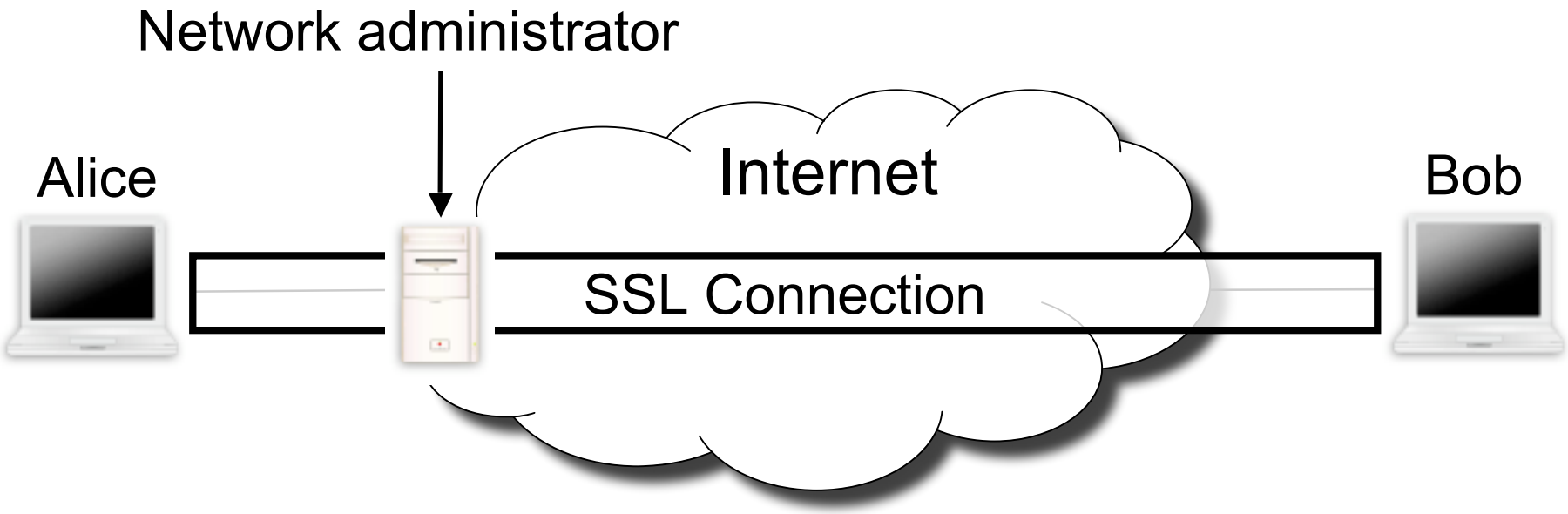
Early Recognition of Encrypted Applications

Laurent Bernaille

with Renata Teixeira

Laboratoire LIP6 – CNRS
Université Pierre et Marie Curie – Paris 6

Can we find the application inside an SSL connection?



Network administrator: profiling, QoS, policies

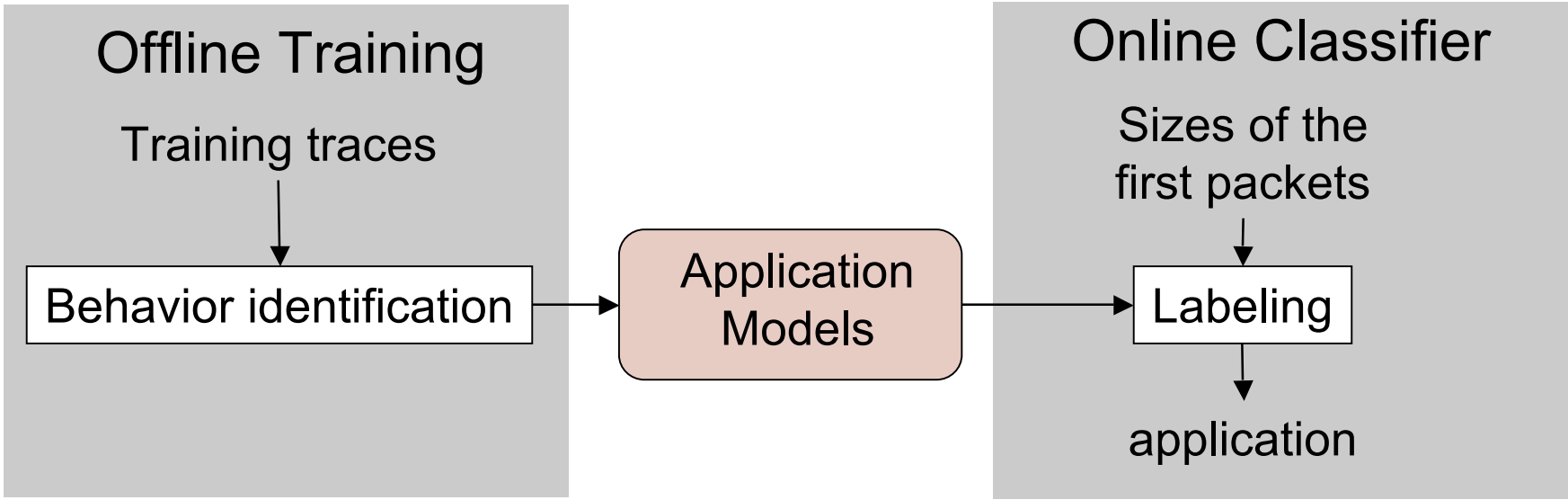
Alice and Bob: privacy issue

Possible identification methods for unencrypted traffic

- Port-based classification
  - Map standard IANA ports to applications (Ex: 443/HTTPS)
  - Unfortunately, this method is inaccurate
- Content-based approaches
  - Search for signatures that identify the application
  - Unfortunately, not possible with encrypted traffic
- Behavior-based methods
  - Model applications with connection statistics
  - Promising for encrypted traffic (not using port or content)

Early Application Identification

- Identify applications using the **sizes of the first application packets** in a TCP connection

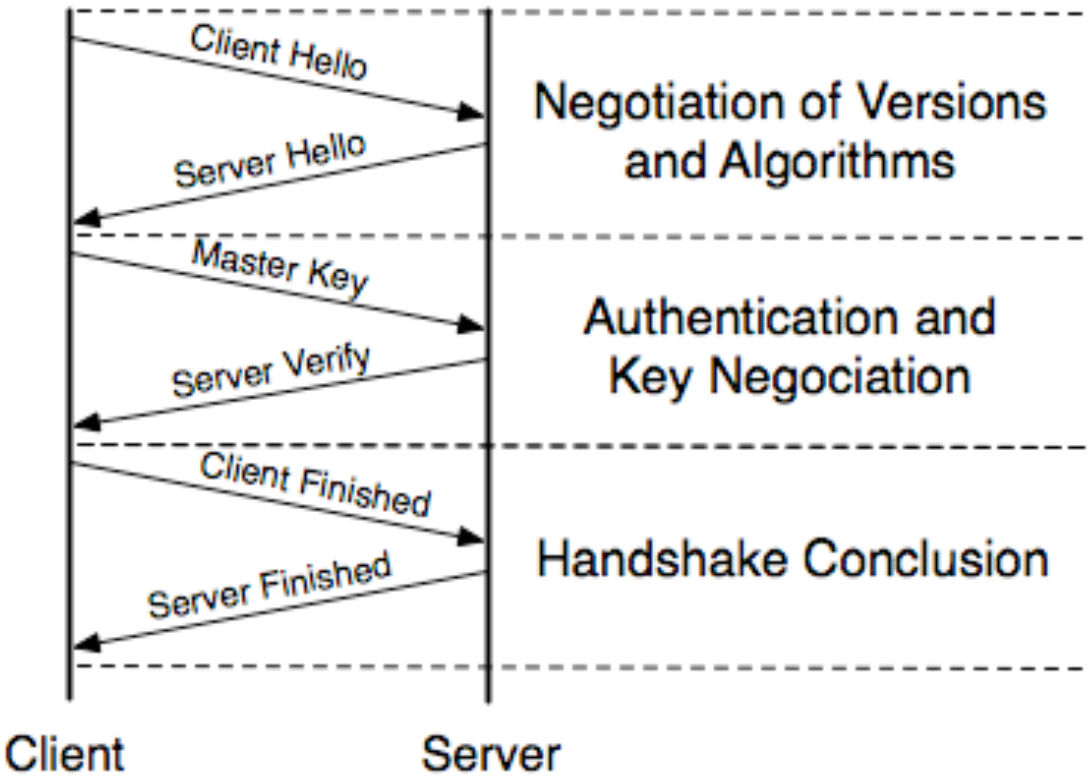


Can this work with encrypted connections?

Can we find the sizes of the first application packets in SSL?

- SSL mechanisms
  - SSL connections begin with a handshake
  - After handshake
    - SSL payload = encrypted application packet
- Challenges for Early Application Identification
  - Can we identify application packets?
  - Can we infer the unencrypted sizes of these packets?

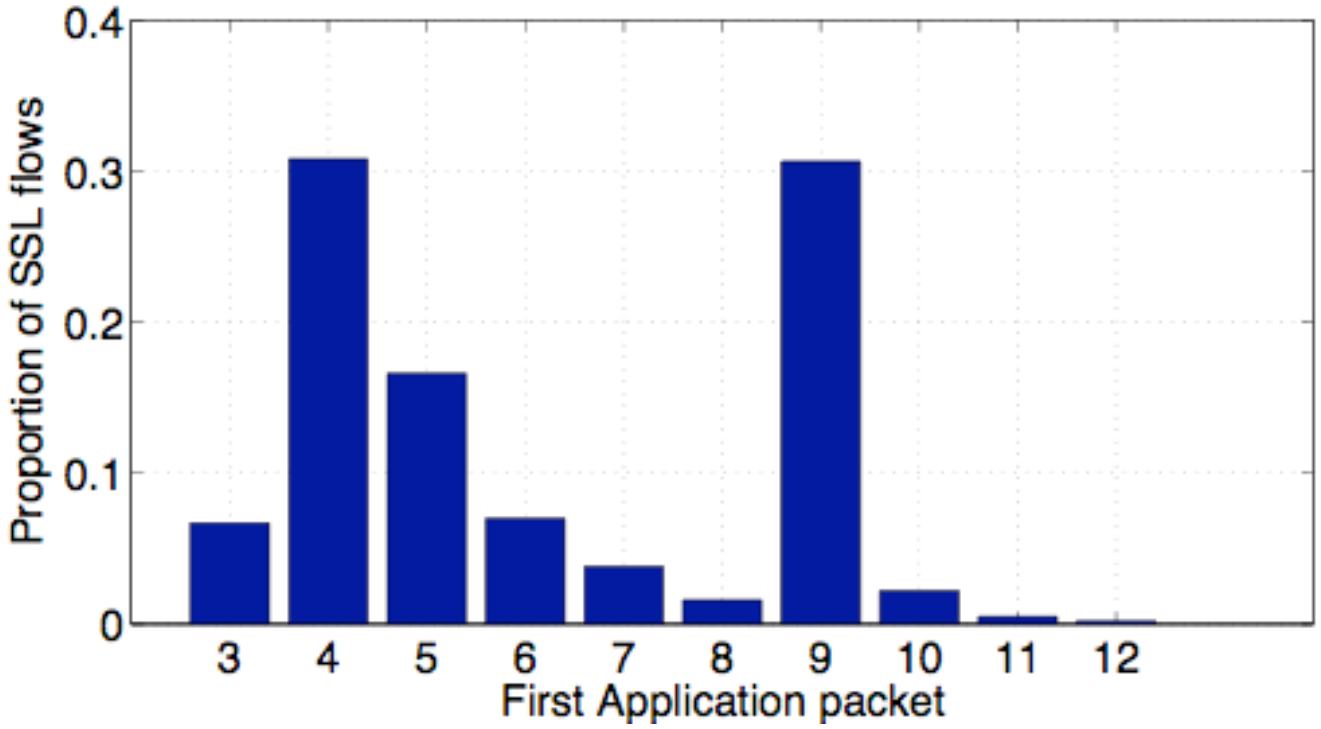
SSLv2 handshake



- SSLv2 negotiation
  - 4 or 6 packets
  - Identification through inspection of SSL headers

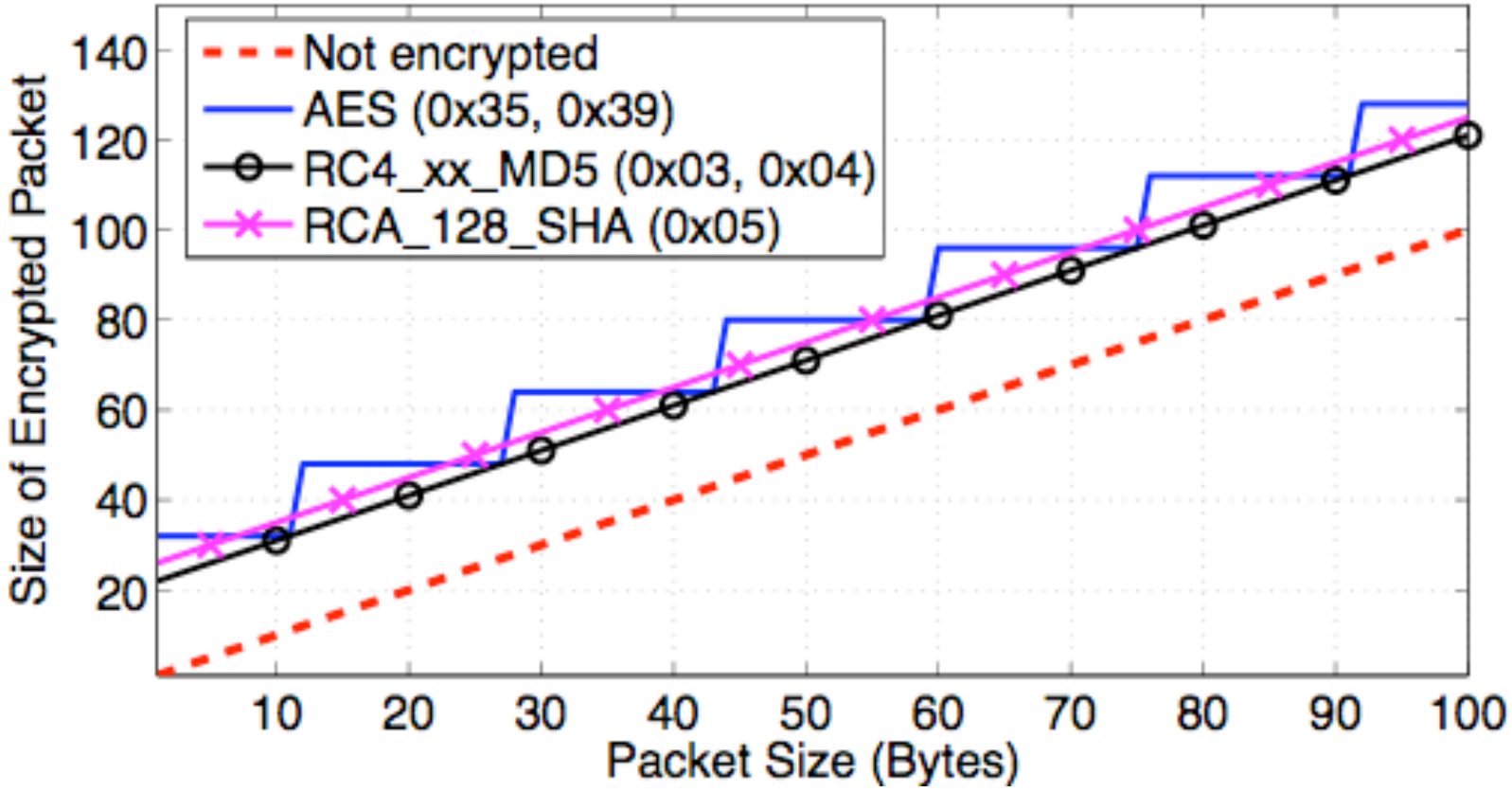
SSLv3 handshake

First application packet in P6 trace



- SSLv3 Negotiation
  - Variable number of packets (implementation)
  - Identification through inspection of SSL headers

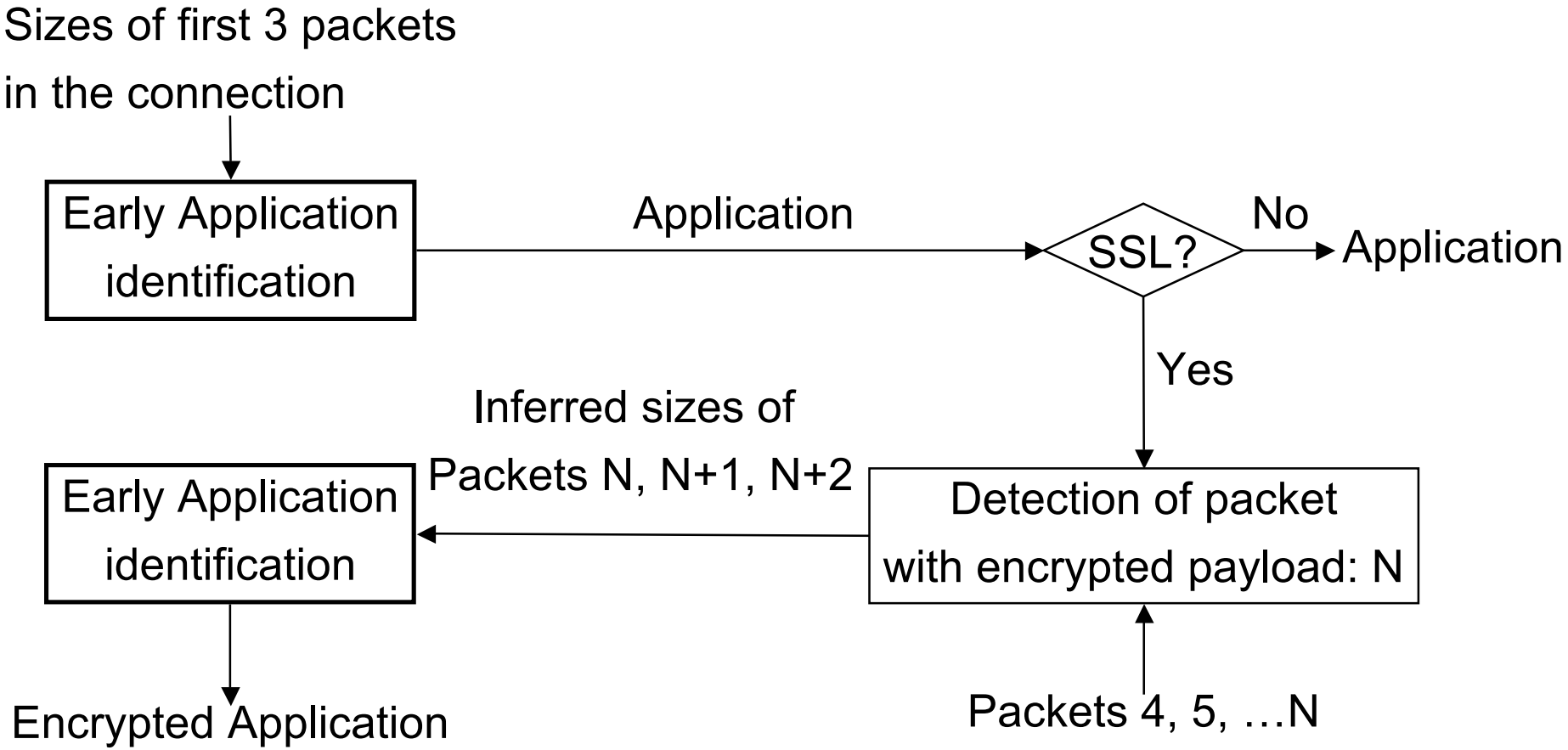
Influence of ciphers on packet size



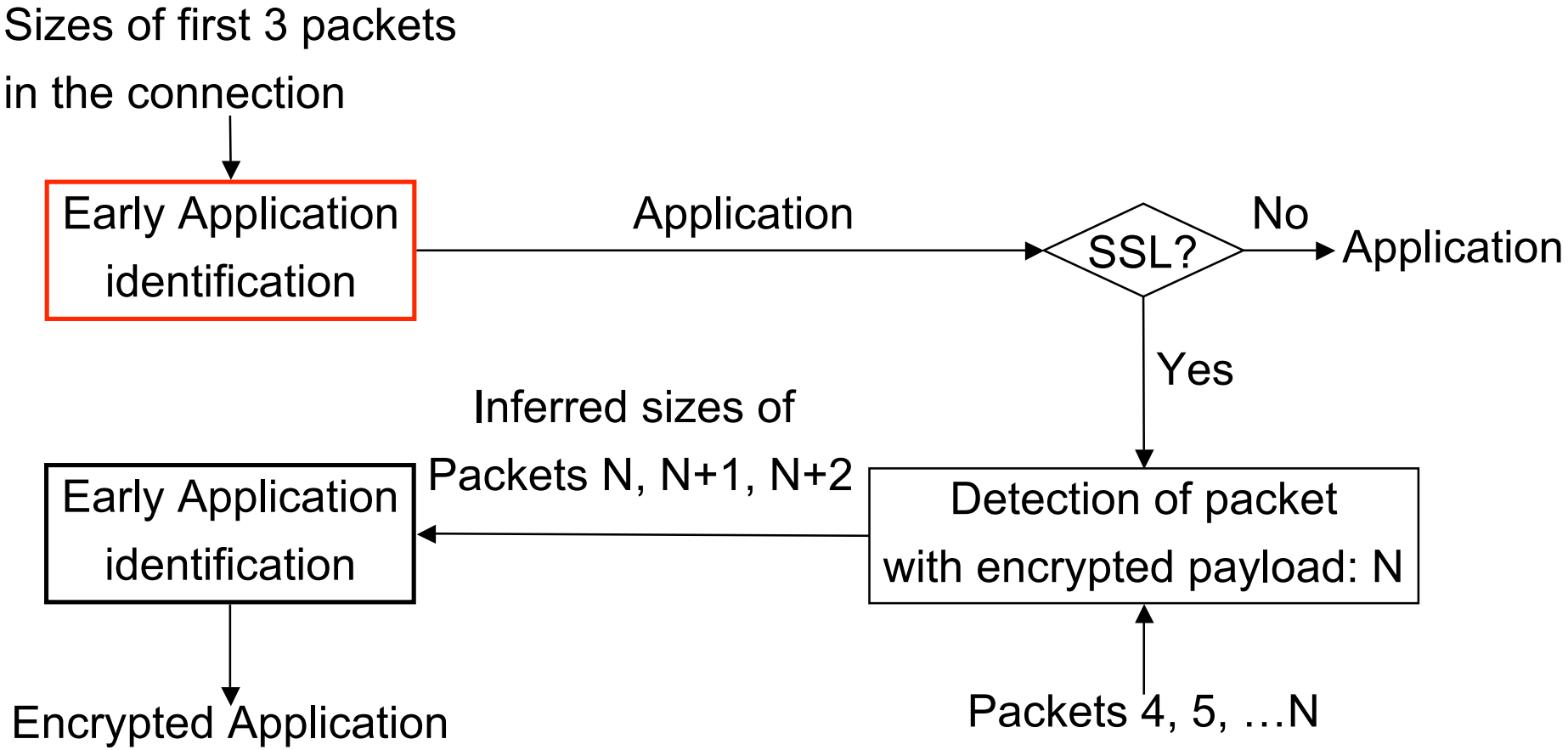
Applicability of Early Application Identification to SSL traffic

- We can identify the first application packet
  - Trough the analysis of SSL headers
  - Need to start inspection at the third packet
- We can infer unencrypted sizes
- Proposed method
  1. Identify SSL using the sizes of first 3 packets
  2. For SSL traffic, find the packet with application data
  3. Identify the application in SSL using the inferred sizes of the first application packets

Classification mechanism



Detecting SSL connections



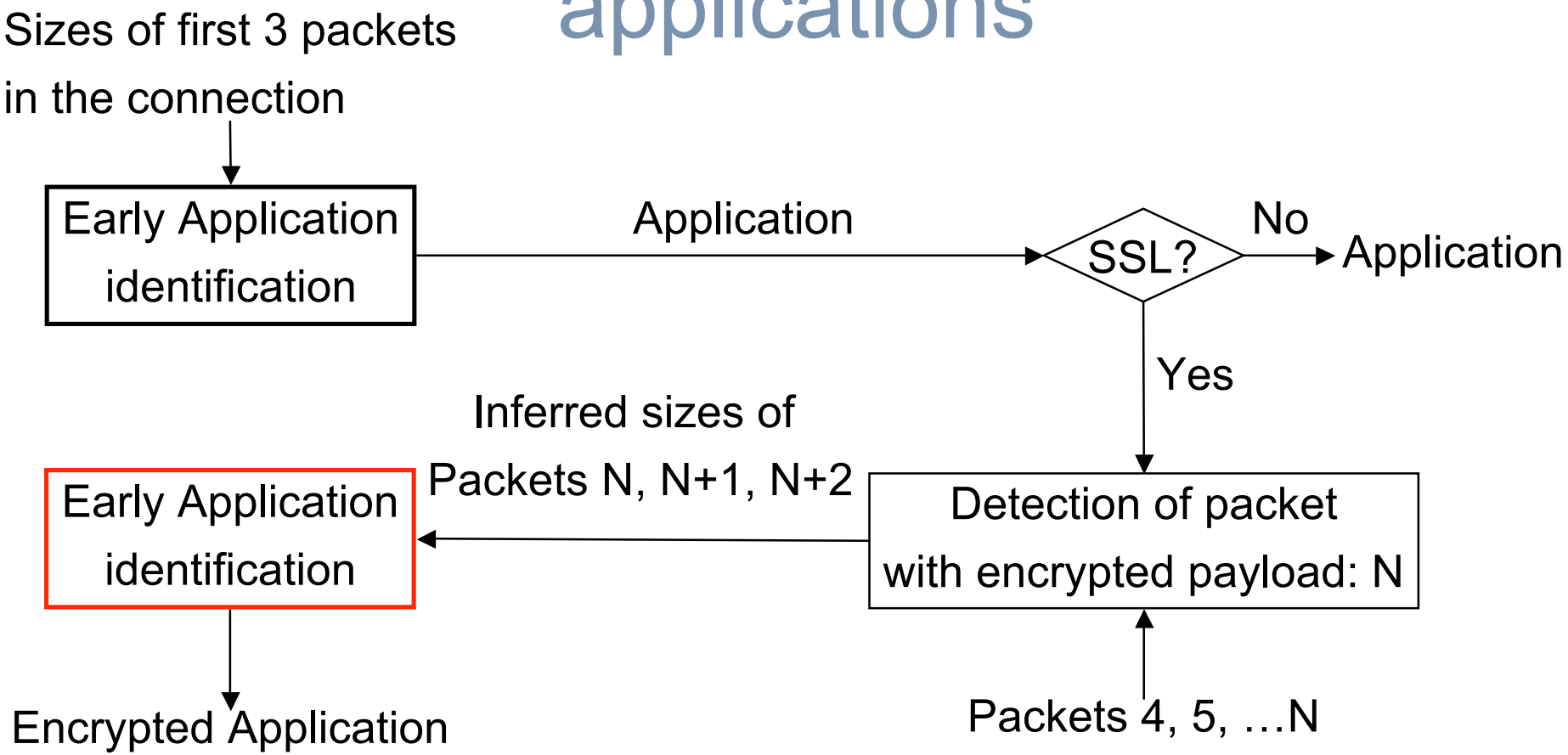
Detecting SSL connections: Evaluation methodology

- Data sets: packet traces
  - UMass campus, Paris 6 network
- Ground truth
  - Non-SSL traffic: content-based classification
  - SSL traffic: identification based on analysis of SSL headers
- Parameters for Early Application Identification
  - Using the payload size of the first 3 packets
  - Training set: 5500 connections from 11 applications

Accuracy of SSL detection

- Test set
  - 50k connections from Paris 6 network
  - more than 2000 connections for each application
- Results
  - SSL traffic: > 85% labeled SSL
  - Other applications: accuracy >95%

Identification of encrypted applications



Method to find ground truth for encrypted traffic

- From packet traces collected at Paris 6
  - Filtered traffic to well-known HTTPS and POP3S servers (IP addresses and ports)
- Manual encryption of traffic
  - Replay connections over an SSL tunnel
  - Applications: bittorent, edonkey, FTP

Accuracy of identification of applications in SSL connections

- Paris 6 traces

	Accuracy
HTTP	99.9%
POP3	98.5%

- Manually Encrypted Traffic

	Accuracy
FTP	92.5%
Bittorent	86.5%
Edonkey	96.5%

Conclusion and Perspectives

- We can identify the application encrypted with SSL
  - Using only the sizes of the first packets
  - With a high accuracy
- Future work: IPsec and SSH
  - Challenge: Finding the start of TCP connections
- Implementation
  - Available at http://rp.lip6.fr/~bernaill/earlyclassif.html

Description of SSL Traffic

Trace	Connections	SSL	SSLv2	SSLv3.0	TLS
P6 2004	500k	4.6%	0.6%	81%	18.4%
P6 2006	1000k	8.6%	0.2%	53.2%	46.6%
UMass	1700k	1.2%	0.0%	48%	52%

Trace	SSL port not SSL	SSL on non-SSL port
P6 2004	1.9%	1.1%
P6 2006	1.1%	4.2%
UMass	5.0%	1.5%

Ciphers

Cipher	Proportion (2004)	Proportion (2006)
RC4_xx_MD5	81.7%	68.1%
AES	6.9%	24.0%
RC4_128_SHA	9.7%	7.0%
Other	<2%	<1%