Network services and traffic engineering methods for supporting applications on the VTHD experimental gigabit network

Philippe CINQUIN[1], Yves DEVILLERS[2], Annie GRAVEY[3], Elodie LARREUR[4]

Abstract

VTHD is a high-performance IP experimental network. This network and associated research projects have been partially funded by the French government through the French Research Network for Telecommunications (RNRT) in order to support the development of leading-edge network services on the one hand, and test a wide-scale deployment of advanced Internet applications on the other hand. This paper describes the network services that were deemed necessary to support the deployment of innovative applications, as well as several of the applications that have been experimented on the network. It also presents a selection of the traffic engineering methods and experiments that have been developed in the course of the VTHD related research projects. This article describes the collective works of members of the project partners, which are represented by the set of authors for the present paper.

Key words: Teletraffic, Telecommunication network, High rate, IP, Experimentation, Research program, Telecommunication service, Telecommunication application, Peer to peer networking, Active telecommunication network, Distributed processing, Telemedecine, Teleconference, Multiple service network, Telecommunication traffic control, Network routing.

SERVICES RÉSEAU ET MÉTHODES D'INGÉNIERIE DU TRAFIC POUR SUPPORTER DES APPLICATIONS SUR LE RÉSEAU EXPÉRIMENTAL GIGABIT VTHD

Résumé

VTHD est un réseau IP expérimental à très haut débit. Ce réseau et les projets de recherche associés ont reçu un financement public par l'intermédiaire du Réseau National de Recherche en Télécommunications (RNRT) dans le but de permettre à la communauté scientifique française de développer des services réseau innovants et de tester le déploiement à grande échelle d'applications Internet avancées. Cet article décrit les services réseau nécessaires pour supporter le déploiement d'applications innovantes, ainsi que les applications qui ont été expérimentées sur le réseau VTHD. Enfin, l'article présente une sélection des

[1] TIMC-IMAG, UMR 5525 UJF-CNRS, Faculté de Médecine ; 38706 La Tronche cedex, France.
[2] INRIA, Domaine de Voluceau-Rocquencourt, BP 105 ; 78153 Le Chesnay cedex, France.
[3] GET/ENST Bretagne, Technopôle Brest Iroise, CS 83818 ; 29238 Brest cedex, France.
[4] France Telecom, R&D Division, 38, rue du Général-Leclerc ; 92794 Issy les Moulineaux cedex 9, France.

méthodes et des expérimentations en matière d'ingénierie de trafic développées lors des projets de recherche associés à VTHD. Cet article décrit les travaux réalisés collectivement par de nombreux membres des institutions partenaires, qui sont représentées ici par les auteurs cosignant l'article.

Mots clés : Télétrafic, Réseau télécommunication, Haut débit, Protocole Internet, Expérimentation, Programme recherche, Service télécommunication, Application télécommunication, Communication égal à égal, Réseau télécommunication actif, Traitement réparti, Télémédecine, Téléconférence, Réseau multiservice, Gestion trafic télécommunication, Routage réseau.

Contents

I. INTRODUCTION

The origin of the VTHD [20] network lies in the 1999 call for projects by the RNRT (Réseau National de la Recherche en Télécommunications), which is the French Research Forum for Telecommunications. Major academic French institutions (GET[1], INRIA[2], IMAG[3], Eurecom[4]) joined with France Télécom[5] in order to propose the implementation of a Gigabit IP network for developing and validating leading-edge network services on the one hand, and enabling and testing the wide-scale deployment of advanced Internet applications on the other hand.

VTHD is the (French) acronym for «Vraiment Très Haut Débit» (Really High Bit rate). The research community had identified the need for a French experimental broadband IP network, similar to the Abilene network [12] in the USA, which would support networking and application development R&D activities in France.

The initial VTHD project was launched in 1999, and the completion date of its sequel VTHD++ is December 2004. The two projects have been funded both by the projects' partners and the Direction Générale de l'Industrie, des Technologies de l'Information et des Postes: DiGITIP (Industry, Information Technologies and Post Department) in the French Ministère de l'Économie, des Finances et de l'Industrie (Ministry for the Economy, Finance and Industry).

[1] The GET or the «Groupe des Ecoles des Télécommunications» (Group of Telecommunications Institutes) includes the ENST, the ENST Bretagne and the INT.
[2] Six INRIA research laboratories (located in Nancy, Rennes, Lyon, Grenoble, Rocquencourt and Sophia-Antipolis) have participated to the VTHD project.
[3] The IMAG Institute (Applied Computing and Mathematics in Grenoble) is a federating institute within the CNRS (National Centre for Scientific Research).
[4] The EURECOM is a joint Institute founded by the ENST and the EPFL (Federal Engineering School in Lausanne).
[5] Six France Telecom (Research & Development Division) laboratories have contributed to the project through six locations (Lannion, Rennes, Caen, Issy-Les-Moulineaux, Grenoble, Sophia-Antipolis).

As part of the project, a countrywide high-performance IP backbone network has been deployed by France Telecom, and can be accessed in most major French cities by the project's partners. The VTHD network is a "platform", i.e. an experimental network intended for testing application deployment in realistic conditions. This has implied the definition of specific policies for making the network available to new partners such as e.g. e-Toile [15], another French research project.

The main objective of a RNRT Platform project, such as VTHD and its sequel VTHD++, is to support the development by French R&D networking communities of advanced applications and multimedia services. The VTHD partners have thus tested several applications, such as telemedicine, code coupling, distributed computing, telepresence wall, community communications, etc... Very early in the projects lifetimes those applications have generated large amount of relevant traffic that was monitored by network experts. Some of these applications are presented in Section II.

One of the initial objectives of VTHD was to develop and test IP over WDM (wavelength division multiplexing) architectures and technologies in order to achieve high-speed packet forwarding. The VTHD network is thus built using standard routing and switching equipment, but it is intended for testing leading-edge network services, and is not a production network. It is therefore possible e.g. to provoke link or node failures in order to understand the tuning of failure recovery mechanisms, to engineer overloading conditions in order to measure and analyze how differentiated services mechanisms selectively handle traffics, to dynamically configure the network for specific classes of applications. Indeed, a major part of the developments during the project has focused on designing methodologies for studying advanced traffic engineering principles and implementing some of them as described in Section III.

Section IV of this paper presents part of the traffic engineering studies that have been carried out by the projects partners since 1999. These methods include dynamic resource allocation, routing, QoS measurement, support of IPv6 applications, and multicasting. Some of those have been used to tune the implementation of the networking protocols used in the VTHD network, others are the basis for proposing contributions to standards.

This article describes the collective works of members of the project partners, which are represented by the set of authors for the present paper. The list of project members that participated to this article is found in [20].

II. VTHD SUPPORT OF APPLICATIONS

Since it was launched, the VTHD network has been used to test the support of many broadband applications by a gigabit WAN (wide area network). This Section describes briefly some of these applications and shows how these experiments have facilitated their development.

II.1. Distributed File Storage

The peer-to-peer (P2P) paradigm consists in considering that all workstations play similar roles. It differs from the client-server paradigm where some workstations (the "servers") are dedicated to serve the others (the "clients"). The most popular types of P2P applications are file-sharing applications, where each workstation can download (music or video) files located on another workstation. However, the P2P paradigm is more general, and is well suited to distributed applications. The distributed file storage application that has been experimented on VTHD is one such example.

Distributed file storage consists in partitioning a file in fixed-length data blocks, which are individually stored, and in block replication, so that data can be recovered in case of peer failure. This method is a lightweight approach to data storage, and since it relies on an efficient distribution and replication of data blocks, it makes it very resilient to both peer and router failures.

The first prototype developed in the VTHD++ project used a standard Chord P2P lookup layer [30], which is well fitted for a LAN environment, but has a high overhead if deployed on a WAN. Therefore, a second prototype used a hierarchical P2P lookup that limits the communication overhead. The storage application was almost unmodified, thanks to a careful layered protocol design.

The distributed file storage application is composed of three layers: application, storage and P2P lookup. The application layer gets the file from user input through a graphical interface. The storage layer divides the file in blocks, and assigns a key to each block using a hash function. Then it is up to the P2P lookup layer to decide in which peer a block is stored.

Creating multiple keys for a block allows for replication on a number of peers, so that data can be recovered in case of peer failure. The storage layer is warned of failing peers by the lookup layer through a monitoring system. This way the storage layer can create new copies of blocks to maintain a constant number of replicas. Multiple replicas allows for increased data retrieval performance using parallel download techniques [28].

The hierarchical lookup [10] layer is a two-tier P2P system that uses Chord in the top level and CARP [29] inside groups. As a consequence of this two-tier organization, the system is designed to best manage several groups of reduced size.

This corresponds to the deployment scenario on VTHD for which a Java prototype of our system was deployed on several hosts in five participating sites. The prototype that has been developed demonstrates the advantages of a hierarchical organization and allows comparing the performance of the flat versus hierarchical approaches.

II.2. High-performance programmable and active networks

Programmable and active networks enable dynamical deployment of value-added services in network equipment. Active networking applies both to core network services and to services offered in layers located above the network services, i.e. to services offered by peers connected to the core network (see for example the special issue of Annals of Telecommunications on Active networks, vol. 59, n°5-6, May-June 2004).

VTHD access and core routers are on-the-shelf equipment and do not implement facilities for active networking. On the other hand, computers connected to the VTHD network can indeed be configured to act as routers. Therefore, the active networking software environment that has been developed within the VTHD projects configures services delivered by active nodes located on the edge of the VTHD backbone.

Achieving high-performance in active networks is one of the most challenging tasks; indeed, complex packet handling procedures may not be easily performed at line speed. The Tamanoir [11] suite is a complete software execution environment that allows the deployment of active services on specific layers: kernel modules for lightweight services, user space Java programs for middleware services, and distributed infrastructure for resource (CPU, memory...) consuming services.

Tamanoir active nodes have been deployed on various partners'sites. Experiments show that a Tamanoir cluster-based software active node is able to support gigabit bandwidth with customized services.

During the VTHD++ project, two active services using Tamanoir have been intensively experimented:

- a reliable multicast service [22], which is based on active routers which efficiently manage NACKs (negative acknowledgement) aggregation, detect packet losses and dynamically elect nodes for retransmitting packets.
- a service (QoSinus) used in grid networking (see Section II.3) that allows the grid's nodes to dynamically map application requirements on existing QoS classes by marking individual packets.

II.3. Dynamic Grids establishment and operation

In the last decade the availability of high-performance networks and large numbers of widely distributed computational resources has motivated significant research for coordinating disparate and heterogeneous resources as a single very powerful computing resource i.e. a "grid".

Several VTHD partners are also members of the e-toile RNTL project [15]. The e-Toile RNTL project is an experimental wide-area grid test bed funded by French Ministry of Research. The VTHD network interconnects the e-Toile grid nodes with access links of 1 to 2 Gbit/s bit rate.

Several grid-related applications have thus been tested for deployment on VTHD.

The first one enables Grid users and applications to transparently access and optimally use the available DiffServ capabilities offered by the underlying network (VTHD). The dynamic mapping of the flows QoS specification to the existing IP QoS facilities is addressed by a service architecture that combines flow aware and infrastructure aware components. As QoS resources are scarce and/or costly, this distributed service finely manages the QoS service allocation at the edge the wide-area network (WAN) to best fit all the various demands coming from heterogeneous grid flows.

The second consists in offering a middleware layer between final applications and the grid. This middleware allows the final application to specify its requirements, and maps these requirements to available grid resources. The Distributed Interactive Engineering Toolbox (DIET) provides such a middleware: a server daemon is attached to each available computer resource, clients request service via a DIET client interface, and a hierarchy of distributed agents provides services such as server selection and data management in an efficient and scalable manner. A tool for the rapid, automatic and large-scale deployment of DIET has been developed, thus enabling the exploration of a variety of distributed configurations of DIET on a large number of VTHD resources.

II.4. Code coupling

Distributed processing consists in using a number of computers to run a single application. Programs can then run faster since more CPU is made available. This allows one to conduct either more realistic simulations or simulations that are impossible to realize due to the CPU limitation of a single computer.

A major difficulty lies in dividing a program in such a way that separate computers can execute different portions without interfering with each other. Another is in providing a communication network between the computers that can support high bandwidth data transfer with limited latency. It is this last aspect that has been experimented on VTHD: it is assumed that two clusters of computers execute parallel codes that require data transfer between the clusters. Both clusters are connected to the VTHD network.

The problem can be stated as efficiently transferring a set of data distributed over an N nodes cluster to a distant M nodes cluster. Since the N nodes work in parallel inside a cluster, they can transmit their data synchronously to the M distant nodes. The communication time starts when the first sender node begins to send and finishes when all data has been received. Since codes typically behave iteratively (i.e. using the same communication patterns with different values) it is possible to rely on information concerning previous transfer in order to optimize future transfers.

Although WANs in general, and VTHD in particular, are now able to provide a high bandwidth, this bandwidth is still much lower than the aggregate bandwidth all the nodes from a cluster can generate (a hundred nodes cluster using Fast Ethernet cards can generate bit rates up to 10 Gbit/s of data and up to 100 Gbit/s using Gigabit Ethernet cards). Classical flow regulation techniques are not applicable because the parallel flows are not independent and also because there is a load balancing issue: the N sender nodes may not have the same amount of data to send and/or the M receiver nodes may not be waiting for an equal quantity of data. So, there is a clear need for techniques to regulate the transfers and to avoid congestion on the access points.

Two different approaches have been tested on VTHD. The first one aims at controlling the bandwidth allocated to each individual flow through TCP send buffer tuning [21]. The available bandwidth is distributed to the parallel flows in order to reach the same transfer duration.

The second approach is based on the concept of parallel objects/components [18] and aims at computing a communication schedule using two types of information: a communication matrix, provided by some redistribution libraries and the performance of the network (latency, bandwidth) that should be provided by an information service and should be guaranteed by a QoS policy. The experiments have demonstrated the possibility to ensure a sustained communication bandwidth of 2 Gbit/s between two parallel applications running on two distinct clusters of 24 nodes, distant from 1000 km and interconnected with VTHD.

II.5. Telemedicine

Telemedicine is one application that can benefit from a high bandwidth reliable network. An efficient sharing of medical imaging is already implemented in medical centers thanks to high bandwidth LANs. However, this type of file sharing has no real-time requirements. The application that has been tested on VTHD consists in a real-time ultrasound observation of a patient by a physician located at a remote site. Ultimately this type of facility would make specialists more accessible to patients since the physicians would not have to travel to the patient's bedside and could then give a tele-consultation.

A Robotized Tele-echography (RTE) application has been developed as part of the VTHD++ project.

When conducting such experiments, the first issue is to design the terminals. The RTE system is shown in Figure 1. The terminal on the patient side (slave robot) should be simple to operate and should not lead to any discomfort to the patient. On the physician side, the terminal (master console) should be able to command the distant terminal, and should display the exam's results.

The physician moves the virtual probe placed on a haptic device (PHANTOM® Desktop) [19] to control the real echographic probe placed on the slave robot, receiving in real-time medical images while discussing via videoconferencing facility. The nurse close to the patient typically sets up the robot and hands over the control of the robot to the remote radiologist who establishes diagnostics using the robot. The videoconferencing is activated for conversations and visualization of the global scene.

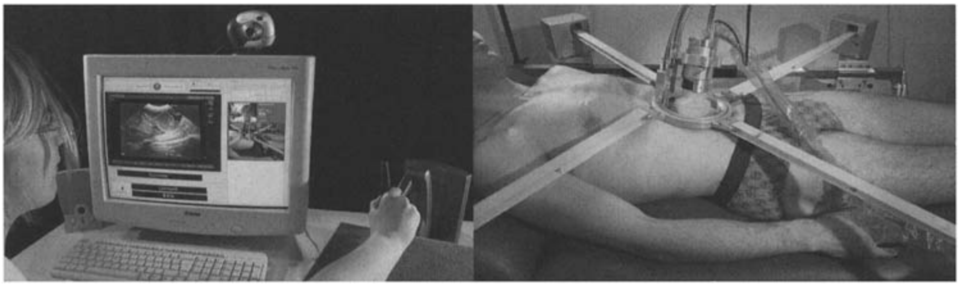


FIG. 1 – RTE Master Console (left) and Slave Robot (right).

Télé-échographie robotisée: Poste médecin (gauche) et poste patient (droite).

The robot is equipped with a feedback actuator. The use of a remotely controlled robot implies a short roundtrip delay (< 500 ms) with minimal jitter (< 5 ms). Furthermore, the jitter between the different flows should also be limited (the limit, 10 ms, is set by the requirements of haptic flows). Lastly, the feedback loop of the robot has to be very reliable (with a disruption time smaller than 40 ms).

These stringent service constraints required that the flows be transported using either a single connection or a bundle of connections over the same route with a 1 + 1 protection. Depending on the video quality (image definition and number of images per second), either a 512 kbit/s ISDN connection or a 10 Mbit/s VTHD flow can be used. The technical limitation lies more with CPU limitations than with network bandwidth.

The tests on VTHD allowed the partners to identify the clinical situations in which image compression was acceptable and those in which no compression was acceptable.

The RTE prototype was available for pre-clinical validation in June 2003, in the timeframe of a specifically developed clinical research protocol, under the French law "Loi Huriet" that rules how clinical tests are run. These experiments showed that the architectural choices, the controllability, and the compatibility with the VTHD network were very adequate. However, some minor modifications had to be made (for instance, straps could in some cases apply an excessive pressure of the probe on the abdomen). The corresponding modifications were integrated, and made fully available in May 2004 for new pre-clinical tests. In May and June 2004, 4 volunteers tested the system. They did not report any discomfort, and radiologists felt enthusiastic to begin the evaluation, planning to involve 100 patients using VTHD connections between Brest and Grenoble hospitals (1125 km) for remote echographic examination of patients with abdominal aortic aneurysms (AAA).

II.6. Advanced Teleconferencing

Teleconferencing is now widely used as an alternative to face-to-face meetings. However, teleconferencing has some drawbacks such as the setting up of the conference and the inferior interpersonal interactions. The inferior quality of communication is often due to bandwidth limitation, and the complexity of the setting up process to the need for reserving facilities (either a studio or bandwidth). The huge bandwidth available in a gigabit WAN allows to develop innovative teleconference systems that would allow participants to experience good interactions in terms of image and sound.

The "Telepresence Wall" is such a system, which works as a permanent informal "meeting window" (see Figure 2).

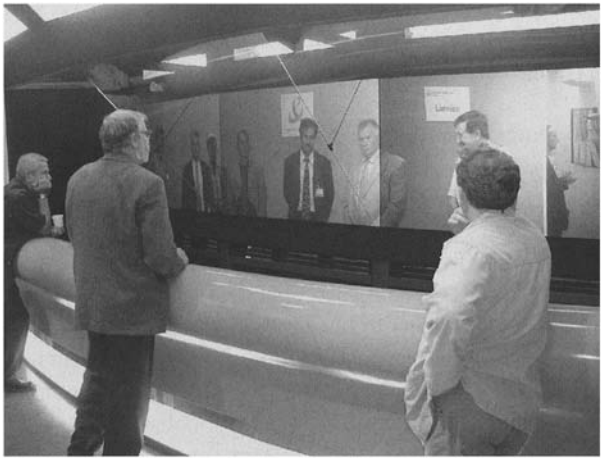


FIG. 2 – The Telepresence wall.

Mur de téléprésence.

Since the meeting window is always on, meeting participants do not have to start and stop the teleconferencing facilities, which is usually still fairly complex. Moreover, cameras film the whole area in front of the Telepresence Wall, and the participants do not have to specifically look at a given camera. They just stand, or sit, in front of the "meeting window" to watch and listen to (and reciprocally!) a life-size image of their remote meeting partners. The Telepresence Wall can thus be used for impromptu meetings of up to 20 people at any time. The terminal offers a large high-definition image of the distant site and remote meeting partners can really visually interact (looking into each other's eyes) which is not possible with current teleconferencing techniques. Furthermore, a spatial audio system has been developed that allows the different sounds perceived by the users to be precisely localized, just as in a movie theatre, further reinforcing the sensation of physical interaction. The Telepresence Wall relies on the econf technology [13] and requires between 1 and 3 Mbit/s full-duplex bandwidth.

VTHD has been used to fine-tune the development of the Telepresence Wall. Since UDP is employed for both video and audio signals in order to ensure a high interactivity, an excellent QoS is mandatory. This system has already been deployed in several France Telecom R&D labs and is currently being deployed in international locations such as China, Poland, UK and the USA.

The future applications of the Telepresence Wall are varied, from multi-site teleconferencing application to distance learning and virtual travel agency [14].

II.7. Communication in large communities

Whereas teleconferencing is intended to replace face-to-face meetings of small groups, the application that is described hereafter aims at simulating interpersonal interactions within large groups of participants. The major problem here is to dynamically manage the group of participants one can directly interact with at a given time. Virtual-Eye (V-Eye) is a 3D virtual world application that implements a Very Large Scale Virtual Environment [23]. Each participant can move in a virtual world, can send an audio or video stream, or text messages, and can receive the streams emitted by the people located in his vicinity.

V-eye aims at providing a tool for facilitating the set-up of experiments related to multicast transmission in heterogeneous networks. The objective is to use a multimedia application involving a large number of multicast sources/receivers inside a very high-bit rate multicast environment without impacting network performance. Scalability is achieved through the use of a large set of multicast groups. A centralized agent maintains in real-time a mapping of the geographic area into cells, each cell being associated with a multicast group. Each participant can then restrict its incoming streams to his zone of interest. The V-Eye experimentation in VTHD involved several partners and required inter-AS multicast configuration.

In addition to interpersonal communications, the virtual world contains a movie theatre that allows – provided that sufficient bandwidth is available - to watch a movie multicast in Digital Video (DV) format (see Figure 3). This feature supports 3D-perspective display of the video content, as well as mixing of the audio soundtrack with the other participants' audio streams. All the audio streams can also be rendered in 3D, so as to provide realistic indications of the source position.

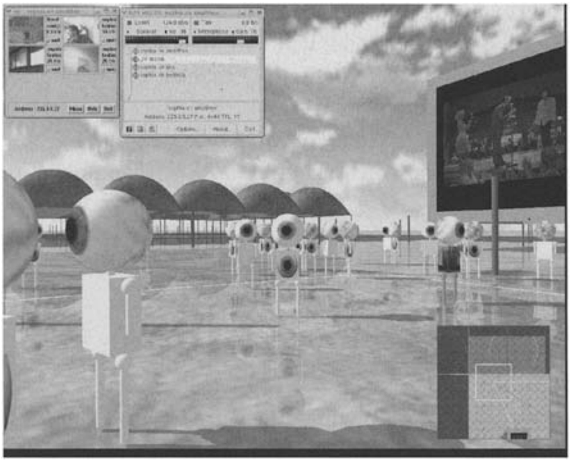


FIG. 3 – A view of the v-eye application: the virtual world with a virtual movie
The small windows corresponds to the vic video tool and the rat audio tool.

Une vue de l'application v-eye : monde virtuel et film virtuel
Les deux petites fenêtres correspondent aux outils vic (vidéo) et rat (audio).

III. VTHD NETWORK SERVICES

The VTHD network shown in Figure 4 comprises 10+ routers from different vendors [20]. They are connected by 2.5 Gbit/s optical channels operated by France Telecom. The 25+ access routers are connected to the VTHD backbone mostly via gigabit Ethernet interfaces

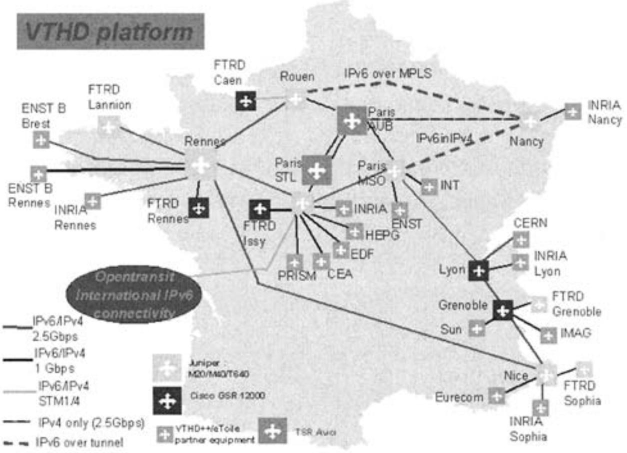


FIG. 4 – The VTHD network in 2004.
Le réseau VTHD en 2004.

(some partners are connected via 2.5 Gbit/s access links). Access routers aggregate the traffics of VTHD partners. The network is operated and managed by France Telecom, and peers with other networks in France and internationally thanks to a public addressing plan and a public autonomous system numbering.

The present Section describes the main network services that have been experimented in order to be deployed in VTHD. They are currently used by the applications supported by the VTHD network.

III.1. Multiservice support

As applications with stringent QoS requirements (e.g. multimedia streaming) are introduced and deployed over IP networks, network operators introduce service differentiation capabilities in order to support traffic flows that do not rely on end-to-end congestion control. Although service differentiation mechanisms are standardized, their practical implementation is still a challenge because of the large number of configuration options and operational parameters. VTHD, as an experimental network, has been used to fine-tune various routers' configuration parameters in order to e.g. study and experiment the service differentiation, and in particular the DiffServ protocol.

When implementing Diffserv in a multi-vendor network such as VTHD, it is mandatory to select options that can be supported in all equipments. The main blocking point was originally the small number of queues available in some of the core routers. This has resulted in defining a very small number of QoS classes. A synthesis of the requirements and the analysis of the traffic on the network lead us to identify five classes of service:

- "Network Control" that is used by routing protocol messages
- "Expedited Forwarding" for delay and jitter sensitive traffic
- "UDP Assured Forwarding" and "TCP Assured Forwarding" for loss sensitive data traffic
- "Best Effort" for traffic that has no specific QoS requirement.

Defining two assured forwarding classes allows the protection from starvation of the "assured TCP traffic" due to non-TCP friendly "assured UDP traffic".

VTHD code points have been defined that can be easily translated into standardized DiffServ code points. Internal routers mechanisms (token buckets, schedulers, queue management policies...) are configured according to the VTHD service classes. Extensive experiments have allowed analyzing the impact of congestion on various applications and validating the VTHD service differentiation strategy.

III.2. Multicast implementation

IP Multicast is an attractive technology, which relies on the data network to provide packet replication for point-to-multipoint or multipoint-to-multipoint communications; it can be used by multimedia applications such as audio or video with large audiences.

Multicast transmission consists in sending single multicast packets addressed to all subscribed recipients instead of broadcasting packets to everybody. This allows to use efficiently resources, and not to replicate packets on a given interface. The tool for supporting such functions is multicast routing, which involves the dynamic management of "multicast groups".

IPv4/IPv6 Multicast service is currently deployed in the multi-vendors VTHD network. Both ASM (Any Source Multicast) and SSM (Source Specific Multicast) models have been implemented for intra-AS and inter-AS connectivity.

In particular, the VTHD IPv6 multicast service gives the opportunity to a VTHD IPv6 multicast source to send multimedia traffic which can be displayed by the French Network for Academic Research Renater receivers thanks to the M6Bone multicast streaming which is available via a peering with Renater.

III.3. Supporting IPV6

Many partners of the VTHD project are involved in IPv6 developments. It was therefore a requirement on the VTHD network to offer an efficient support of the IPv6 protocol stack, and therefore a support of applications developed over IPv6 (High-definition video streaming, videoconference tool, 3D visualization).

France Telecom received an IPv6 prefix from RIPE in January 2001. The decision to offer a native IPv6 network service in VTHD followed in June 2001.

At that time the vendors did not implement the IPv6 stack in their core routers. It was then decided to deploy additional IPv6 (actually dual-stack) edge routers at network's border which would be interconnected in a first step through MPLS tunnels using proprietary encapsulation, and then in a second step, using a mechanism discussed at IETF. Using MPLS encapsulation technology was shown to be more efficient than the use of IPv6 in IPv4 tunnels ([1], see also Section IV.5).

Nevertheless the final objective of the partners was to implement dual stack architecture in the core of the network: this dual stack deployment mainly occurred in 2002 in the context of VTHD++ project. In July 2004, most of the core routers of VTHD are dual-stack except two tera-routers that still lack an IPv6 implementation and thus require the use of the MPLS encapsulation method.

This non-congruent topology (IPv4 and IPv6 topologies are not similar) mandates for the general case the use of IS-IS Multi Topology as a single internal gateway routing protocol. This IS-IS extension (used in VTHD network) allows the two IPv4 and IPv6 topologies being treated in separated routing tables within a single routing process.

Partners' sites are connected through IPv6 e-BGP sessions and may optionally receive the IPv6 BGP full routing (as VTHD is connected to the IPv6 international network).

Regarding operating an IPv6 backbone network, there are still some operational problems. Indeed, IPv6 networks' MIBs are currently being standardized. Therefore, Router's vendors do not yet offer a full native management support of the IPv6 stack.

Project members from the Nancy INRIA Laboratory have developed a native IPv6 management tool exploiting the management features currently implemented by vendors. These tools have been adapted and deployed in the VTHD Back-Office (the France Télécom operated VTHD management platform), in addition to an initial IPv6 compatible HP Openview version.

III.4. Scalable Firewalling

Security is one of the mandatory services to be deployed by network operators, either for protecting their own architecture or on behalf of their customers.

IP networking is sensitive to security attacks. Firstly, the network itself supports network control protocols; any major transfer plane degradation directly jeopardizes both control and management planes. Secondly, Denial of Service (DoS) attacks are easily launched in connectionless networks, implying that transfer plane degradations can be engineered fairly simply. These problems, which are inherent to IP networking, are exacerbated for gigabit networks since decisions have to be taken at line speed. Therefore, implementing security policies in gigabit networks relies on adaptable and scalable mechanisms. Indeed, packet filtering may involve a large number of rules that may change due to transient conditions (e.g. attacks).

The VTHD network has been the testbed for experimenting efficient security policies against common attacks (Trojan, DDoS, ...). The IFT (IP Fast Translator) table, based on the "trie" memory [24], analyzes IPv6 traffic and filters the traffic according to a specified security policy. Using this security policy, the CaraHD6 compiler [4] implements a classification algorithm designed in such a way that packet analysis time is independent from the number of rules specified by the security policy. The maximum packet analysis time for IPv6 is about 1.5 microseconds (~ 260m of optical fiber), which has shown that the couple CaraHD6/IFT is well adapted to network operators' requirements.

III.5. On-line network service activation

A central Back Office (BO) offers the means of managing the VTHD network. The BO offers tools dedicated to network management such as: a dual stack IPv4/IPv6 DNS for managing the VTHD domain (vthd.prd.fr), Fault and Performance management, topology discovery and dynamic service configuration.

This last aspect is significant in the sense that some tools have been developed to offer "on-line services" to the different users of the VTHD network. This is particularly interesting for an experimental network, since partners can modify some network configuration features according to the requirements of their own experiments.

Most of the dynamic IPv4 configuration actions are performed with the Metasolv tool "IP Service Activator" (IPSA) [17]. IPSA is a stand-alone system that operates exclusively on IPv4 interfaces. Its role is to provide a user-friendly web interface for the user (the network operator or a network's user) to specify the requested network service, and then to deploy the configuration on the VTHD network.

VTHD partners can use this web interface to specify their requirements, and the IPSA tool translates the requirements into vendor specific commands in order to configure the VTHD equipment (IPSA configures IP services on an existing MPLS network). The BO can also use these tools to counteract some performance degradations due to experiments in extreme conditions.

The following services are currently offered:

- On-demand activation of Virtual Private Networks (VPN). The VPN topology can be both "full mesh" or "hub and spoke". The routing protocol between the VTHD Provider Edge router (PE) and the Customer Edge router can be one of the following protocols (static, OSPF, eBGP, RIP).
- Access Control Lists (ACL) instantiation. Since the VTHD network is partially open to the Internet, some users may wish to protect their own network or even the VTHD network from specific threats. The ACL may be a combination of the following elements: Protocol type, Port number, Source and/or destination, Permit or deny traffic.
- Dynamic QoS adaptation. VTHD QoS classes are statically defined. It may be useful, for some experiments, to transiently modify the QoS configuration on selected interfaces. A tool coupling the performance monitoring software used by the BO with IPSA allows the reconfiguration of router interfaces when some specific performance indicators go above thresholds. The router configuration is restored to the standard configuration when the performance indicator stays below the threshold for a given period of time (typically one minute).

Lastly, a middleware platform (OSSIP) has been developed to help adapting the set of QoS requirements defined by the partners to the network level services and QoS statically defined in VTHD. The middleware generates adequate management rules to be applied exclusively to the edge routes (CPEs) through which the partner's sites access the VTHD backbone (the configuration of the core routers is not modified by OSSIP).

The partner's application profiles together with the network's topology and capabilities are stored and are made accessible to the middleware through an XML interface between the back-office VTHD and the middleware. This is illustrated in Figure 5.

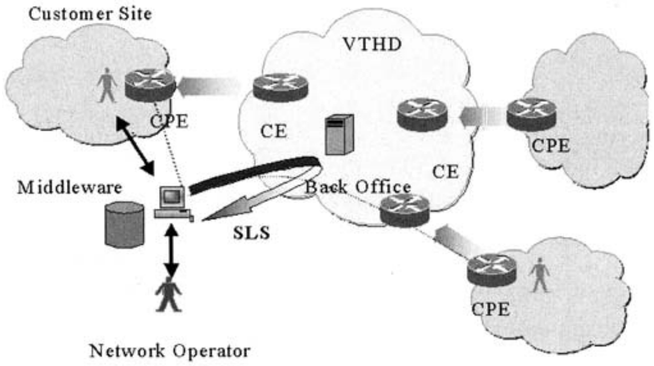


FIG. 5 – Adaptation of the partner's QoS requirements to the static VTHD QoS profiles.

Adaptation des requêtes du client aux profils statiques de QoS de VTHD.

IV. TRAFFIC ENGINEERING EXPERIMENTS AND DEPLOYMENT

IV.1. MPLS Fast Reroute

Network availability is certainly the most important network performance parameter. The availability of a network portion is the percentage of time during which network operations are correctly functioning in that portion, with at least "five 9s" (99.999%) availability being the carrier standard. Achieving such a high availability relies on quickly detecting a fault on a path (the fault may be due either to a router failure, or to a link failure), and quickly switching over another path. The current technology in transmission networks is based on SDH rings and achieves a recovery time smaller than 50ms.

IV.1.1. Network Availability in IP/WDM networks

Traditional SDH-based protection switching architectures are very expensive, as they require dedicating links for protection (i.e. only 50% of link capacity can be allocated to protected traffic). Furthermore they do not protect traffic in case of router failure.

On the other hand, current recovery delays measured with IP link state routing protocols (OSPF, IS-IS) are over one second. Several enhancements are currently being developed in order to go below one second, but these recovery delays are not strictly guaranteed, as convergence is realized by the control plane, and implies a distributed procedure.

Another approach relies on MPLS functions to provide "Fast Reroute" [7], i.e. protection switching facilities. Indeed, MPLS, being connection oriented, enables the provisioning of backup paths that can be used in case of failure. However, MPLS offers currently no support for reporting remote failures. Therefore, the fast reroute mechanisms rely on local failure indications (i.e. of an attached link or of a neighboring node).

The present work presents an approach for providing link and router recovery mechanisms in the IP/MPLS layer.

IV.1.2. Case study

VTHD is carrying highly sensitive traffic, such as voice, video, and telemedicine. This brings the need for high network availability; in particular, these applications require guaranteed recovery delays, within 50ms, in case of link or router failure.

A good way to protect IP links and nodes with a sub-50 ms guaranteed recovery delay consists in deploying an MPLS local protection mechanism, better known as MPLS Fast Reroute [7]. Real-time traffic is routed into primary MPLS tunnels that are protected locally by pre-established backup tunnels. Rerouting, after failure, is performed in the data plane, and relies on a local recovery procedure performed by the nodes adjacent to the failure. This allows guaranteeing a sub-50ms recovery delay.

MPLS Fast Reroute Node protection has been deployed within VTHD to protect telemedicine traffic between Brest and Grenoble Hospitals (see Section II.5). Figure 6 illustrates the protection of the Brest-Grenoble path to a failure of the Paris router.

Primary MPLS tunnels are set up (one in each direction, as MPLS tunnels are unidirectional) between the aggregation routers located in Rennes and Grenoble. The primary tunnels carry all telemedicine traffic between Brest and Grenoble hospitals. These primary tunnels are locally protected against Paris-STL node failure, by two pre-established backup tunnels that bypass Paris-STL. In case of Paris-STL failure, the adjacent nodes Paris-MSO and Rennes detect the failure thanks to SDH alarms. Then they immediately splice traffic onto the backup tunnels. Information on backup interfaces and labels is pre-installed in the MPLS forwarding table, which allows guaranteeing sub-50 ms recovery delays.

Several experiments have been launched on VTHD to validate the Fast Reroute Node protection mechanisms. They were actually the first MPLS Fast Reroute Node protection experiments on an operational multi-vendor network, which indeed allowed the validation of FRR inter-working procedures across several vendors' equipment.

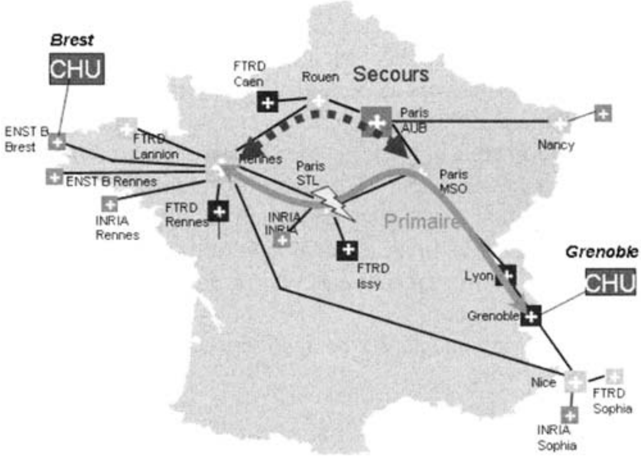


FIG. 6 – MPLS Fast Reroute on the Brest-Grenoble paths.

Reroutage rapide MPLS sur les chemins Brest-Grenoble.

Measured recovery delays in case of link or node failure are around 15ms, which perfectly fit with telemedicine requirements. This highlights the advantages of the MPLS Fast Reroute technology, which significantly improves IP network availability and thus allows for transport of real-time traffic.

Some of the procedures developed at that time in order to evaluate Fast Reroute performances are currently being standardized within IETF [8].

IV.2. Dynamic SLA management

Conceptually, the control of network dynamics can be thought as a feedback control system composed of a demand system (i.e. the traffic), a constraint system (i.e. the interconnected elements) and a response system (i.e. the network protocols and control mechanisms) [2].

Traffic Engineering (TE) defines the parameters and points of operation for the network, as well as the mechanisms that control its return to the defined point of operation when the demand and/or the constraint systems vary.

TE objectives are generally referred as being "traffic-oriented" (if they aim at offering differentiated quality of service (QoS) to the individual users and applications) or "resource-oriented" (if they aim at achieving efficient network resources usage). A TE mechanism is termed as "rational" if it aims at reaching specific traffic-oriented objectives, while optimizing resource-oriented objectives [2]. Network operators would obviously prefer this last type of TE mechanisms.

The different TE mechanisms available to the operator are applied within appropriate timescales and scopes. For example, TE mechanisms applicable on short timescales (e.g. admission control) are also likely to be applied on local scopes (e.g. interfaces). Similarly, TE mechanisms applicable to longer timescales (e.g. dimensioning) are likely to be applied on global scopes (e.g. network domain). As such, each timescale constitutes a control loop, where the observation of the network is done through metrology techniques at that timescale. As represented in Figure 7, three control loops can be identified.

The role of the long-term control loop (i.e. the outer loop) is to establish an optimal point of operation with respect to the resource-oriented objectives, i.e. to calculate an optimal layout with respect to the actual costs of operation and maintenance.

The inner loops accommodate small demand variations, measured at shorter timescales by using some mid- to short-term TE techniques such as admission control or load balancing, without having to change the layout.

Whenever the traffic variations cannot be accommodated by the inner control loops, a new optimal layout is calculated.

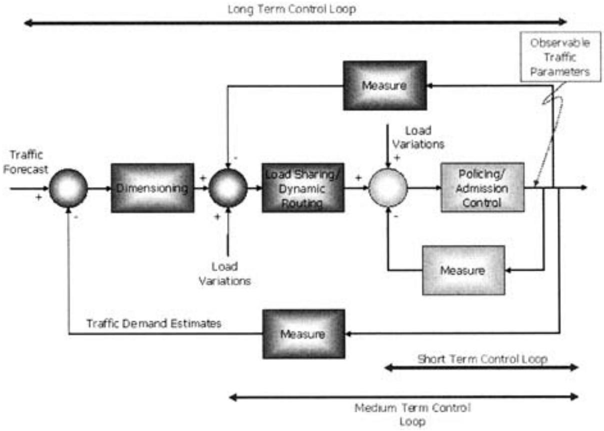


FIG. 7 – TE control loops at various timescales.

Les boucles de contrôle TE à différentes échelles de temps.

In the sequel, we propose a realistic problem formulation taking into account the actual costs of layout deployment and redeployment under dynamic traffic conditions, while providing some end-to-end QoS guarantees.

IV.2.1. Problem Formulation

For long timescales, the off-line computation of an optimized layout for a given traffic matrix allows the operator to globally optimize the usage of network resources by correctly dimensioning the network. The set-up of this optimal layout requires that the network supports advanced TE mechanisms such as those provided by MPLS [26, 27].

Cost functions that are generally used in the optimization problem consider minimizing the total delay throughout the network. However, under a given threshold, applications are insensitive to a further improvement in the end-to-end delay. The approach in [3] includes the per-path delay guarantees as constraints in the problem formulation, while letting the objective function to represent the actual costs of operation and maintenance of the layout.

In the static case (computation of an optimized layout) for a given set of demands, a cost function is proposed, which corresponds to the number of paths that the management system needs to set up to accommodate the traffic demands. In the dynamic case (computation of a modification of a previously optimal layout), the cost function takes into account both the number of paths the system needs to set up, and also the number of paths that have to be redeployed in the new layout (reconfiguration). Indeed, when reconfiguring a network, it is preferable not to tear down existing paths if it can be avoided.

Therefore, the general cost function given by equation (1) includes a term considering the cost of layout deployment through a binary variable indicating if the path is being used or not in the new layout, and a second term considering the cost of reconfiguration through a binary variable indicating if the path is new to the layout or not. Each path is weighted according to the cost model assumed by the operator (in the present case, the weights are calculated as the number of hops the path is traversing). In this way, the total number of paths is being minimized in the new layout, while minimizing the total number of path additions from the old layout. The binary variable $s^k_{q,(t+1)} = (h^k_{q,(t+1)} - h^k_{q,(t)})\, h^k_{q,(t+1)}$ penalizes indeed any new path not being used in the previous layout.

Given A, C, $D_{(t+1)}$, $H_{(t)}$, Θ, W

(1) Minimize $$\alpha \sum_{q=1}^{Q} \sum_{k=1}^{K_q} w^k_q h^k_{q,(t+1)} + v \sum_{q=1}^{Q} \sum_{k=1}^{K_q} w^k_q S^k_{q,(t+1)}$$

(2) Subject to $$\sum_{q \in Q,\, k \in K_q} b^k_{q,(t+1)} a^k_{q,i} \le C_i \qquad \forall\, i \in M$$

(3) $$\sum_{k \in k_q} b^k_{q,(t+1)} = d_q \qquad \forall\, q \in Q$$

(4) $$h^k_{q,(t+1)} \sum_{i=1}^{M} \frac{\lambda a^k_{q,i}}{C_i - x_{i(l+1)}} \le \theta_q \qquad \forall\, k \in K_q;\ q \in Q$$

(5) $$h^k_{q,(t+1)};\ S^k_{q,(l+1)} \in \{0, 1\} \qquad \forall\, k \in K_q;\ q \in Q$$

The multi-commodity flow allocation problem described above can be formulated as a MINLP (Mixed Integer Non-Linear Program) problem. The arc-path incidence matrix A of the graph representing the network, the capacity vector C of the component links, the new demands D(t + 1), the previously operational layout H(t), the delay constraint matrix Θ for each path in the layout, and the vector of weights associated to each path W are given.

Besides the contribution of the objective function, the end-to-end delay constraints introduced in [3] have not been used so far in the literature, although they seem very natural in a realistic operational context.

IV.2.2. Dimensioning and Reconfiguration numerical results

The MINLP is known to be NP-complete. In order to obtain results for large networks it is necessary to develop heuristics, which yield only approximate solutions. In [3] two types of heuristics are proposed: a generic heuristic based on Taboo Search techniques, and an ad-hoc heuristic based on the flow deviation algorithm. The Modified Flow Deviation Algorithm (MFD) is the latter heuristic algorithm and is based on the well-known Flow Deviation Algorithm by Frank and Wolfe [9]. Numerical results have shown that MFD largely outperforms the Taboo Search-based heuristics, which leads to prefer the MFD-based method as described below.

MFD shares with the original method the idea of shifting flow from non-optimal paths to optimal paths. This aims at ensuring that the new layout shares as many paths as possible with the previous one.

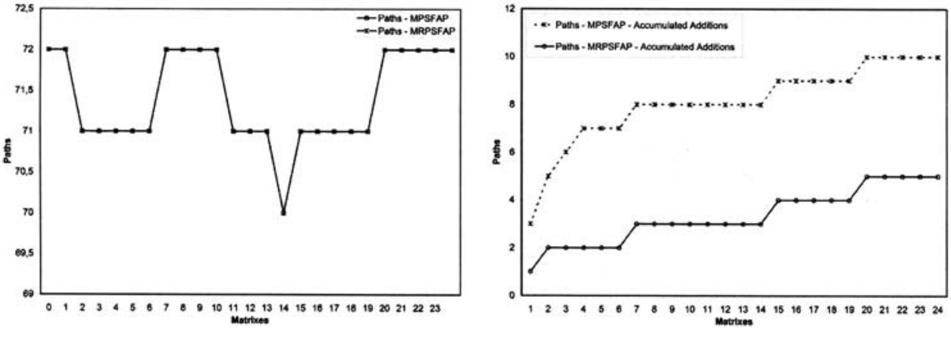


FIG. 8 – Modification of an optimal layout.

Modification d'une configuration optimale.

Figure 8 shows results obtained for the VTHD core network (9 nodes and 24 links) by using the MFD algorithm, for a series of 25 traffic matrixes obtained by applying a transformation of an initial traffic matrix (e.g. by changing the demand for a set of 40% of the nodes); this represents major traffic shifts. Two strategies are compared: the first strategy "Minimum Path Set and Flow Allocation Problem" (MPSFAP) consists in considering each traffic matrix independently of the others, and to compute the optimal dimensioning according to the static policy. The second "Minimum Reconfiguration and Path Set Flow Allocation Problem"(MRPSFAP) consists in using the reconfiguration policy for each new traffic matrix.

The figure on the left side shows the number of paths available at a given step. The fact that both curves are identical in this figure shows that the reconfiguring method achieves the optimal dimensioning in limiting the number of path for each traffic matrix.

The figure on the right side shows the total number of generated paths during the experiment, and shows that the reconfiguration policy indeed achieves a better overall performance than the static policy. Indeed, the number of paths that need to be added at the end of the cycle of 25 matrices is significantly reduced with respect to the number of paths being added by the dimensioning only objectives [3].

These numerical results, obtained on a realistic scenario, allow us to conclude on the interest of introducing sophisticated reconfiguration policies when computing optimal layouts on networks of the size of VTHD or larger, since these policies do limit resource usage while providing the required end-to-end QoS guarantees.

IV.3. Routing topology discovery

The retrieval of the backbone topological information is a major building block in many processes such as network dimensioning, logical topology design or traffic engineering implementation.

The main issue for realizing this function is to limit the configuration and signaling overheads, while not interfering with operational network processes and remaining efficient in terms of reactivity and robustness. Current solutions proposed by vendors in their OSS are based on SNMP and suffer from slow reactivity due to the update delay of counters in Management Information Bases (MIBS).

A "control plane" approach (as opposed to the SNMP "management plane" traditional approach) has been tested in VTHD in order to infer network topology. A topology server based on OSPF has been developed. This server offers a simple and efficient solution for discovering a backbone's topology.

The objective of the tool is to retrieve and extract topological information that is implicitly included in the link-state protocol paradigm. The concept of "Virtual Link" has been extended in order to define a "Monitoring Link". The network is configured as a single backbone area, where routers exchange and synchronize their link-state information. The "Monitoring Link" enables two distant machines to establish an OSPF neighboring through a transit area using any subsequent routing information. Hence, in order to retrieve the backbone topology, a local domain machine establishes a Monitoring Link with one backbone router. As shown in Figure 9, the monitoring machine thus acquires a complete dynamic view of the backbone topology.

This new tool has been tested on VTHD. In order not to interfere with forwarding decisions, the tool has been designed for OSPF. While IS-IS is the internal routing protocol used in the VTHD network, VTHD also supports OSPF but does not use it for forwarding traffic.

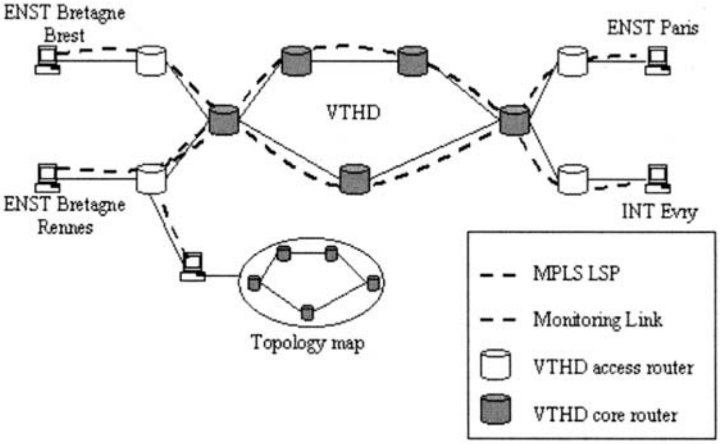


FIG. 9 – Monitoring network topology using link state routing information.

Surveillance de la topologie d'un réseau en utilisant l'information du routage à état de liens.

The topology server deployed on VTHD provides an explicit list of routers that are used by a traffic engineering signaling protocol, like RSVP-TE, to establish LSPs (Label Switched Paths). Moreover, network events such as failures or reconfigurations, are captured by the server in a timely manner, thus constituting an efficient tool for monitoring network state. Finally, the server enables to estimate and evaluate the routing convergence time after a failure, which could help in evaluating the performance of rerouting strategies. VTHD has offered an ideal demonstration field for the OSPF topology server, together with the opportunity to perform scalability tests on an operational network.

IV.4. Metrology at gigabit speed

Metrology consists in measuring achieved network performance, bandwidth usage and QoS parameters. It is a major issue for network operators in general and Internet Service Providers in particular both for technical issues (planning and dimensioning) and for marketing issues (customer relationship, pricing).

Several metrology strategies have been introduced in VTHD in order to study network performance and to analyze application behavior. The network performance delivered by VTHD is analyzed with SNMP based procedures. End-to-end service evaluation is achieved using commercially available metrology equipment provided by Ipanema [16] that performs passive end-to-end and an active end-to-end metrology tool, Saturne [5] that has been developed during the VTHD projects.

Passive measurement techniques consist in identifying individual packets related to the application currently evaluated, storing relevant events concerning these packets and processing this information to obtain global measures. This requires the capacity to both analyze an application flow at link speed and store the captured information for each measured packet. This explains why the few equipment that offer these features can analyze applications requesting only a limited percentage of backbone link bandwidth.

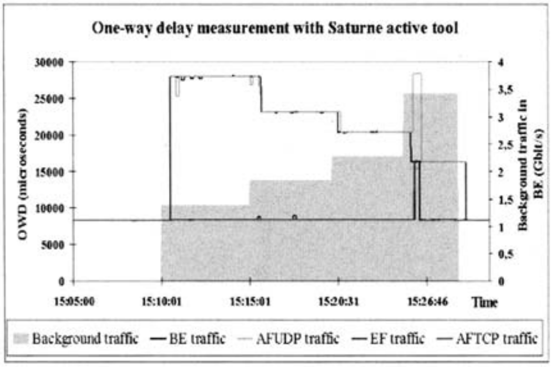
When using passive measurement techniques, it is often necessary to trade-off between measurement precision and available resources. Precision depends on measurement frequency, which is directly correlated to resources in terms of CPU, of memory, etc. The Ipanema passive measurement infrastructure installed in VTHD can scan a flow at link speed (Gbit/s) and analyzes up to 300 Mbit/s.

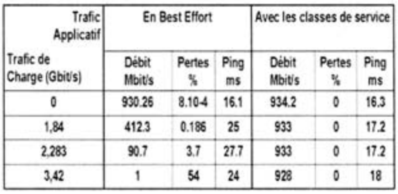
Active measurement is a lightweight alternative to passive measurement. It consists in generating a probe flow that is aggregated to the application flow in a router [5], and to estimate the QoS delivered to the application by measuring the QoS received by the probe flow. Aggregating the probe flow to the application flow ensures that the probe packets will receive the same treatment as the application packets.

The trade-off to be achieved regarding active measurement techniques is between measurement precision and disruption of the network service offered to the application. On the one hand, the probe flow must be small enough to ensure that it does not impact on network state and in particular does not generate congestion; on the other hand, it must be large enough to yield statistically significant measurements. It is shown in [6] that active measurement techniques can offer an accurate image of broadband network behavior.

Figure 10 shows a sample of the measurement results that can be obtained with the Saturne tool. The Saturne metrology platform offers a panel of measurement capabilities: local scope versus end-to-end scope, estimation of the entire bandwidth or of a portion of it, with or without service differentiation. Measurement frequency can be defined and the sampling law used to generate the probe packet is configurable.

The metrology facilities available with VTHD offer a remarkable performance estimation platform for any application or protocols to be deployed in a broadband WAN. In particular, this platform has been used to verify DiffServ implementation in VTHD.





Trafic Applicatif Trafic de Charge (Gbit/s)	En Best Effort			Avec les classes de service		
	Débit Mbit/s	Pertes %	Ping ms	Débit Mbit/s	Pertes %	Ping ms
0	930.26	8.10-4	16.1	934.2	0	16.3
1,84	412.3	0.186	25	933	0	17.2
2,283	90.7	3.7	27.7	933	0	17.2
3,42	1	54	24	928	0	18

FIG. 10 – One delay measurement using Saturne.

Mesure de délai unidirectionnel en utilisant l'outil Saturne.

IV.5. Testing the efficiency of IPV6 support

As pointed out in Section III.3, it was a primary requirement on VTHD to offer an efficient support of the IPV6 protocol stack. However, although experimental IPV6 backbones (6bone, G6bone in France) have been available for several years, back in 2002, none of those networks supported IPV6 beyond 1 Gbit/s. This requirement was therefore not straightforward to fulfill.

Some of the equipments deployed in VTHD were not even IPV6 compatible at that date. Moreover, others were compatible, but it was known that they had potentially been designed (optimized) to explicitly forward IPV4 packets and could only offer a poor IPV6 connectivity. Indeed, it has been verified during the VTHD projects that some available IPV6 forwarding implementations were software-based, which implied that they could achieve only 5% of the nominal interface throughput.

It was then deemed necessary to implement transition mechanisms to bypass existing equipment, and thus to test the three available alternatives:

- Classical IPV6 in IPV4 tunnel where IPV6 packets are encapsulated in IPV4 packets,
- 6 Provider Edge tunneling (6PE) which is a transition solution allowing to encapsulate IPV6 packets at the edge of the network and to switch them efficiently in a MPLS/IPV4 core network. The principles of these mechanisms are close to the L3 VPN MPLS architecture defined in RFC2547 bis.
- A dual stack approach in which each router interface runs both IPV4 and IPV6.

The test architecture is represented in Figure 11. The test devices are Smartbit600 with POS OC48c interfaces. A Juniper M40 router in Paris sends IPV6 traffic to a Juniper T640 router located in Rennes. The two Juniper routers are shipped with Tunnel Physical Interface Cards; these hardware cards are designed to support IPV6 in IPV4 tunneling functions with a good level of performance.

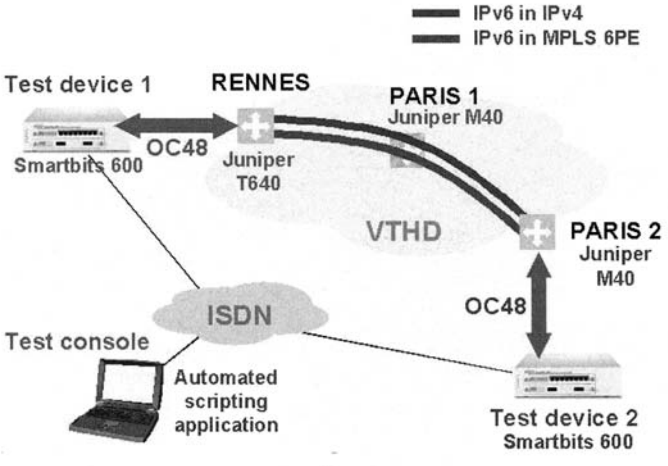


FIG. 11 – IPV6 testing architecture in VTHD.

Architecture de test de l'implémentation IPV6 de VTHD.

The testing methodology is designed according to the standard benchmarking methodology described in [25]. The test console runs TCL scripts for computing frame loss rate and achieved using the Application Programming Interfaces (API) provided by the test devices. Measurements are performed for several packet sizes as small packets may lead to degraded performance.

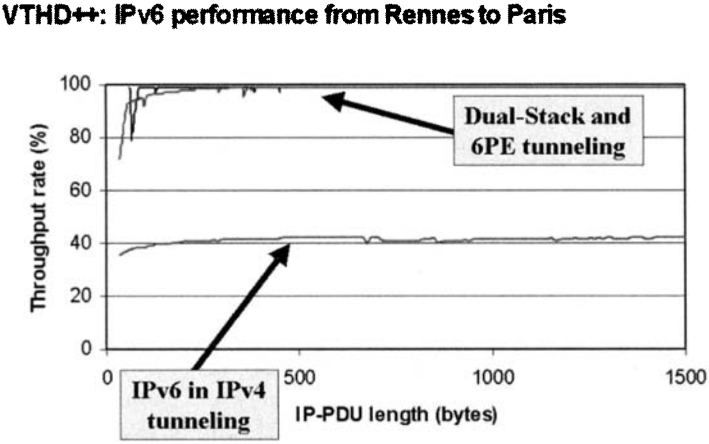


FIG. 12 – Comparing alternatives for supporting IPv6 on 2.5 Gbit/s interfaces.

Comparaison des solutions permettant de supporter IPV6 sur des interfaces à 2,5 Gbit/s.

Measurement results are shown in Figure 12. Both dual-stack and 6PE solutions achieve almost wire speed performance, whereas IPv6 in IPv4 tunneling showed a quite good (if compared to classical software tunneling) but limited performance in spite of the use of dedicated hardware. These basic tests have been completed with additional ones in which security related features such as packet filtering using access lists or an anti-spoofing mechanism based on Unicast Reverse Path Forwarding (URPF). The additional tests showed no impact due to these additional features.

These test results led to the operational strategy described in Section III.3, and has been successfully exploited by several IPv6 applications running over VTHD.

V. CONCLUSION

This paper has proposed an overview of the results obtained within the VTHD and VTHD++ projects. These projects and the deployment of the VTHD network, an experimental gigabit IP/MPLS network, have made possible realistic experimentations of both advanced services and cutting-edge applications requiring high-performance network services not currently available on existing production networks.

The ability to modify networking protocols settings, to implement experimental traffic engineering principles and to generate and control realistic loading conditions have enabled the VTHD partners to gain an expertise that could not have been obtained using only production IP backbones. Indeed, when a network has operational commitments, the network operator must avoid these types of experiments in order to protect the network from potential collapse.

Moreover, these projects have brought together networking experts with advanced application designers, who rarely interact. This cross-fertilization process is necessary for optimally designing networks that will be able to support new communication paradigms.

Lastly, the academic partners have been able to check their theoretical proposals on commercially-available networking products, in a wide-area network; this has led to findings concerning e.g. the scalability of solutions and the robustness of applications in realistic settings. On the other hand, France Telecom has had the opportunity to study and test advanced features (e.g. fast reroute, MPLS VPN....) of new equipment (giga and tera routers), the opportunity to design and to experiment in real conditions new advanced services, while benefiting of this collaboration with academic experts.

The members of the VTHD community are well aware of the exceptional opportunities offered by a nationwide experimental gigabit network within partnerships such as VTHD and VTHD++. All of these results have been made possible by the partial public funding of the project for 5+ years. Although this public funding might still be available in the future, the VTHD community is also considering the opportunity to evolve from short-term partnership to a more permanent and more extended structure.

There are numerous studies that can be further envisaged on the VTHD platform such as e.g. transport layer adaptation for GRID computing applications (TCP tuning, network-aware applications) or advanced network services for GRID and video streaming applications, or even MPLS and traffic engineering improvements. Furthermore, most of the network services available in VTHD have been deployed in a single provider environment, and testing the support of these services in a multi-operator context is a challenge in itself.

Manuscrit reçu le 31 juillet 2004
Accepté le 7 septembre 2004

REFERENCES

[1] ADAM (Y.), FILLINGER (B.), ASTIC (I.) LAHMADI (A.), BRIGANT (P.), Deployment and Test of IPV6 Services in the VTHD Network, *IEEE Communications Magazine*, January 2004, IEEE.

[2] AWDUCHE (D.), MPLS Traffic Engineering in IP Networks, *IEEE Communications Magazine*, December 1999, IEEE.

[3] BEKER (S.), Optimization Techniques for the Dimensioning and Reconfiguration of MPLS Networks, *Thesis Report*, April 2004, ENST.

[4] CORNILLEAU (J.-M.), TADONKI (P), GOMBAULT (S.), BERNIER (JL.), Contrôle d'accès dans les réseaux haut débit. *SETIT'03, Conférence internationale Sciences électroniques, technologies de l'information et des télécommunications*, IEEE, mars 2003.

[5] CORRAL (J.), TEXIER (G.), TOUTAIN (L.), End-to-End Active Measurement Architecture in IP Networks (SATURNE). *Passive and Active Measurement Workshop Proceedings,* La Jolla, CA, USA, pp. 241-247, April 2003.

[6] CORRAL (J.), INCERA (J.), TEXIER (G.), TOUTAIN (L.), HEU (JR), Saturne active measurement tool within the metrology context. *IEEE ROC&C'2003, Reunión de Otoño Comunicaciones, Computación,* Acapulco, Mexique, 2003.

[7] PAN (P.), SWALLOW (G.), ATLAS (A.), Fast Reroute Extensions to RSVP-TE for LSP Tunnels, *draft-ietf-mpls-rsvp-lsp-fastreroute-06.txt, work in progress.*

[8] PORETSKI (S.), KHANNA (R.) PAPNEJA (R.), RAO (S.), LE ROUX (J.L.), Benchmarking Methodology for MPLS Protection Mechanisms, *draft-poretsky-mpls-protection-meth-03.txt, IETF draft, work in progress.*

[9] FRATTA (L.), GERLA (M.), KLEINROCK (L.), The Flow Deviation Method: An Approach to Store and Forward Communication Network Design, *Networks,* (3), p. 97-133, 1973.

[10] GARCES-ERICE (L.), BIERSACK (E.W), ROSS (K), FELBER (P), URVOY-KELLER (G.), Hierarchical Peer-to-Peer Systems, *In the Special issue of the Parallel Processing Letters (PPL),* **13** (4), p. 643-657, December 2003

[11] GELAS (JP.), EL HADRI (S.), LEFEVRE (L.), Towards the Design of an High-performance Active Node, *Parallel Processing Letters journal,* **13**, n° 2, pp. 149-167, June 2003

[12] http://abilene.internet2.edu/

[13] http://www.rd.francetelecom.com/en/brevets/e-conf/econf.php

[14] "Innovation Gallery" at http://www.francetelecom.com/en/

[15] http://www.urec.cnrs.fr/etoile/

[16] http://www.ipanematech.com/

[17] http://www.metasolv.com/

[18] http://www.irisa.fr/paris/Paco++/welcome.htm

[19] http://www.sensable.com/products/phantom_ghost/phantom.asp

[20] http://www.vthd.org/

[21] JEGOU (Y), Performance Analysis of Code Coupling on a Long Distance High Bandwidth Network. *EuroPar 2002, Parallel Processing, volume 2400 of Lect. Notes in Comp. Science,* pp. 753-756, August 2002

[22] MAIMOUR (M.), PHAM (C.), DYRAM: an Active Reliable Multicast framework for Data Distribution, *Journal of Cluster Computing,* 7(2), pp. 163-176, April 2004.

[23] PARMENTELAT (T.), GOURDON (A.), TURLETTI (T.), LARREUR (E.), V-eye, a very large scale virtual environment for multimedia conferencing, *INRIA,* RT Number 0296, May 2004.

[24] PAUL (O.), LAURENT (M.), GOMBAULT (S.), A Full Bandwidth ATM Firewall. *Proceedings of ESORICS'2000,* Lecture Notes in Computer Science, LNCS 1895, Springer-Verlag, October 2000.

[25] BRADNER (S.), MAC QUAID (J.), Benchmarking Methodology for Network Interconnect Devices. *RFC1944,* May 1996, IETF.

[26] ROSEN (E.), VISWANATHAN (A.), CALLON (R.), Multiprotocol Label Switching Architecture, *RFC3031,* January 2001, IETF.

[27] LE FAUCHEUR (F.), DAVIE (B.), DAVARI (S.), VAANEN (P.), KRISHNAN (R.), CHEVAL (P.), HEINANEN (J.), Multiprotocol Label Switching (MPLS) Support for Differentiated Services, *RFC3270,* May 2002, IETF.

[28] RODRIGUEZ (P.), BIERSACK (E.W), Dynamic Parallel-Access to Replicated Content in the Internet. *IEEE/ACM Transactions on Networking,* **10** (4), p. 455-464, August 2002.

[29] ROSS (K), Hash-routing for Collections of Shared Web Caches, *IEEE Network Magazine,* **11**(7), p. 37-44, Nov.-Dec. 1997.

[30] STOICA (I.), MORRIS (R.), KARGER (D.), KAASHOEK (M), BALAKRISHNAN (H.), Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications. *In Proceedings of the 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications,* San Diego, California, USA, ACM Press, p. 149-160, August 2001.

Acronyms

ACL	Access Control Lists
AF	Assured Forwarding
API	Application Programming Interfaces
AS	Autonomous System
ASM	Any Source Multicast
BGP	Border Gateway Protocol
BO	Back Office
CPU	Central Processing Unit
DDOS	Distributed Denial of Service
DOS	Denial of Service
DV	Digital Video
EF	Expedited Forwarding
FRR	Fast ReRoute
IFT	IP Fast Translator
IPSA	IP Service Activator
IS-IS	Intermediate System to Intermediate System
LSP	Label Switched Path
MFD	Modified Flow Deviation
MIB	Management Information Base
MINLP	Mixed Integer Non-Linear Program
MPLS	Multi Protocol Label Switching
MPSFAP	Minimum Path Set and Flow Allocation Problem
MRPSFAP	Minimum Reconfiguration and Path Set Flow Allocation Problem
NACK	Negative ACKnowledgement
OSPF	Open Shortest Path First
OSS	Operational Support System
P2P	Peer-to-peer
QOS	Quality of Service
RIPE	Réseaux IP Européens
RTE	Robotized Tele-echography
RSVP-TE	ReSerVation Protocol with Traffic Engineering extensions
SNMP	Simple Network Management Protocol
SSM	Source Specific Multicast
TCL	Tool Command Language
TCP	Transmission Control Protocol
TE	Traffic Engineering
UDP	User Datagram Protocol
VPN	Virtual Private Networks
VTHD	Vraiment Très Haut Débit
WAN	Wide Area Network
WDM	Wavelength division multiplexing