

ICT-317756

TRILOGY2

Trilogy2: Building the Liquid Net

Specific Targeted Research Project
FP7 ICT Objective 1.1 – The Network of the Future

D4.2 Public record of the Trilogy 2 operators workshop and other selected dissemination activities

Due date of deliverable: 31 December 2014
Actual submission date: 15 January 2015

Start date of project	1 January 2013
Duration	36 months
Lead contractor for this deliverable	Telefónica, Investigación y Desarrollo, S.A.U.
Version	v1.0 , 15 January 2015
Confidentiality status	“Public”

Abstract

This deliverable documents the Trilogy2 Operator Workshop and provides a report on standardisation activities of Trilogy2. The Operator Workshop was co-located with the Spanish Network Operator Meeting ES-NOG in Madrid on the 31st of October, 2014. Standardisation activities concentrated on the IETF and ETSI.

Target Audience

The target audience is anyone interested in the activities performed by the Trilogy 2 project to disseminate and standardise the project results.

Disclaimer

This document contains material, which is the copyright of certain TRILOGY2 consortium parties, and may not be reproduced or copied without permission. All TRILOGY2 consortium parties have agreed to the full publication of this document. The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the TRILOGY2 consortium as a whole, nor a certain party of the TRILOGY2 consortium warrant that the information contained in this document is capable of use, or that use of the information is free from risk, and accept no liability for loss or damage suffered by any person using this information.

This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content.

Impressum

Full project title

TRILOGY2: Building the Liquid Net

Title of the workpackage

WP4 Dissemination and Standardisation

Editor

Pedro A. Aranda, TID

Project Co-ordinator

Marcelo Bagnulo Braun, UC3M

Copyright notice

© 2015 Participants in project TRILOGY2

Executive Summary

This deliverable documents the Trilogy2 operators workshop held in Madrid, Spain on the 30th of October 2014. The workshop format was a one day seminar held immediately before the 14th meeting of the Spanish Network Operators Group (ES-NOG). These meetings, known as GORE meetings - GORE stands for *Grupo de Operadores de Red Españoles* (i.e. Spanish Network Operators Group) - are held twice a year and the audience is mainly the operator community in Spain, irrespective of their size. We include the presentations made² by Telefónica, I+D; NEC; OnApp and U3CM.

Additionally, this deliverable reports on the standardisation activities of members of the consortium in different Standards Defining Organisations (SDOs) during **2014**.

²in order of appearance

List of Authors

Authors	Pedro A. Aranda Gutiérrez, Diego López, Felipe Huici, John Thomson, Bob Briscoe
Participants	Telefónica, I+D; NEC; OnApp; UC3M; BT
Work Package	WP4 - Dissemination and Standardisation
Security	Public (PU)
Nature	R
Version	v1.0
Total number of pages	86

Contents

Executive Summary	3
List of Authors	4
List of Figures	6
List of Tables	7
1 Introduction	8
2 The Trilogy2 operators' workshop	9
2.1 Attendance	9
2.2 Presentations and reactions	9
3 Standardisation activities	11
3.1 ETSI related standardisation	11
3.2 Standardisation activities at the Internet Engineering Task Force and Internet Research Task Force	12
3.2.1 Multipath TCP (MPTCP)	12
3.2.2 Service Function Chaining (SFC)	13
3.2.3 VNFPOOL	13
3.2.4 NFVRG	13
3.2.5 Encryption of TCP streams	13
3.2.5.1 TCP Increased Security (TCP Inc.) Working group charter	14
3.2.5.2 TCP Increased Security (TCP Inc.) Progress	15
3.2.6 Congestion Exposure (ConEx) Working Group (WG)	16
3.2.7 Transport Area WG (TSVWG)	16
3.2.8 Transmission Control Protocol (TCP) Maintenance (TCPm) WG	16
3.2.9 Data Centre Latency Control (DCLC) Proposed Research Group	17
4 Conclusion	18
A The workshop presentations	19
A.1 General project introduction	19
A.2 SDN-NFV: An introduction	23
A.3 Towards the Superfluid (Network) Cloud	43
A.4 Road to the Federated Market and beyond	63
A.5 Data Centre (DC) Topologies for Network Function Virtualisation (NFV)	77
References	84

List of Figures

List of Tables

2.1 Attendance to the GORE-14 and the T2 workshop 9

2.2 Organisations present in the event 9

1 Introduction

In this document, we document the Trilogy2 operator workshop. Additionally, we also include information regarding the standardisation activities at the Standards Defining Organisations (SDOs), where Trilogy2 related activities have served as a base for proposals and contributions:

- European Technical Standards Institute (ETSI) (Network Function Virtualisation (NFV) Industry Specification Group (ISG))
- Internet Engineering Task Force (IETF)
- Internet Research Task Force (IRTF)

The document is structured as follows:

- in Chapter 2 we describe the operators' workshop
- in Chapter 3 we describe the standardisation activities
- in Appendix A we include the different presentations at the workshop

2 The Trilogy2 operators' workshop

The Trilogy2 operators workshop was held in Madrid, Spain on the 30th of October 2014. The workshop format was that of a one day seminar held immediately before the 14th meeting of the Spanish Network Operators Group (ES-NOG). These meetings, known as Grupo de Operadores de Red Españoles Spanish Network Operators Group (GORE) are held twice a year and the audience is mainly the operator community in Spain, irrespective of their size.

ES-NOG documents their GORE meetings in their Web site [31]. The project's operator workshop was announced as an SDN-related workshop as part of the GORE-14 meeting. The whole event is documented in [36].

2.1 Attendance

Table 2.1 shows a summary of the attendance to both days. The full list is available at [35]. The participants who indicated affiliation belong to 30 organisations, as shown in Table 2.1. Out of these, 20 can be singled out as network operators. We had 4 academic network operators, marked in bold in the table.

Event	Attendance
GORE-14	46
Workshop	30
Workshop and GORE-14	21
Only workshop	9
Only GORE-14	25
Either workshop or GORE-14	55

Table 2.1: Attendance to the GORE-14 and the T2 workshop

ADAMO TELECOM	AGRAMON	PROYECTOS
	PROGRAMAS	E INVER-
	SIONES S.L.	
Acens	Alturna Networks	
Atract consulting	Atractivo consulting	
BICS Spain	British Telecom	
CESGA	CSUC	
Espanix	Eticom Somos conexión	
Fon	Genetsis	
Gigas.com	IBERMATICA	
IP Broker Spain	Interoute	
L&M Data Communications	Medialab-Prado	
Mercado IT	NEUTRA (NEO)	
Nimbus Concept S.L.	ONO-VODAFONE	
OnApp	RedIRIS	
Telefónica I+D	Xtratelecom	
inAsset NixMad	www.fcsc.es	

Table 2.2: Organisations present in the event

2.2 Presentations and reactions

We include the presentations made by Telefónica, I+D; NEC; OnApp and U3CM¹ in Appendix A. The main workshop language was Spanish and, as such, the fact that three out of the four presentations were done in it was widely appreciated and kept the audience well concentrated on them.

¹in order of appearance

After each of the presentations, there was space for questions and debate. Questions to the first presentation included if it is possible to identify a clear the boundary between SDN and NFV in the liquid network; what the situation of OpenFlow and traditional network equipment vendors is and how they were adapting to the emergence of NFV. The debate took a more general turn when questions on how could organisations start to apply these techniques. First experiences where exchanged in the technical field and initial reactions on the operational aspects were shared.

As NEC's talk had a highly technical content, reactions and questions to it were in the same tune. The audience wanted to know more abbot the different technologies mentioned in the talk (e.g. MiniOS, VALE), how they compare and inter-work with other state-of-the-art (e.g. docker or Linux containers). A lively debate developed on the use of small Virtual Machine s (VMs) versus the use of containers.

OnApp's talk also stimulated a good debate on the Software Defined Networking (SDN) technologies used by them. In the more general front, people wondered how to enable contents distribution by users using liquidity would impact on the overall stability. Questions on their plans for expansion Latin America were asked, given the small presence shown in the presentation.

UC3M's talk showed stimulated a debate on the traffic patterns registered at the operators data centers, how current topologies can handle the new model, and the outlook of the impact of Network Function Virtualisation (NFV) on datacentre topologies.

3 Standardisation activities

Trilogy2 has been highly active in the European Technical Standards Institute (ETSI), the Internet Engineering Task Force (IETF) and the Internet Research Task Force (IRTF) during 2014. This chapter gives an overview of these activities, with a special stress on:

- The second year of phase 1 of the ETSI NFV Industry Specification Group (Industry Specification Group (ISG))
- The new IETF working groups on Service Function Chaining (Service Function Chaining (SFC)) and Virtual Network Function (VNF) Pools driven by the introduction of NFV
- The new IRTF NFV research group
- The new IETF Transmission Control Protocol (TCP) INCreased security (tcpinc)
- Analysis of residual threats to MPTCP as part of the completion of the IETF's experimental specifications of Multipath TCP (MPTCP)
- Completion of the IETF's experimental specifications of Congestion Exposure (ConEx), including work on congestion feedback in the TCP maintenance (TCP Maintenance & Minor Modifications (TCPM)) Working Group (WG)
- New work on transport protocol extensibility in the IETF's TCPM WG
- The proposed new IRTF research group on data centre latency control (DCLC).

3.1 ETSI related standardisation

2014 has been the year of consolidation of the NFV concepts and activities, not only focused on the delivery of the first release of the ETSI NFV ISG documents, but also in the practical demonstrations through the NFV Proof-of-Concept (POC) framework.

The final version of the first release of ETSI NFV ISG specifications was finished and made ready for formal approval by the group along the NFV#8 meeting held in November in Scottsdale (Arizona, USA). The documents include contributions made by the partners in areas related to security, performance and portability, infrastructure, and management and orchestration. In particular, the most salient contributions were focused on multi-tenancy considerations, best practices to guarantee a predictable performance of virtualised network functions, and a layered approach to service and infrastructure orchestration in order to support flexible deployment and operational patterns.

The specific published documents contributed to by partners have been:

- NFV Infrastructure Architecture Overview (Rapporteur: Andy Reid) [4] and the other outputs of the Architecture of the Virtualisation Infrastructure WG;
- NFV Infrastructure; Methodology to describe Interfaces and Abstractions (Co-author: Andy Reid) [1]
- NFV Security Problem Statement (Rapporteur: Bob Briscoe) [5]
- NFV Security; Security and Trust Guidance (Co-author: Diego López) [29]
- NFV Performance and Portability Best Practises (Expert Group chair and coauthor: Fco. Javier Ramón) [2]
- NFV Proof of Concepts; Framework (Expert Group Chair and co-author: Fco. Javier Ramón) [3]
- NFV Management and Orchestration (Co-authors: Gerardo García and Fco. Javier Ramón) [45]

Being part of the NFV leadership team partners have also contributed presentations too numerous to list. Partners also contributed considerably to the series of three (so far) joint carrier white papers, giving annual updates on the progress of the ISG. The third edition "Network Operator Perspectives on Industry Progress" gives a very useful summary of the technical progress of all the working groups, as well as the commercial status of NFV technology and prospects for the ISG into 2015 [46].

Also an implementation of the best practices on performance have been contributed to OpenStack, in the framework of the recently formed group on NFV within the project.

The POC framework was overseen by a member of the TID team. Three POCs directly connected with the activity in the IRTF NFV Research Group (NFV-RG) are worth highlighting:

- VNF Router Performance with Distributed Denial-of-Service (DDoS) Functionality [54].
- VNF Router Performance with Hierarchical Quality of Service (QoS) Functionality [55]
- Virality based content caching in NFV Framework [53]

At the same meeting, the NFV ISG defined the structure for the so-called Phase 2, already approved by the ETSI Board. The working groups have been restructured to focus on producing interface definitions at all reference points in the NFV Reference Architecture (IFA WG), explore the evolution of NFV (EVE WG), guarantee security (SEC WG) and reliability (REL WG), and guide implementation and testing (TST). We foresee to continue with a high implication in SEC, and mostly contribute to EVE and TST. A representative of TID was re-elected as chair of the NFV Technical Steering Committee for this Phase 2 of the NFV endeavour.

Finally, in terms of ETSI activities, it is worth noting the launch of a new ISG, called Mobile-Edge Computing (MEC) and focused on "interoperable and deployable specifications that will allow the hosting of third-party applications in a multi-vendor Mobile-edge Computing environment". Being these objectives so much aligned with the essential liquidity concepts (and with the NFV approach as well) we foresee an interesting field for additional standardisation activities along the coming year.

3.2 Standardisation activities at the Internet Engineering Task Force and Internet Research Task Force

3.2.1 Multipath TCP (MPTCP)

Trilogy2 partners continued to actively contribute to the Multipath TCP working group (MPTCP) during 2014. Since the publication of RFC6824 [33], the working group has shifted its energy to improve the experimental protocol specification in order to move to standards track.

The main document that is being produced within the MPTCP working group is the revision of RFC6824 [34]. This document is lead by project partners and several revisions have been produced in 2015. Several other internet drafts have been submitted and presented at the IETF :

- `draft-barre-mptcp-tfo` [12] describes how the TCP Fast Open extension [30] can be supported by Multipath TCP. The extension proposed in this draft has already been implemented in the Linux kernel implementation of Multipath TCP maintained by the project.
- `draft-bonaventure-mptcp-experience` [16] is one of the documents required in the charter of the MPTCP working group. It describes the lessons that have been learned by implementing Multipath TCP and using it in the real Internet. This document has now been accepted as a working group document [17].
- `draft-bonaventure-mptcp-rst` [18] proposes a new MPTCP option that can be used in RST segments to provide additional information on the reasons for a RST. With regular TCP, when a RST segment is sent/received, the TCP connection disappears. With Multipath TCP, a RST segment can be sent/received on a subflow and the Multipath TCP connection continues. Since a host could send a RST segment for different reasons (bad performance, middlebox interference, policies, ...), it is important to inform the remote host about the reason for the termination of a subflow. The MPTCP working group has adopted this proposed extension which has been included in the last revision of [34].
- `draft-bonaventure-mptcp-timestamp` [15] proposes a new option to encode timestamps for Multipath TCP. This draft was motivated by the publication of RFC7323 [19] that redefines the Timestamp option proposed for TCP in [38] and makes them mandatory in all segments. In regular TCP, this is motivated by the utilisation of the timestamps to protect against problems with wrapped sequence numbers [19]. Since Multipath TCP uses 64 bits sequence numbers, these problems cannot occur and there is no need to use a regular TCP timestamp in each Multipath TCP segment. The proposed timestamp options gives new opportunities to improve the delay estimations in Multipath TCP.

- `draft-paasch-mptcp-control-stream` [49] shows that it is possible to extend Multipath TCP to divide the bytestream into a data stream and a control stream. This technique would allow Multipath TCP to use options that contain up to 64 KBytes of data and are transmitted reliably.

We also performed a residual threat analysis of MPTCP [11]. The actual work was reported as part of D1.1 in Year 1. We promoted the work in the MPTCP working group in the IETF and it was accepted as a working group draft. The current status is that the draft is in Internet Engineering Steering Group (IESG) review to become an Request For Comments (RFC), hopefully during 2016. The draft as well as its status in the IETF can be found at: <https://datatracker.ietf.org/doc/draft-bagnulo-mptcp-attacks/>. Furthermore, several drafts have proposed solutions to improve the security of Multipath TCP.

- `draft-bonaventure-mptcp-tls` [14] builds upon earlier work described in [48] and proposes to better integrate Multipath TCP and Transport Layer Security (TLS) together.
- `draft-bagnulo-mptcp-secure` [10] discusses how `tcpcrypt` could be combined with Multipath TCP.

3.2.2 SFC

SFC is another working group where Trilogy2 activities have been reported. It provides the project a stage to present and discuss their thoughts on how to control liquidity. A draft regarding the requirements on Operations, Administration, and Maintenance [7] (OAM) requirements [40] was submitted for the IETF'90. In a liquid scenario, where network functions are virtualised, it makes sense to introduce mechanisms for tracing the chains of virtualised network functions that implement a given service or service function. This draft has finally been merged with another draft proposing a similar approach [6] for the IETF'91.

3.2.3 VNFPOOL

Regarding the VNFPOOL activities in the IETF, we have been working on getting the working group set up. However, there seems that there has not been enough momentum for this at the meetings in London and Toronto in 2014. The working group does not meet in the Honolulu meeting and there is a consensus to retry forming the working group during 2015.

In the mean time, we have contributed to several Internet drafts:

- the problem statement for the VNFPOOL working group [56].
- a use case centered around the virtual Content Distribution Network (vCDN) use case [8], where we explore at resilience requirements in a virtualised Content Distribution Network (CDN) deployment.

3.2.4 NFVRG

During this year, a research group on NFV has been proposed to the IRTF [47]. The group has met along two of the recent IETF meetings (IETF90 in Toronto, and IETF91 in Honolulu), gathering a high interest among the community. The proposed RG co-chairs are preparing a final charter proposal to be formally approved by the IRTF chair. One of the members of the TID team is co-chairing NFVRG and two Internet drafts connected with the research aspects identified by the group charter have been submitted and presented:

- An Open NFV Architectural Framework for Virality Based Content Caching [41].
- NFVIaaS Architectural Framework for Policy Based Resource Placement and Scheduling [39].

3.2.5 Encryption of TCP streams

The conclusion of the residual threat analysis for MPTCP is that in order to secure MPTCP we need to protect the payload. In order to standardize a solution for encrypting the payload of MPTCP we need first to define a general solution for encrypting TCP streams. One of such approach is `tcpcrypt`, designed in the previous Trilogy (I) project.

We decided then to promote the work on opportunistic encryption of TCP streams in the IETF. In order to do so, we proposed the creation of a working group in this topic and the proposal was adopted. The TCP Increased security (TCPINC) working group was created and had its first meeting in the IETF meeting in July. The co-chair and proponent of the WG is Marcelo Bagnulo from T2. The TCPINC WG is aligned with the

reactions the IETF community is proposing after the Snowden revelations on pervasive monitoring captured in [32]. A large number of WG participants are interested in designing a general solution for encryption by default in the Internet.

The proposed and adopted charter is attached below.

3.2.5.1 TCP Increased Security (TCP Inc.) Working group charter

The TCP Inc. WG will develop the TCP extensions to provide unauthenticated encryption and integrity protection of TCP streams. The WG will define an unauthenticated key exchange mechanism. In addition, the WG will define the TCP extensions to utilize unauthenticated keys, resulting in encryption and integrity protection without authentication. This is better than plain-text because it thwarts passive eavesdropping, but is weaker than using authenticated keys, because it is vulnerable to man-in-the-middle attacks during the initial unauthenticated key exchange. This work is part of the IETF effort to harness the Internet architecture given the latest events of pervasive monitoring (see [32]).

The goal of this WG is to provide an additional security tool that complements existing protocols at other layers in the stack. The WG will be looking for the designs that find the right tradeoff spot between conflicting requirements: to provide reasonable security for the majority of connections. Because we are dealing with unprotected connections, we are more focussed on improving from baseline of no security than achieving the high standard of security that is already available to users of TLS. Providing unauthenticated encryption and integrity protection at the TCP layer will provide a set of features that cannot be achieved with existing tools, namely, encryption and integrity protection without modifications to the upper layers (no API changes), encryption and integrity protection with forward secrecy with a per-connection granularity, simple NAT and firewall traversal capabilities, key rollover without significant impact to the TCP connection, lower overhead compared to solutions relying in stacking multiple protocols to achieve different features, no manual configuration required. A more detailed description of the motivations for TCP-based solutions can be found in draft-bellare-tcpsec-01 [13] and in RFC5925 [51].

The working group is looking to produce experimental documents specifying the required TCP extensions and any additional documents needed.

The high-level requirements for the protocol for providing TCP unauthenticated encryption and integrity protection are:

- It should work over the vast majority of paths that unmodified TCP works over, in particular it must be compatible with NATs (at the very minimum with the NATs that comply with BEHAVE requirements as documented in RFC4787 [9], RFC5382 [37] and RFC5508 [50]);
- The protocol must be usable by unmodified applications. This effort is complementary to other security protocols developed in the IETF (such as TLS) as it protects those applications and protocols that are difficult to change or may even not be changed in a backward compatible way. It also provides some protection in scenarios where people are unwilling to do any change just for the sake of security (e.g., like configure encryption in an application).
- The protocol must provide cryptographic algorithm agility.
- Must gracefully fall-back to TCP if the remote peer does not support the proposed extensions
- When encryption is enabled, it must at least provide protection against passive eavesdropping by default,
- Should attempt to use the least amount of TCP option space, especially in SYN segments.
- Must not require any authentication or configuration from applications or users. However, hooks for external authentication must be made available. The WG will not work on new authentication mechanisms.
- The protocol must have acceptable performance, including acceptable latency and processing overheads. For example, the protocol may try to re-use existing cryptographic material for future communication between the same endpoints to avoid expensive public key operations on connection set up.

When encryption is enabled, then the protocol:

- must always provide forward secrecy.

- must always provide integrity protection of the payload data (it is open for discussion for the WG if the TCP header should or not be protected)
- must always provide payload encryption.
- must not provide extra linkability. When encryption is enabled the TCP traffic should not give a third party observer any extra way to associate those packets with the specific peers beyond information that would have been present in a cleartext session.
- must allow the initiator of the connection to avoid fingerprinting: some initiators may want to avoid appearing as the same endpoint when connecting to a remote peer on subsequent occasions. This should either be the default or some mechanism should be available for initiators to drop or ignore shared state to avoid being fingerprintable any more than would be present for a cleartext session.

Security features at the TCP-level can benefit other TCP extensions. For example, both Multipath TCP and TCP Fast Open require proof that some connections are related. Session resumption and Message Authentication Codes (MACs) can provide this evidence. The working group should identify synergies and design the security protocol in such a way that other TCP efforts can benefit from it. Of course, TCP extensions that break must be identified too, and kept to a minimum.

The working group will produce the following documents:

- A framework for unauthenticated encryption and integrity protection of TCP connections. This document will describe basic design considerations, including the motivation and the applicability of the proposed mechanism, the interaction with other security mechanisms in different layers of the stack, the interaction with external authentication mechanisms, the expected protection, privacy considerations and residual threats.
- Definition of the unauthenticated key exchange mechanism and the extensions to current TCP to utilize unauthenticated key to provide encryption and integrity protection. This covers all the protocol changes required. This will be an experimental document.
- An extended API describing how applications can obtain further benefits of the proposed extensions. In particular, the hooks for supporting external authentication will be defined in this document. This will be an informational document.

3.2.5.2 TCP Increased Security (TCP Inc.) Progress

In the context of increasing TCP security, the T2 project has performed several related activities, namely:

- The work on securing MPTCP (both using tcpcrypt and TLS) reported in D2.4 and outlined in the above section on MPTCP;
- The architectural framework for opportunistic encryption in the Internet described in D2.3, which formed the background to the chartering of the TCP Inc. WG;
- the Inner Space proposal described in D2.4, which makes extension of TCP feasible, including cryptographic extensions as outlined below.

At the Nov 2015 IETF, we presented a proposal to decompose the tcpinc protocols into a framing layer and a layer for cryptographic control options. The Inner Space protocol provides the appropriate framing and layering. We have specified this approach for both candidate tcpinc solutions: tcpcrypt or the TLS-option. We pointed out that a downgrade attack on the current candidate tcpinc solutions would be unremarkable, given any new TCP options are already blocked over a large minority of Internet paths. Therefore, currently it would be easy for any government agency to disable the tcpinc protocol whenever it wanted, simply by making it appear as if a middlebox did not support the new protocol. We showed how our solution based on Inner Space would solve the middlebox traversal problem, so that a real downgrade attack would be obvious. Our proposal also considerably simplifies tcpcrypt, reducing the number of new TCP suboptions from 18 to 9 and removing a number of protocol states.

Our proposal based on Inner Space also removes the extra handshake latency introduced by all the candidate tcpinc solutions. The latency aspects of Inner Space are being investigated in the EU FP7 Reducing Internet Transport Latency (RITE) project, while the Trilogy 2 project is focusing on the base Inner Space protocol.

3.2.6 Congestion Exposure (ConEx) WG

Section 3.2 of Trilogy 2 Deliverable “D2.4 Advanced Tools for Controlling Liquidity” (Jan 2015) gives the rationale behind our work on Congestion as a Metric and its centrality to control of pooled (liquid) resources, including a more recent shift away from techniques that require standardisation.

Subsection 3.2.3 of D2.4 describes the wrapping up of the standards specifications defined in the IETF Congestion Exposure (ConEx) WG that was initiated by the first Trilogy project. That discussion will not be repeated here, instead the standards outputs will simply be listed.

Beyond the pre-Trilogy-2 RFC describing ConEx Concepts and Use-Cases (RFC6789 [21]), we have completed:

- “ConEx Concepts & Abstract Mechanism” [43], which is the main technical ConEx draft. Since Jul 2014, it has been passing through the steering group review process on its way to RFC status.

Although no Trilogy 2 partners are co-authors, we have also given thorough reviews of the two other ConEx drafts that are now completed and starting on the steering group review process towards RFC.

Three of our Internet Drafts remain as individual submissions, and D2.4 describes the process we are going through to find a new home for two of them in the transport area working group (TSVWG—see below), now that the ConEx WG is closing. All three have already been generalised and are not specific to ConEx:

- “Network Performance Isolation using Congestion Policing” [24];
- “Network Performance Isolation in Data Centres using Congestion Policing” [28];
- “Reusing the IPv4 Identification Field in Atomic Packets” [25]. This draft is a proposed way to encode ConEx signalling in the IPv4 header (ConEx is only chartered for IPv6), and it is generalised to allow the IPv4 header to be extended for other purposes.

Trilogy 2 partners have presented on the congestion policing and performance isolation in data centres drafts to the ConEx WG in 2014. Progress on the abstract mechanism draft has also been regularly presented.

3.2.7 Transport Area WG (TSVWG)

In Nov 2013, “Byte and Packet Congestion Notification” was published as Informational RFC 7141 [20]. It concerns whether it is correct to measure congestion in units of bytes or packets. This work was actually done in the first Trilogy project (in fact the very first draft was written before that) via the IETF’s TSVWG, but it is nonetheless still very relevant to Trilogy 2.

Note that the work on Congestion Policing mentioned above is continuing in the TSVWG, particularly given its heavy reliance on tunnelling, which is becoming an important area of work for the TSVWG (again, see D2.4 for details).

3.2.8 TCP Maintenance (TCPm) WG

The work of Trilogy 2 partners in the IETF’s TCPm WG falls into two main categories:

- Accurate and Trustworthy congestion feedback, including a test for receiver compliance and the greater accuracy of ECN feedback needed for Congestion Exposure (ConEx) and Data Centre TCP (DCTCP). Respectively Section 2.4 and Subsection 3.2.4 of Trilogy 2 Deliverable “D2.4 Advanced Tools for Controlling Liquidity” (Jan 2015) give the rationale and details for this work.
- Extensibility of Transport Control (the Inner Space protocol). Section 2.5 of Trilogy 2 Deliverable “D2.4 Advanced Tools for Controlling Liquidity” (Jan 2015) gives the rationale and details of this work.

The standards contributions and their status will simply be listed here, rather than repeating the discussion already given in D2.4:

- “Problem Statement and Requirements for More Accurate ECN Feedback” [42]. Following numerous revisions, this requirements draft has now completed the working group phase and is working through the steering group review phase on its way to becoming an RFC;

- “More Accurate ECN Feedback in TCP (AccECN)” [27]. This is a fully specified candidate solution to above requirements statement. It is not yet adopted as IETF WG business, because we have asked for it to be put on hold to see if the Inner Space work is adopted, which would then greatly simplify all additions to TCP, including this one;
- “A Test To Allow TCP Senders to Identify Receiver Non-Compliance” [44]. This is an individual submission that we have recently reintroduced after a long hiatus.
- “TCP SYN Extended Option Space Using an Out-of-Band Segment” [52]. This individual submission included two ideas for how to extend the space for transport control, each contributed by different co-authors. We have since developed preferred ideas and withdrawn our part of this draft, whereas our co-authors are continuing to maintain the idea in their part (which is very unlikely to traverse firewalls);
- “Extended TCP Option Space in the Payload of an Alternative SYN” [22]. We published a couple of revisions of this individual submission, before we decided it should be superseded by the Inner Space protocol.
- “Inner Space” [23]. We have published two revisions of this fully specified draft, which is still an individual submission.
- “The Echo Cookie TCP Option” [26]. This draft was split out from the Inner Space draft. It has the potential to act as a signalling channel between the end-systems and middleboxes. It is also still an individual submission.

The Accurate Explicit Congestion Notification (ECN) work has been presented in both the TCPM WG and the Data Centre Latency Control Research Group (RG), given its relevance to Data Centre TCP (DCTCP) as well as ConEx.

The feedback non-compliance testing draft has been presented in TCPM.

The transport extensibility work has been presented in a special sub-session of the TCPM WG, given our work had prompted other candidate solutions. Our work on applying Inner Space specifically to TCP INCreased security (tcpinc) led to discussions that consumed nearly the whole Nov’14 session.

3.2.9 Data Centre Latency Control (DCLC) Proposed Research Group

DCLC is a proposed new IRTF research group that Trilogy 2 partners were peripherally instrumental in creating. It has run for two sessions, with some interesting talks. However, one of the primary aims of the group—to encourage more discussion of operational experience—has not been achieved, due mainly to widespread operator confidentiality.

Trilogy 2 partners have presented twice. Once in a session on Traffic Management and Performance Isolation in Data Centres, and once on Accurate ECN for Data Centre TCP and ConEx.

4 Conclusion

This document has presented the Trilogy 2 operator workshop held in Madrid in October, 2014 and the standardisation activities in the IETF, IRTF and ETSI.

The operator workshop was co-located with the Spanish Network Operator group and was well attended. It gave the opportunity to publicise the project's activities to a very interested community.

A lot of standards-related Trilogy 2 activity has involved project partners promoting Trilogy 2 outcomes in different Standards Defining Organisations (SDOs). This document provides an overview and complements the project management reports, which include more comprehensive listings of the different proposals brought forward at the different SDOs.

A The workshop presentations

A.1 General project introduction

<http://trilogy2.eu>



Building the liquid Internet

Timescale: Start January 2013, 36 months

Resources:

- Total cost: 5.1m Euros
- EC contribution: 3.7m Euros

Project identification: INFSO-ICT-317756



trilogy 2

Trilogy 2 participants

Partners

- **Operator**
 - BT (WP leader)
 - Telefónica (WP leader)
- **Vendors**
 - NEC (WP leader)
 - OnApp (WP leader)
 - Intel
 - Nextworks
- **Academia**
 - UC3M (Coordinator)
 - UCL-UK
 - UCL-BE
 - UPB
 - UCAM

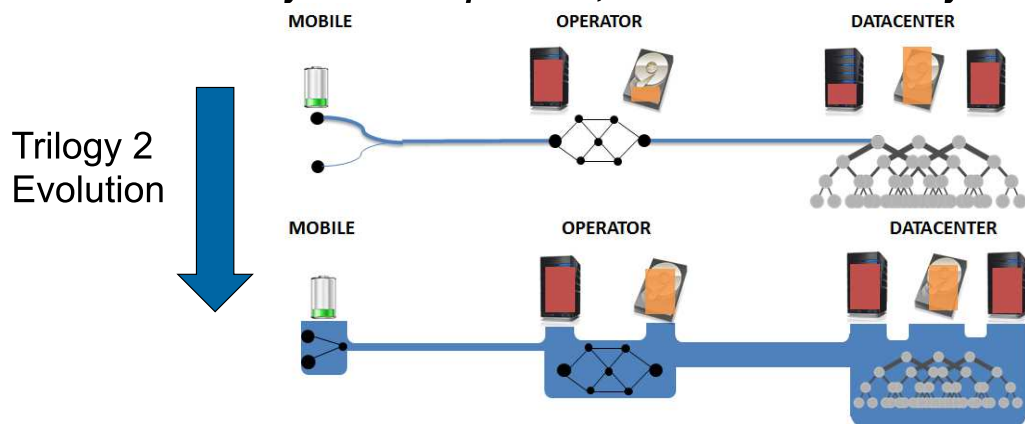
Key People

- **Operators**
 - Bob Briscoe
 - Pedro Aranda, Diego Lopez
- **Vendors**
 - Felipe Huici
 - Julian Chesterfield, John Thomson
 - George Milescu, Valentin Ilie
 - Giacomo Bernini
- **Academia**
 - Marcelo Bagnulo, Francisco Valera
 - Mark Handley
 - Olivier Bonaventure
 - Costin Raci
 - Jon Crowcroft, Anil Madhavapeddy

Trilogy 2 vision

- **Vision: The Internet should behave as a liquid network.**

A liquid system allows resources including bandwidth, storage and processing to be used by any application whether they are contributed by network operators, data centers or end systems



Building the liquid Internet

Trilogy 2 Overview 3

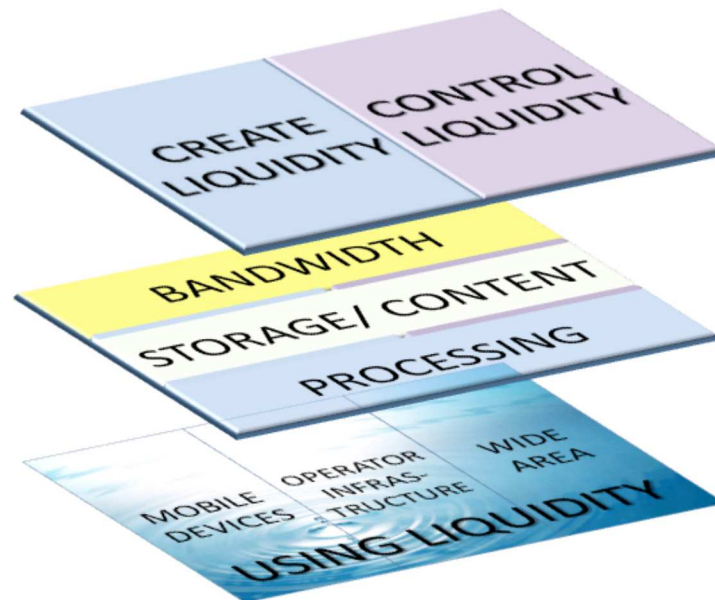
Objectives

1. Develop a unified architecture for the Internet for generalized resource pooling and trading between different types of resources including bandwidth, processing and storage in a scalable, dynamic, autonomous and robust manner to local operational and business requirements.
2. Research, develop, implement and evaluate new technical solutions for the creation of resource pools and trading off between them, in the areas of bandwidth, storage and processing.
3. Research, develop and evaluate mechanisms to control the created liquidity. Liquidity needs to be controlled, in order to be able to isolate different tenants, to make quality of service guarantees and to trade off resources in an economically efficient manner.
4. Implement and perform validation trials of the proposed architecture and mechanisms for specific use cases such as mobile scenarios, network-as-a-service and data centers.

Building the liquid Internet

Trilogy 2 Overview 4

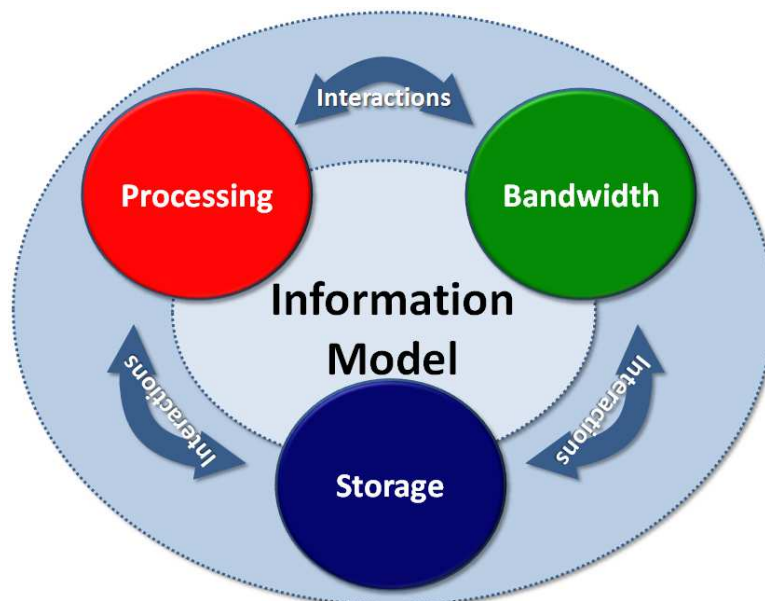
Conceptual schema



Building the liquid Internet

Trilogy 2 Overview 5

Trilogy 2 Framework



Building the liquid Internet

Trilogy 2 Overview 6

Results

- **Publications on scientific venues**
- **Standardization**
 - IETF
 - ETSI NFV
- **Software**
 - MPTCP
 - Polyversal TCP
 - Irminsule
 - Trevi
 - GRIN

A.2 SDN-NFV: An introduction



We are evolving towards a **Hyper Connected and Intelligent Digital World***

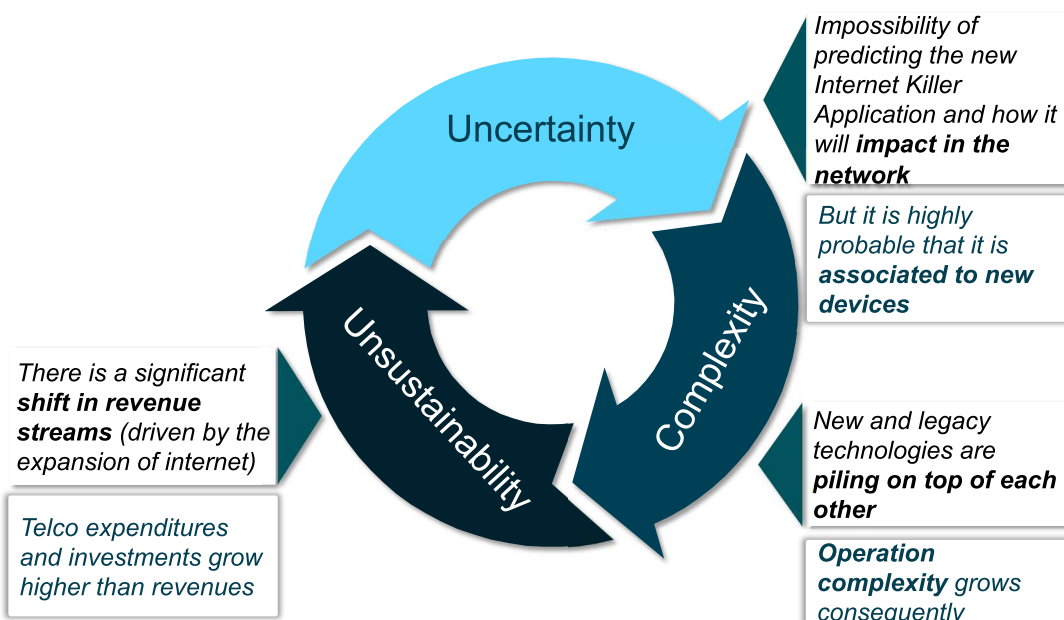


BE MORE_
DISCOVER, DISRUPT, DELIVER

(*) César Alierta, TEF CEO.
TEF Digital Investor's Day 2012

Telefónica

This digital world is introducing **relevant challenges** for telecom operators...



BE MORE
DISCOVER, DISRUPT, DELIVER

Telefónica

Beyond evolution: evolution is mandatory to keep in the market. Transformation is the only way to lead...

Network evolution is reasonably under control...



...from 3G to 4G
...from copper to fibre

The challenge is transforming the network and its operation taking into account the inertia of its legacy



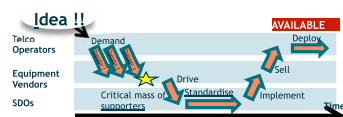
BE MORE
DISCOVER, DISRUPT, DELIVER

Telefónica

What are the current limitations of Telco's networks?

Long innovation cycles (2-6 years)

- Long **standardization cycles**
- **Scale is needed** to introduce innovations



Hardware and Software vertically integrated



- **Capacity** is tied to a **function**
- **Vendors lock in** (it is difficult to switch from one vendor to another when deployments are made)

Complex Network Management

- **Small changes in a network element requires an adaptation of the EMS** (Element Management System)
- **Complex stitching** of network functions across segments and technologies, since **network nodes are tightly coupled to the network segment and technology**

Difficult IoT

- Interoperability tests required per protocol and node



BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

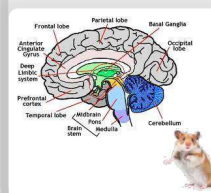
We need to adapt and define the change to lead in the Digital era

Telco players



- Very intensive in **hardware**
- **Capital intensive**
- Software is not a core

Internet players



- Very intensive in **software**
- Can have **global impact** with not too much capital
- Hardware is a support, and is located in the network periphery

HARDWARE

SOFTWARE

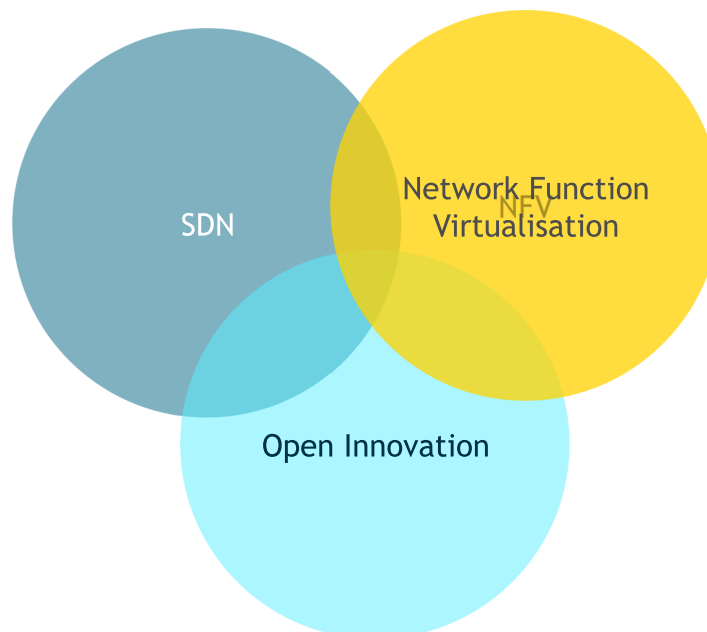


BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica



Components for a “liquid network”



Or maybe by non-example

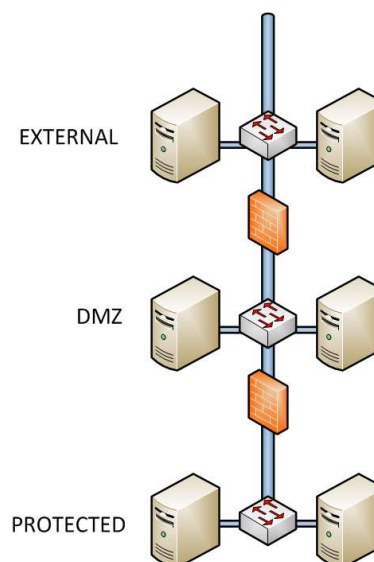
SDN BY EXAMPLE

BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

IaaS in a data centre: per user requirements

- Well-known design
 - Chaining FWs to increase the level of protection
 - DMZ to place resources that need to be connected to the Internet with some level of protection
- This is current best practice that is implemented on separate boxes nowadays
- Users expect to have the same level of protection when outsourcing this infrastructure to an IaaS provider



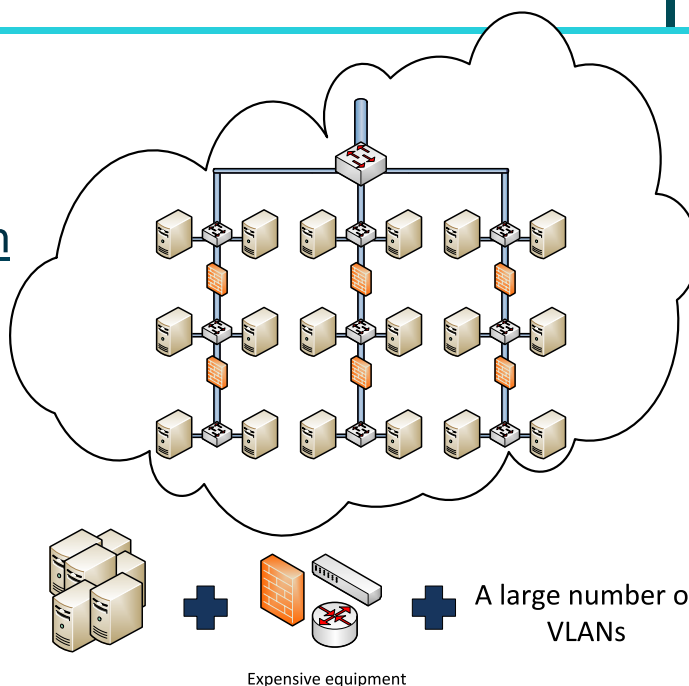
BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

The datacenter replicates the logical IaaS structure for every client...

... and this challenge has limitations with the current paradigm

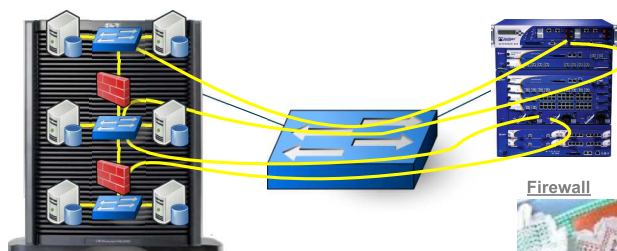
Fine grained connectivity & isolation require extensive use of VLANs & firewall rules (and combine them!)



BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

“Client stitching”... or how to combine a VM server with a common networking infrastructure



- First solution
 - Provide virtualised FW functionality
 - Provide network isolation using VLANs
 - Provide isolated switching realms
- And this for each client...
- Since there is virtualisation and the system was controlled by a “software”, the vendor claimed this was

SDN



BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

The acronym war



BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

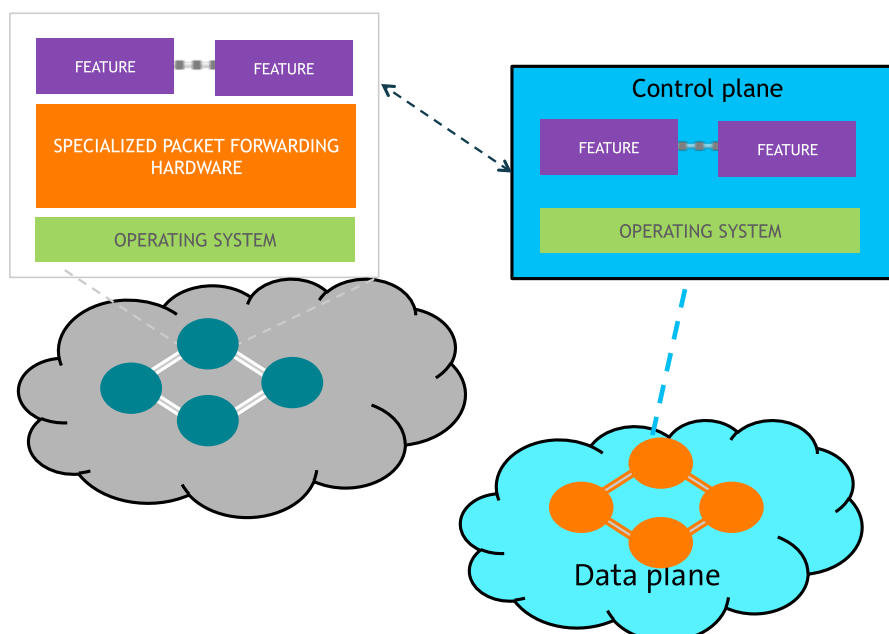
So... What is SOFTWARE DEFINED NETWORKING?

- Software Defined Networking is recognising that
 1. The network is not a shapeless entity AND network shape matters
 2. Network nodes don't need a massive amount of intelligence for bringing packets from port A to port B
 3. Distributed is nice, but not a DOGMA

BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

How do you SDN ?



BE MORE
DISCOVER, DISRUPT, DELIVER

Telefónica

What are the implications?

- The data plane can be simplified
 - Best case scenario is using commercial off-the-shelf boxes
 - The x86 architecture
 - Is known to provide significant throughput
 - Provides a lot of interesting features to make the network flexible (virtualisation...)
- The control plane
 - Also benefits from the advances in the x86 architecture
 - Better control of the control plane features means
 - More overall stability
 - More flexibility, when wisely used

BE MORE
DISCOVER, DISRUPT, DELIVER

Telefónica

But this is complex, right?

- Of course... Did I ever say it would be easy
- However, the process is worth the gain
- A significant community has been working on this for the last couple of years
 - Network Operators
 - Hardware suppliers
 - Software suppliers
 - ...

ETSI NFV-ISG

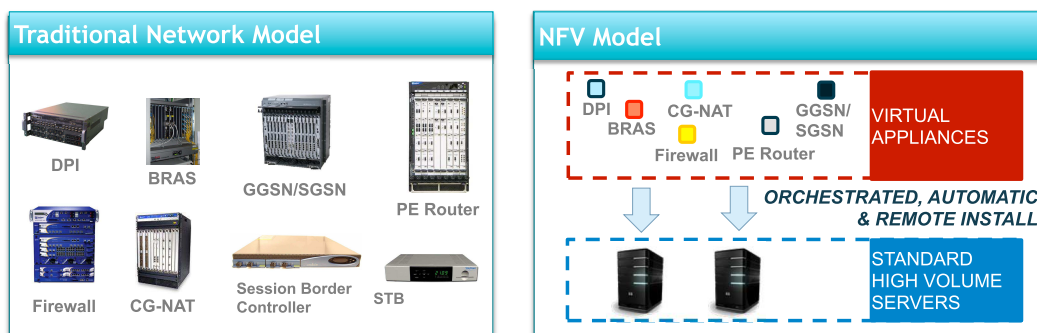
BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

NFV ISG vision & objectives...

NFV ISG vision:

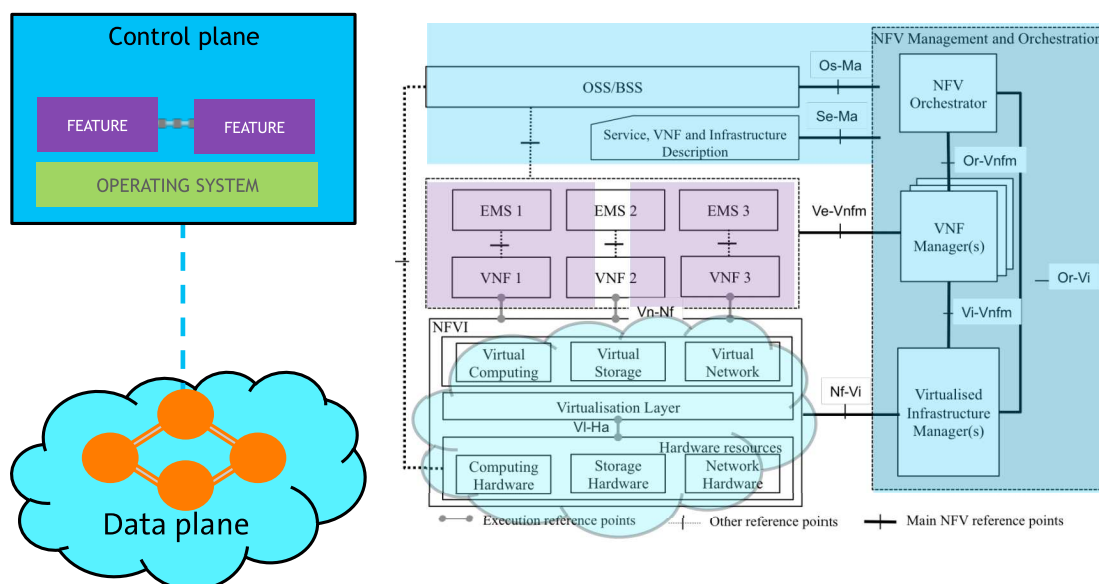
"Leverage standard IT virtualisation technology to consolidate many network equipment types onto industry standard high volume servers, switches, and storage"



BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

The ETSI NFV Reference Architecture



BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

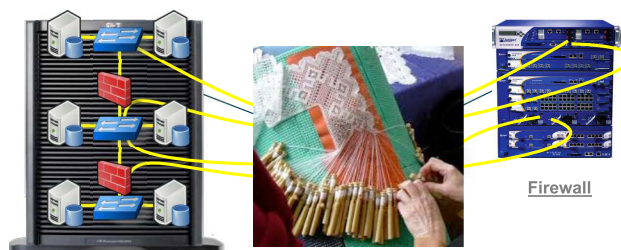
Bringing packets from A to B

SOLVING THE NETWORKING PART

BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

Getting back to packet handling...



- Handling packets can be part of the headache
- However, virtualisation can help us getting rid of it

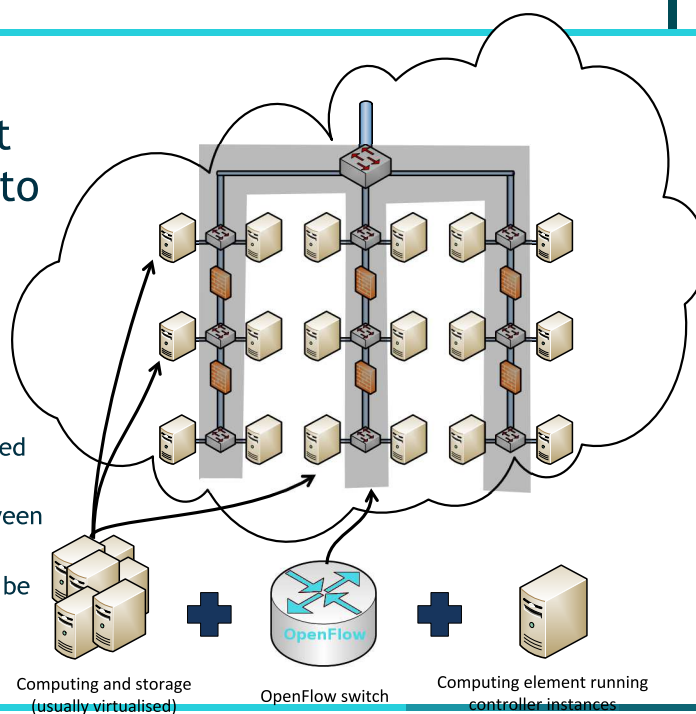
BE MORE
DISCOVER, DISRUPT, DELIVER

Telefónica

One candidate for packet handling can OpenFlow

OF places packet handling logic into a centralised controller

- Easy to manage
- Some network functions become a program executed in the controller
- Smooth coordination between network and computing
- OpenFlow switch role can be played by the own hosts! (Open vSwitch)



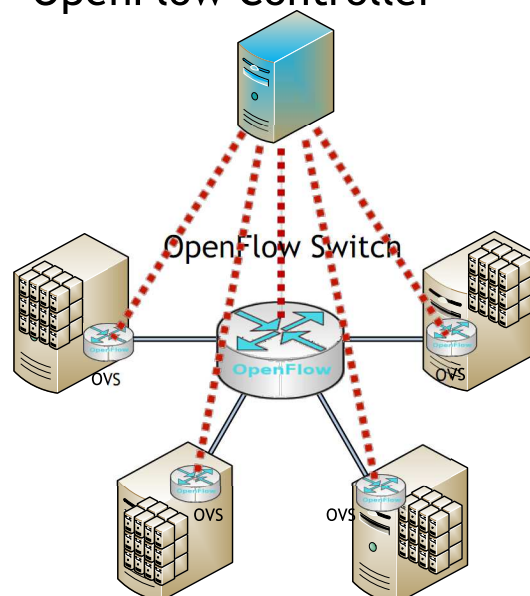
BE MORE
DISCOVER, DISRUPT, DELIVER

Telefónica

OpenFlow in green-field deployments

- OpenFlow in the switching infrastructure
- OpenFlow integrated in the server
 - Open vSwitch is already built-in in the commonest virtualisation environments and the latest Linux kernels (3.3)
- Other SDN control protocols also applicable

OpenFlow Controller

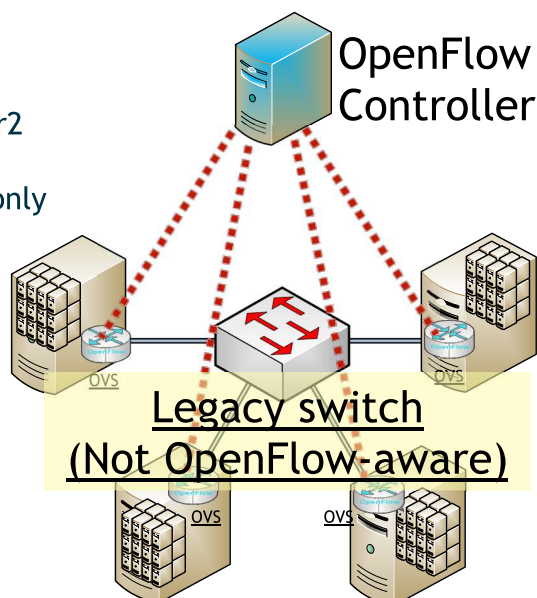


BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

... or in evolutionary scenarios, where legacy switching elements can be preserved

- Maximise reuse
 - Rely upon existing Layer2 or Layer3 connectivity
 - Use OVS in the servers only
- Nicira Networks' approach



BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

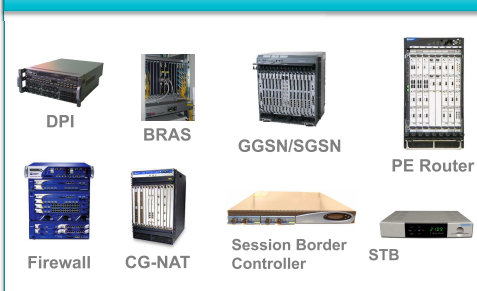


Performance & Portability are required to fully accomplish NFV ISG objectives...

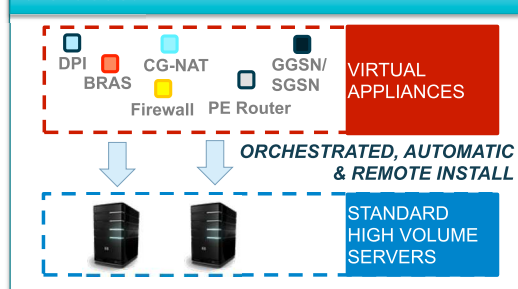
NFV ISG vision:

"Leverage standard IT virtualisation technology to consolidate many network equipment types onto industry standard high volume servers, switches, and storage"

Traditional Network Model



NFV Model



TO FULLY REALISE THIS VISION:

Virtualised network appliances should provide **high performance...**
... while being **portable** between servers (& hypervisors)

BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

... while providing the telco ecosystem actors a more predictable and manageable environment

VIRTUAL NETWORK FUNCTIONS PROVIDERS

Would not need to be aware beforehand of the infrastructure server on which their SW would be deployed in the end...
... but still can provide realistic performance estimations for different sets of HW (& hypervisor) setups.

HARDWARE (& HYPERVISOR) PROVIDERS

Could describe their equipment in objective terms, suitable for automated network operation
Would not need to be aware beforehand of the virtual network functions which might be deployed in their servers.

NETWORK OPERATORS

Define a set of requirements for network functions to be deployed and their target performance
Might be partially unaware of low-level details of each network function's HW requirements: **Provision & management can be uniform & automated.**

BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

MANO NFV Orchestration Overview

L2/L3 Network Fabric Components & Network Connectivity Functions

Switches*
Routers
Server NIC
Virtual Network Overlays
vNIC
vswitch

NOTES:
*includes OpenFlow Switches and Controllers
** It could be single Tier or Multi-Tier App

Network Functions

Firewalls
Load Balancer
Traffic Steering
NAT
L4-L7 Routing
WAN routers
CDN
EPC
...

Compute

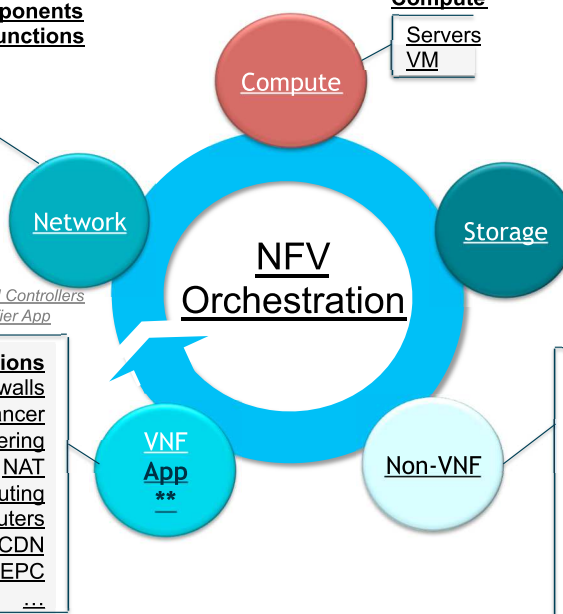
Servers
VM

Storage

Storage
Arrays
Local
Storage
SSD
SAN
NAS
LUN

Network Functions

Firewalls
Load Balancer
Traffic Steering
NAT
L4-L7 Routing
WAN routers
CDN
EPC
...

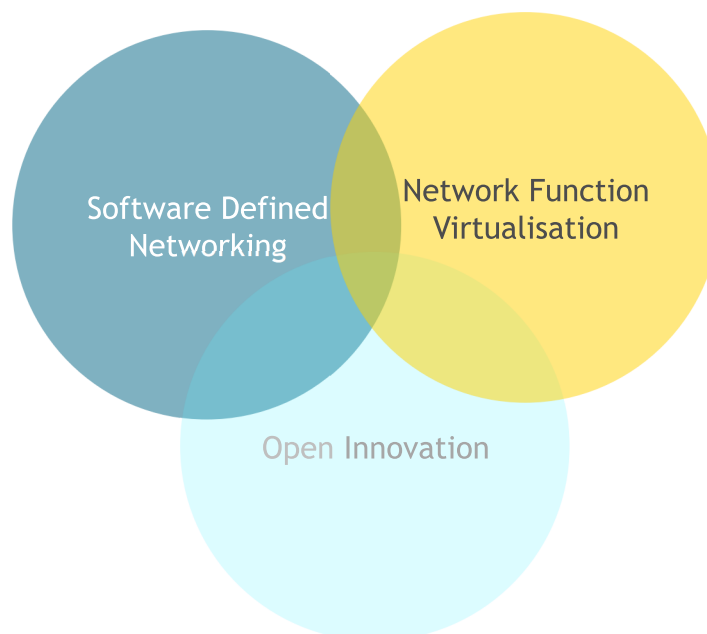


BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica



Components for a “liquid network”



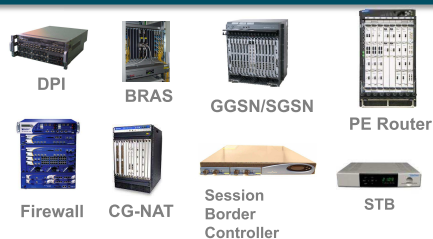
What is the promise of Network Virtualisation?

It is an opportunity to **build moudable Networks** and **redefine the Architecture**:

- Makes the **infrastructure uniform**
- **Reduces IoT complexity**
- Improves **management of risk** in a changing and ambiguous environment
- Introduces **capacity in an easy and flexible way**
- Fosters **competition** (new entrants) and **innovation**
- Prevents **hardware scale** from being an entry barrier

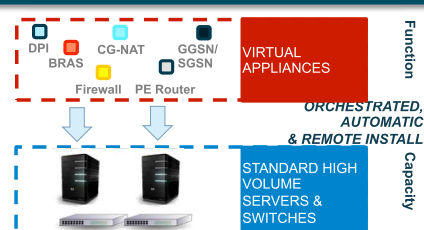


Traditional Network Model: APPLIANCE APPROACH



- Network functionalities are **based on specific HW** with **specific SW** linked to **HW vendors**
- **One physical node per role**

Virtualised Network Model: VIRTUAL APPLIANCE APPROACH

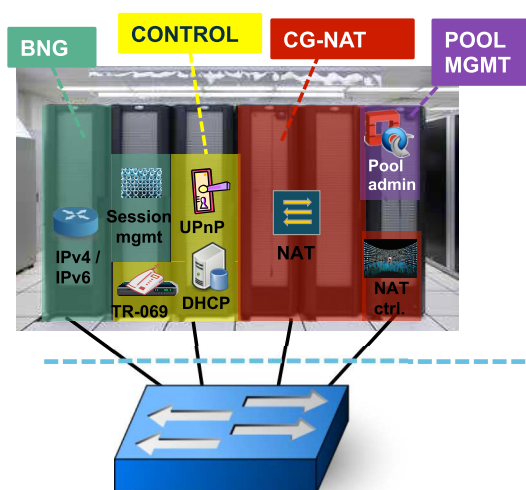


- When possible, network functionalities are **SW-based over COTS HW**
- **Multiple roles over same HW**

BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

A simple equation to define Network Virtualisation:
NV = NFV + SDN



NFV

SW-defined network functions

- Separation of HW and SW
- No vertical integration
 - HW vendor ≠ SW vendor ≠ Mgmt vendor
- Once network elements are SW-based, HW can be managed as a pool of resources

SDN

Interconnecting Virtual Network Functions (a.k.a. backplane)

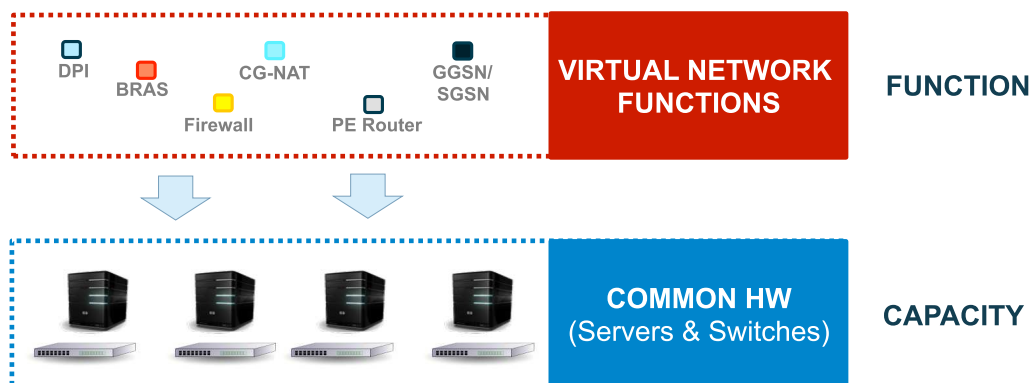
- Separation of control and data plane
- Easy orchestration with SW domain

BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

Network Virtualisation provides a mean to make the network more flexible, taking for granted a common HW layer

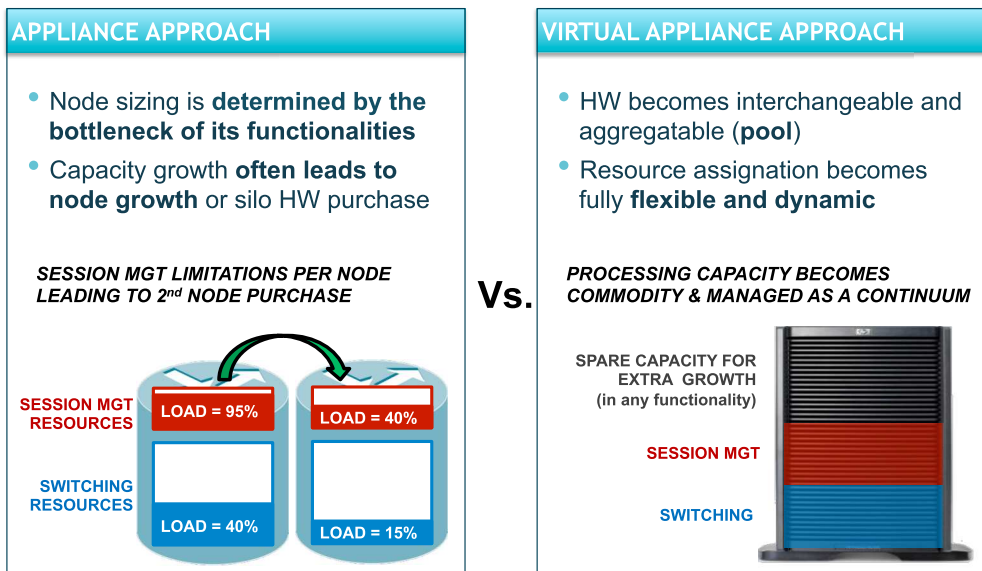
Network functions are fully defined by SW, minimising dependence on HW constraints



BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

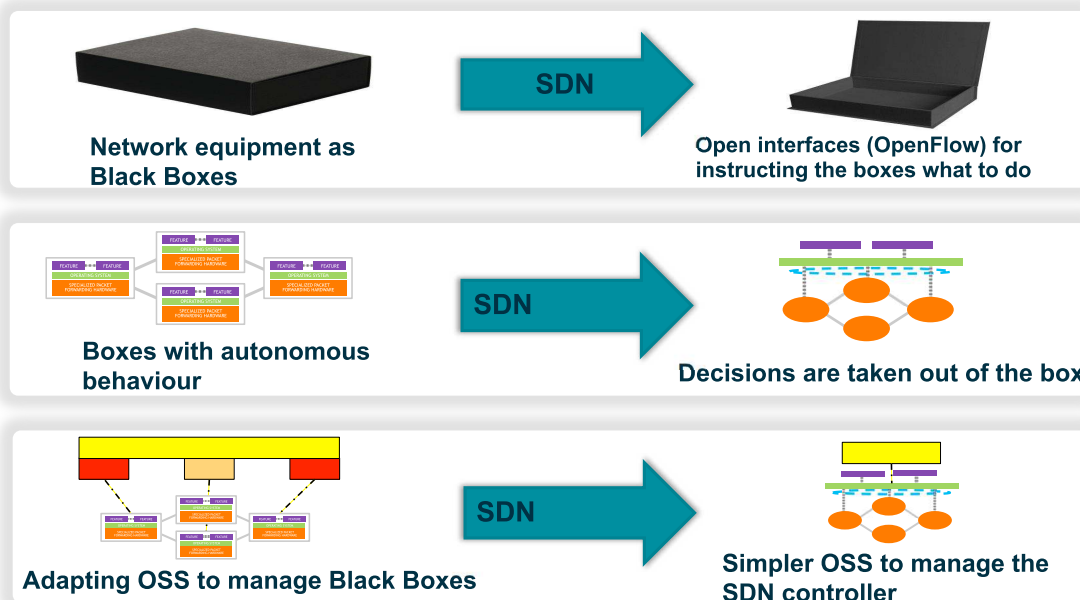
Network virtualisation helps reducing network management complexity, as HW can be treated as a pool of resources



BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

Software Defined Networking provides a first mean to improve operation and control of networks



BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

This new network model will help us to deeply transform our factory

Network Paradigm Change

- **Computing principles** used in IT world are beginning to be applied in telecoms by the means of **Network Virtualization IP** as common language for **all services**, included traditional Telco ones
- **Network virtualisation** enabling network re-programmability & agile service creation

Operation Model Change

Global E2E vision instead traditional silo model, not linked to monolithic OSS

Organization Model Change

Breaking the traditional model mapping isolated network domains

BE MORE
DISCOVER, DISRUPT, DELIVER

Telefonica

DISCOVER_

DISRUPT_

DELIVER_

BE MORE_

Telefonica

BE MORE_

A.3 Towards the Superfluid (Network) Cloud



Empowered by Innovation

NEC

Towards the Superfluid (Network) Cloud

Felipe Huici
[felipe.huici@neclab.eu]

NEC Laboratories Europe

Motivation

Virtualization and cloud deployments have brought great benefits

- OPEX/CAPEX reduction (fewer servers, lower cooling and power costs)
- Faster deployment
- Better disaster recovery
- Flexibility through migration
- Isolation, multi-tenancy

Can we improve things further, making the cloud more “fluid”?

- High consolidation (Hundreds? Thousands of VMs?)
- On-the-fly service instantiation (in milliseconds)
- Fast migration (hundreds of milliseconds?)
- High throughput (10-40+ Gb/s)

Talk Overview

Novel technologies and optimizations

1. ClickOS: High performance NFV
2. Minicache: Virtualized content caches
3. VALE: High performance, modular, energy efficient SW switch
4. Massive consolidation: thousands of VMs on a single server

Check out our open source portal!

- <http://cnp.neclab.eu/>

Cloud Networking Performance Lab

Experimenting with Flexible, High-Speed
Network Functions for the Cloud

[Learn more](#)

[Download](#)

Modular VALE: A Blazingly Fast Software Switch

With our VALE extensions and contributions you get over 200 Gbps of switching capacity and even allowing to extend it with your own lookup and filtering functions. Check it out!

[View details >](#)

Streamlined, High-Speed Virtualized Packet I/O

Our Xen optimizations result in 10 Gbps throughput for almost all packet sizes on a single CPU core, scaling up to 40 Gbps on an hypersensitive old server. Experience one of the most efficient packet I/O pipes in a virtualization technology.

[View details >](#)

Tiny, Agile Virtual Machines for Network Processing

The ClickOS Xen VM requires only 5 MB to run, boots in just ~30 milliseconds and over a hundred of them can be concurrently run on a single, hypersensitive old server. Massive and nimble consolidation at your fingertips!

[View details >](#)

1. ClickOS: High Performance NFV*






*ClickOS and the Art of Network Function Virtualization
NSDI 2014*

NFV: Shifting Middlebox Processing to Software

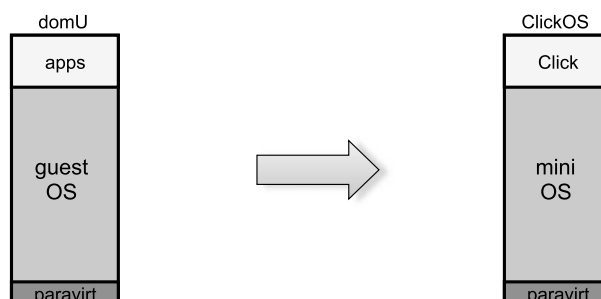
- Can share the same hardware across multiple users/tenants
- Reduced equipment/power costs through consolidation
- Safe to try new features on a operational network/platform
- But can it be built using commodity hardware while still achieving high performance?
- ClickOS: tiny Xen-based virtual machine that runs the Click modular router software

From Thought to Reality - Requirements

ClickOS

- | | |
|----------------------|--|
| ■ Fast Instantiation |  < 20 msec boot times |
| ■ Small footprint |  5MB when running |
| ■ Isolation |  provided by Xen |
| ■ Performance |  10Gb/s line rate*
45 µsec delay |
| ■ Flexibility |  provided by Click |

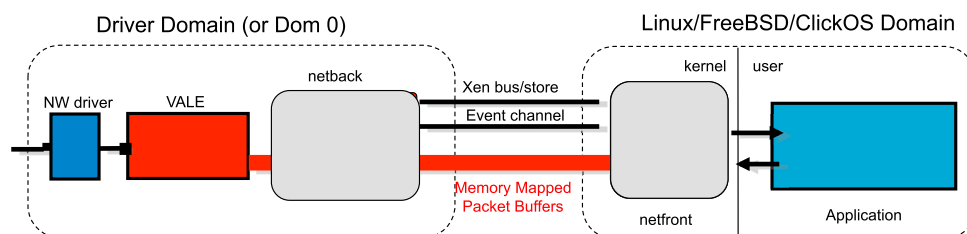
What's ClickOS?



Work consisted of:

- Build system to create ClickOS images
- Emulating a Click control plane over MiniOS/Xen
- Reducing boot times
- Optimizations to the data plane
- Implementation of a wide range of middleboxes

Data Plane Optimizations



Introduce VALE/netmap as backend switch in XEN

- Same switch is available also for KVM/QEMU

Permanently map grants with backend (not once per packet)

Bypass kernel network stack for high speed packet I/O

Larger I/O request batches

Split interrupts for transmission and receipt

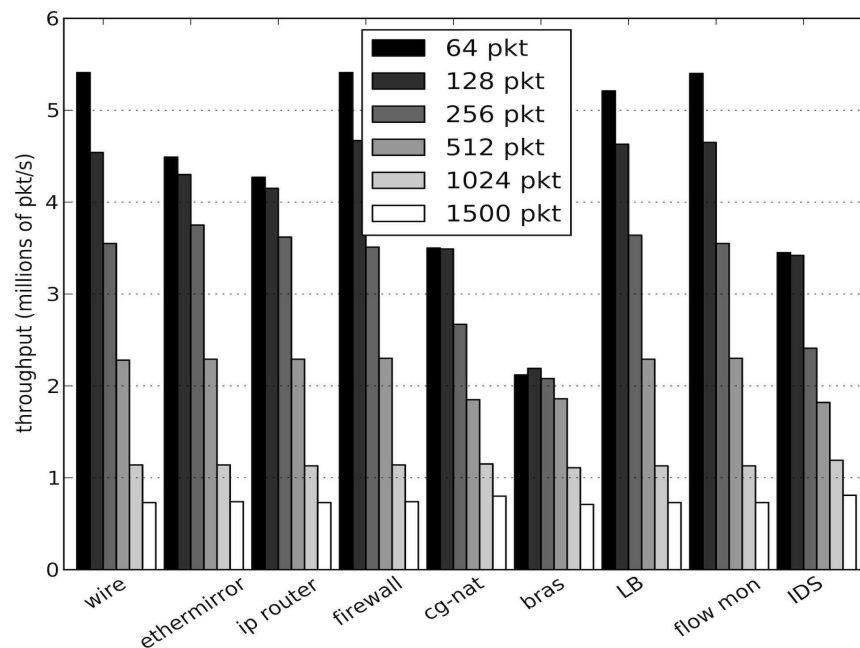
Optimizations result in 10Gb/s line rate for almost all packet sizes

Experiment Setup

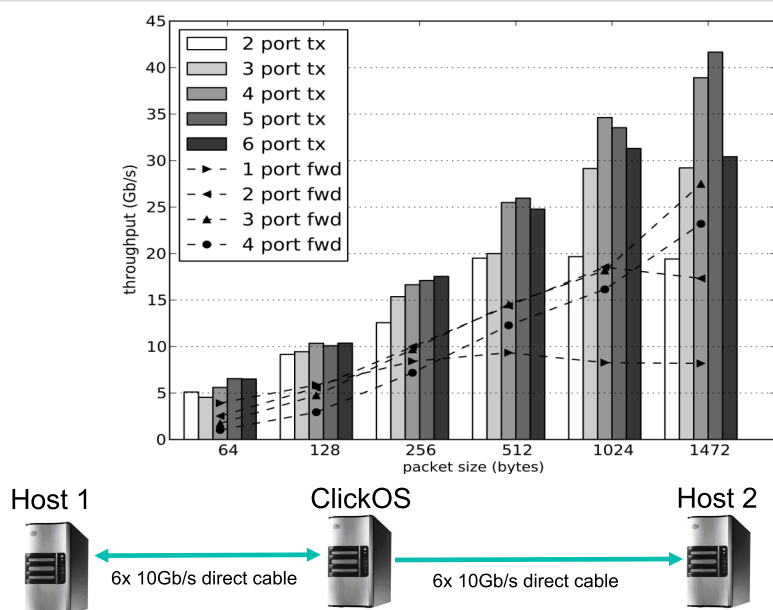


Intel Xeon E1220 4-core 3.2GHz (Sandy bridge)
16GB RAM, 2x Intel x520 10Gb/s NIC.
One CPU core assigned to Vms, 3 CPU cores Domain-0
Linux 3.6.10

Middlebox Performance (single VM)

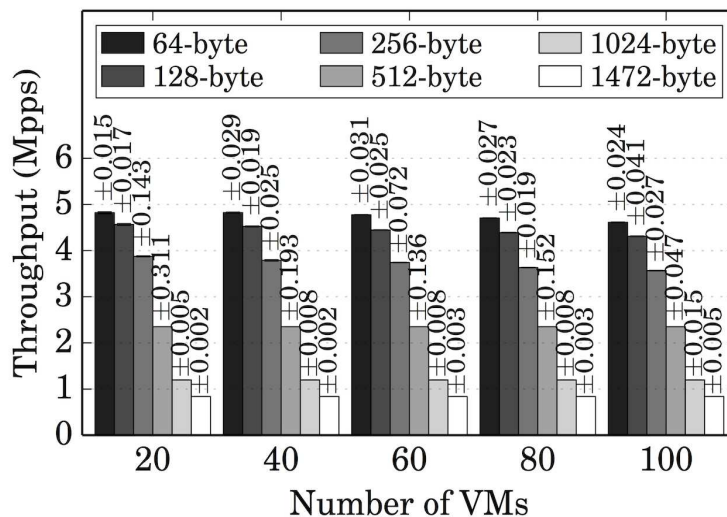


Scaling out – Multiple NICs/VMs



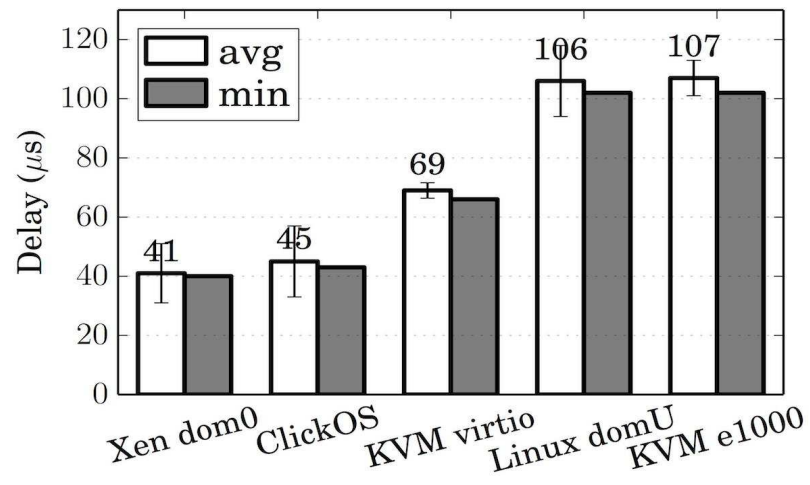
Intel Xeon E1650 6-core 3.2GHz, 16GB RAM, dual-port Intel x520 10Gb/s NIC.
3 cores assigned to VMs, 3 cores for dom0

Scaling Out – 100 VMs, Aggregate Throughput



Intel Xeon E1650 6-core 3.2GHz, 16GB RAM, dual-port Intel x520 10Gb/s NIC.
3 cores assigned to VMs, 3 cores for dom0

ClickOS Delay vs Other Systems



2. minicache: Virtualized Content Caches*

** Towards Minimalistic, Virtualized Content Caches with Minicache
CoNEXT Hot Middlebox 2013*

Overview – Virtualizing CDNs

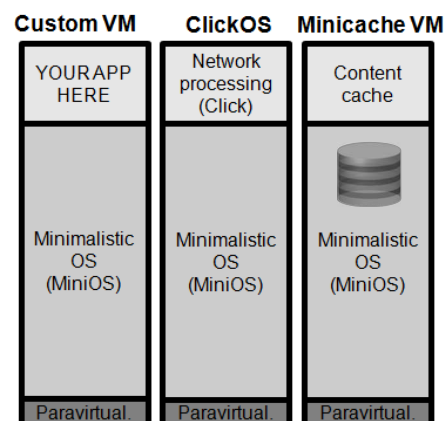
- Current trend: Internet is becoming a “videonet”
 - 57% of Internet traffic today is video
 - 1/3 of peak traffic is the US is Netflix
 - These numbers will continue to grow
- Large majority of videos are delivered by CDNs (e.g., Akamai)
 - CDN performance is dependent on distance between content and users
 - Deploy content caches in operator networks
- More recently, trend towards renting infrastructure at the network's edge
 - Micro DCs at PoPs
 - Mobile Edge Computing (e.g., next to base stations)

What's Minicache?

- Minimalistic VM for serving (video) content (CDN node)
 - Based on MiniOS
 - Uses lwIP (1.4.1) as network stack
 - Simple hash-based filesystem (SHFS)
 - Simple HTTP server
 - Interactive Shell (uSh)

- Idea: create virtual CDNs as needed, no need for upfront investments

- Added bonus: a more general VM than ClickOS, can support other types of processing



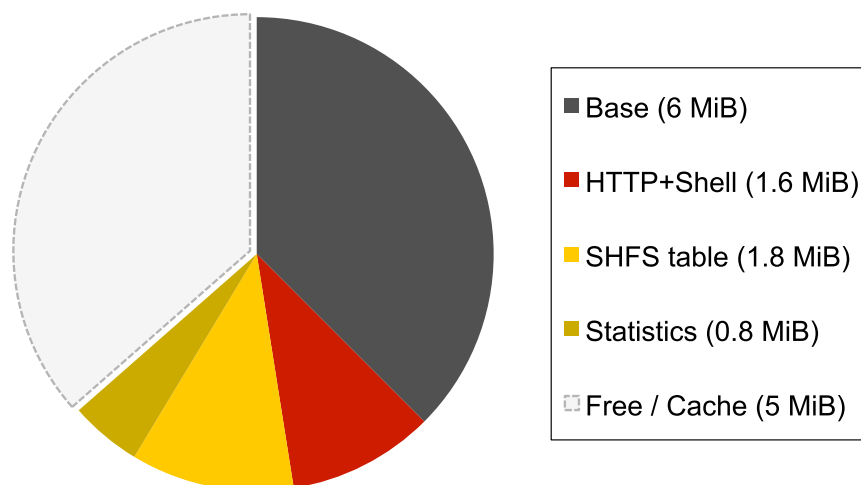
Memory Footprint

- Minimum: 8MB
 - SHFS mount adds extra memory:

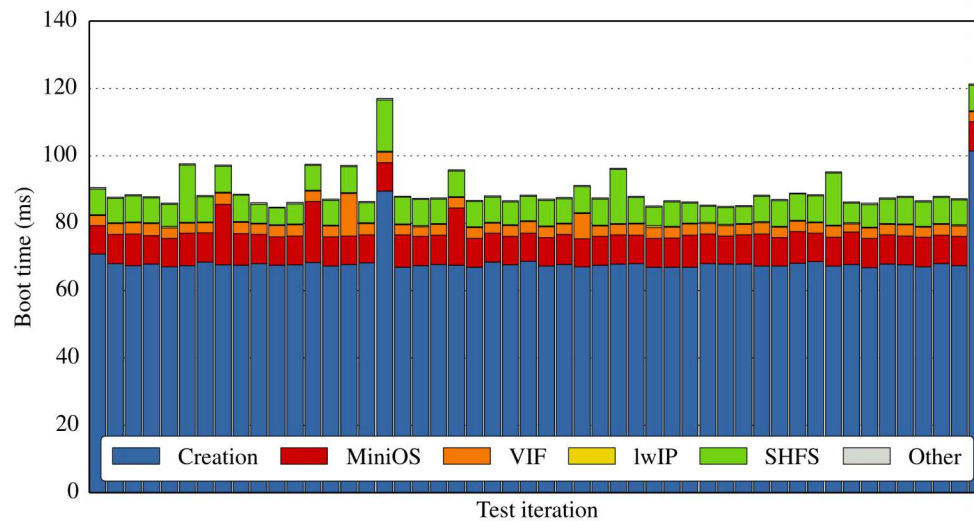
#Entries	SHFS Table size	Allocation in RAM (without stats)
512	128 KiB	230 KiB
1024	256 KiB	460 KiB
2048	512 KiB	922 KiB
4096	1 MiB	1.8 MiB
8192	2 MiB	3.6 MiB
16384	4 MiB	7.2 MiB
32768	8 MiB	14.4 MiB
65536	16 MiB	28.8 MiB

Memory Footprint - Breakdown

- 16MB Minicache VM
- SHFS mounted with 4K entries



Boot-up Times

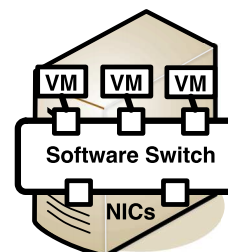


3. VALE: a High Performance, Modular, Software Switch

Motivation

Software switches play an increasingly important role

- Interconnection between VMs and NICs
- SDN, Network Function Virtualization (NFV)



Requirements

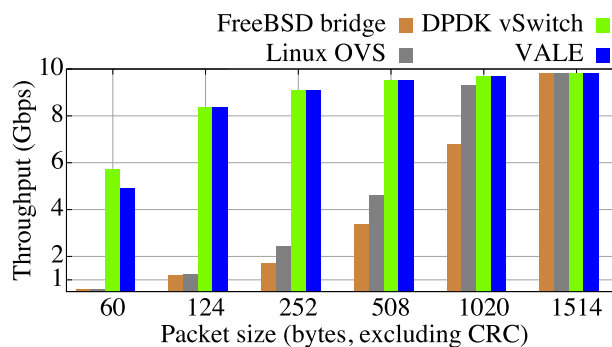
- Throughput (e.g., 10 Gbps)
- Scalability (e.g., 100 ports)
- Flexibility (i.e., forwarding decision and packet processing)
- Reasonable CPU utilization

Do existing software switches satisfy these requirements?

Existing Software Switches

OS standard switches lack high throughput

- Small packets are common (e.g., TCP SYNs, ACKs)



Recent switches lack scalability, flexibility and/or reasonable CPU utilization

	Throughput	CPU Usage	Density	Flexibility
FreeBSD switch	×	✓	✓	×
Linux switch	×	✓	✓	×
Open vSwitch	×	✓	✓	✓
Hyper-Switch	×	✓	×	✓
DPDK vSwitch	✓	×	×	✓
CuckooSwitch	✓	×	×	×

Our Contribution

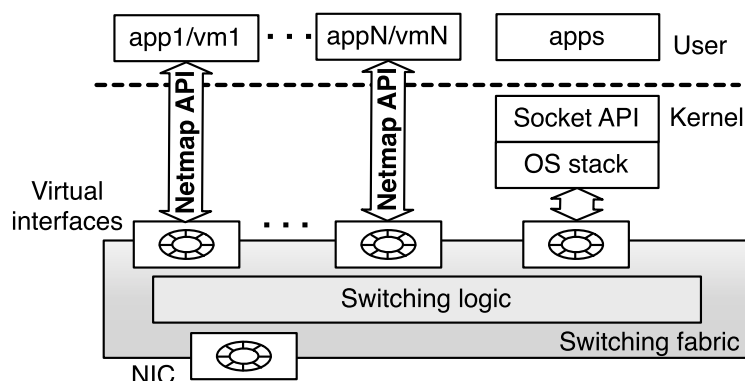
A scalable, modular software switch

- Ideal as a virtualization backend

Scalable packet forwarding algorithms

- Tens to hundreds of destination ports
- Concurrent senders to a common destination port

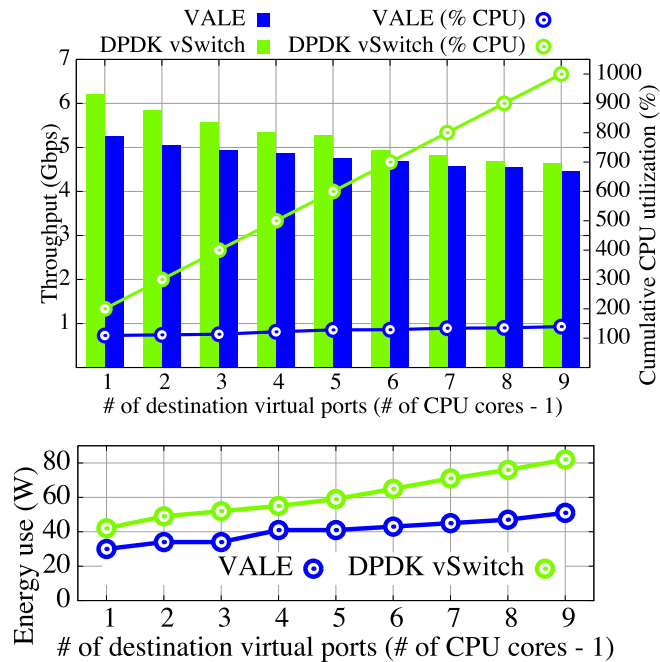
System Architecture



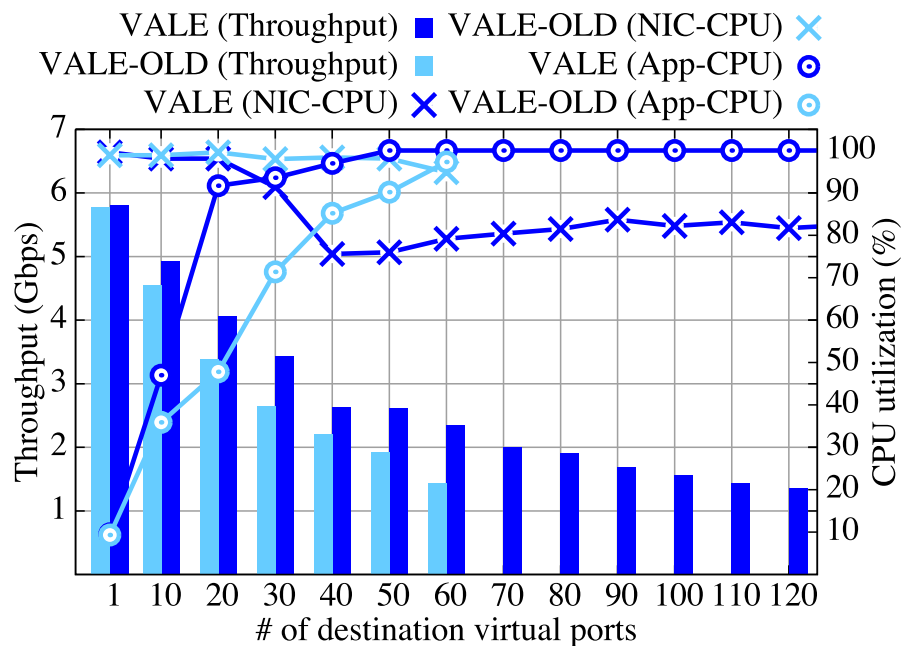
Switching fabric moves packets efficiently among ports → part of the system

Switching logic decides packet's destination → the user develops this

CPU utilization and Power Consumption, VALE vs OVDK



Port Scalability



4. Massive Consolidation*

**Towards Massive Server Consolidation
Xen Developer Summit, 2014*

Wouldn't it be Nice if...

- Thousands of guests on a single server, up to 100K
- Extremely fast domain creation, destruction and migration
 - Tens of milliseconds
 - Constant as number of guests increases

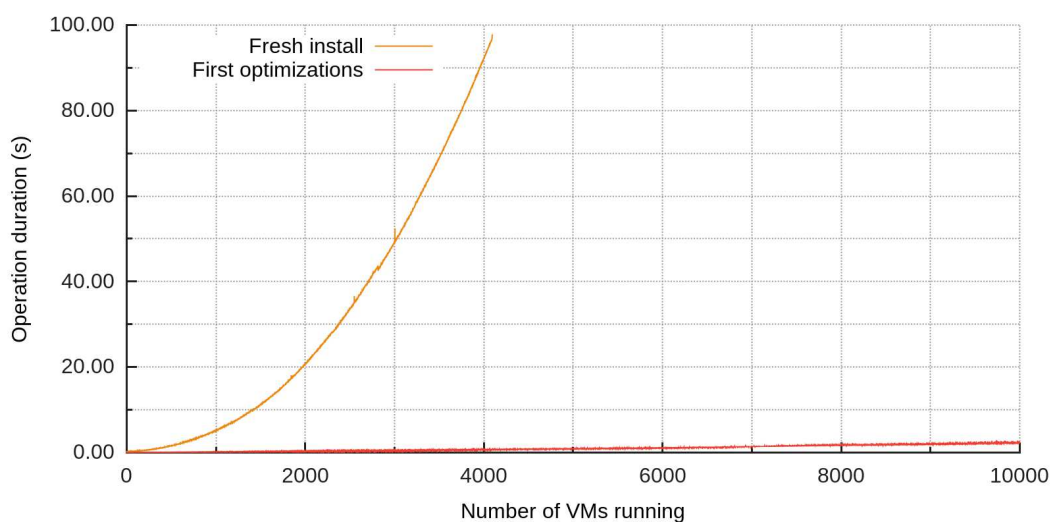
Two Types of Problems

- Hard limitations
 - Prevent guests from booting correctly
 - Only ~300 guests fully usable
- Performance limitations
 - Decreasing system performance
 - System (dom0) unusable after just a few hundred guests

First Optimizations

- Increase number of file descriptors in Linux
 - fixes console issues
- Increase number of PTYs in Linux
 - fixes console issues
- Upgrade to Xen 4.4 + Linux 3.14, kernel with NR_CPUS=4096
 - fixes # of event channels limit
- Use multiple instance of back-end switch
 - fixes # of virtual ports limitation

First Optimizations - 10K VM Boot Times



Server: 64 Cores @ 2.1GHz [4 x AMD Opteron 6376]
128GB RAM DDR3 @ 1333MHz

With Optimizations...

- Improvement: system is still usable after 10K guests
 - Although domain creation time is far from ideal
- However...
 - xenstored still CPU heavy
 - xenconsole still CPU heavy

Current Status

- Usable system running 10K guests
 - 10K guests actually working...
 - ...although idle most of the time
- Lower domain creation times
 - First domain: < 10ms
 - With 10K domains: < 100ms
- Currently working on
 - Xenconsole: switch from poll to epoll: CPU util down to 10% max
 - Improved XenStore (lixs, **Lightweight XenStore**)
 - Simplified control toolstack (xcl: **XenCtrl Light**)

Will it Work up to 100K VMs? Remaining Issues

- Improve lixs and Xenstore protocol
- Have guests doing useful work
- Scheduling
 - Number of guests much bigger than number of cores
 - With that many guests we'll have scheduling issues
- Reducing Memory Usage
 - Smaller image sizes
 - Share memory between guests booting same image

Wrap-Up

Conclusions

Introduced a number of technologies and technologies in support of a more “fluid” network cloud

- Massive consolidation
- On-the-fly service instantiation (in milliseconds)
- Fast migration (hundreds of milliseconds)
- High throughput (10-40+ Gb/s)

Tailor-made operating system, supports

- Network processing functions (e.g., firewall, tunnel endpoint, etc.)
- Content caching (MiniCache)
- **Your application!**

Ongoing and Future Work

- Integration with OpenStack/Neutron
- Started porting to KVM (OSv & MiniOS)
- Support for ARM platforms
 - Cubietruck already working
 - ARM64 when available



Cubietruck

- We're looking for operators for PoCs/trials...

Questions?

Cloud Networking Performance Lab

<http://cnp.neclab.eu>

felipe.huici@neclab.eu

Orchestrating a brighter world

NEC brings together and integrates technology and expertise to create the ICT-enabled society of tomorrow.

We collaborate closely with partners and customers around the world, orchestrating each project to ensure all its parts are fine-tuned to local needs.

Every day, our innovative solutions for society contribute to greater safety, security, efficiency and equality, and enable people to live brighter lives.

Empowered by Innovation

NEC

A.4 Road to the Federated Market and beyond



Road to the Federated Market and beyond

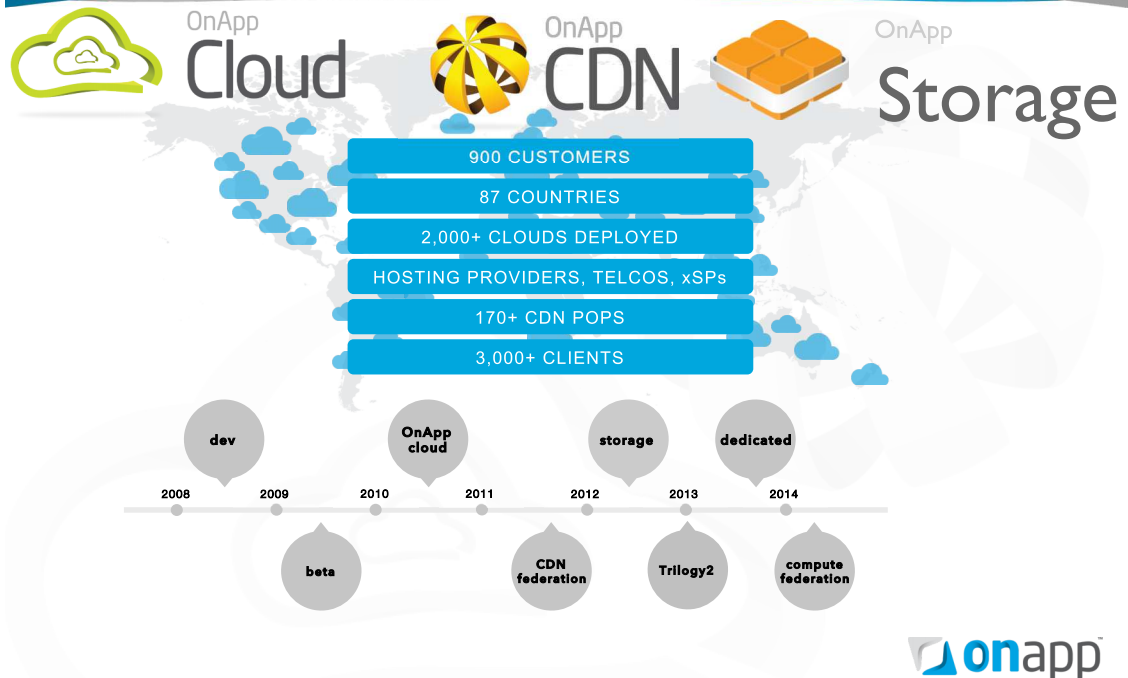
By John Thomson, 30th October 2014

Overview

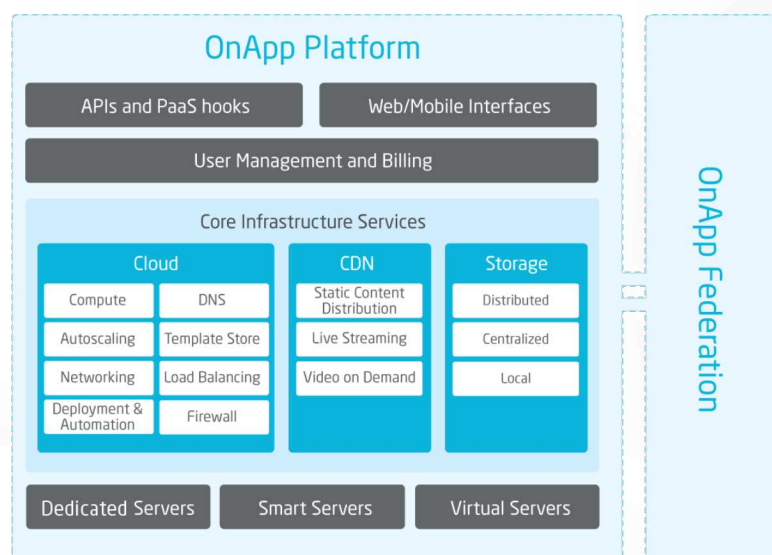
- Intro to OnApp
- Cloud / CDN / Storage
- Liquid Net
 - Trilogy2
- Federation
 - Market place
- Beyond



Who is OnApp



How OnApp provides an IaaS platform



Who are our customers?

Main customer base was

CLOUD PROVIDERS



Who are our customers?

TELCOs & CARRIERS



CLOUD PROVIDERS



MSPS & INTEGRATORS



CDN PROVIDERS



Vital statistics

1 in 3
public clouds

2000+
cloud deployments

3000+
global clients



CDN

Now have an established market of Cloud and CDN providers

Content and location providers can benefit from co-operating



<http://onapp.com/federation/locations/>



Why OnApp CDN?

- End-customers want globally responsive applications
- BUT**
- Global CDN too expensive to build for hosters
- Margins tight when reselling existing CDN solutions
- AND**
- Cloud Providers not able to fully monetise spare cloud infrastructure



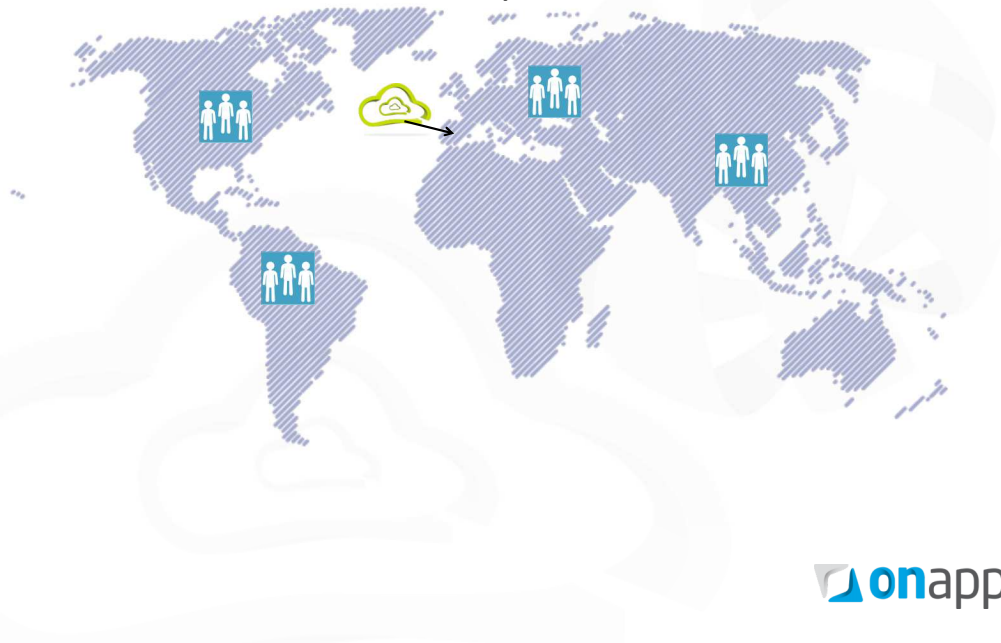
CDN – quick example

As an operator you may have a datacenter in Madrid with content



CDN – quick example

Customers from other locations may want the data



CDN – quick example

Expensive to go into new markets



CDN – quick example



With OnApp Federated CDN – choose local edge providers



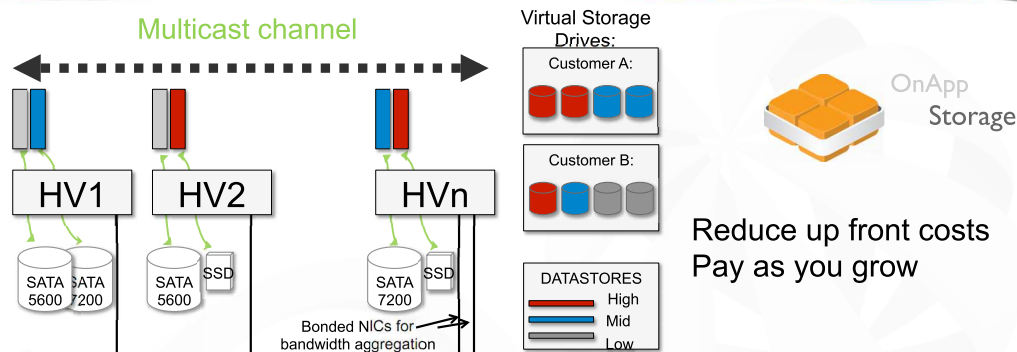
CDN – quick example



Edge providers can sell excess capacity



Integrated Storage and Cloudboot



Cloudboot



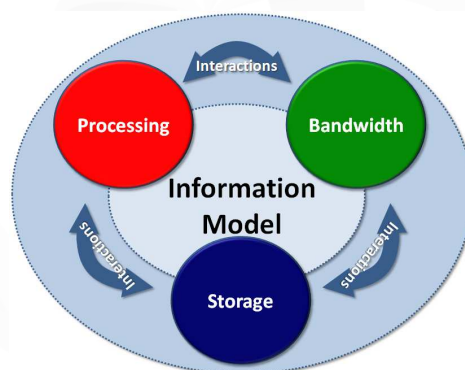
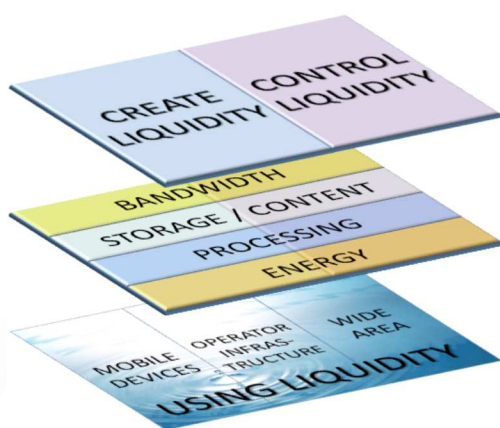
- Fast provisioning of Xen & KVM HV's, with no pre-installation
- Server boots over network (no local storage required) as a fully configured hypervisor, ready to host VMs
- Diskless boot enables full storage controller hardware passthrough to storage layer

Creating and managing liquid resources



trilogy 2

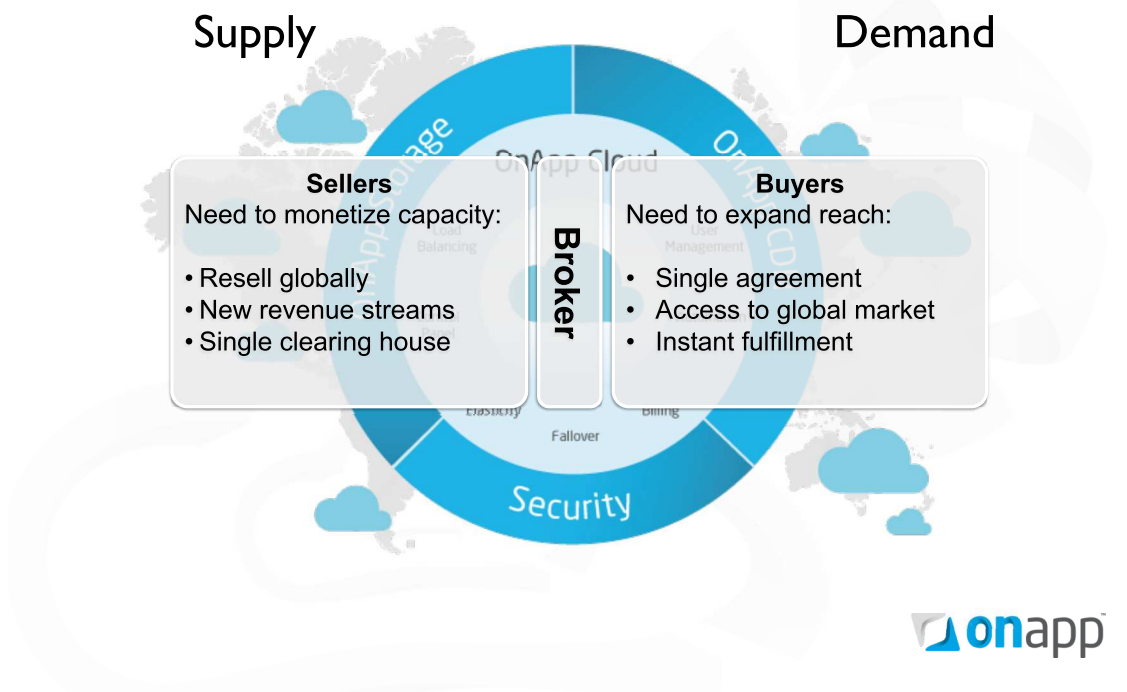
'Trilogy 2: Building the Liquid Net' is aimed at developing a new Internet architecture based on the concept of the liquid network



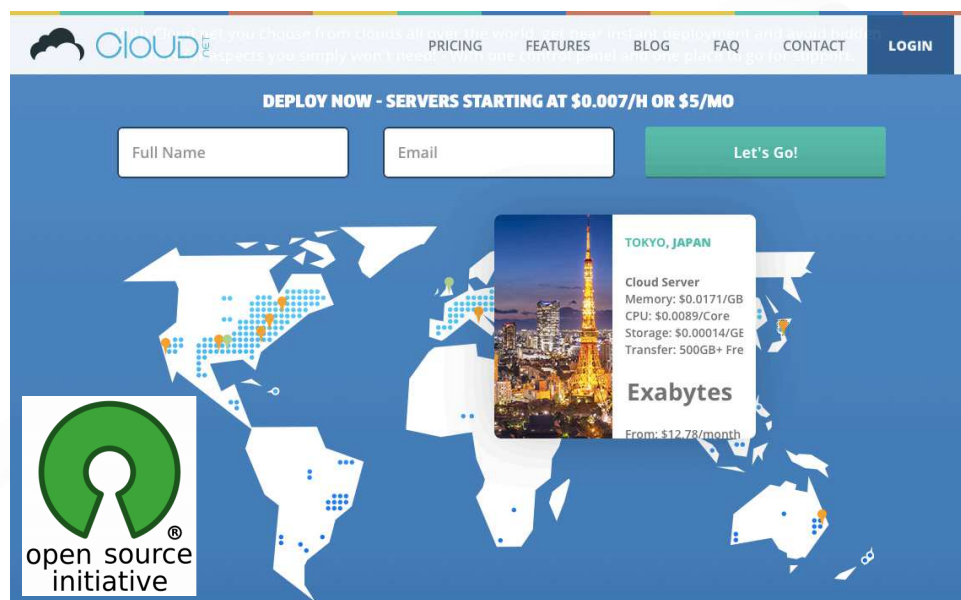
<http://trilogy2.eu/>



Liquidity through a Global Marketplace



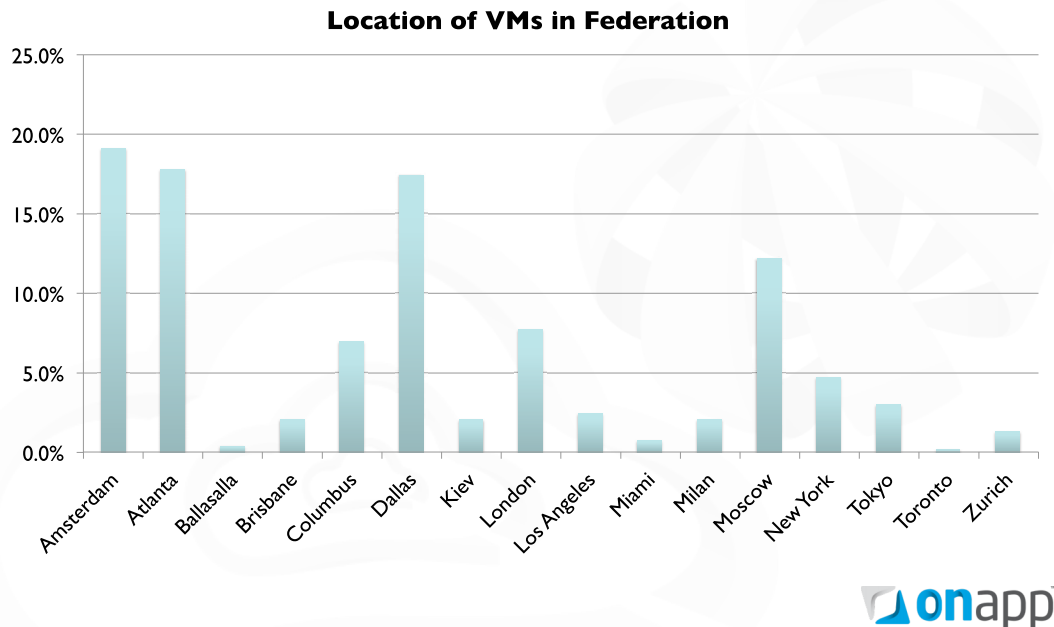
Trilogy2 – Cloud.Net



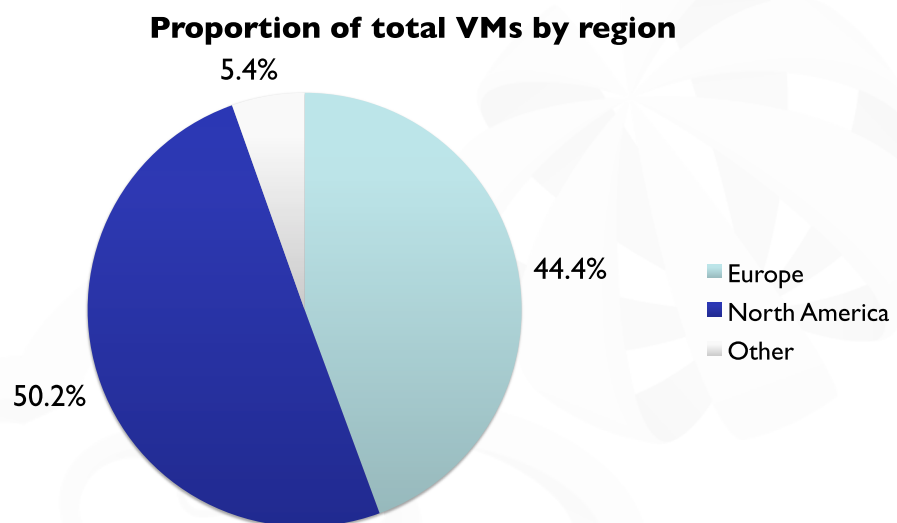
Cloud.Net – Website is live



Location of VMs in Federation



Cloud.net stats – early days



Advantage for end-users

- Provides choice
- Ability to select provider based on;
 - SLA, location, latency, cost, etc.
- Flexibility
 - Avoidance of walled gardens
 - Vendor lock in
- Scale up / scale to demand
- Change from CAPEX to OPEX
 - Reduce up-front costs



Advantage for service providers

- Advertise and sell spare capacity to an additional market/channel with service demand
- Market services at very low cost
 - Cost Per Acquisition for new customers is very low
- Provide global footprint to customers as their own



What else is OnApp doing?

OnApp acquires SolusVM to bolster federated cloud network

September 16, 2014 Written by Business Cloud News

Print

OnApp, which provides cloud orchestration and federation software for infrastructure as a service providers, has acquired rival SolusVM for an undisclosed sum. Kosten Metreweli, chief commercial officer at OnApp told BCN the move will strengthen the company's ambitious federated cloud network due to go live later this year.

London-based OnApp made its name developing a federated CDN service that leverages a connected network of service providers to feed up content to customers most proximal to those providers. Now it's trying to do the same thing for infrastructure as a service, offering its cloud orchestration and federation software (a kind of proprietary OpenStack) to telcos, managed service providers and others looking to get into the IaaS game.



OnApp has acquired SolusVM in a bid to strengthen its cloud federation



<http://www.businesscloudnews.com/2014/09/16/onapp-acquires-solusvm-to-bolster-federated-cloud-network/>



What else is OnApp doing?

OnApp acquires SolusVM to bolster federated cloud network

September 16, 2014 Written by Business Cloud News

Print

OnApp buys SolusVM to boost the demand side of its federated marketplace

GIGAOM

by David Meyer SEP. 16, 2014 - 12:44 AM PDT

2 Comments

AV A

SUMMARY: SolusVM will give OnApp a more bare-bones deployment option for those who want it, but more importantly it's intended to increase demand for the spare capacity rattling around in the OnApp marketplace.



<http://www.businesscloudnews.com/2014/09/16/onapp-acquires-solusvm-to-bolster-federated-cloud-network/>

<https://gigaom.com/2014/09/16/onapp-buys-solusvm-to-boost-the-demand-side-of-its-federated-marketplace/>





- OnApp acquired SolusVM – 16th September
- Way of seeding the Federated Market
- 2000 additional providers
 - Growing the providers to encourage demand
- SolusVM - leader in entry level VM hosting
- Large deployed footprint
 - Low cost of entry for basic, cloud hosting services

<http://onapp.com/solusvm/>



Looking ahead

- Emergence of Virtual Service Providers
 - Mobile virtual service provider
- True resource liquidity is around the corner
 - Application mobility between locations
- New market model where end-users are more aware and mobile
 - Dynamic market
- Managing a cloud of clouds



Conclusion

- Managing infrastructure is a hard problem
 - Solutions such as those from OnApp allow companies to solve this and focus on core business
- The market is changing – Federation is growing
 - Will go GA at the end of 2014
- Cloud resources are becoming more liquid



Thanks!

John Thomson

John.thomson@onapp.com



A.5 Data Centre (DC) Topologies for NFV

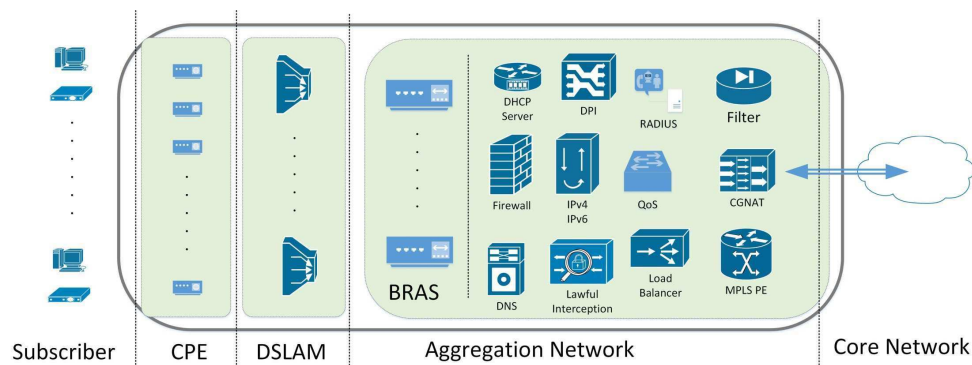
DC Topologies for NFV

Juan Brenes

jbrenes@it.uc3m.es

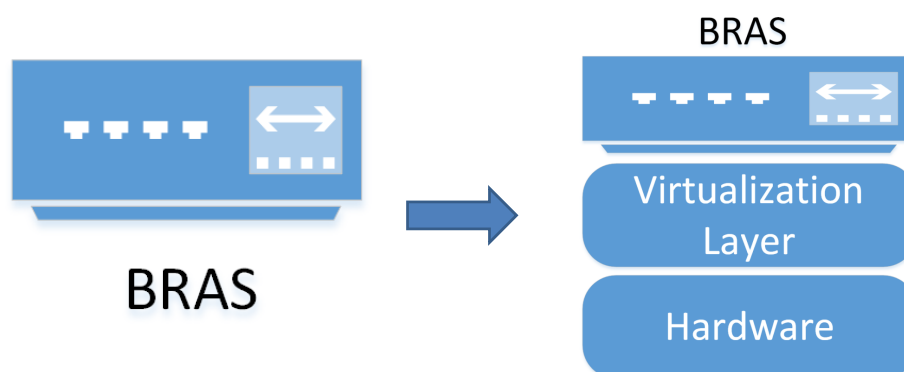
Universidad Carlos III de Madrid

What are we going to talk about?



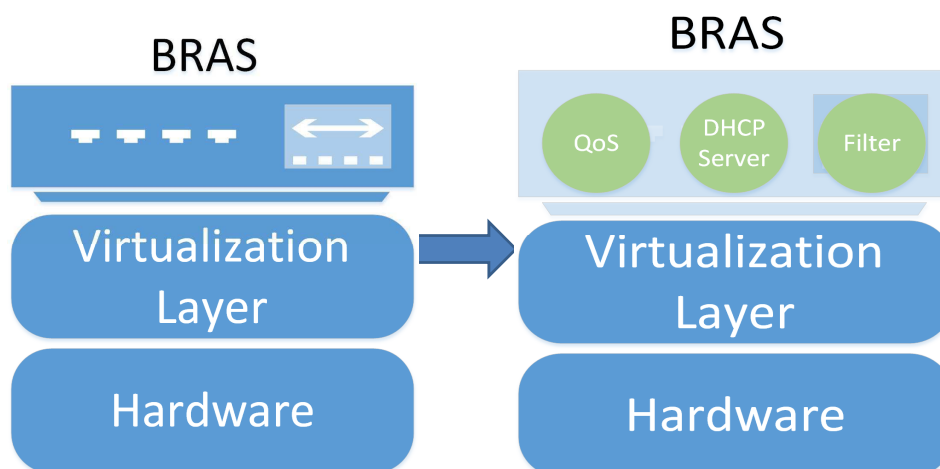
How is NFV going to affect the traffic?

First Step : Virtualize the whole equipment as one Virtual Machine



How is NFV going to affect the traffic?

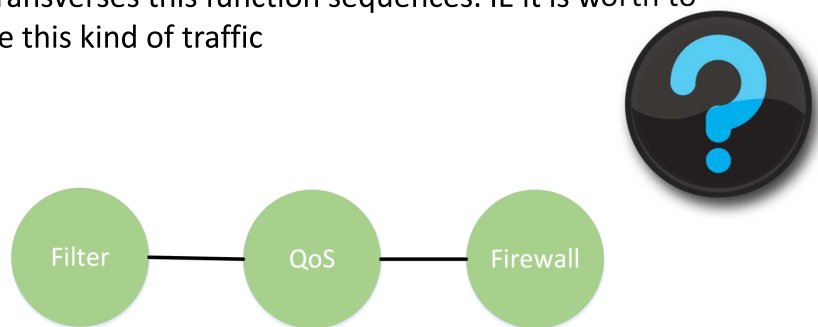
Second Step: Split the VM in independent VM representing the different functions



How do this VM functions interact?

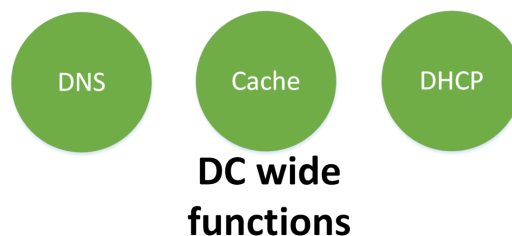
We consider that most of the functions are executed in the same order every time, generating sequences of functions.

As a design hipotesis we say that a significant amount of the traffic transverses this function sequences. IE it is worth to optimize this kind of traffic

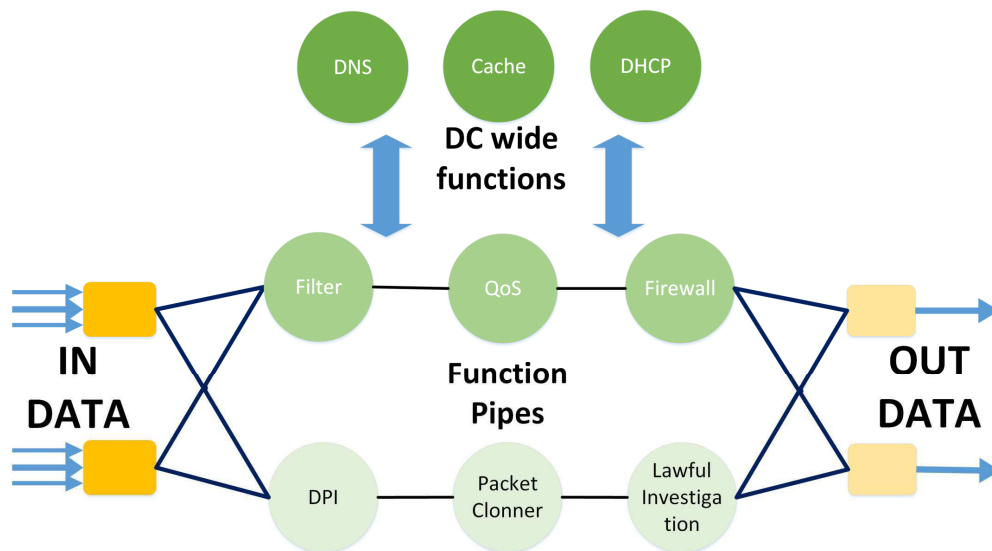


DC Wide Functions

Besides the functions that build the sequences, we consider there are some other functions to be modeled.

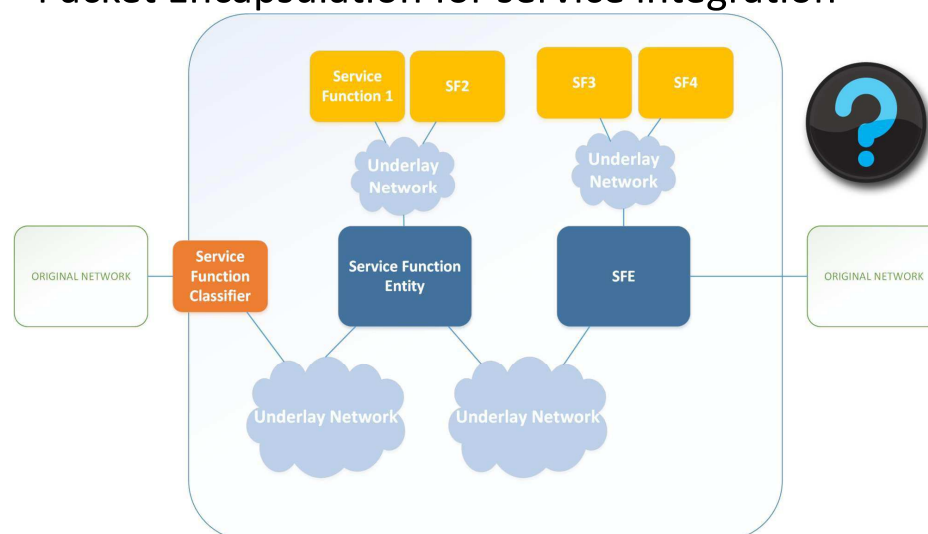


Data Center Model

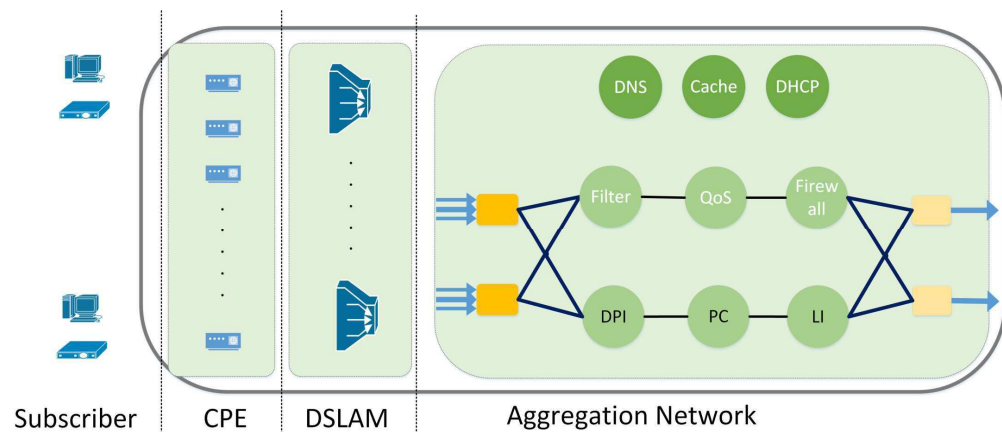


Service Function Chaining (SFC)

Packet Encapsulation for service integration



Including the model in the Architecture



Are current topologies efficient for the new constraints?

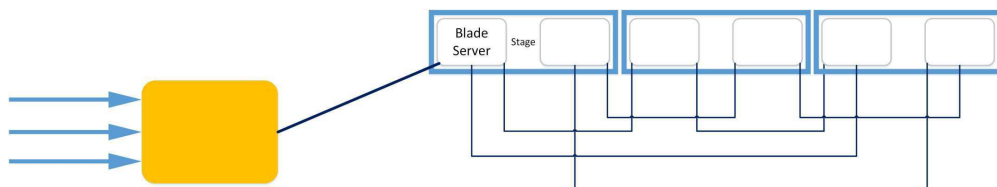
We argue that current topologies are not efficient to handle the linearity of the new traffic pattern.

New topology constraints

- 2 Different structures with different requirements:
 - A sequential structure, to couple with the linearity of the function chains.
 - A non sequential structure, to access the functions that are outside the pipes, communicate between chains and to allow stage jumps.

Sequential structure constraints

- All in all out.
- Capacity reutilization.
- Easy to grow stages.
- Stage redundancy



Key Questions - Chains

Which are the key functions and which are the main service chains ?

Key Questions - Scalability

Scale Horizontal vs Vertical ?

Which are today requirements ?

Bibliography

- [1] Network Functions Virtualisation; Infrastructure; Methodology to describe Interfaces and Abstractions. Group Specification ETSI GS NFV-INF 007, ETSI NFV ISG, October 2014.
- [2] Network functions virtualisation (nfv); nfv performance and portability best practises. Group Specification RGS/NFV-PER001ed112, ETSI NFV ISG, dec 2014.
- [3] Network functions virtualisation (nfv); proof of concepts; framework. Group Specification RGS/NFV-PER002ed112, ETSI NFV ISG, dec 2014.
- [4] Network Functions Virtualisation; Part 1 Infrastructure Architecture; Sub-part 1 Overview. Group Specification ETSI GS NFV-INF 001-1 v0.3.12, ETSI NFV ISG, November 2014. (work in progress).
- [5] Network Functions Virtualisation; Security; Problem Statement. Group Specification ETSI GS NFV-SEC 001, ETSI NFV ISG, October 2014.
- [6] Sam Aldrin, Ram Krishnan, Nobo Akiya, Carlos Pignataro, and Anoop Ghanwani. Service Function Chaining Operation, Administration and Maintenance Framework. Internet-Draft draft-aldrin-sfc-oam-framework-01, IETF Secretariat, October 2014. I-D Exists.
- [7] L. Andersson, H. van Helvoort, R. Bonica, D. Romascanu, and S. Mansfield. Guidelines for the Use of the "OAM" Acronym in the IETF. Technical Report 6291, Internet Engineering Task Force, June 2011.
- [8] Pedro Aranda, Daniel King, and Masaki Fukushima. Virtualization of Content Distribution Network Use Case. Internet-Draft draft-aranda-vnfpool-cdn-use-case-00, IETF Secretariat, October 2014. I-D Exists.
- [9] F. Audet and C. Jennings. Network Address Translation (NAT) Behavioral Requirements for Unicast UDP. Technical Report 4787, Internet Engineering Task Force, January 2007.
- [10] Marcelo Bagnulo. Secure MPTCP. Internet-Draft draft-bagnulo-mptcp-secure-00, IETF Secretariat, February 2014.
- [11] Marcelo Bagnulo, Christoph Paasch, Fernando Gont, Olivier Bonaventure, and Costin Raiciu. Analysis of MPTCP residual threats and possible fixes. Internet-Draft draft-ietf-mptcp-attacks-02, IETF Secretariat, July 2014. IESG Evaluation::Revised I-D Needed.
- [12] Sebastien Barre, Gregory Detal, and Olivier Bonaventure. TFO support for Multipath TCP. Internet-Draft draft-barre-mptcp-tfo-00, IETF Secretariat, July 2014.
- [13] Steven Bellovin. Problem Statement and Requirements for a TCP Authentication Option. Internet-Draft draft-bellovin-tcpsec-01, IETF Secretariat, July 2007.
- [14] Olivier Bonaventure. MPTLS : Making TLS and Multipath TCP stronger together. Internet-Draft draft-bonaventure-mptcp-tls-00, IETF Secretariat, October 2014. I-D Exists.
- [15] Olivier Bonaventure. Multipath TCP timestamp option. Internet-Draft draft-bonaventure-mptcp-timestamp-00, IETF Secretariat, October 2014. I-D Exists.
- [16] Olivier Bonaventure and Christoph Paasch. Experience with Multipath TCP. Internet-Draft draft-bonaventure-mptcp-experience-00, IETF Secretariat, July 2014.
- [17] Olivier Bonaventure, Christoph Paasch, and Gregory Detal. Experience with Multipath TCP. Internet-Draft draft-ietf-mptcp-experience-00, IETF Secretariat, September 2014. I-D Exists.
- [18] Olivier Bonaventure, Christoph Paasch, and Gregory Detal. Processing of RST segments by Multipath TCP. Internet-Draft draft-bonaventure-mptcp-rst-00, IETF Secretariat, July 2014.

- [19] D. Borman, B. Braden, V. Jacobson, and R. Scheffenegger. TCP Extensions for High Performance. Technical Report 7323, Internet Engineering Task Force, September 2014.
- [20] B. Briscoe and J. Manner. Byte and Packet Congestion Notification. Technical Report 7141, Internet Engineering Task Force, February 2014.
- [21] B. Briscoe, R. Woundy, and A. Cooper. Congestion Exposure (ConEx) Concepts and Use Cases. Technical Report 6789, Internet Engineering Task Force, December 2012.
- [22] Bob Briscoe. Extended TCP Option Space in the Payload of an Alternative SYN. Internet-Draft draft-briscoe-tcpm-syn-op-sis-03, IETF Secretariat, October 2014. Replaced by draft-briscoe-tcpm-inner-space.
- [23] Bob Briscoe. Inner Space for TCP Options. Internet-Draft draft-briscoe-tcpm-inner-space-01, IETF Secretariat, October 2014. I-D Exists.
- [24] Bob Briscoe. Network Performance Isolation using Congestion Policing. Internet-Draft draft-briscoe-conex-policing-01, IETF Secretariat, February 2014.
- [25] Bob Briscoe. Reusing the IPv4 Identification Field in Atomic Packets. Internet-Draft draft-briscoe-intarea-ipv4-id-reuse-04, IETF Secretariat, February 2014.
- [26] Bob Briscoe. The Echo Cookie TCP Option. Internet-Draft draft-briscoe-tcpm-echo-cookie-00, IETF Secretariat, October 2014. I-D Exists.
- [27] Bob Briscoe, Richard Scheffenegger, and Mirja Kuehlewind. More Accurate ECN Feedback in TCP. Internet-Draft draft-kuehlewind-tcpm-accurate-ecn-03, IETF Secretariat, July 2014.
- [28] Bob Briscoe and Murari Sridharan. Network Performance Isolation in Data Centres using Congestion Policing. Internet-Draft draft-briscoe-conex-data-centre-02, IETF Secretariat, February 2014.
- [29] Mike Bursell(rap.) and Kurt Roemer(rap.). Network functions virtualisation(nfv); nfv security; security and trust guidance. Group Specification DGS/NFV-SEC003, ETSI NFV ISG, dec 2014.
- [30] Y. Cheng, J. Chu, S. Radhakrishnan, and A. Jain. Tcp fast open. Technical Report 7413, IETF Secretariat, February 2014.
- [31] Es-nog. <http://www.esnog.net>, Oct 2014.
- [32] Stephen Farrell and Hannes Tschofenig. Pervasive Monitoring Is an Attack. Internet-Draft draft-farrell-perpass-attack-06, IETF Secretariat, February 2014.
- [33] A. Ford, C. Raiciu, M. Handley, and O. Bonaventure. TCP Extensions for Multipath Operation with Multiple Addresses. Technical Report 6824, Internet Engineering Task Force, January 2013.
- [34] Alan Ford, Costin Raiciu, Mark Handley, and Olivier Bonaventure. TCP Extensions for Multipath Operation with Multiple Addresses. Internet-Draft draft-ietf-mptcp-rfc6824bis-03, IETF Secretariat, October 2014. I-D Exists.
- [35] Gore-14. <http://www.esnog.net/gore14/lista.html>, Oct 2014.
- [36] Gore-14. <http://www.esnog.net/gore14.html>, Oct 2014.
- [37] S. Guha, K. Biswas, B. Ford, S. Sivakumar, and P. Srisuresh. NAT Behavioral Requirements for TCP. Technical Report 5382, Internet Engineering Task Force, October 2008.
- [38] V. Jacobson, R. Braden, and D. Borman. TCP Extensions for High Performance. Technical Report 1323, Internet Engineering Task Force, May 1992.

- [39] Ram (Ramki) Krishnan, Norival Figueira, Dilip Krishnaswamy, Diego Lopez, Steven Wright, and Tim Hinrichs. NFVaaS Architectural Framework for Policy Based Resource Placement and Scheduling. Internet-Draft draft-krishnan-nfvrg-policy-based-rm-nfvias-03, IETF Secretariat, November 2014. I-D Exists.
- [40] Ram (Ramki) Krishnan, Anoop Ghanwani, Pedro Gutierrez, Diego Lopez, Joel Halpern, Sriganesh Kini, and Andy Reid. SFC OAM Requirements and Framework. Internet-Draft draft-krishnan-sfc-oam-req-framework-00, IETF Secretariat, July 2014.
- [41] Ram (Ramki) Krishnan, Dilip Krishnaswamy, Diego Lopez, Peter Willis, and Asif Qamar. An Open NFV Architectural Framework for Virality Based Content Caching. Internet-Draft draft-krishnan-nfvrg-open-nfv-virality-00, IETF Secretariat, July 2014. I-D Exists.
- [42] Mirja Kuehlewind, Richard Scheffenegger, and Bob Briscoe. Problem Statement and Requirements for a More Accurate ECN Feedback. Internet-Draft draft-ietf-tcpm-accecn-reqs-07, IETF Secretariat, July 2014. Waiting for Writeup.
- [43] Matt Mathis and Bob Briscoe. Congestion Exposure (ConEx) Concepts, Abstract Mechanism and Requirements. Internet-Draft draft-ietf-conex-abstract-mech-13, IETF Secretariat, October 2014. Approved-announcement to be sent::Revised I-D Needed.
- [44] Toby Moncaster, Bob Briscoe, and Arnaud Jacquet. A TCP Test to Allow Senders to Identify Receiver Non-Compliance. Internet-Draft draft-moncaster-tcpm-rcv-cheat-03, IETF Secretariat, July 2014.
- [45] Nfv management and orchestration. Group Specfic DGS/NFV-MAN001, ETSI NFV ISG, dec 2014.
- [46] Network Functions Virtualisation (NFV): Network Operator Perspectives on Industry Progress. http://portal.etsi.org/Portals/0/TBpages/NFV/Docs/NFV_White_Paper3.pdf, October 2014. SDN and OpenFlow World Congress, Darmstadt, Germany.
- [47] Proposed network function virtualisation research group (nfvr). <https://datatracker.ietf.org/rg/nfvrg/charter/>, 2014. note.
- [48] Christoph Paasch and Olivier Bonaventure. Securing the MultiPath TCP handshake with external keys. Internet-Draft draft-paasch-mptcp-ssl-00, IETF Secretariat, October 2012.
- [49] Christoph Paasch and Olivier Bonaventure. A generic control stream for Multipath TCP. Internet-Draft draft-paasch-mptcp-control-stream-00, IETF Secretariat, February 2014.
- [50] P. Srisuresh, B. Ford, S. Sivakumar, and S. Guha. NAT Behavioral Requirements for ICMP. Technical Report 5508, Internet Engineering Task Force, April 2009.
- [51] J. Touch, A. Mankin, and R. Bonica. The TCP Authentication Option. Technical Report 5925, Internet Engineering Task Force, June 2010.
- [52] Joseph Touch and Ted Faber. TCP SYN Extended Option Space Using an Out-of-Band Segment. Internet-Draft draft-touch-tcpm-tcp-syn-ext-opt-01, IETF Secretariat, September 2014. I-D Exists.
- [53] Virality based content caching in nfv. http://nfvwiki.etsi.org/index.php?title=Virality_based_content_caching_in_NFV_framework, 2014. note.
- [54] Vnf router performance with ddos functionality. http://nfvwiki.etsi.org/index.php?title=VNF_Router_Performance_with_DDoS_Functionality, 2014. note.
- [55] Vnf router performance with hierarchical quality of service functionality. http://nfvwiki.etsi.org/index.php?title=VNF_Router_Performance_with_Hierarchical_Quality_of_Service_Functionality, 2014. note.
- [56] Ning Zong, Linda Dunbar, Melinda Shore, Diego Lopez, and Georgios Karagiannis. Virtualized Network Function (VNF) Pool Problem Statement. Internet-Draft draft-zong-vnfpool-problem-statement-06, IETF Secretariat, July 2014.