`tracerate`: a non-intrusive method for measuring the hop-by-hop capacity of a path

Bandwidth Estimation Workshop

Mathieu Goutelle and Pascale Primet
INRIA RESO team, LIP laboratory (ENS Lyon, France)

9-10 december 2003

Overview

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

- Introduction ;
- Bandwidth measurement in IP networks ;
- The Packet Pair method ;
- Our proposition: `tracerate` ;
  - Method principles ;
  - Data Analysis.
- Results and validations ;
- Conclusion.

Introduction

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

- **Simplicity** of IP networks: no control channel, few informations provided by equipments ;

Introduction

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

- **Simplicity** of IP networks: no control channel, few informations provided by equipments ;
- Need of an external mean to evaluate the end-to-end performances:
  - delay, loss rate: classical and easy (`ping`, `traceroute`);
  - **Capacity**: Maximal available rate between two nodes ;
  - **Available rate**: Accessible rate between two machines given an utilization on the followed path ;

Introduction

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

- **Simplicity** of IP networks: no control channel, few informations provided by equipments ;
- Need of an external mean to evaluate the end-to-end performances:
  - delay, loss rate: classical and easy (`ping`, `traceroute`);
  - **Capacity**: Maximal available rate between two nodes ;
  - **Available rate**: Accessible rate between two machines given an utilization on the followed path ;
- Use of the delay between two machines $\rightarrow$ not enough to evaluate the duration of a data transfer.

Bandwidth measurement in IP networks

- A “rate” knowledge gives a more realistic view ;
  - to estimate a transfer duration estimation ;
  - to schedule transfer in grid computing ;
  - to choose a data source or a data mirror ;
  - ...

Bandwidth measurement in IP networks

- A “rate” knowledge gives a more realistic view ;
  - to estimate a transfer duration estimation ;
  - to schedule transfer in grid computing ;
  - to choose a data source or a data mirror ;
  - ...
- **Available rate**: intrusive measurements (`iperf`, `MRTG`, `NWS`) or non-intrusive (`pathload` [JD02]);

Bandwidth measurement in IP networks

- A “rate” knowledge gives a more realistic view ;
  - to estimate a transfer duration estimation ;
  - to schedule transfer in grid computing ;
  - to choose a data source or a data mirror ;
  - ...
- **Available rate**: intrusive measurements (`iperf`, `MRTG`, `NWS`) or non-intrusive (`pathload` [JD02]);
- **Total capacity**:

	Method	Type of measure	Measure	Protocol	Receiver
`pathchar` [Jac97]	*Variable Packet Size*	hop-by-hop	slow	UDP, ICMP	no
`tailgater` [LB00]	*Packet Tailgating*	end-to-end	fast	TCP, ICMP	no
`pathrate` [DRM01]	*Packet Pair*	end-to-end	slow	UDP	yes

Bandwidth measurement in IP networks

- A “rate” knowledge gives a more realistic view ;
  - to estimate a transfer duration estimation ;
  - to schedule transfer in grid computing ;
  - to choose a data source or a data mirror ;
  - ...
- **Available rate**: intrusive measurements (`iperf`, `MRTG`, `NWS`) or non-intrusive (`pathload` [JD02]);
- **Total capacity**:

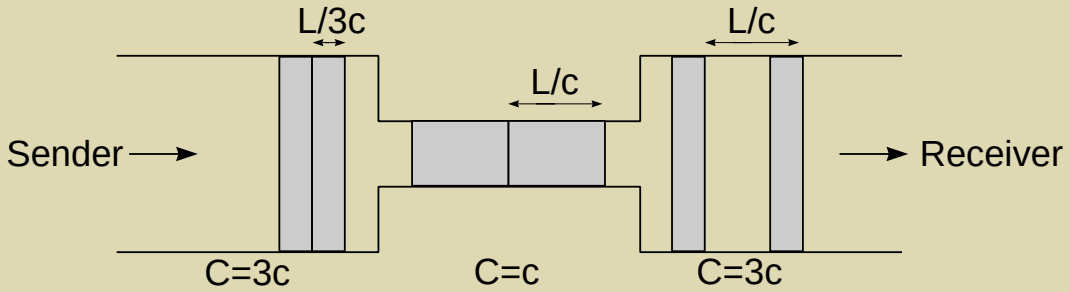
	Method	Type of measure	Measure	Protocol	Receiver
`pathchar` [Jac97]	*Variable Packet Size*	hop-by-hop	slow	UDP, ICMP	no
`tailgater` [LB00]	*Packet Tailgating*	end-to-end	fast	TCP, ICMP	no
`pathrate` [DRM01]	*Packet Pair*	end-to-end	slow	UDP	yes

- **Issues**: high-performance network, bottleneck localization, low intrusivity.

The Packet Pair method (1)

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

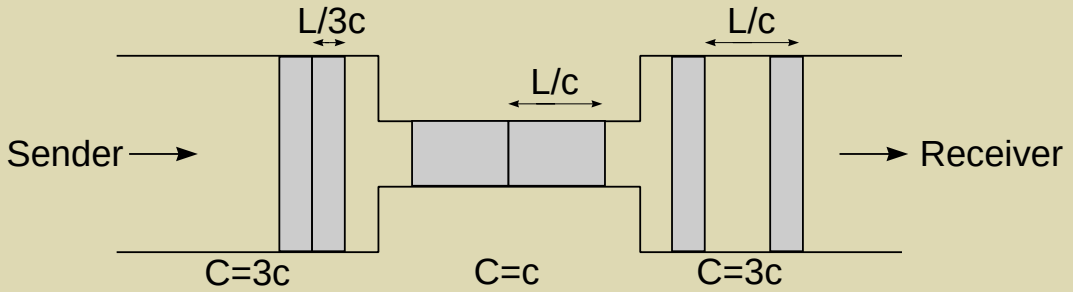
- Quite old principle (Van Jacobson, 1988 [Jac88]) ;
- A path is considered as a succession of delays (queue waiting time, transmission time, *etc.*) ;
- Capacity evaluation through the dispersion (inter-packet delay) measurement of two packets sent back-to-back ;
- This delay is the consequence of the smallest link on the path:



The Packet Pair method (1)

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

- Quite old principle (Van Jacobson, 1988 [Jac88]) ;
- A path is considered as a succession of delays (queue waiting time, transmission time, *etc.*) ;
- Capacity evaluation through the dispersion (inter-packet delay) measurement of two packets sent back-to-back ;
- This delay is the consequence of the smallest link on the path:

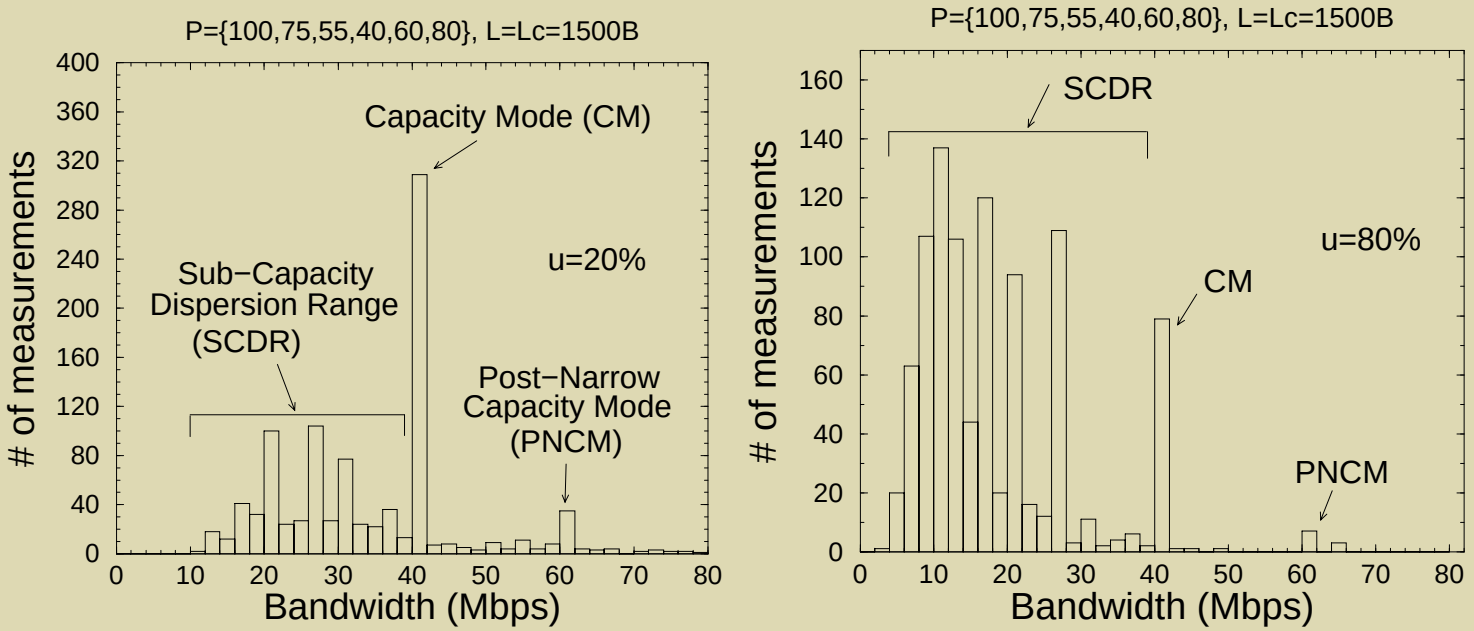


- **Hypothesis**: No concurrent traffic!
- Otherwise, concurrent traffic may cause the measure to under- or overestimate the real path capacity.

The Packet Pair method (2)

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

- Due to concurrent traffic, the measurement distribution is multimodal [DRM01]:



  - **under-estimation** (SCDR): A packet has spaced the two probe packets ;
  - **over-estimation** (PNCM): The first probe has waited for the second in an non-empty queue.

Our proposition: `tracerate`

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

- **Objectives**: to propose a method little intrusive to measure and to localize the bottleneck of a path. It must work in a high-performance environment and without cooperation of the destination ;

Our proposition: `tracerate`

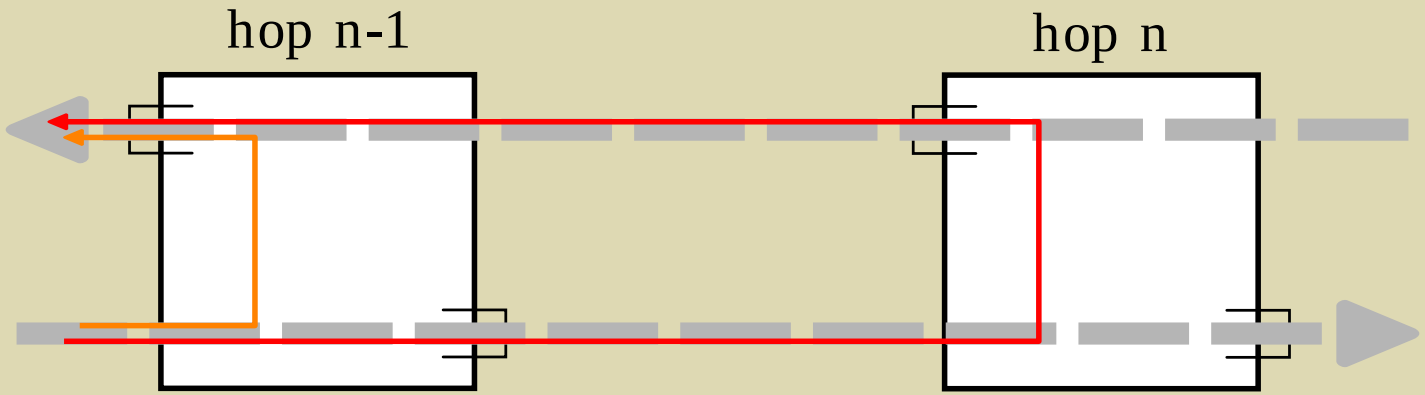
Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

- **Objectives**: to propose a method little intrusive to measure and to localize the bottleneck of a path. It must work in a high-performance environment and without cooperation of the destination ;
- **Proposition**:
  - We use a Packet Pair because it is more robust regarding the presence of invisible nodes [PDM03] ;
  - We measure the hop-by-hop capacity (and delay and loss) up to the path bottleneck ;
  - We eliminate “topology” parasitic modes with a better hop-by-hop knowledge of the topology (like `traceroute`) ;
  - We will be able to evaluate the hop-by-hop utilization rate up to the path bottleneck ;

Topology discovery

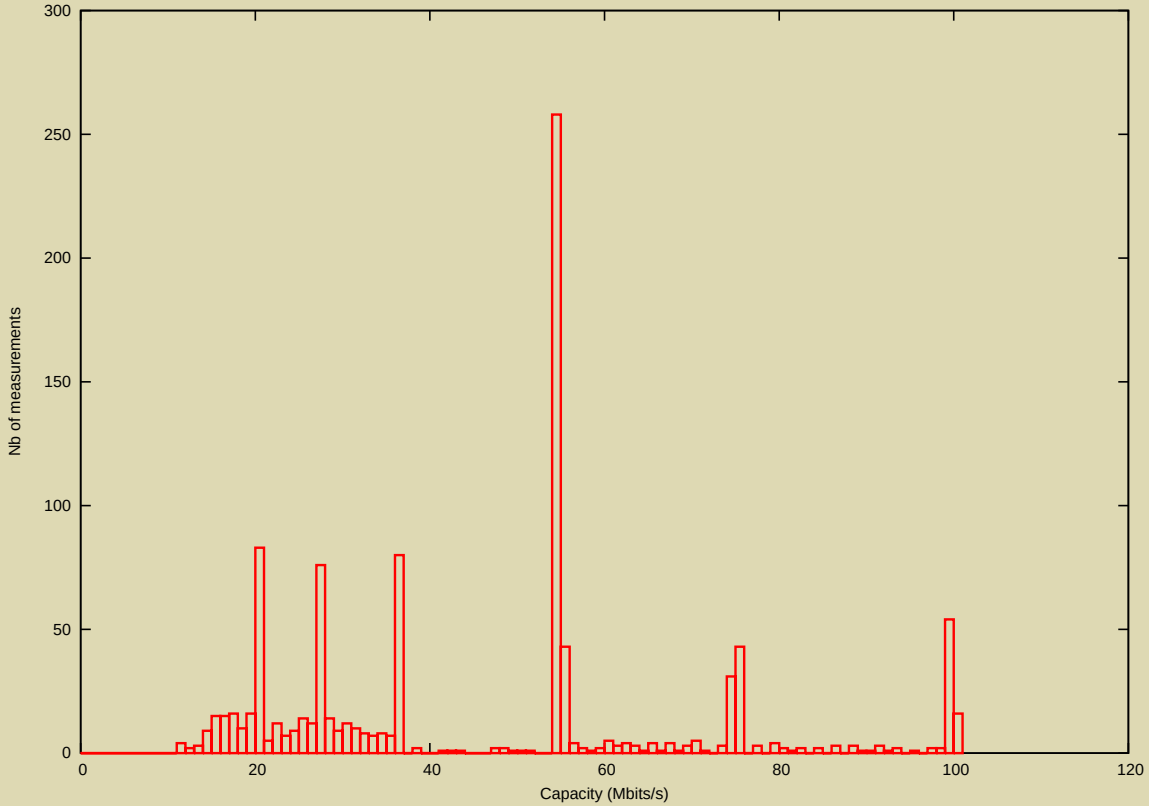
Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

- TTL (*Time To Live*): field in the IP header. It indicates the remaining number of equipments a packet can go through:
  - If an equipment receives a packet with a zero value TTL, it sends this packet back to the sender ;
  - Otherwise, it decreases this value and sends the packet to the next hop.
- With this mechanism, you can discover the topology with increasing TTL loops.



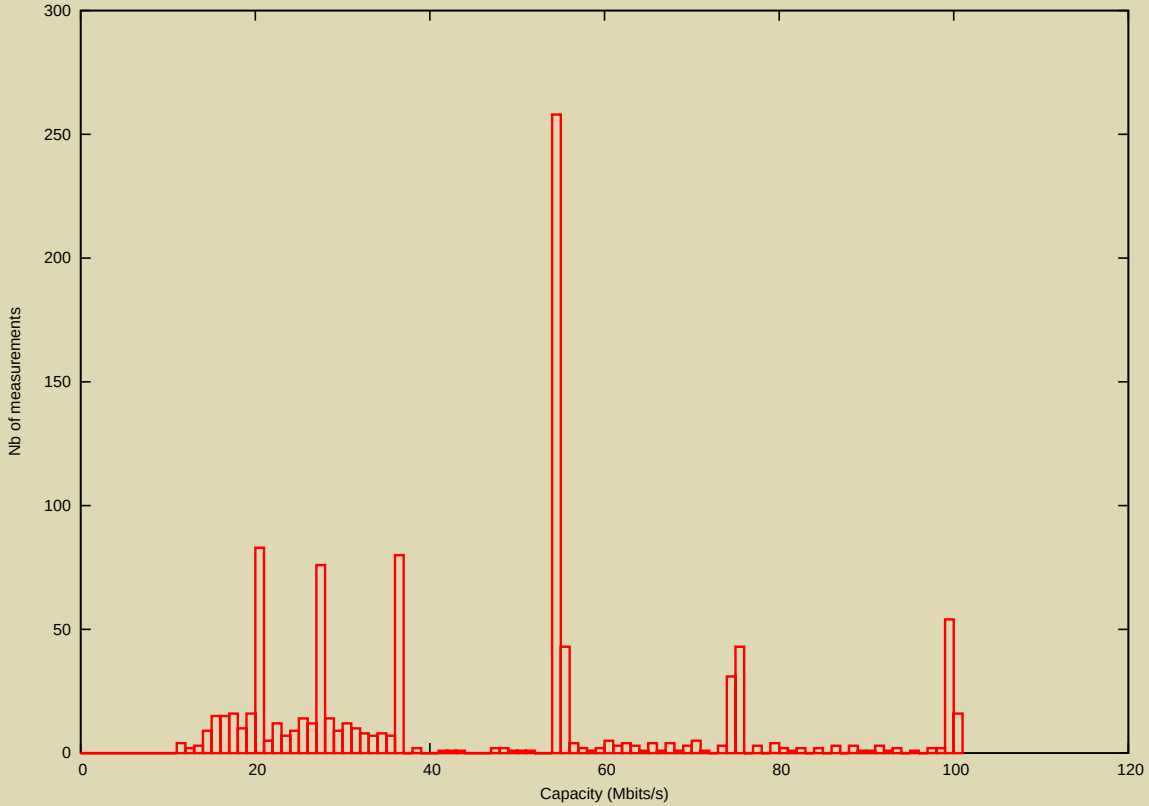
Method principles

- Measurements gathering ;
- Distribution analysis: extraction of the capacity mode.



Method principles

- Measurements gathering ;
- Distribution analysis: extraction of the capacity mode.

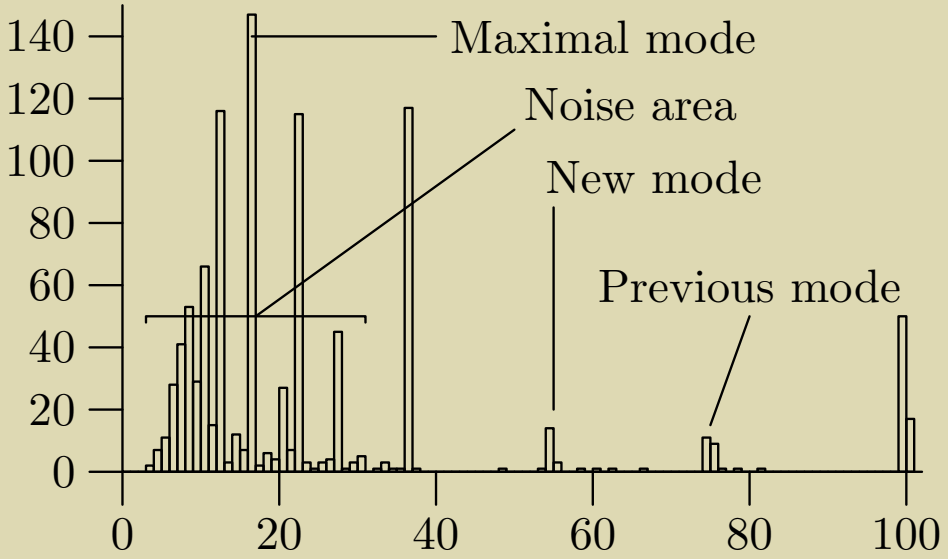


- At step $n+1$, we already have the capacity value for the loops up to n:
  - If there is no relatively acute mode below the previous capacity mode, the bottleneck (up to hop $n+1$) is in the previous loop ;
  - Otherwise, a mode below the previous capacity mode has been detected and the link between hop n and $n+1$ is the new bottleneck.

Data Analysis

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

- Mode detection (increase up to a maximum and then decrease) ;
- Determination of four characteristics of the distribution:



- Capacity mode extraction depending on the position and the population of the characteristics.

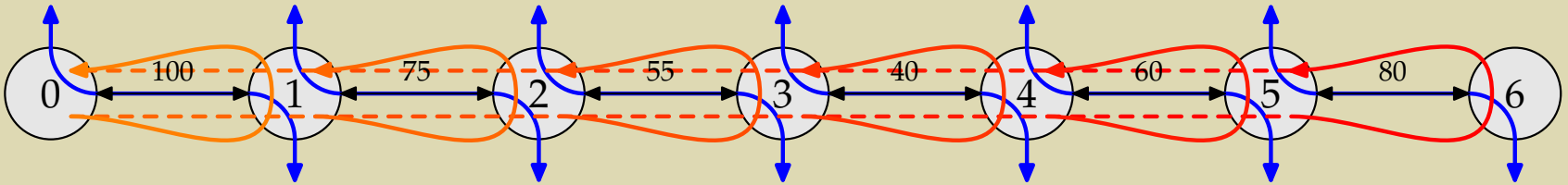
Capacity mode extraction

for all hop on the path **do**
  Compute the measurements distribution
  Determine the new, previous and maximal modes and the noise area
  if max_mode = prev_mode
    or (max_mode = new_mode and new_mode not in noise_area) **then**
    capacity_mode $\leftarrow$ max_mode
  else if $1.1 \times |\text{new\_mode}| \geq |\text{prev\_mode}|$
      and new_mode not in noise_area **then**
    capacity_mode $\leftarrow$ new_mode
  else if (max_mode not in noise_area or $|\text{max\_mode}| \geq 0.6 \times$ total_pop)
      and $|\text{max\_mode}| \geq 1.25 \times |\text{prev\_mode}|$ **then**
    capacity_mode $\leftarrow$ max_mode
  else
    capacity_mode $\leftarrow$ prev_mode
  end if
end for

Validations

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

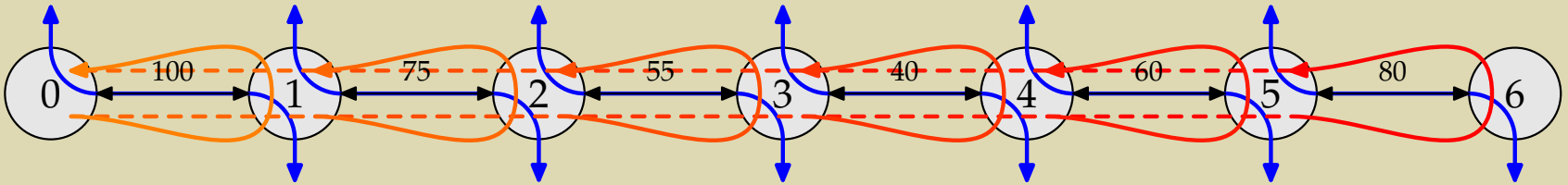
- Validations in simulations (NS-2) in a controlled environment (capacity, delay):
  - Behaviour consistent with the one expected ;
  - Accuracy validation of the analysis method ;
  - Robustness validation regarding the network conditions (path length, load).



Validations

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

- Validations in simulations (NS-2) in a controlled environment (capacity, delay):
  - Behaviour consistent with the one expected ;
  - Accuracy validation of the analysis method ;
  - Robustness validation regarding the network conditions (path length, load).



- Experimentation in a high-performance environment (DataTAG platform, `http://www.datatag.org`).

Accuracy study

- 100 simulations with a variable utilization rate from 0 up to 100% ;
- Measure of the relative error between the real capacity and the measured value for each hop:

Accuracy study

- 100 simulations with a variable utilization rate from 0 up to 100% ;
- Measure of the relative error between the real capacity and the measured value for each hop:

Relative error	hop 1	hop 2	hop 3	hop 4	hop 5	hop 6
$u \leq 0{,}5$	0,1%	0,1%	1,1%	2,5%	4,8%	6,9%
$u \leq 0{,}75$	0,1%	1,4%	4,6%	7,1%	5,9%	8,3%
$u \leq 1$	0,1%	12,4%	14,9%	15,3%	11,5%	13,7%

- Influence of the load and path length, but the quality degradation of the result remains low ;
- The method tries to be conservative: it can detect the bottleneck at one or two steps later.

Robustness study

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

- 100 simulations with a random load and link capacities ;

Robustness study

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

- 100 simulations with a random load and link capacities ;

	Correlation	Avg relat. err.
6 hops ($u < 0{,}5$)	0,82	0,14
6 hops ($u < 1$)	0,58	0,28
10 hops ($u < 0{,}5$)	0,88	0,16
10 hops ($u < 1$)	0,62	0,37

- The method is robust regarding the path length ;
- The network load may be a diffi culty if it becomes high.

Experimental validation

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

- Implementation in Linux based on `tcptraceroute` ;
- Experimentation on a real platform (DataTAG):
  - It works!
  - Measure up to 1Gbit/s ;
  - Some problems with ICMP in routers: limitation due because the ICMP path is different from the normal path ;
  - it needs some extra tests to validate `tracerate`...

Experimental validation

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

- Implementation in Linux based on `tcptraceroute` ;
- Experimentation on a real platform (DataTAG):
  - It works!
  - Measure up to 1Gbit/s ;
  - Some problems with ICMP in routers: limitation due because the ICMP path is different from the normal path ;
  - it needs some extra tests to validate `tracerate`...
- Non-intrusivity:

Tool	Short path (4 hops)	Long path (11 hops)
`pathchar`	11,562	31,782
`clink`	6,002	16,400
`pchar`	11,732	32,417
`nettimer`	982	6,663
`tracerate`	4,000	11,000

Utilization rate evaluation

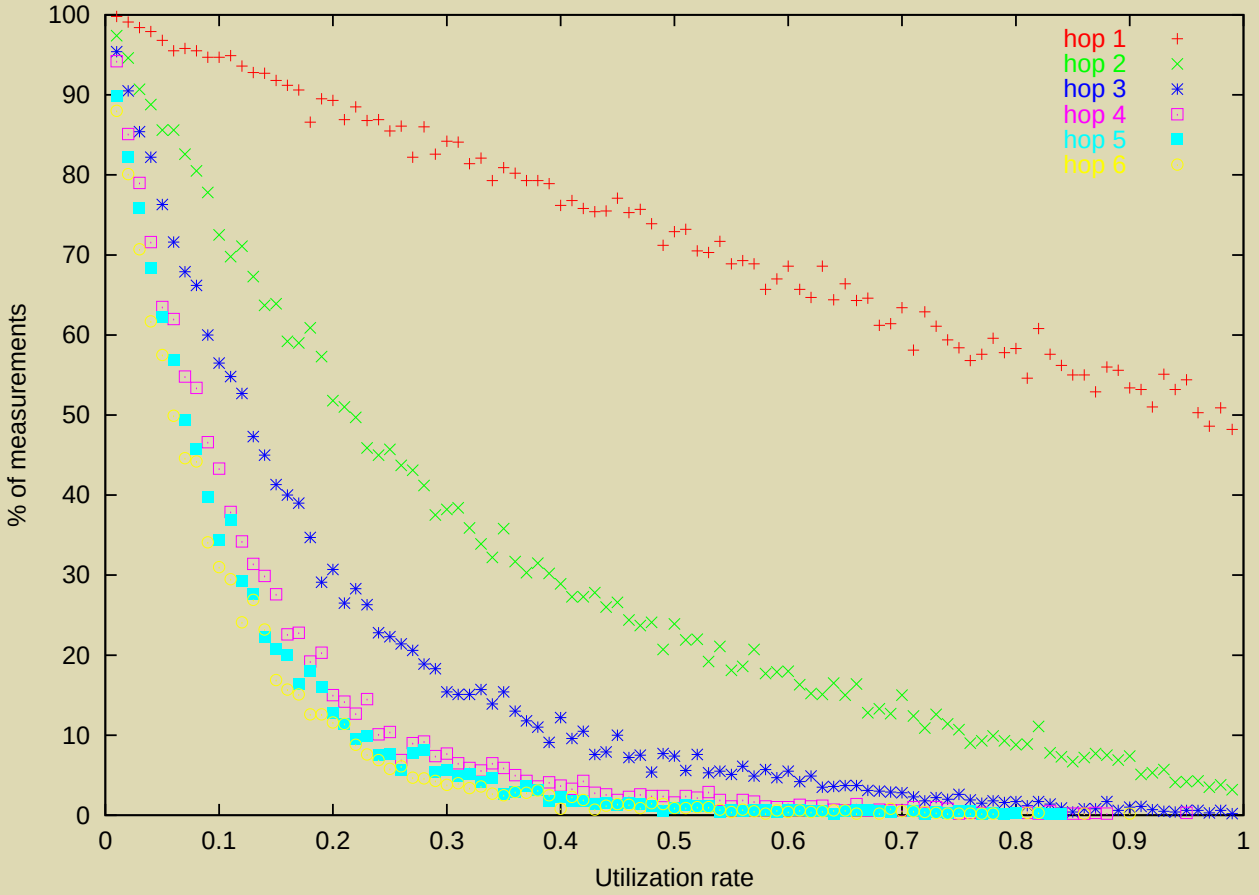
Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

- When you know both the capacity and the utilization rate of a path, you can deduce almost all its characteristics ;
- Given the previous measurements (distribution for each hop), we want to find a relation between the **population in the capacity mode** and the **utilization rate**.

Utilization rate evaluation

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

- When you know both the capacity and the utilization rate of a path, you can deduce almost all its characteristics ;
- Given the previous measurements (distribution for each hop), we want to find a relation between the **population in the capacity mode** and the **utilization rate**.



Conclusion

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

- **Non-intrusive** method to evaluate the capacity: determination of the bottleneck and its **localization** on the path ;
- Validations in simulations ;
- Linux implementations, working in a **high-performance** environment → `tracerate` ;
- **Promising future work:** utilization rate evaluation and finalization of the implementation ;

Conclusion

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

- **Non-intrusive** method to evaluate the capacity: determination of the bottleneck and its **localization** on the path ;
- Validations in simulations ;
- Linux implementations, working in a **high-performance** environment → `tracerate` ;
- **Promising future work:** utilization rate evaluation and finalization of the implementation ;

More details in this research report:
`http://www.inria.fr/rrrt/rr-4959.html`

References

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

[CM01] James Curtis and Tony McGregor. Review of bandwidth estimation techniques. In *New Zealand Computer Science Research Students' Conference*, University of Canterbury, New Zealand, April 2001.

[DRM01] Constantinos Dovrolis, Parameswaran Ramanathan, and David Moore. What do packet dispersion techniques measure? In *Proceedings of INFOCOM'01*, pages 905–914, 2001.

[Jac88] Van Jacobson. Congestion Avoidance and Control. In *ACM SIGCOMM '88*, volume 18, pages 314–329, Stanford, CA, August 1988.

[Jac97] Van Jacobson. pathchar - a tool to infer characteristics of internet paths. MSRI talk, April 1997.

[JD02] Manish Jain and Constantinos Dovrolis. Pathload: A measurement tool for end-to-end available bandwidth. In *Passive and Active Measurements (PAM) Workshop*, March 2002.

[LB00] Kevin Lai and Mary Baker. Measuring link bandwidths using a deterministic model of packet delay. In *SIGCOMM*, pages 283–294, August 2000.

[PDM03] Ravi S. Prasad, Constantinos Dovrolis, and Bruce A. Mah. The effect of layer-2 store-and-forward devices. In *Proceedings of INFOCOM '03*, San Fransisco, CA, April 2003.

Overview
Introduction
BW mes. in IP networks
The Packet Pair method
Our proposition
Topology discovery
Method principles
Data Analysis
Capacity extraction
Validations
Accuracy study
Robustness study
Experimental validation
Utilization rate evaluation
Conclusion
References

Questions?