploitation

The following objectives have been pursued by the project and supported by WP8: i) open source contributions, ii) initial proof of concepts, iii) contributions to standardisation organisations related to 5G, iv) organisation of special sessions at events like OpenStack Summits and Xen Hackathons, v) "Marketing-oriented" presentations at industry events, vi) Discussions with policy makers, social and environmental organisations regarding project results, vii) Cooperation with other peer 5G-PPP projects in the framework of the 5G-PPP overall programme, viii) Scientific publications at top tier conferences, journals and workshops, ix) Wiki-style website allowing easy retrieval of main project data (including software repositories), as well as social networks presence, x) Press releases targeting the public at large to allow for wide dissemination of project concepts and results.

Moreover, an innovation and management plan has been pursued, including the following activities:

- Overall project responsibility for identifying and driving the innovation and commercial potential for the technical work and goals of the project.
- Ensure of a common approach for Innovation and Commercial Exploitation in all work packages while maintaining the overall vision of the Project.

- Identification of potential innovation and commercial exploitation inhibitors in and between the work packages while implementing the technical work.
- Preparation of proposals for the Management Board on innovation concepts. Preparation of summaries for the periodic reports and annual reports, and for the reviews.
- Delivering of relevant presentations both internally and externally to the industry.

The deliverables D8.2 and D8.3 document this work. Their contents is not reported here, as they are already written in the form of a report and are public.

4.2 Current achievements: demonstrations

In this section, we report the work performed as regards demonstration activities.

The project plan was to start integration activity in month 16 (October 2016). However, the consortium decided to anticipate the integration work, to show a first practical demonstration of project achievements at the next review, planned for September 2016.

Thus, partners developed sub-components and are now in the process of integrating them, with some integrated functionality already available as of today.

It is also worthy of note that the project decided to design and implement another test-bed, in addition to those envisaged in the Description of Work. As a matter of fact in the DoW we planned to have: i) stand-alone experiments deployed at some partner's premises for dedicated assessment work, involving validation of subsets of functions, scalability and stress tests, etc.; ii) an integrated trial to be deployed in France, at Nokia premises, allowing consortium-wide remote access for integration purposes and for the execution of experiments. During the project we decided to implement another integrated test-bed consisting of high-performance servers and related operating software, located at BT premises and also accessible by all partners; its cost has been shared among partners on a pro quota basis, as agreed by the project General Assembly. In the following, the two integrated test-beds will be referred to as test-bed 1 (@Nokia) and test-bed 2 (@BT).

Test-bed 1:

The testbed is located at Nokia France premises in Villarceaux close to Paris (France). It is a hardware and wireless platform allowing to demonstrate some innovations (e.g. Cloud RAN, RFB decomposition,...) conducted in the project. First of all, a secure remote access will allow the project partners to upload, compile and if possible test, their own software contributions. Following the security policy of Nokia, the connection to the testbed will be realised using OpenVPN. According to the architecture and design work performed in other work packages, especially in WP3, we already have a preliminary idea of the SUPERFLUIDITY testbed initial setting. This setting will evolve in the second year to accommodate all the requirements from the early implementation. The testbed is composed by 4 servers, 3 Ethernet switches (Omniswitch 6860(E)), two USRP B210 cards, and a laptop acting as RRH gateway to convert CPRI connection coming from the Remote Radio Head to Ethernet. The wireless connection will be ensured through the use of LTE dongles. Another possibility offered in the testbed is the over the air transmission using very low transmission power in order to not interfere with commercial deployment. Two servers will play the role of a compute node. For that reason, we selected a high performance configuration (4CPU, core i7). Two different Linux distributions (Ubuntu, Fedora) are installed in order to have a heterogeneous environment. The two others servers will play the role of orchestrator and controller nodes. In addition, the testbed integrates different Ethernet switches to support SUPERFLUIDITY work on the fronthaul network, which is mainly conducted by Nokia Fr and Eblink. The fronthaul network is between the Radio site and the edge cloud which is represented in the testbed as a compute node 2. The fronthaul network will integrate Eblink development for next steps of the project. Figure 8 shows a picture of the planned initial setting of the testbed #1.

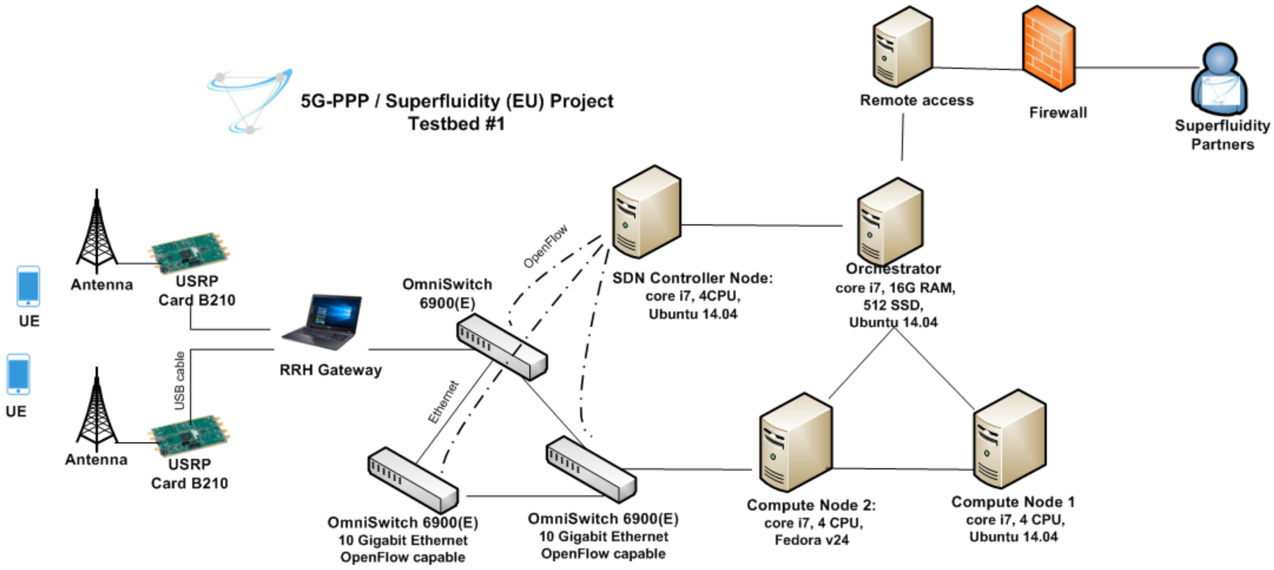


Figure 8: Initial setting for the SUPERFLUIDITY testbed #1

Test-bed 2:

Test-bed 2 is a hardware platform that allows flexible virtualisation experiments and demonstrations to be run by all partners. It consists of 5 Cisco servers (UCSC-C220-M4S) with Intel X540 Dual Port 10GBase-T Adapters plus a Nexus switch (Nexus 3172-T). Four of the servers are installed with CentOS7 (potentially later this will be changed to RHEL7) and one with Ubuntu 14.04, in order to arrange different experiments and for example the installation of Citrix Hammer to generate traffic. Partners can access the testbed, which is on BT premises, by using Open VPN. The diagram below shows the set-up (Figure 9):

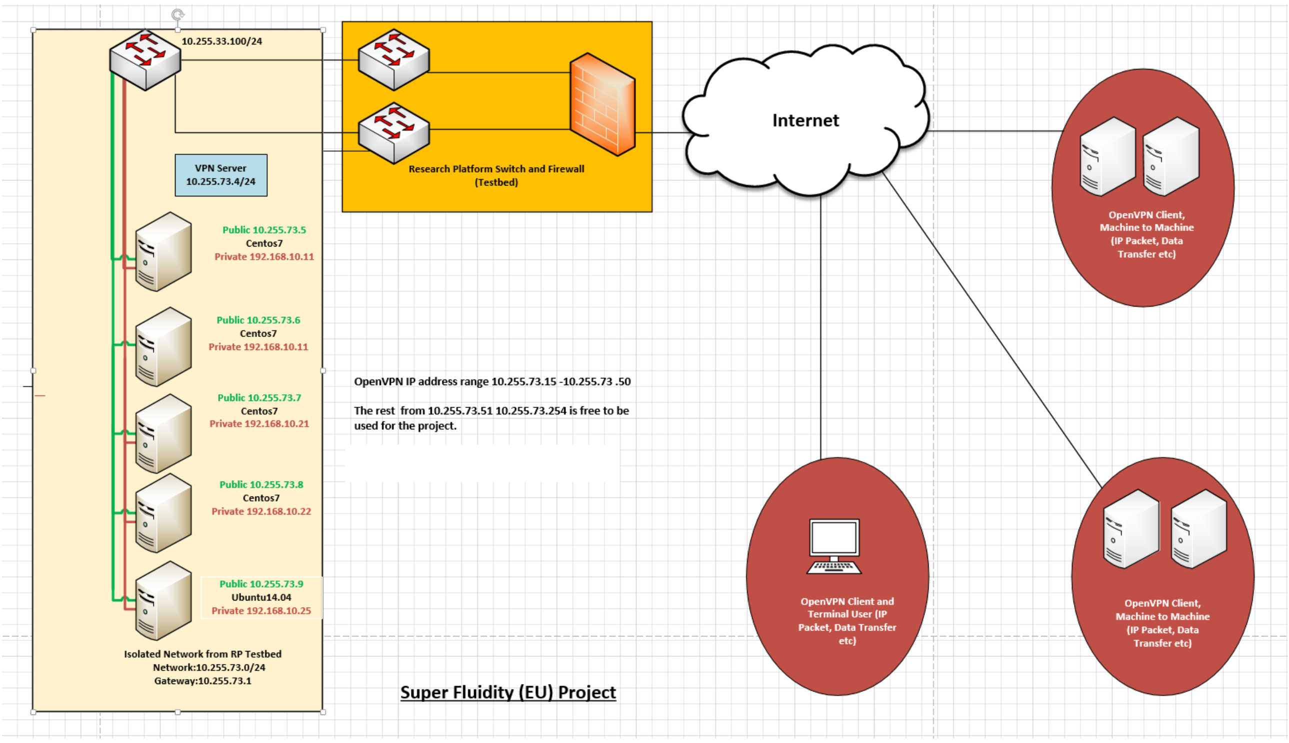


Figure 9: Initial setting for the SUPERFLUIDITY testbed #2

As of now the project decided to prepare and show 4 different demonstrations, briefly described below:

4.2.1 Demo 1: Orchestration

Demo coordinator: OnApp

Other involved partners: Nokia IL, CITRIX, Unified Streaming, Intel, Redhat

Description: The Orchestration demo intends to show the improved functionality of a heterogeneous video delivery platform having specific architectural characteristics, including resource re-allocation based on service monitoring, intelligent node deployment and advanced orchestration capabilities. Such features render the system capable of reacting in a real-time manner to changes in networking conditions, increased number of simultaneous user requests and end-user quality of experience deterioration.

To demonstrate such a diverse set of attributes incorporating all the aforementioned characteristics, we integrated the following components into a generic platform: (i) OpenStack with emphasis on the Controller, Neutron and additional Computing Nodes, deployed in different physical nodes, (ii) Unified Origin's Video Streaming Platform, a commercial platform with extended video protocol support (i.e., ABR) and advanced features such as late transmuxing (iii) Hammer Stress Test Tool, an end-to-end network traffic simulator, capable of simulating complex and dynamic network, user and server behaviours. Its intended role in the particular demonstration is to simulate increased network traffic conditions. (iv) Intel Telemetry Solutions including Cimmaron and/or SNAP agents for capturing telemetry data from the necessary interconnected compute nodes.

Demo steps description:

1. Original Setup is presented and explained.
2. The demonstration begins when the subscriber (Bob) makes a request using his equipment for a specific video stream.
3. The system receives the request and after the necessary analysis (i.e. device type, network conditions, node availability), allocates Bob's streaming session to be handled via Compute Node 1.
4. The underlying telemetry platform starts collecting data, corresponding to limited amounts of traffic, since Bob is currently the only user making service requests to Compute Node 1.
5. At some point, Hammer Stress Testing Tool is initiated, after loading the necessary profile which simulates a substantial increase to the number of users trying to connect to Compute Node 1 and make video streaming requests.
6. The telemetry platform collects data consisted to the now compromised service interconnected users are receiving, due to network congestion.
7. The orchestration entity deploys Compute Node 2 to alleviate network congestion and improve quality of experience for end users.
8. Since Bob's quality of service is now inferior to the original one, the orchestrator seamlessly migrates his session to the newly deployed Compute Node 2.
9. Once Bob's migration is concluded, the telemetry platform collects data from Compute Node 2 that verify the elevated levels of service quality he enjoys.

Required HW/SW: Server 1: OpenStack Controller / OpenStack Neutron (no need for separate node) Server 2: Compute Node 1 Server 3: Compute Node 2 Server 4: Hammer Node Switch The hardware configuration of each node is identical to the one described in the BT Lab Bill of Materials. No need for additional/specialised hardware whatsoever, unless a VM including the Front end is deployed separately onto a demo laptop. Ubuntu 14.04 LTS/CentOS 7 as the underlying operating systems for all nodes, Hammer installation ISO, (additional software packages to be added here by each partner ...)

Testbed location to run: test-bed 2 at BT Lab

4.2.2 Demo 2: Software Defined Superfluid Wireless Network Demo

Demo coordinator: NOKIA-FR

Other involved partners: E-Blink

Description: The demo intends to demonstrate a first iteration of cloud RAN PoC named software defined superfluid wireless network. It consists in instantiating an end-to-end network including the core and the RAN using a pure software implementation component and establishes a YouTube connection to run a video streaming service. The core will be composed of four separate RFBs components (MME, PGW, SGW and HSS) deployed as containers. The RAN will be instantiated using *OpenAirInterface* software as a soft modem running also as a container. Between the remote radio head and the edge cloud hosting the base band unit container, we have a fronthaul network composed by three switches which are controlled using an SDN controller in order to select the right path fulfilling the service requirements. This capability is shown in the demo through a dynamic selection of the path between RRH and EDGE cloud.

Demo steps description:

Via a GUI, the service chain of an end to end mobile network is set-up and deployed. After validating the service chain, different steps are executed:

- SDN controller instantiation
- RFB of the EPC instantiation
- BBU instantiation
- Fronthaul registration
- BBU registration
- Selection of a first path for RRH-BBU flow
- UE connection
- New routing path selection for RRH-BBU flow

Required HW/SW: 1 USRP BOARD B210 (Ettus), 4 Laptops Toshiba core i7 (SSD+16Gb), 1 Bandluxe 4G dongle, 1 SMA cable set and LP0410 Antenna (Ettus), 1 Laptop Lenovo core i3 or 4G UE, 1 10 Gb/s OpenFlow Switch (OmniSwitch)

Testbed location to run: Test-bed 1 at Villarceaux, NOKIA-FR

4.2.3 Demo 3: Verification of OpenStack

Demo coordinator: UPB

Other involved partners: Redhat, Nokia IL.

Description: The demo will show a tenant network API in the Horizon GUI in OpenStack and then it will invoke Symnet to do static analysis.

Demo steps description:

- A sample network configuration is shown and explained.
- The SEFL model generator is invoked and Symnet is run on the resulting model.
- The output of Symnet is shown to receivers.
- Possibly change the network configuration (e.g. modify a firewall rule) and goto step 2.

Required HW/SW: Single server with access to the BT OpenStack deployment (need Neutron API). Demo will run remotely, X forwarding needed to show Horizon GUI. We will also create a video of the demo in case there are glitches.

Testbed location to run: test-bed 2 at BT Lab.

4.2.4 Demo 4: Mobile Edge Computing (MEC)

Demo coordinator: ALB

Other involved partners: Unified Streaming

Description: The Mobile Edge Computing (MEC) demo intends to show the basic functionality of an edge computing system, as described by the ETSI MEC ISG. In this scenario, the subscriber is accessing a video streaming service, which is provided from the core. However, in order to improve the service quality and save network resources, the MEC system decides to deploy the service in the edge where the subscriber is located, starting to provide the service from that point. From this moment on, the video stream is transparently delivered from the edge application, without any impact or change on the subscriber (it is transparent). On the other hand, the video streaming server on the edge is exactly the same as in the core, and does not need to deal with the complexity of migration and traffic offloading and redirection. To demonstrate this MEC functionality, we use the OpenEPC software as a testbed in order to emulate the complete mobile network. As Application, we are using the LTM video streaming service, provided by Unified Streaming. The MEC solution is a development that is been performed by the SUPERFLUIDITY project from scratch. The OpenEPC comprises a simulated RAN and a full Core network. The OpenEPC was slightly changed in order to be able to introduce a MEC component (the TOF - Traffic OFfloading) in the data plane, between the eNB and the Core, in order to intercept the traffic on the S1-U interface. Then, the GTP-U packets are decapsulated, striping the subscriber IP packet and, redirecting them to the right edge MEC Application, or forwarding it to the Core, as usual, depending on the provisioned offloading rules.

Demo steps description:

- The subscriber (Bob) uses a video application (e.g. browser) to search for its favourite video, and starts streaming it.
- The video server in the core (App) is contacted, since service FQDN is DNS resolved to that global IP address.
- At some point during the video streaming, the operator decides to start providing the video service from a local edge server, e.g. because there are many other subscribers using the service on that edge that justifies that option.
- Management and Orchestration (MANO) layers instantiate the Application in the edge (MEC App). Note: So far, MANO layers are not fully implemented yet; they will be simulated by scripts.
- After the MEC App instantiation is fully operational, MANO layers provision the TOF in order to start offloading the traffic that regularly go to the core App (global IP) to the local edge server (MEC App) using the local IP.
- From this moment on, the subscriber will start getting the video service from the edge (MEC App). Note 1: As this service is stateless, there is no need to migrate any state between core and edge (MEC) Apps. Note 2: From the subscriber perspective, the migration process is transparent; he is completely unaware.
- At some point during the video streaming from the edge, the operator decides to start providing the video service back from the core, e.g. because there are now a few other subscribers using the service and does not justifies the edge option (too expensive).
- Before the MEC App termination is started, MANO layers provisions the TOF in order to stop offloading the traffic to the edge MEC App (local IP), letting them flow, as regularly, to the core server (App) using the global IP.
- From this moment on, the subscriber will start getting the video service from the core again (App). Note 1: As this service is stateless, there is no need to migrate any state between

edge and core MEC) App. Note 2: From the subscriber perspective, the migration process is again transparent; he is completely unaware.

- After TOF reconfiguration is completed, MANO layers can terminate the local server in the edge (MEC App).

Required HW/SW: No special HW requirements / Ubuntu

Testbed location to run: ALB Cloud Infrastructure

4.3 Scientific and Technological Impact

At the time of writing is rather early to assess SUPERFLUIDITY's impact; thus, this section and the two following ones will be suitably expanded in the next version of this document. However, the work performed during the first year produced some visible and tangible results.

As regards scientific publications, SUPERFLUIDITY is targeting top tier research conferences and journals. These leading venues shape future research in this area and as such provide maximum visibility for the project results. Several partners have an established presence and a strong track record in such venues. The current list of publications is reported in http://superfluidity.eu/results/dissemination/.

Another important activity to exert impact is of course standards. Standards are a key way of ensuring that SUPERFLUIDITY has a real impact, in terms of actually achieving the "softening" of the network. The standards process distils the best technical approach, as well as helping to ensure interoperability and that operators can source components from different / multiple vendors. With critical input from SUPERFLUIDITY people, ETSI have just established a new work item on 'end to end process descriptions' (this builds on some of our ideas about recursion and common interfaces - work is just starting), and there was a meeting between many SDOs and industry groups in order to discuss their various Information Models (to start converging them and understanding how far convergence needs to go). Several SUPERFLUIDITY people play leading roles in standardisation activities: at the ETSI Industry Steering Group on NFV, BT's Andy Reid is Vice Chair of Open Source MANO (OSM), Chair of the End User Advisory Group and Rapporteur for the new Work Item on 'end to end process descriptions'.

4.4 Market Impact

At the time of writing, there are no products that have been developed that benefit from SUPERFLUIDITY. Based on events and discussions there would appear to be a high interest in the potential outcomes of SUPERFLUIDITY. Most of the demonstrations are tied in with expected commercial needs of the partners involved in those demonstrations and as such it is expected that in the revision of this document there will be a significant change in this section. In the research field, the papers that have been submitted have had an impact in their relevant areas, with some of the challenges posed providing new areas for investigation.

4.5 Societal Impact

At the time of writing it is too early to provide measurable results on the societal impact other than in the areas of skills related to 5G. Through research, having active participation in the 5G whitepaper and related 5G phase I discussions, the partners involved in SUPERFLUIDITY have been involved at the forefront of 5G technology discussions.

5 Summary

This document summarizes the plans of SUPERFLUIDITY and the work performed so far, providing in a single document a bird's eye view of the project. The second version of this deliverable, due at the end of the project at month 30, will of course provide a much more complete and clear picture of our achievements.

6 References

[8021Q] IEEE Std. 802.1Qbg-2012, "IEEE Standard for Local and Metropolitan Area Networks — Media Access Control (MAC) Bridges and Virtual Bridged Local Area Networks— Amendment 21: Edge Virtual Bridging," July 5, 2012, http://standards.ieee.org/getieee802/download/802.1Qbg-2012.pdf, p. 191.

[ALIEN] ALIEN "Hardware Abstraction Layer (HAL)" white paper, available on http://www.fp7-alien.eu/files/deliverables/ALIEN-HAL-whitepaper.pdf

[ARN2010] Applied Technology and Innovation Management, Heinrich Arnold, Michael Erner, Peter Möckel, Christopher Schläffer Editors, Springer Heidelberg Dordrecht London New York

[BAN12] M. Bansal, J. Mehlman, S. Katti, and P. Levis, "OpenRadio: A Programmable Wireless Dataplane," in Proc. of HotSDN'12, Aug. 2012.

[BEL2010] Beloglazov, Anton, and Rajkumar Buyya. "Energy efficient resource management in virtualised cloud data centers." 10th IEEE/ACM Int. Conf. on Cluster, Cloud and Grid Computing. IEEE Computer Society, 2010.

[BHA2013] Bhatnagar, Vaibhav, et al. "An FPGA Software Defined Radio Platform with a High-Level Synthesis Design Flow." Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th. IEEE, 2013.

[BHAN10] G. Bhanage, I. Seskar, R. Mahindra, and D. Raychaudhuri, "Virtual Basestation: Architecture for an Open Shared WiMAX Framework," in Proc. of VISA'10, 2010.

[BIA12] G. Bianchi, P. Gallo, D. Garlisi, F. Giuliano, F. Gringoli, I. Tinnirello, "MAClets: Active.

[BIA14] Giuseppe Bianchi, Marco Bonola, Antonio Capone, Carmelo Cascone: OpenState: programming platform-independent stateful openflow applications inside the switch. ACM Comp. Commun. Rev. 44(2): 44-51 (2014)

[BiaArx14] Bianchi, G., Bonola, M., Picierro, G., Pontarelli, S., & Monaci, M. (2013). StreaMon: a data-plane programming abstraction for Software-defined Stream Monitoring. arXiv preprint arXiv:1311.2442.

[BOE1988] Boehm B, "A Spiral Model of Software Development and Enhancement", IEEE Computer, IEEE, 21(5):61-72, May 1988.

[BOS13] P. Bosshart, G. Gibb, H.-S. Kim, G. Varghese, N. McKeown, M. Izzard, F. Mujica, and M. Horowitz, “Forwarding metamorphosis: Fast programmable match-action processing in hardware for sdn,” in ACM SIGCOMM 2013. ACM, 2013, pp. 99–110.

[BOT2012] Botero, Juan Felipe, et al. "Energy efficient virtual network embedding." Communications Letters, IEEE 16.5 (2012): 756-759.

[CELAR] Celar Cloud. http://www.celarcloud.eu

[CHANGE] EU CHANGE project. http://www.change-project.eu

[CIS2012] Cisco Cloud Services Router 1000v Data Sheet. http://www.cisco.com/en/US/prod/collateral/routers/ps12558/ ps12559/data_sheet_c78-705395.html, July 2012.

[CLA2012] Andrew Clappison, “Dissemination, Communication and the social life of research impact”, http://www.researchtoaction.org/2012/01/dissemination-communication-and-the-social-life-of-research-impact/

[CN] CumuloNimbo: Highly Scalable Transactional Multi-Tier PaaS. http://www.cumulonimbo.eu

[CONTR] Contrail project. http://contrail-project.eu

[CP] CoherentPaaS Website. http://coherentpaas.eu

[Cra13] B. Mack-Crane. “OpenFlow Extensions”. In: US Ignite ONF GENI workshop, October 8, 2013.

[CROWD] “Connectivity management for eneRgy Optimised Wireless Dense networks (CROWD),” http://www.ict-crowd.eu/publications.html

[CSA2013] Császár, András, Wolfgang John, Mario Kind, Catalin Meirosu, Gergely Pongrácz, Dimitri Staessens, Attila Takács, and J. Westphal. "Unifying Cloud and Carrier Network." Proceedings of. DCC, Dresden, Germany, to appear Dec(2013).

[CSTK] Apache Cloudstack. Open Source Cloud Computing. http://cloudstack.apache.org

[DIWINE] http://diwine-project.eu/public/

[DON2012] Dong, Yaozu, et al. "High performance network virtualisation with SR-IOV." Journal of Parallel and Distributed Computing 72.11 (2012): 1471-1480.

[DPDK] Intel. Intel DPDK: Data Plane Development Kit http://dpdk.org, September 2013

[DT2012] Deutsche Telekom tests TeraStream, the network of the future, in Croatia. http://www.telekom.com/media/company/168008

[DUN09] R. Duncan and P. Jungck, “PacketC: Language for High-Performance Packet Processing”, IEEE HPCC'09, 2009.

[FAWG] Openflow Forwarding Abstractions Working Group Charter. http: //goo.gl/TtLtw0. Apr. 2013.

[FLA] FLAVIA project. Flexible Architecture for Virtualizable Future Wireless Internet Access. http://www.ict-flavia.eu

[FOS11] Foster, Nate, et al. "Frenetic: A network programming language." ACM SIGPLAN Notices. Vol. 46. No. 9. ACM, 2011.

[FRA2013] Pushing cdn-isp collaboration to the limit. Benjamin Frank, Ingmar Poese, Yin Lin, Georgios Smaragdakis, Anja Feldmann, Bruce Maggs, Jannis Rake, Steve Uhlig, and Rick Weber. SIGCOMM Comput. Commun. Rev., 43(3): 34–44, July 2013.

[GUD13] A. Gudipati, D. Perry, L. E. Li, and S. Katti, “SoftRAN: software defined radio access network,” in ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking (HotSDN’13), ACM SIGCOMM 2013, pp. 25–30.

[GUD2013] Gudipati, Aditya, et al. "SoftRAN: Software defined radio access network." Proceedings of the second ACM SIGCOMM workshop on Hot topics in software defined networking. ACM, 2013.

[Han10] S. Han, K. Jang, K. Park, and S. Moon. "Packetshader: a gpu-accelerated software router", In Proc. of ACM SIGCOMM 2010, pp. 195–206, New York, NY, USA, 2010.

[HAR] Hardware and Network-Enhanced Software Systems for Cloud Computing. http://www.harness-project.eu

[HOLLY] Global Online TV, Video Revenue Forecast to Hit $51.1B by 2020 http://www.hollywoodreporter.com/news/global-ott-revenue-forecast-netflix-802417

[IEEECTN] "On the Road to 5G", IEEE CTN, http://ctn.comsoc.org

[JAI13] R. Jain and S. Paul: "Network Virtualisation and Software Defined Networking for Cloud Computing: A Survey", IEEE Communications Magazine, Nov 2013, pp. 24-31

[JAM2013] Jamakovic, Almerima, Thomas Michael Bohnert, and Georgios Karagiannis. "Mobile Cloud Networking: Mobile Network, Compute, and Storage as One Service On-Demand." The Future Internet. Springer Berlin Heidelberg, 2013. 356-358.

[JAN15] S. Jankowski, B. Feldman, Duval, D. Takyama, T. Hari, D. Lu, I. Matsuhashi, M. Shin, D. Clark, M. Delaney and In Y. Chung, 5G: How 100x faster wireless can the future, Goldman Sachs Equity Research, April, 2016

[JEY13] V. Jeyakumar, M. Alizadeh, C. Kim, and D. Mazieres, "Tiny packet programs for low-latency network control and monitoring," in ACM Workshop on Hot Topics in Networks (HOTNETS 2013), 2013.

[Koh00] E. Kohler, R. Morris, B. Chen, J. Jannotti, and M. F. Kaashoek. The click modular router. ACM Trans. Comput. Syst., 18:263–297, August 2000.

[KOK10] R. Kokku, R. Mahindra, H. Zhang, and S. Rangarajan, "NVS: a virtualisation substrate for WiMAX networks," in Proc. of MobiCom'10, 2010.

[KVM] KVM project homepage http://www.linux-kvm.org/page/Main_Page

[LINC] LINC switch project homepage https://github.com/FlowForwarding/LINC-Switch

[MAD2013] Unikernels: Library operating systems for the cloud. A. Madhavapeddy, R. Mortier, C. Rotsos, D. Scott, B. Singh, T. Gazagnaire, S. Smith, S. Hand, and J. Crowcroft. SIGPLAN Not., 48(4):461–472, Mar. 2013.

[MAD2014] A. Madhavapeddy and D. Scott, "Unikernels: the Rise of the Virtual Library Operating System", ACM CACM, 2014.

[MAE11] Maeder, A., Mancuso, V., Weizman, Y., Biton, E., Rost, P., Perez-Costa, X., & Gurewitz, O., "FLAVIA: Towards a generic MAC for 4G mobile cellular networks", Future Network & Mobile Summit, 2011

[MAI2012] S. Maital, D. V. R. Seshadri, Innovation Management, Response Books

[MAR2014] ClickOS and the Art of Network Function Virtualisation. J. Martins, M. Ahmed, C. Raiciu, V. Olteanu, M. Honda, R. Bifulco, F. Huici. NSDI 2014.

[MEC] "Mobile Edge Computing – Introductory Technical White Paper". https://portal.etsi.org/Portals/0/TBpages/MEC/Docs/Mobile-edge_Computing_-_Introductory_Technical_White_Paper_V1%2018-09-14.pdf

[MICRODC] Benjamin Frank, Ingmar Poese, Yin Lin, Georgios Smaragdakis, Anja Feldmann, Bruce Maggs, Jannis Rake, Steve Uhlig, and Rick Weber. Pushing cdn-isp collaboration to the limit. SIGCOMM Comput. Commun. Rev., 43(3):34–44, July 2013.

[MON13] Monsanto, Christopher, et al. "Composing Software Defined Networks." NSDI. 2013.

[MOR99] Robert Morris, Eddie Kohler, John Jannotti, and M. Frans Kaashoek. 1999. The Click modular router. SIGOPS Oper. Syst. Rev. 33, 5 (December 1999), 217-231. DOI=10.1145/319344.319166 http://doi.acm.org/10.1145/319344.319166

[MOS] mOSAIC. cordis.europa.eu/fp7/ict/ssai/docs/call5-mosaic.pdf

[MOS14] M. Moshref, A. Bhargava, A. Gupta, M. Yu, and R. Govindan. Flow-level State Transition as a New Switch Primitive for SDN. In Proceedings of the ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking (HotSDN'14), August 2014.

[MUE2010] Mueck, Markus, et al. "ETSI reconfigurable radio systems: status and future directions on software defined radio and cognitive radio standards." Communications Magazine, IEEE 48.9 (2010): 78-86.

[NFV007] ETSI GS, "Network Functions Virtualisation (NFV); Infrastructure; Methodology to describe Interfaces and Abstractions"; NFV-INF 007 V1.1.1 (2014-10)

[NFV2014] ETSI Network Functions Virtualisation, http://www.etsi.org/technologies-clusters/technologies/nfv

[NVGRE] M. Sridharan et al., "NVGRE: Network Virtualisation Using Generic Routing Encapsulation," IETF Draft- sridharan-virtualisation-nvgre-03.txt, Aug. 2013

[NVO] Narten et al., "Problem Statement: Overlays for Network Virtualisation," IETF Draft draft-ietf-nvo3-overlay-problem-statement-04, July 31, 2013

[ODLT] OpenDaylight. A Linux Foundation Collaborative Project. http://www.opendaylight.org

[ODPL] OpenDataPlane. http://www.opendataplane.org

[OF1.4] Open Networking Foundation "Openflow specification, ver 1.4", October 2013

[OFRef] OFN reference implementations http://archive.openflow.org/wp/tag/reference-implementation/

[OFSW] ofsoftswitch13 GitHub page https://github.com/CPqD/ofsoftswitch13

[OSTK] OpenStack. Open Source Cloud Computing Software. https://www.openstack.org

[OVS] OpenvSwitch project home page http://vswitch.org/

[Pantou] Pantou project homepage http://archive.openflow.org/wk/index.php/OpenFlow_1.0_for_OpenWRT

[PDNA] L. Deri. http://www.ntop.org/products/pf_ring/dna/.

[PEN13] K. Pentikousis, Y. Wang, and W. Hu, "Mobileflow: Toward software defined mobile networks," Communications Magazine, IEEE, vol. 51, no. 7, pp. 44–53, 2013.

[Per13] P. Peresini, M. Kuzniar, and D. Kostic. "OpenFlow Needs You! A Call for a Discussion About a Cleaner OpenFlowAPI". In Proc. of the EU Workshop on Software Defined Network (EWSDN). 2013.

[PICSIG] PCI-SIG, "Single Root I/O Virtualisation and Sharing 1.1 Specification," http://www.pcisig.com/members/down-loads/specifications/iov/sr-iov1_1_20Jan10.pdf, available only to members.

[RAG2012] Software-defined internet architecture: decoupling architecture from infrastructure. Raghavan, Barath and Casado, Martin and Koponen, Teemu and Ratnasamy, Sylvia and Ghodsi, Ali, Shenker, Scott. s.l. : Hotnets, 2012.

[RAM2009] Ram, Kaushik Kumar, et al. "Achieving 10 Gb/s using safe and transparent network interface virtualisation." Proc. 2009 ACM SIGPLAN/SIGOPS international conference on Virtual execution environments. ACM, 2009.

[RFC6325] R. Perlman et al., “Routing Bridges (RBridges): Base Protocol Specification,” IEEE RFC 6325, July 2011, 99 pages, http://tools.ietf.org/html/

[Riz12] L. Rizzo, “Netmap: a novel framework for fast packet i/o. In Proceedings of the 2012 USENIX conference on Annual Technical Conference”, USENIX ATC’12, USENIX Association, Berkeley, CA, USA, 2012.

[Riz12a] L. Rizzo, M. Carbone, G. Catalli, “Transparent acceleration of software packet forwarding using netmap”, in Proceedings of IEEE INFOCOM’12, 2012.

[SACH08] J. Sachs and S. Baucke, “Virtual radio: a framework for configurable radio networks,” in Proc. of WICON’08, Nov. 2008.

[SEK2012] Design and implementation of a consolidated middlebox architecture. V. Sekar, N. Egi, S. Ratnasamy, M. K. Reiter, and G. Shi. 9th USENIX conf. on Networked Systems Design and Implementation, NSDI 2012.

[SHE2012] A Survey of Enterprise Middlebox Deployments. J. Sherry and S. Ratnasamy. UC Berkeley, Department of Electrical Engineering and Computer Sciences, Technical Report No. UCB/EECS-2012-24

[SHE2012b] J. Sherry, S. Hasan, C. Scott, A. Krishnamurthy, S. Ratnasamy, V. Sekar, “Making middleboxes someone else's problem: network processing as a cloud service”, SIGCOMM, 2012.

[SMART] Smart Cells Revolutionize Service Delivery, Intel white paper. http://www.intel.co.uk/content/dam/www/public/us/en/documents/white-papers/smart-cells-revolutionize-service-delivery.pdf

[SON13] H. Song, “Protocol-oblivious forwarding: Unleash the power of sdn through a future-proof forwarding plane,” 2nd ACM SIGCOMM HotSDN Workshop, 2013, pp. 127–132.

[STT] B. Davie, Ed., J. Gross, “A Stateless Transport Tunneling Protocol for Network Virtualisation (STT),” IETF Draft draft-davie-stt-03.txt, Mar. 12, 2013, 19 pages, http://tools.ietf.org/html/draft-davie-stt-03

[T2] Trilogy2. Building the Liquid Net. http://trilogy2.it.uc3m.es

[TAN11] K. Tan et al., “Sora: High-Performance Software Radio Using General-Purpose Multi-Core Processors,” Communications of the ACM, vol. 54, pp. 99–107, Jan. 2011.

[TERA] Deutsche Telekom tests TeraStream, the network of the future, in Croatia. http://www.telekom.com/media/company/168008

[TIN12] I. Tinnirello, G. Bianchi, P. Gallo, D. Garlisi, F. Giuliano, F. Gringoli, “Wireless MAC Processors: Programming MAC Protocols on Commodity Hardware”, IEEE INFOCOM, March 2012.

[TNOVA] T-NOVA project. Network Functions As-a-Service over Virtualised Infrastructures. http://www.t-nova.eu

[VXLAN] M. Mahalingam et al. “VXLAN: A Framework for Over-laying Virtualised Layer 2 Networks over Layer 3 Net-works,” IETF Draft draft-mahalingam-dutt-dcops-vxlan-04.txt, May 8, 2013.

[VYA2012] Vyatta, the Open Source Networking Community. http://www.vyatta.org/, July 2012.

[WAZ2014] http://www.wazoku.com/ [April 2014]

[WER13] Werthmann, Grob-Lipski, Proebster: Multiplexing Gains Achieved in Pools of Baseband Computation Units in 4G Cellular Networks. PIMRC 2013, London

[WU2014] Wu, Wenfei, et al. "PRAN: Programmable Radio Access Networks." 13th ACM HotNet Workshop, 2014.

[XEN] Xen project homepage http://www.xenproject.org/

[Yad11] N. Yadav and D. Cohn. “OpenFlow Primitive Set”, available online at http://goo.gl/6qwbg. July 2011.

[YAP10] K.-K. Yap et al., “Blueprint for introducing innovation into wireless mobile networks,” Proc. of VISA’10, 2010.

[ZAK10] Y. Zaki, L. Zhao, C. Grg, A. Timm-Giel, “A Novel LTE Wireless Virtualisation Framework,” MONAMI’10.