Towards the design of an active network architecture supporting throughputs of gigabit networks

«Vers la conception d'une architecture de réseaux actifs apte à supporter les débits des réseaux gigabits»

Friday, December 5th, 2003

Jean-Patrick Gelas

RESO/LIP - INRIA/UCBL/CNRS/ÉNS de Lyon

Université Claude Bernard Lyon I - FRANCE

Overview

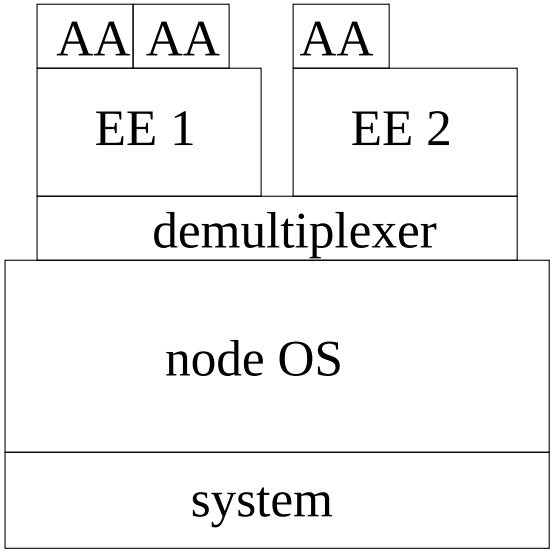
- Introduction to Active Networks
- Propositions for an High Performance Active Network
- The Tamanoir project
- Experimentations
- Conclusions and future works

Clever use of the network . . .

"Active networks allow individual user, or groups of users, to inject customized programs into the nodes of the network" (DARPA)

- Concept born in mid-90's, MIT [Tennenhouse and Wetherall 96]
- Wide research area (USA, Europe, Asia), academic and industrial

Active Node Architecture (DARPA)



- Application Active (AA) also called *Active Service*.
  - data plan
  - control plan
  - management plan
- Execution Environment (EE).
- Demultiplexer used to direct active packets to the required EE.

Two different approaches

- Integrated approach
- capsule : code and data in the same stream (in-band code injection)

Service CODE	DATA

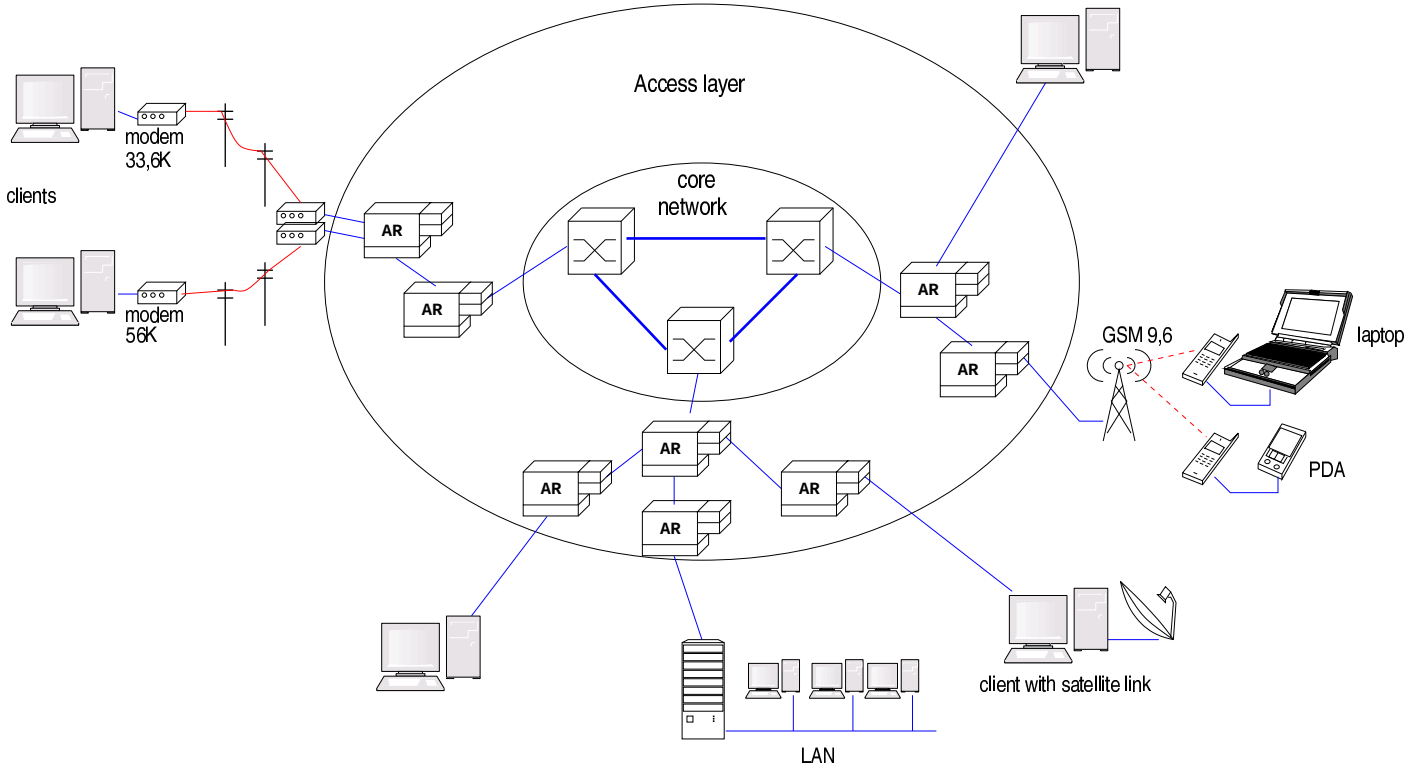
[Active IP, Smart Packets, Switchware, ANTS, . . .]

- Discrete approach
- active packet : just a reference to a service (out-band service deployment)

Service Id.	DATA

[Protocol Boosters, Flexinet, . . .]

Why do we need Active Networks ?



- Better heterogeneity support
  - applications requirements, networks features, clients capabilities.
- Experimental platform for new services

Issues

Two points prevent the deployment of the Active Networks technology :

- Security
  - service deployment
  - ressource allocation
- **Performance**
  - minimize overhead transmission
  - efficient EE

High Performance Active EE

- Software approach :
  - C language, kernel or user space : PAN [Nygren et al.,MIT,99]
  - Cluster based : CLARA [Ott et al., NEC,00]
  - Kernel services : in Solaris [Bhattacharjee et al., 97]
- Dedicated Hardware approach :
  - ANN : Active Network Node [Decasper et al. U.Washington and ETHZ, 99]
  - NP : Network Processor [Wolf et al.,01]
  - FPGA : P4 project [Hedzic et al., U.Penn,97], APE [Takahashi et al., NTT,02]

Contributions

- Design an active node architecture
  - high performance
  - software based
  - service oriented : horizontal and vertical deployment
  - providing solutions for all plan
- Implement this architecture
  - open, available and easy to deploy
- Deployment and evaluation over Gigabits networks and provide additional tools.

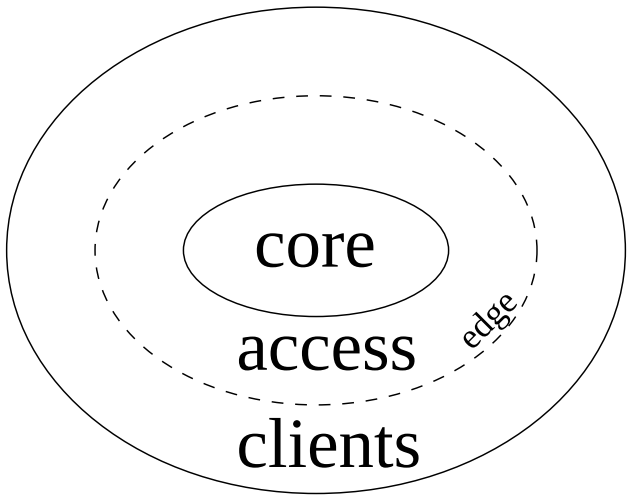
Overview

- Introduction to Active Networks
- **Propositions for an High Performance Active Network**
- The Tamanoir project
- Experimentations
- Conclusions and future works

Where for High Performance ?

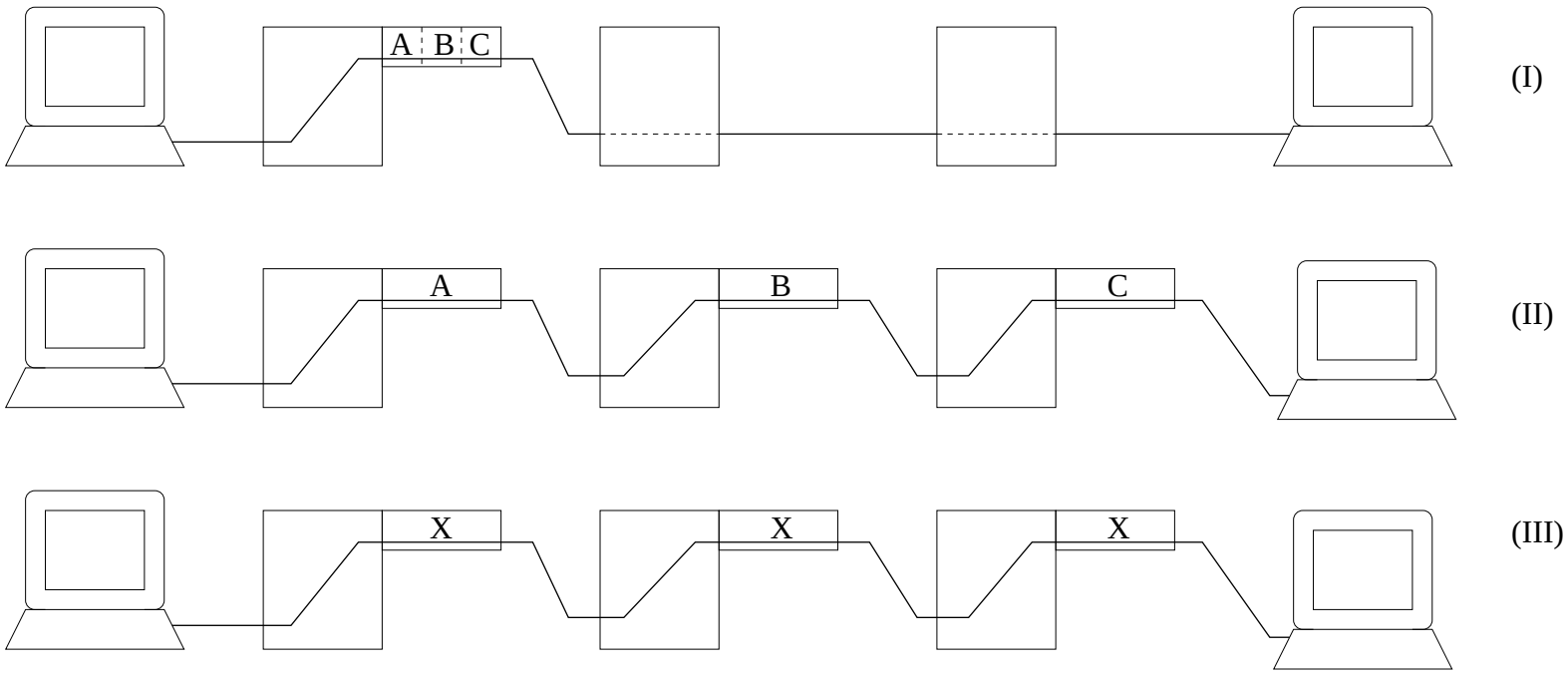
- Active Routers location
- Services in the Active Network : Horizontal deployment
- Services in the Active Node : Vertical deployment

Active Routers location



- core network
  - pros: good knowledge of the network status
  - cons: optical to electronic conversion required for processing
- access network and edge routers
  - pros : slow throughputs and heterogeneity, process only active packets, good location in the network
  - cons : edge router have a limited knowledge of the network

Horizontal deployment (or Services urbanisation)

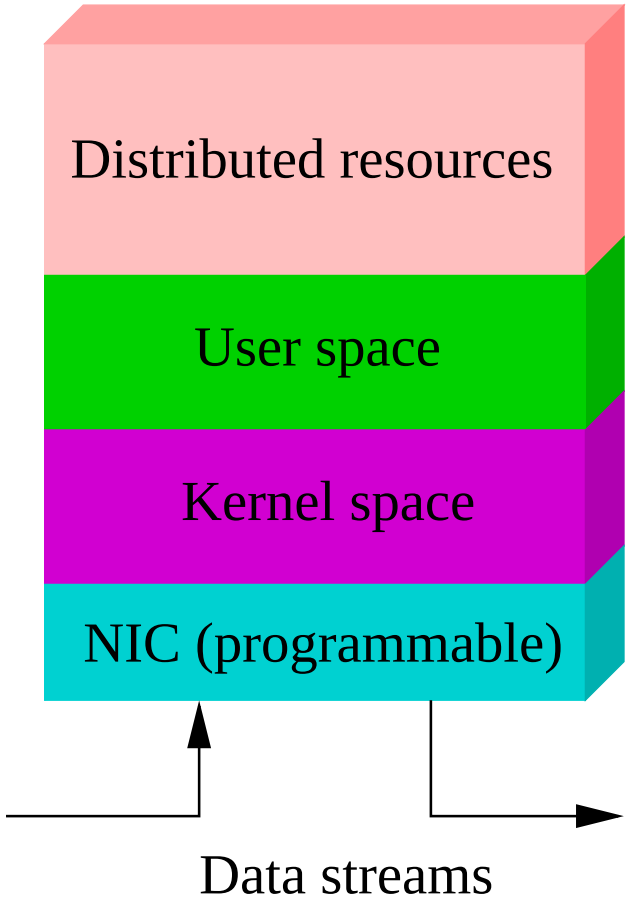


- composed services
- opportunistic deployment
- pipelined processing, recycled services

Service classification

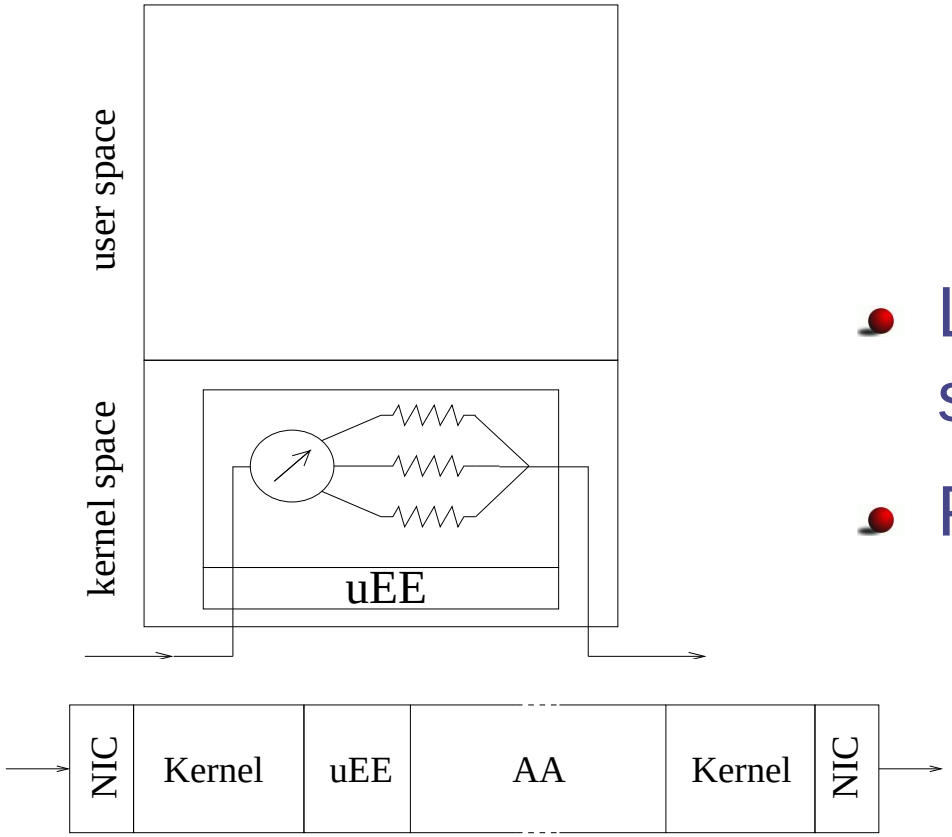
- Hyper lightweight : consume few CPU cycles, no memory space required
- Lightweight : consume few CPU cycles, few space memory
- Middle : required a rich environment for complex processing, can access all resources
- Heavy : CPU or space memory consumer. Must be distributed.

Vertical deployment



- NIC: close to the link but few memory space and limited processing capabilities
- Kernel space: general purpose processor, limit packet ascent
- User space: general purpose processor, easy development but far from the link
- Distributed ressources: parallelism, large space memory

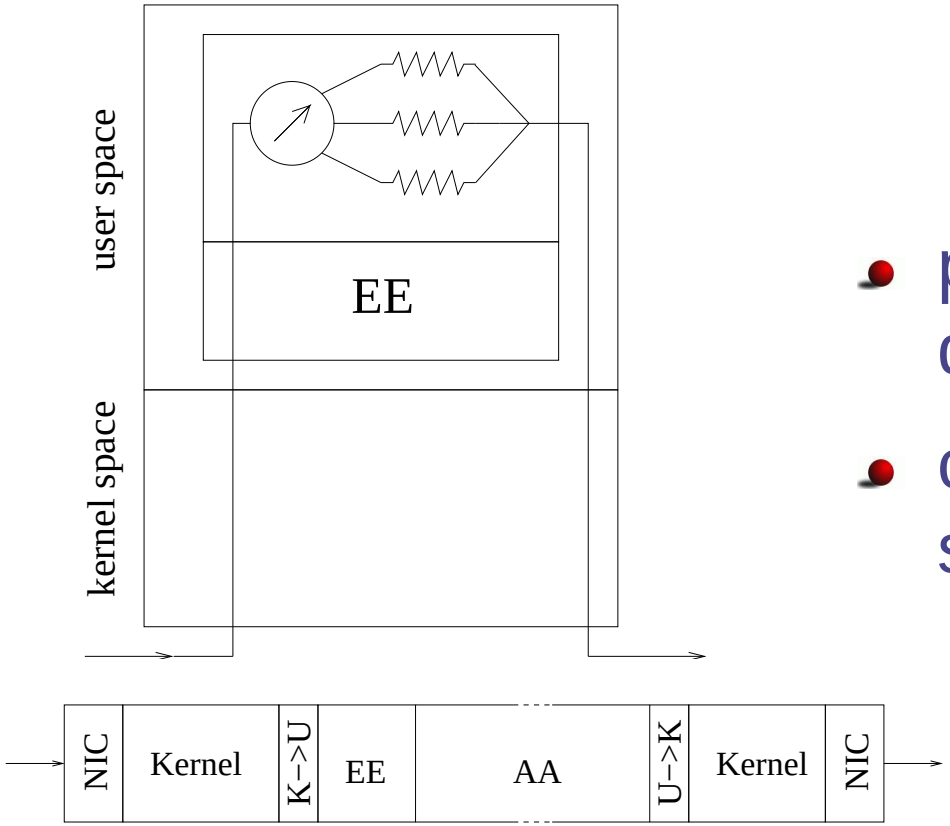
Kernel space level



- Limit packet ascent in the system
- Packet processing near the link

$$T(t_{AA}) = 2(T_{NIC} + T_K) + T_{\mu EE} + t_{AA}$$

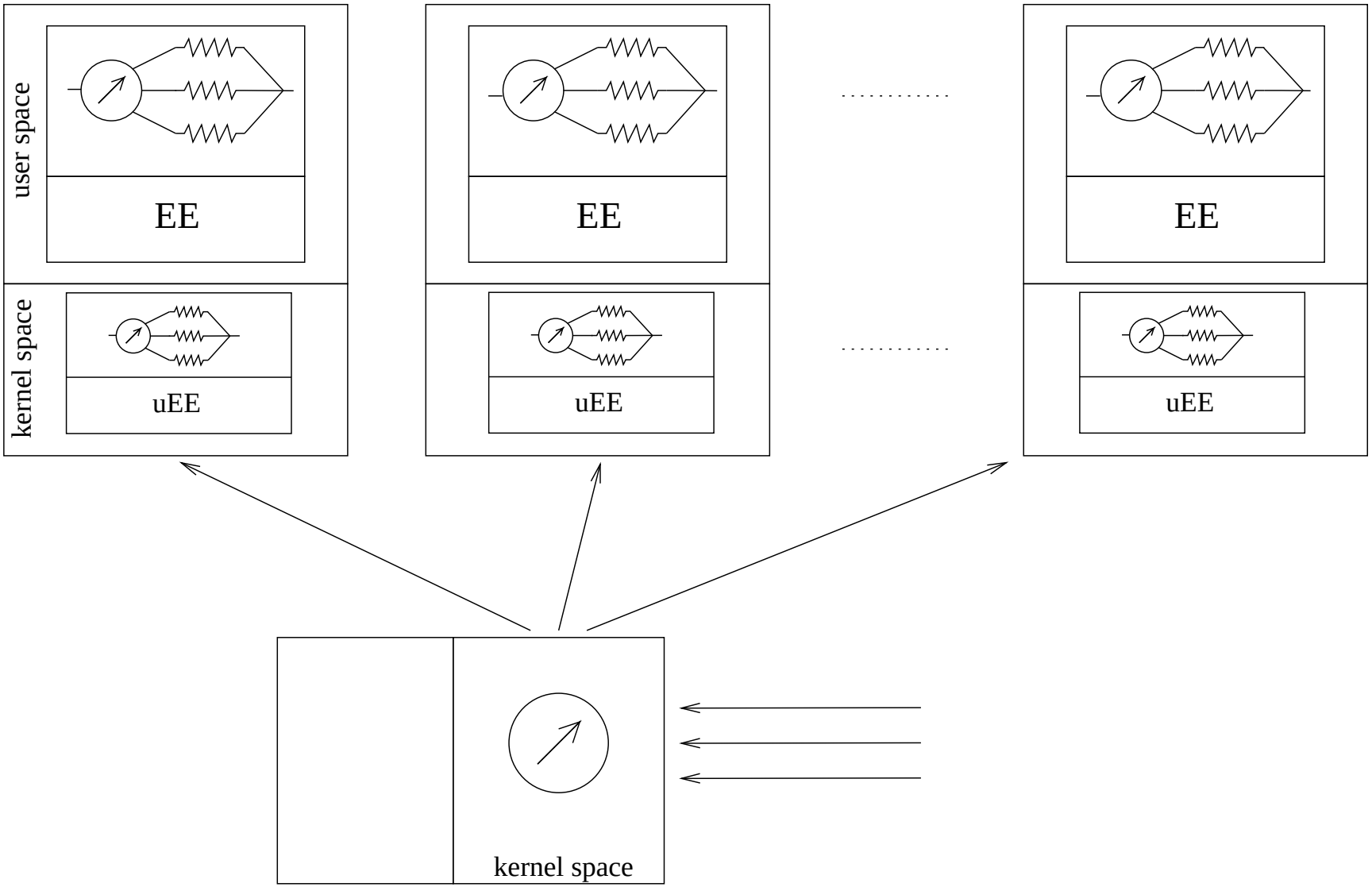
User space level



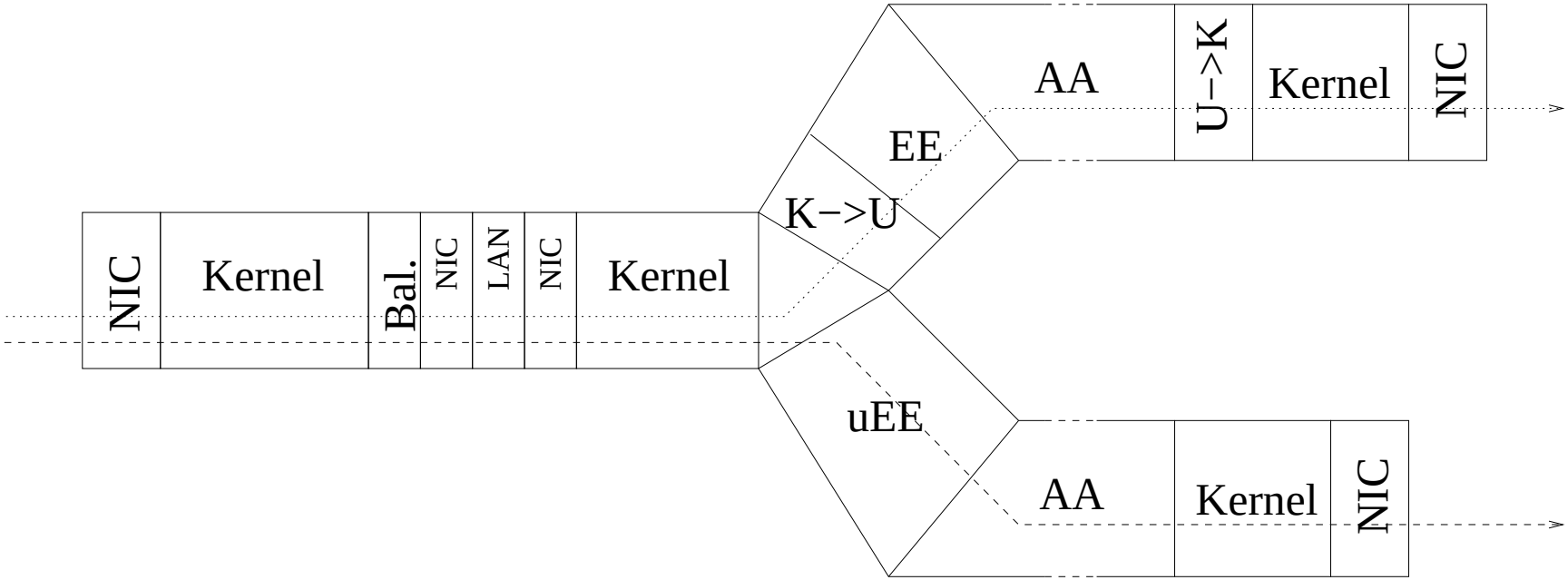
- pros: service deployement and development easier
- cons: copy from kernel to user space

$$T(t_{AA}) = 2(T_{NIC} + T_K + T_{U \to K}) + T_{EE} + t_{AA}$$

Replicated EE



Replicated EE



$$C = T_{NIC} + T_K + T_{Bal} + T_{NIC} + T_{LAN}$$

$$\Theta_u(t_{AA}, n, N) = n(C + \frac{2(T_{NIC}+T_K)+T_{EE}+2T_{U\rightarrow K}+t_{AA}}{N})$$

$$\Theta_k(t_{AA}, n, N) = n(C + \frac{2(T_{NIC}+T_K)+T_{\mu EE}+t_{AA}}{N})$$

Θ : Time to cross a TAN for a finite size stream

n : number of packets, N : number of Backends

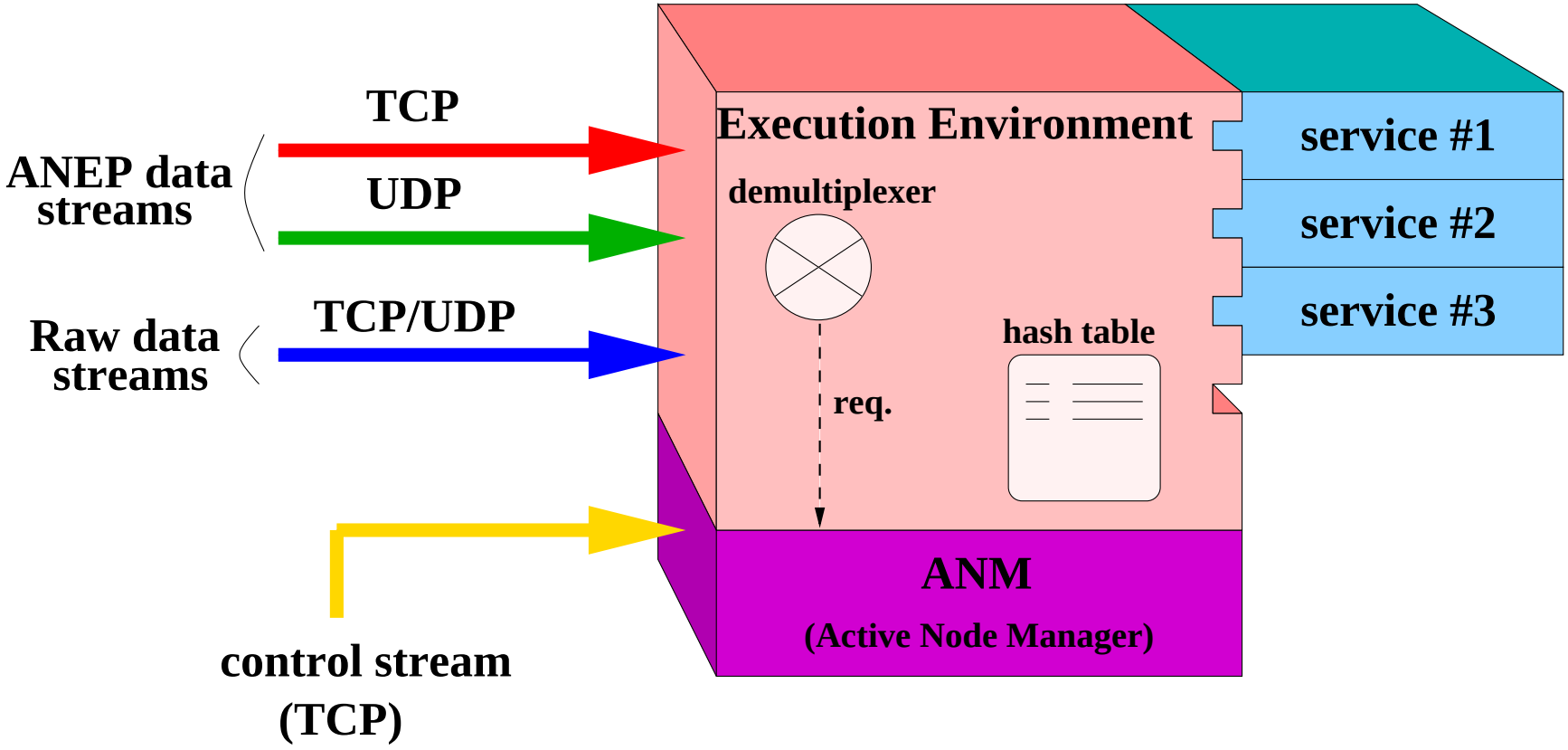
Overview

- Introduction to Active Networks
- Propositions for an High Performance Active Network
- **The Tamanoir project**
- Experimentations
- Conclusions and future works

The Tamanoir project

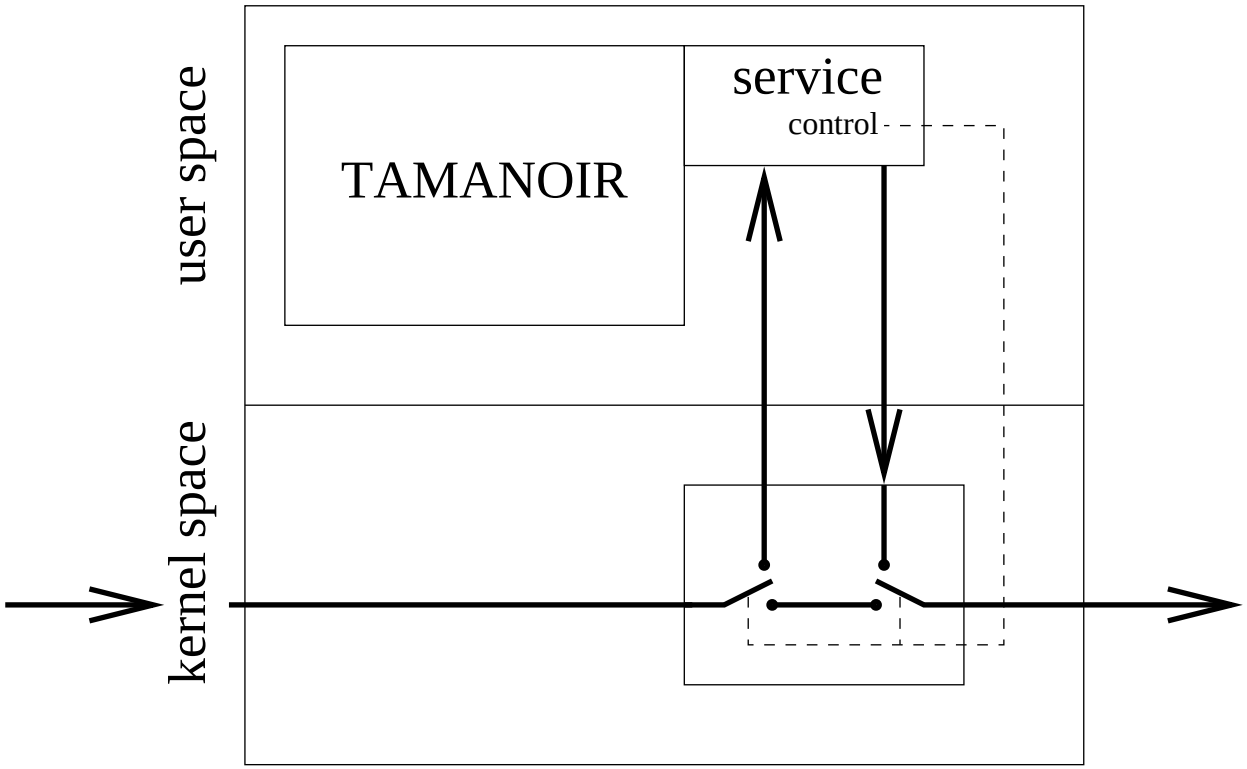
- Tamanoir Active Node (TAN)
  - Netfilter for kernel space
  - Java for user level
  - Linux Virtual Server (LVS) for distributed level
- Service deployment mechanism
- Supplemental tools
  - Distributed active streams generator : Echidna
  - Active Network Moniroring Tools : Pangolin

Tamanoir Active Node in user space



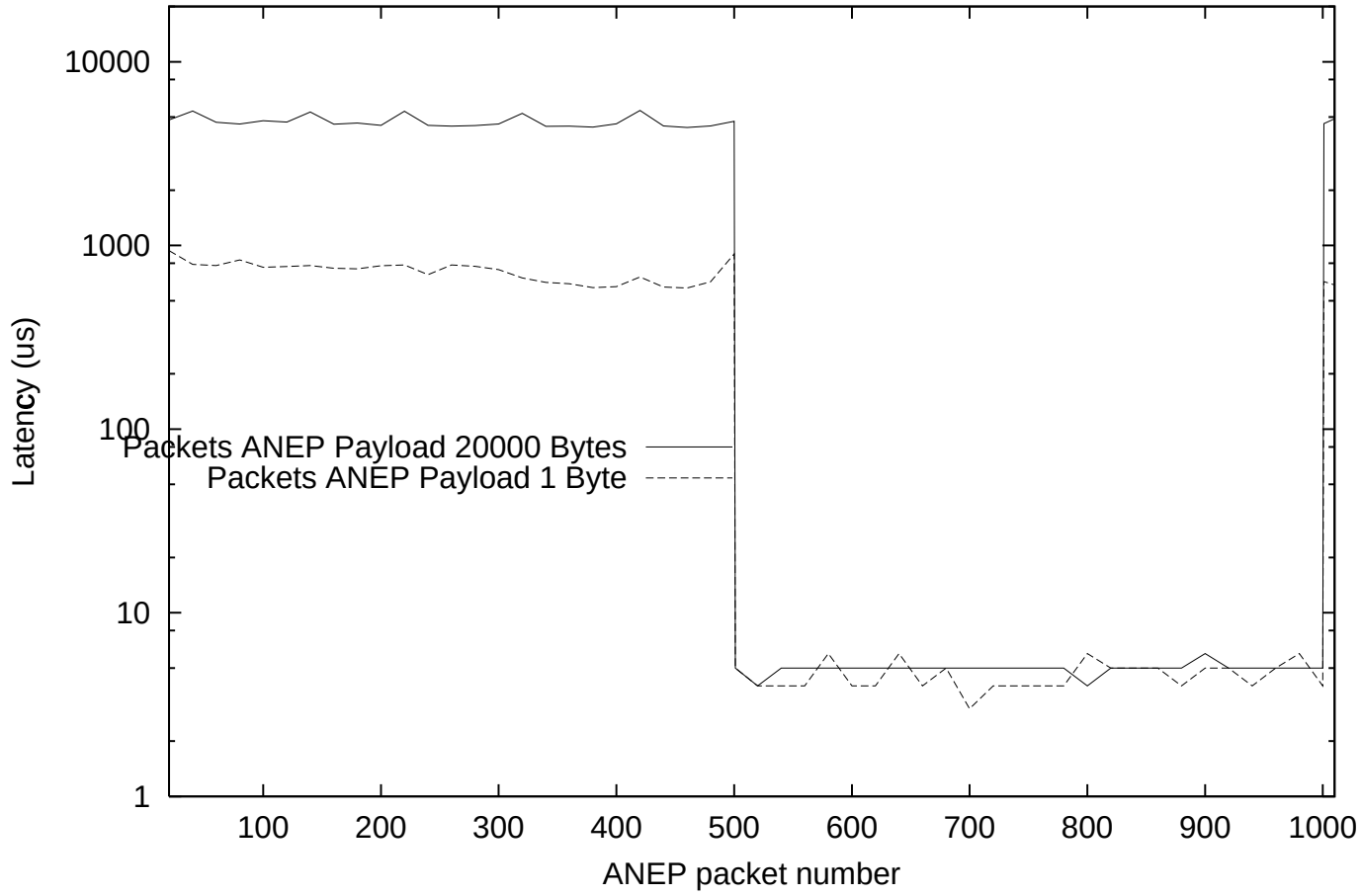
- An Execution Environment EE, a manager (ANM) and services dynamically loaded.
- Muli-thread, Java, UDP and TCP

Communication kernel ↔ user space



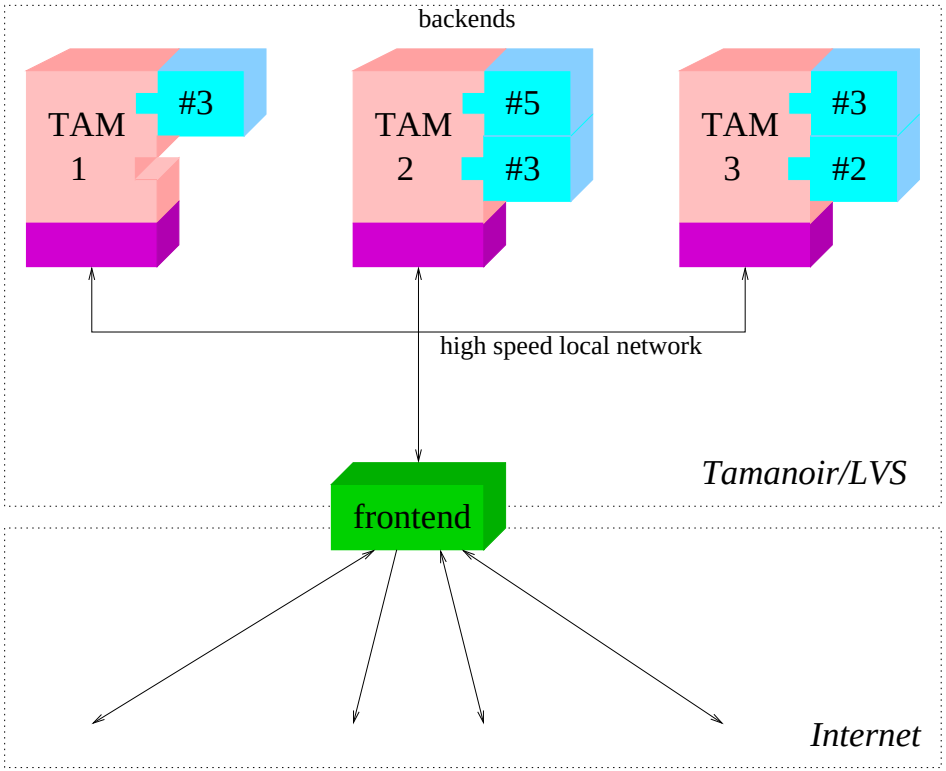
- Associate active services to a Netfilter hook.

Latencies - Kernel vs. User space



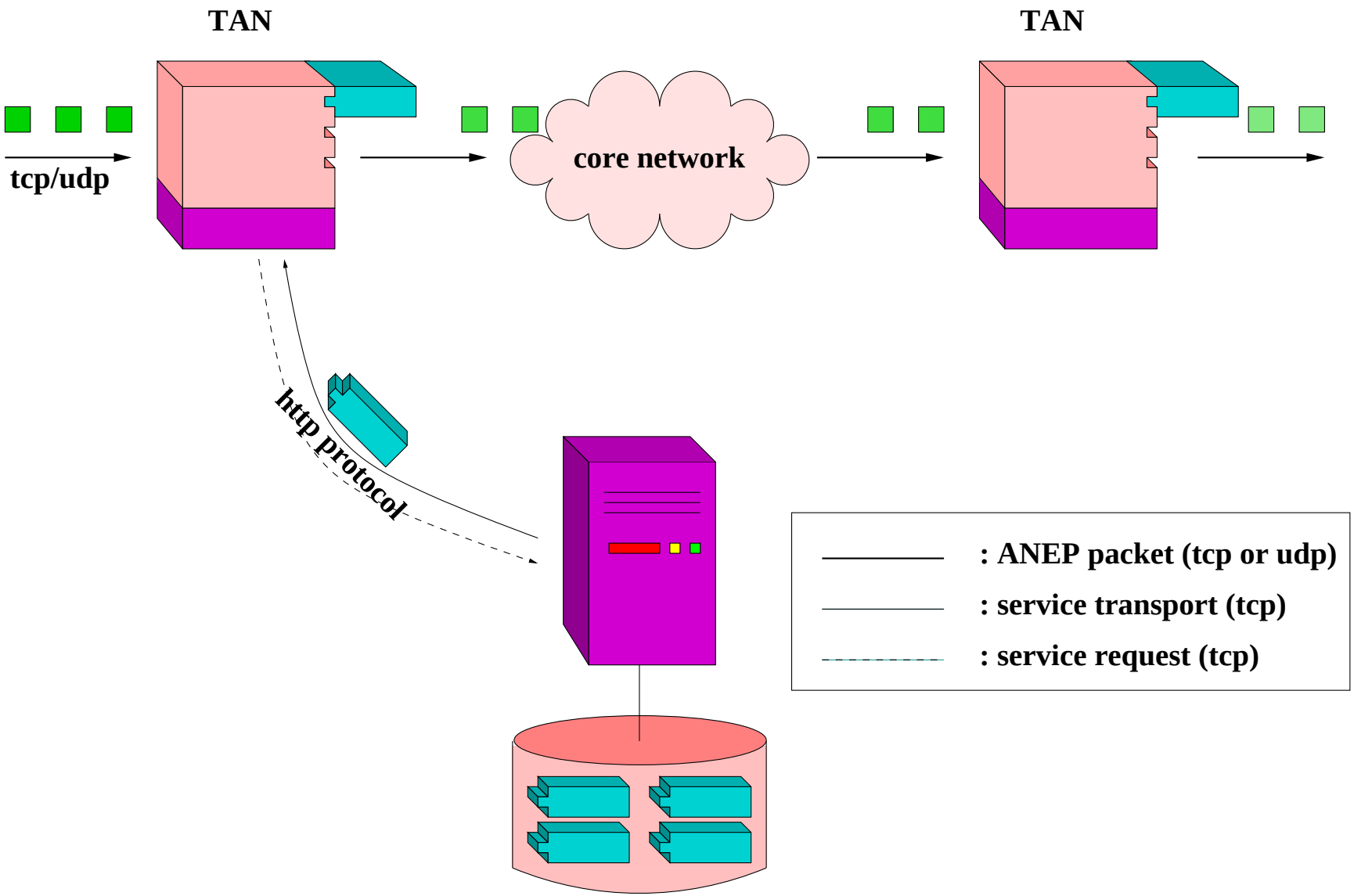
- A speedup of 1000.

Distributed resources - replicated EE

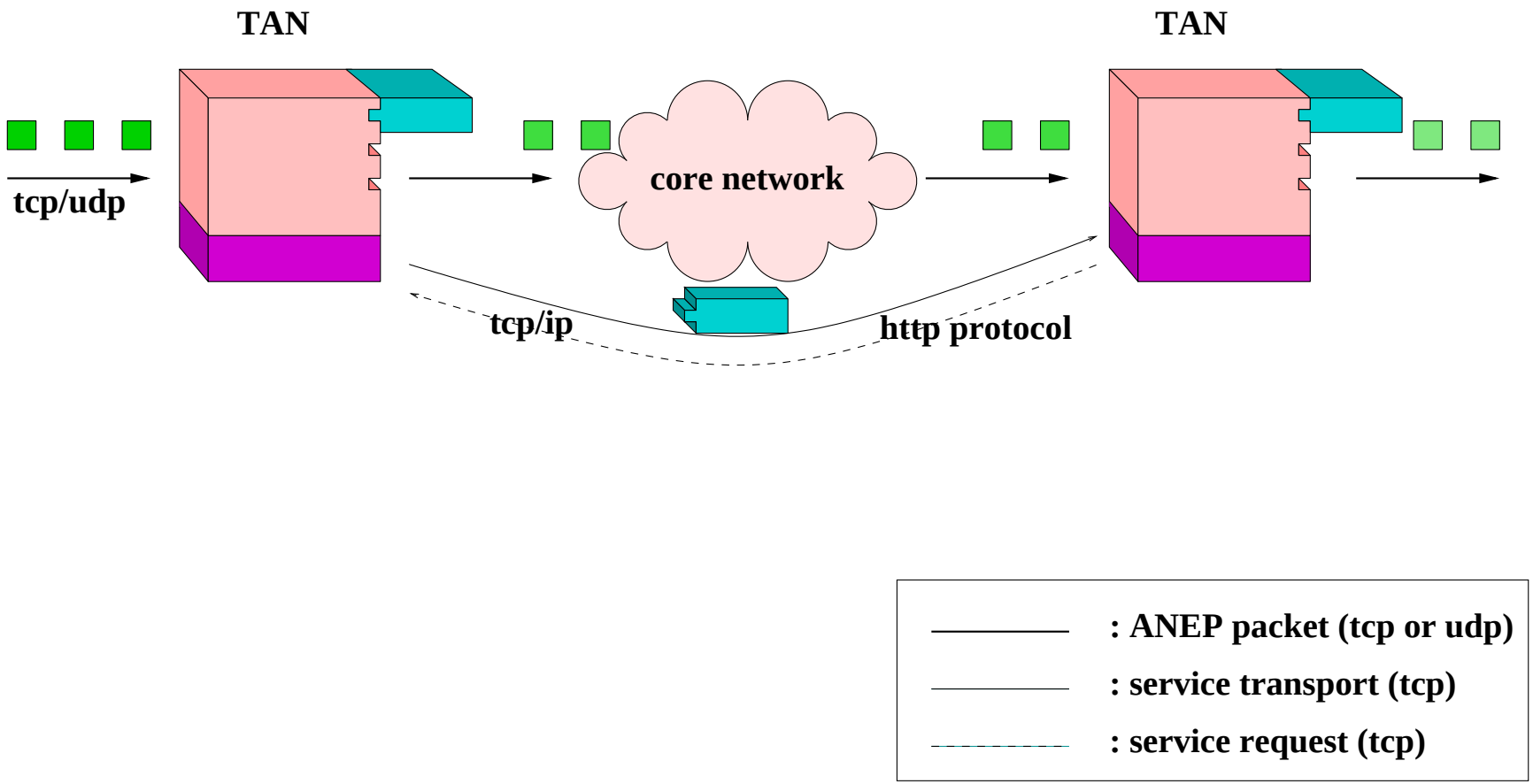


- Based on LVS (Linux Virtual Server) project
- Frontend (director) used to distribute connections
- Load balancing policy
- Stream(tcp) or packet(udp) granularity
- Backend (BE) service deployment

Service deployment - service repository



Service deployment - step by step



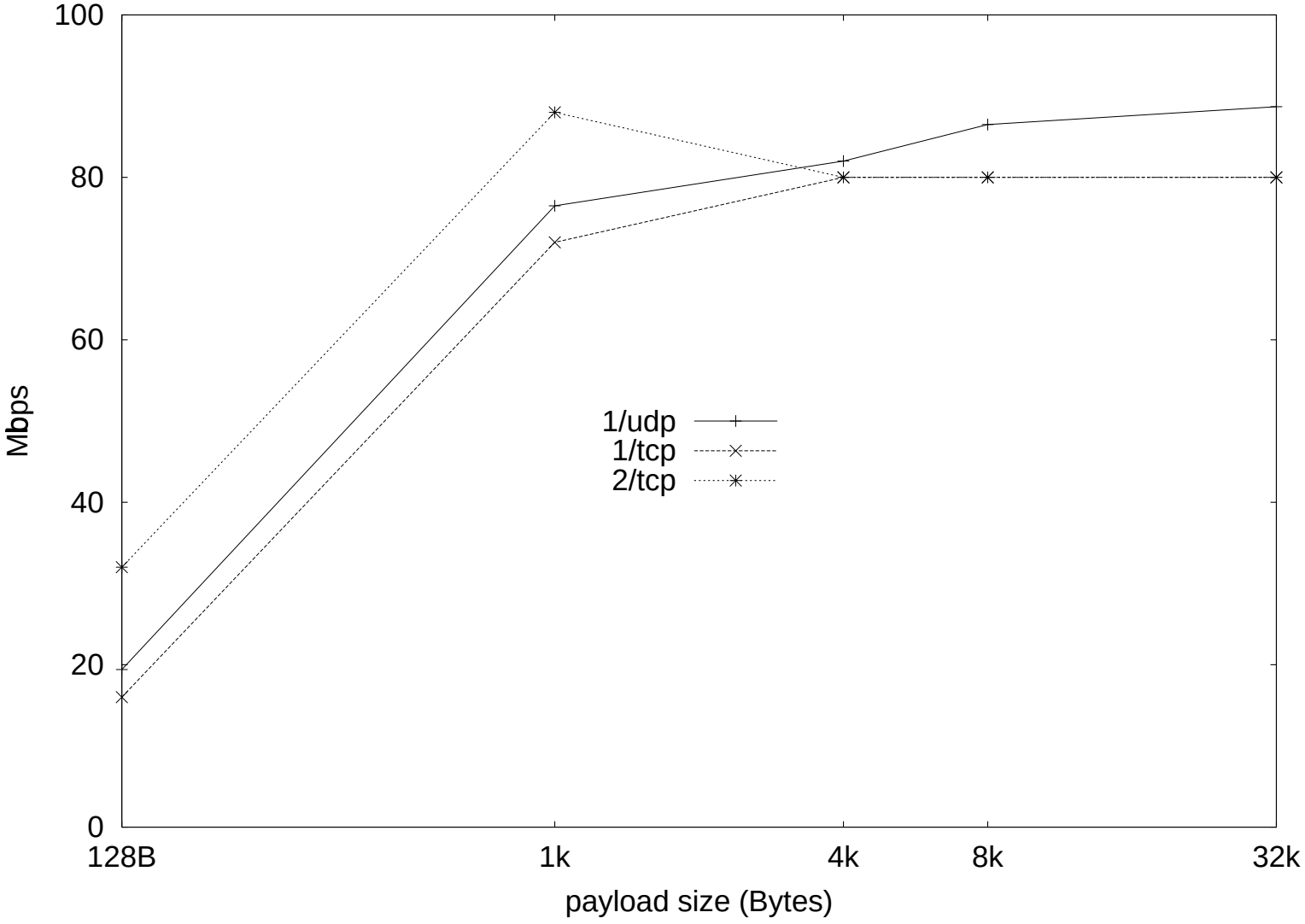
Overview

- Introduction to Active Networks
- Propositions for an High Performance Active Network
- The Tamanoir project
- **Experimentations**
- Conclusions and future works

Experimental context

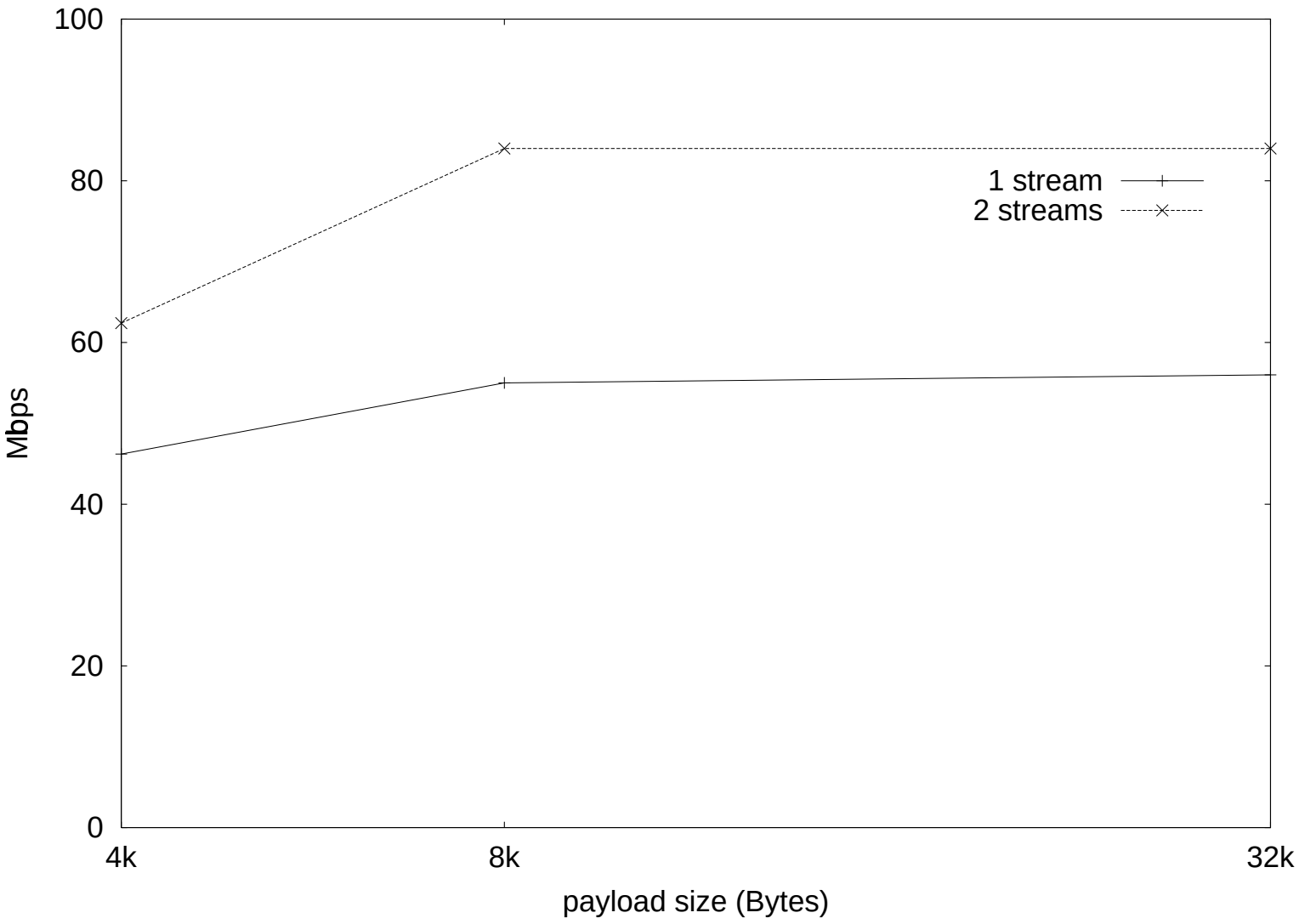
- Local : Fast Ethernet, Giga Ethernet (1Gbps), Myrinet (>1.8Gbps)
- Wide : VTHD (RNRT VTHD++ project) (Giga Ethernet)
- Dual-processor, 1.4GHz, Pentium III, 66 MHz PCI bus
- Linux 2.4.x, JDK IBM 1.3
- Ligthweight and Heavy service scenario :
  - Packet counting and time stamping
  - On the fly data compression

Fast Ethernet — lightweight service



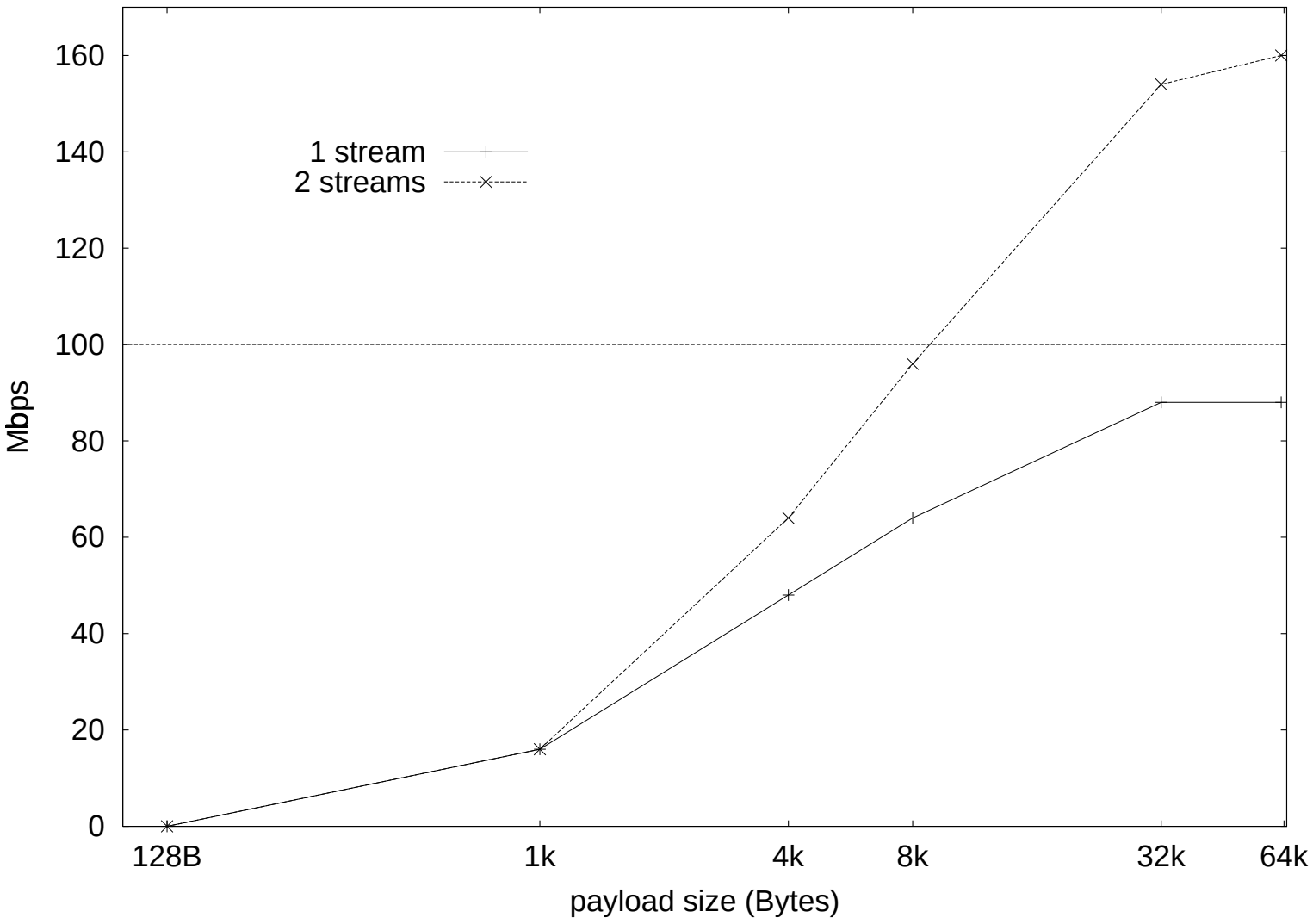
- Saturation of a Fast Ethernet link in user space with Java

Fast Ethernet — heavy service/TCP



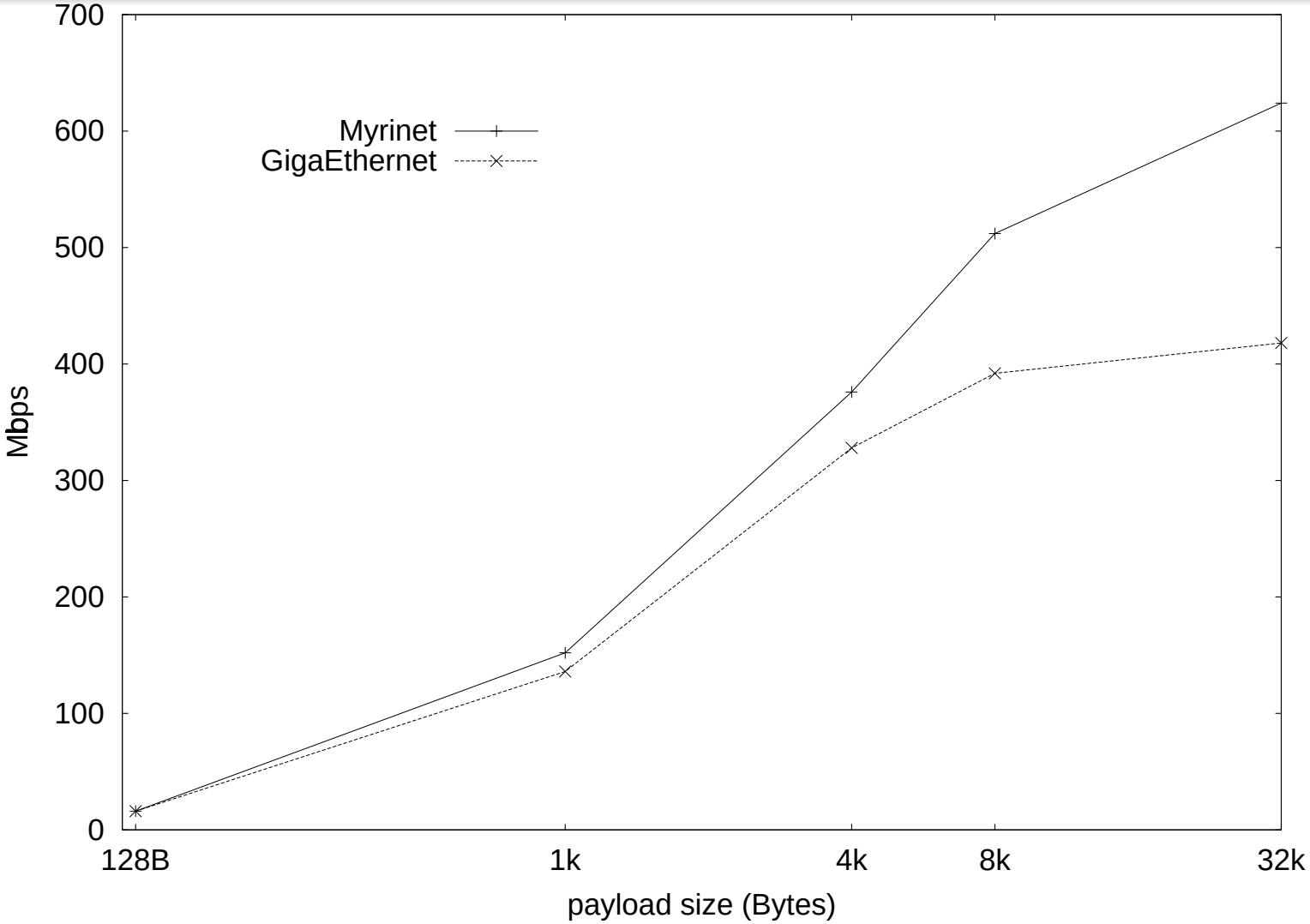
- Take advantage of SMP architecture

Giga Ethernet — heavy service/TCP



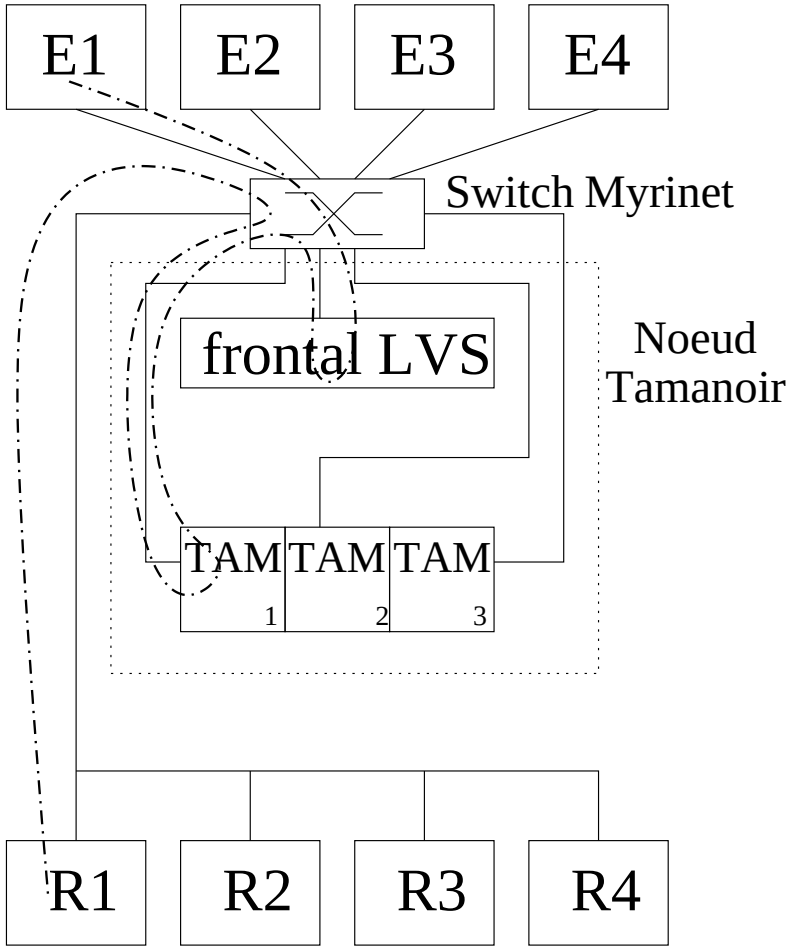
- Big packets are well suited. Take advantage of SMP.

Giga Networks — lightweight service/TCP



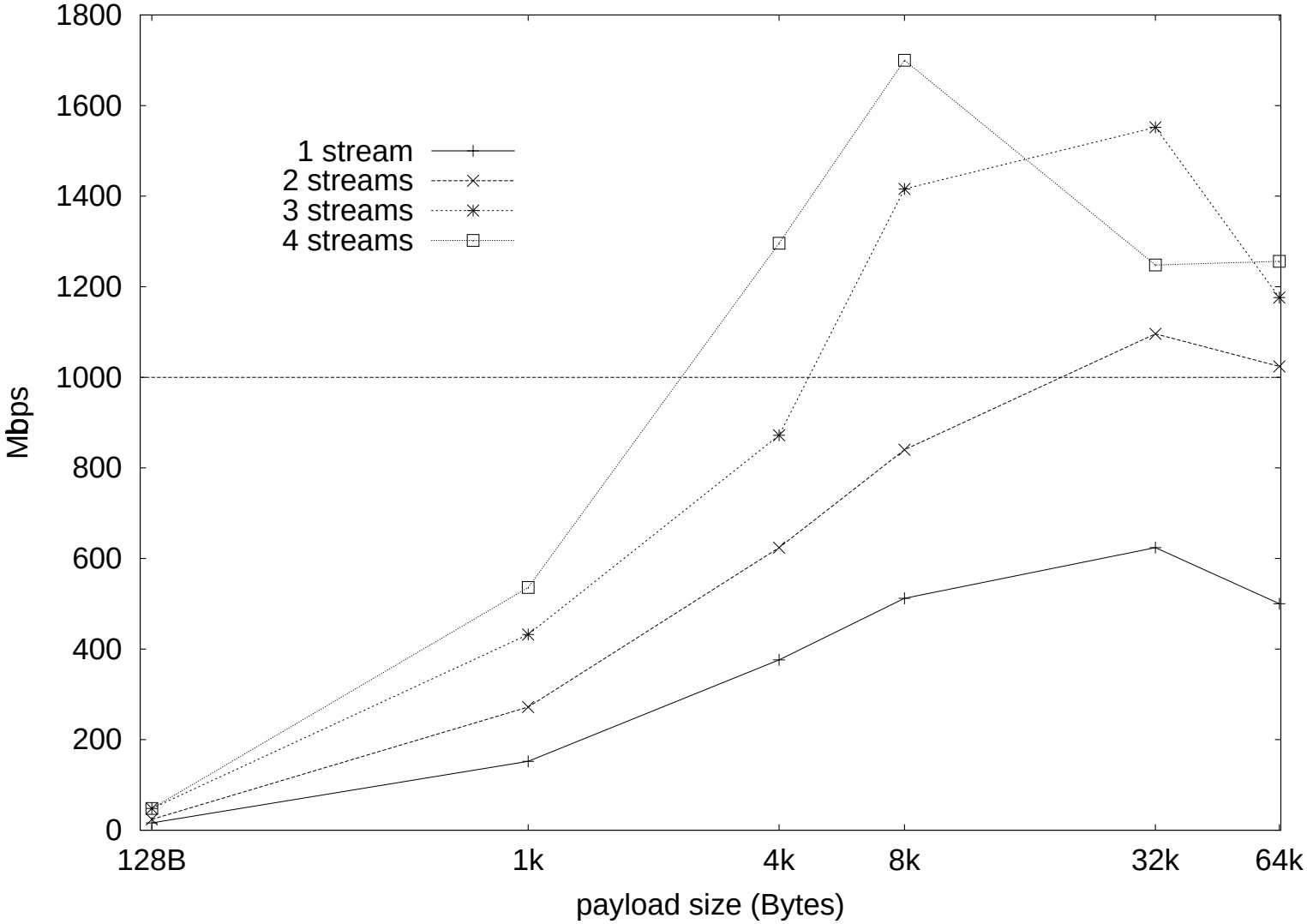
- Need big packets. Does not saturate the link.

Tamanoir cluster



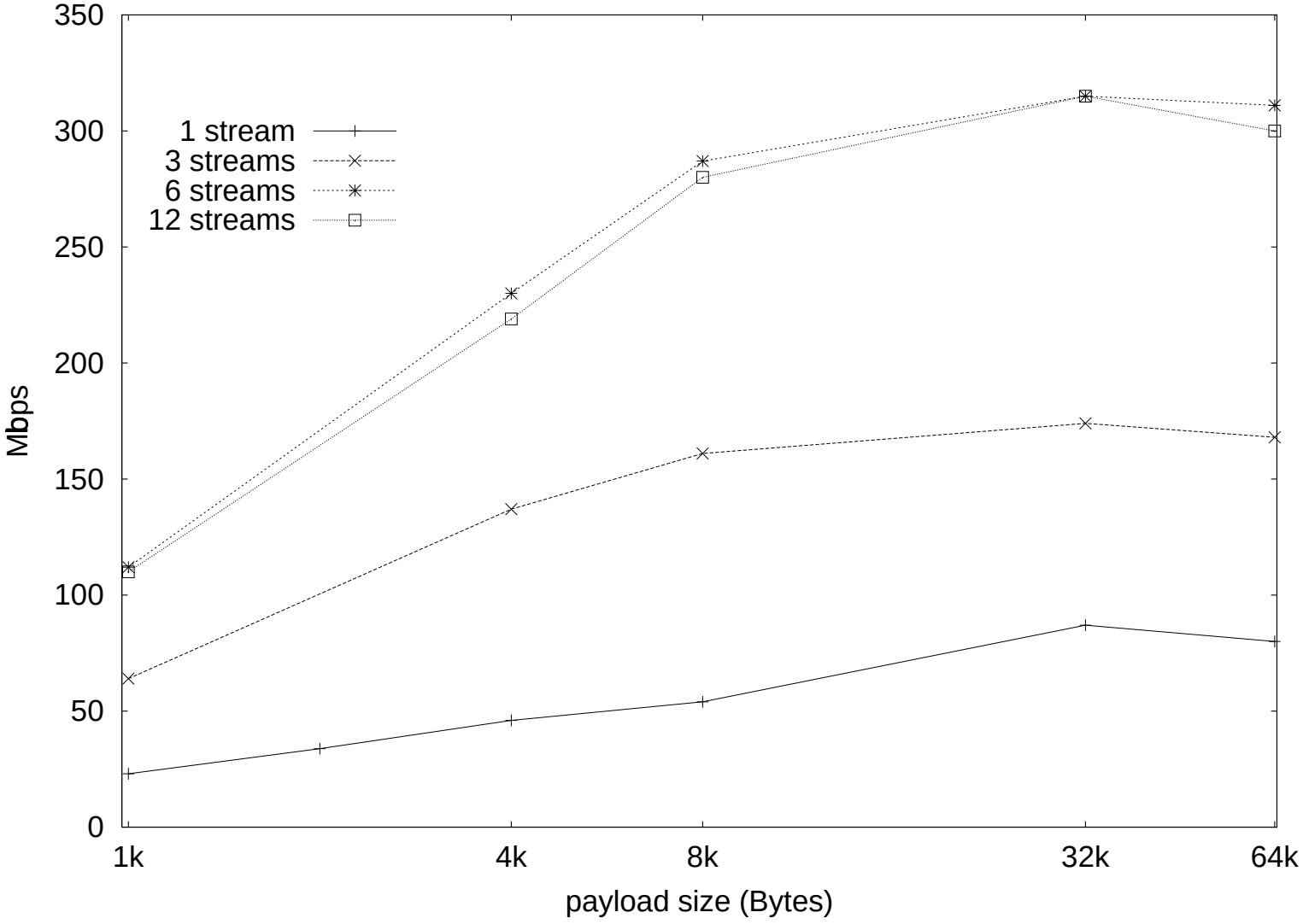
- Every nodes are connected through a Myrinet switch
- Each packet cross the switch three times

Tamanoir cluster — lightweight service/3BE



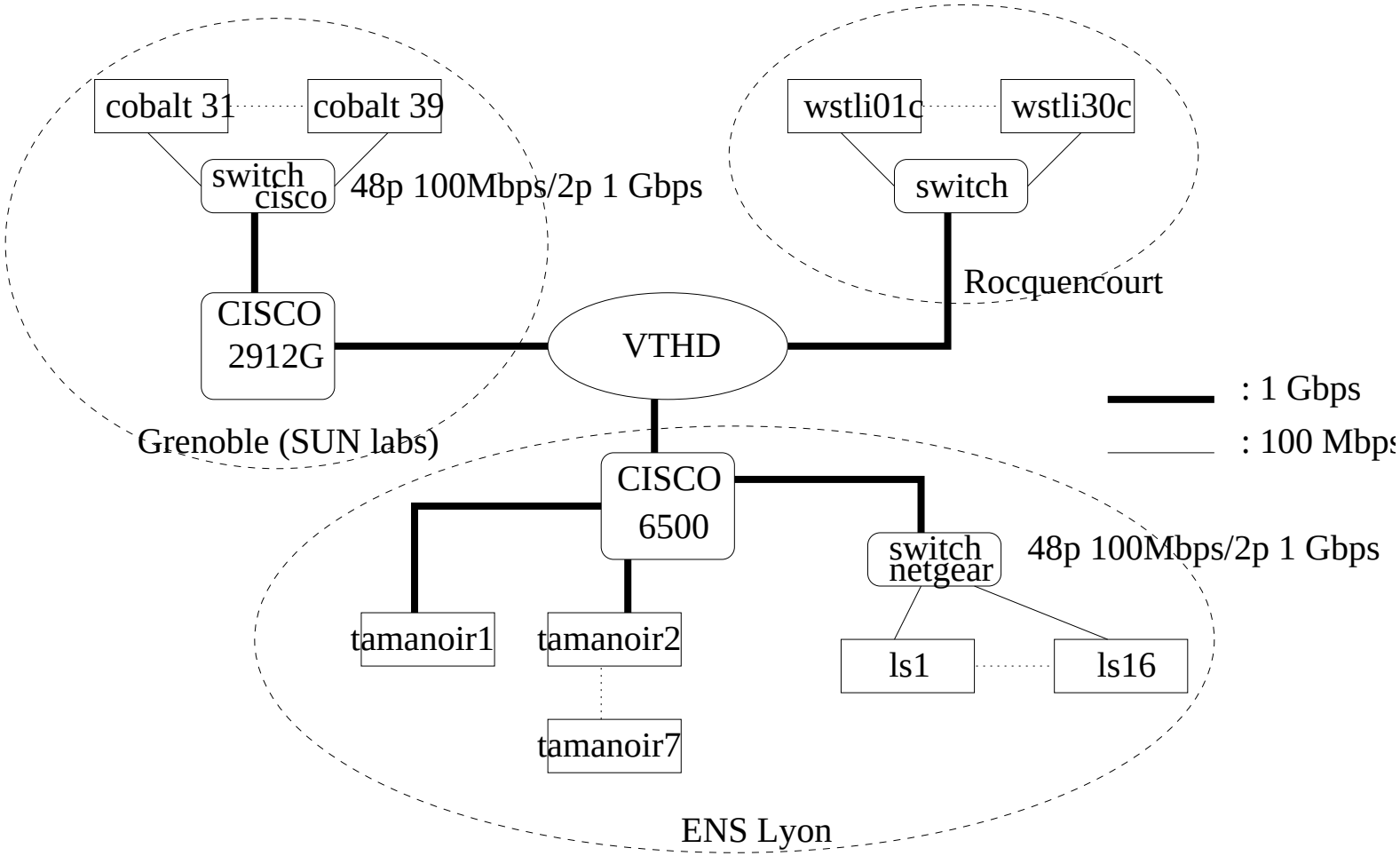
- Very high throughput.

Tamanoir cluster — heavy service/3BE



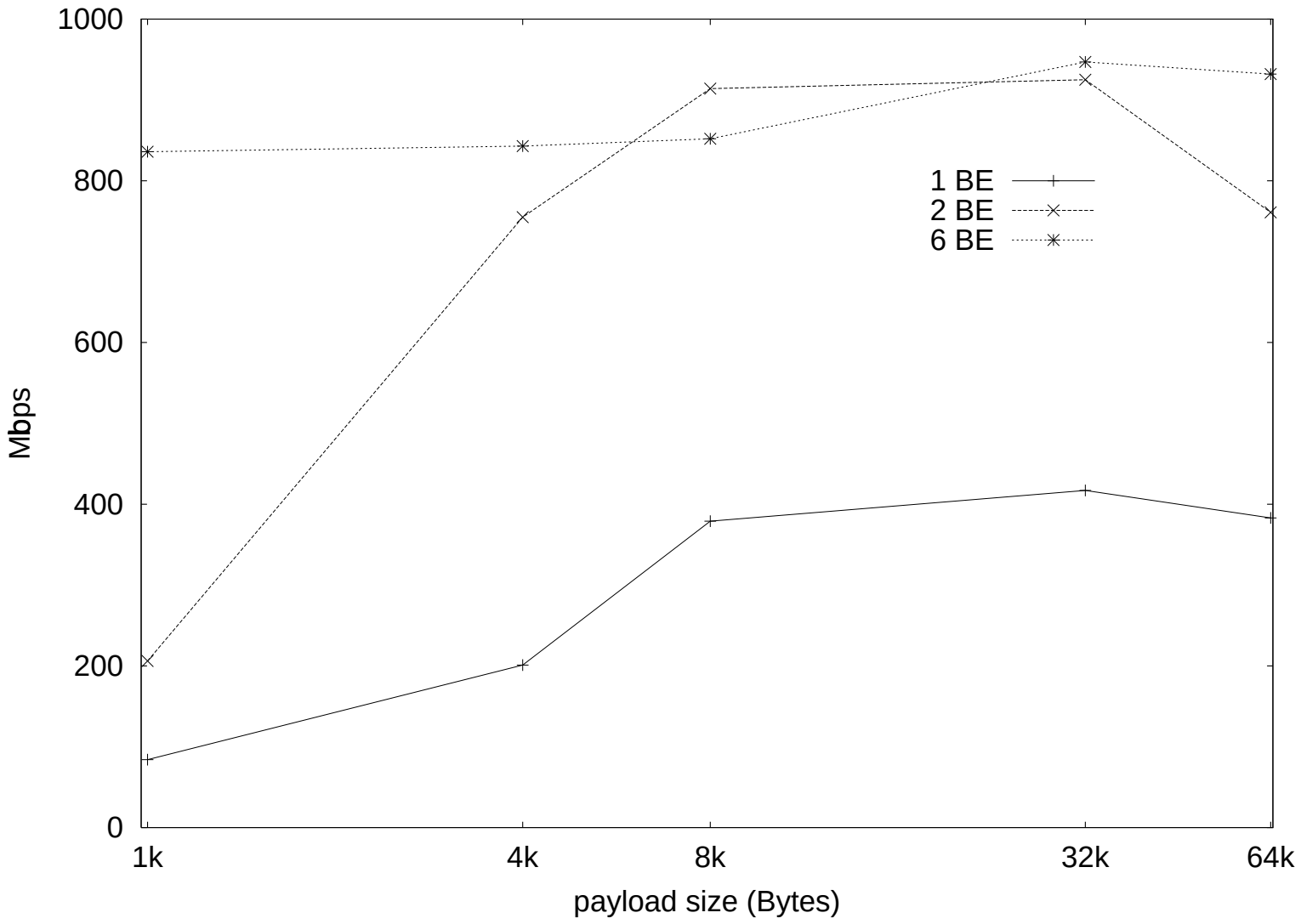
- 6 streams on 3BE reach the processing limits.

RNRT VTHD++



- VTHD : Vraiment Très Haut Débit, French experimental Gigabit network.

VTHD++ — 12 streams (lightweight service)



- Need of more backends for small packets.

Overview

- Introduction to Active Networks
- Propositions for an High Performance Active Network
- The Tamanoir project
- Experimentations
- **Conclusions and future works**

Conclusions

- An High Performance active architecture
  - multilayered adapted to service classification
- Tamanoir, an efficient implementation
  - dynamic service in kernel space
  - High Performance value added function in user space
  - replicated EE cluster based active node
- Experimentations/Validations
  - Software Gigabit active node on a standard Linux box whithout dedicated hardware

Application and usage of Tamanoir

- Internal project
  - QoSINUS : active QoS [Chanussot, Primet/Vicat-Blanc, 03].
  - DyRam : reliable multicast protocol [Maïmour, Pham, 03]
- External project
  - Laboratoire LIRIS/INSA Lyon : proxy cache in active routers [Sid Ali Guebli et al.03]
  - Laboratoire LAAS Toulouse : FPTP (Fully Programmable Transport Protocol) [Ernesto Exposito et al.03]
  - Labo LoCI, University of Tennessee, USA : IBP (Internet Backplane Protocol) [Micah Beck et al.99]

Future Works

- Support faster throughput : need of hardware assisted active equipment
- Large scale validation
- Mixing High Performance and security issues for industrial requirements

Questions ?