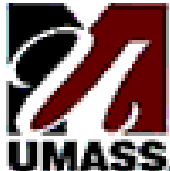


Exploiting the IPID field to infer network path and end- system characteristics

W. Chen, Y. Huang, B. Ribeiro, K. Suh,
H. Zhang, E. de Souza e Silva (UFRJ,
Brazil), J. Kurose, D. Towsley
presented by Kyoungwon Suh



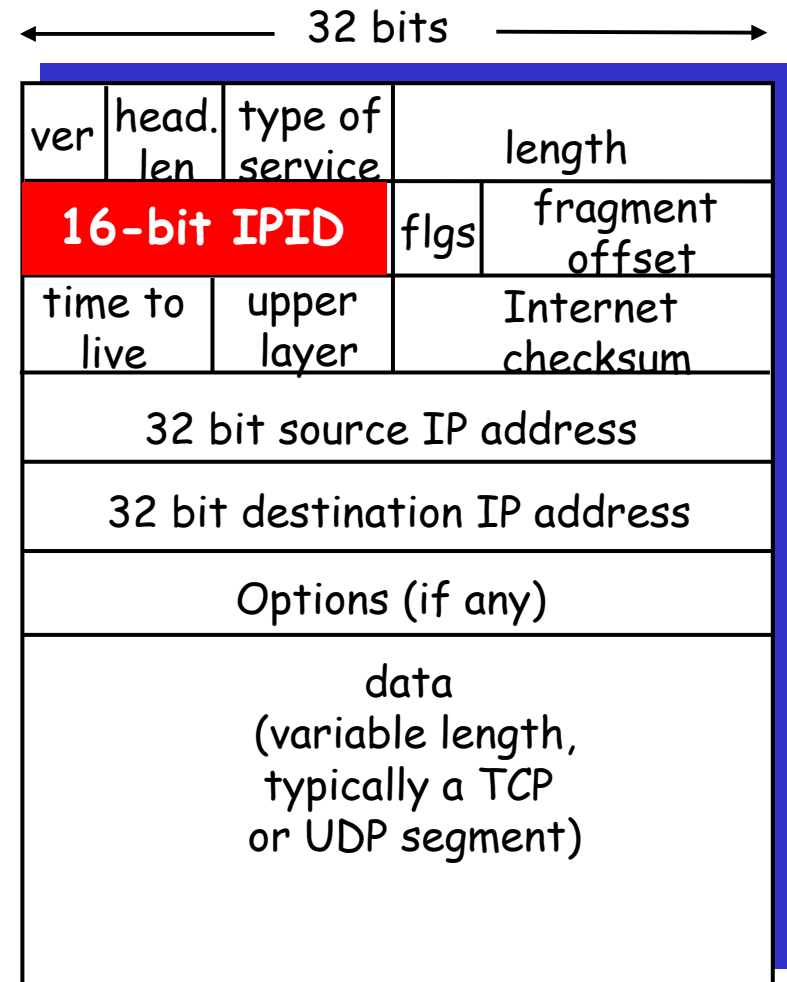
University of Massachusetts
Department of Computer Science

Outline

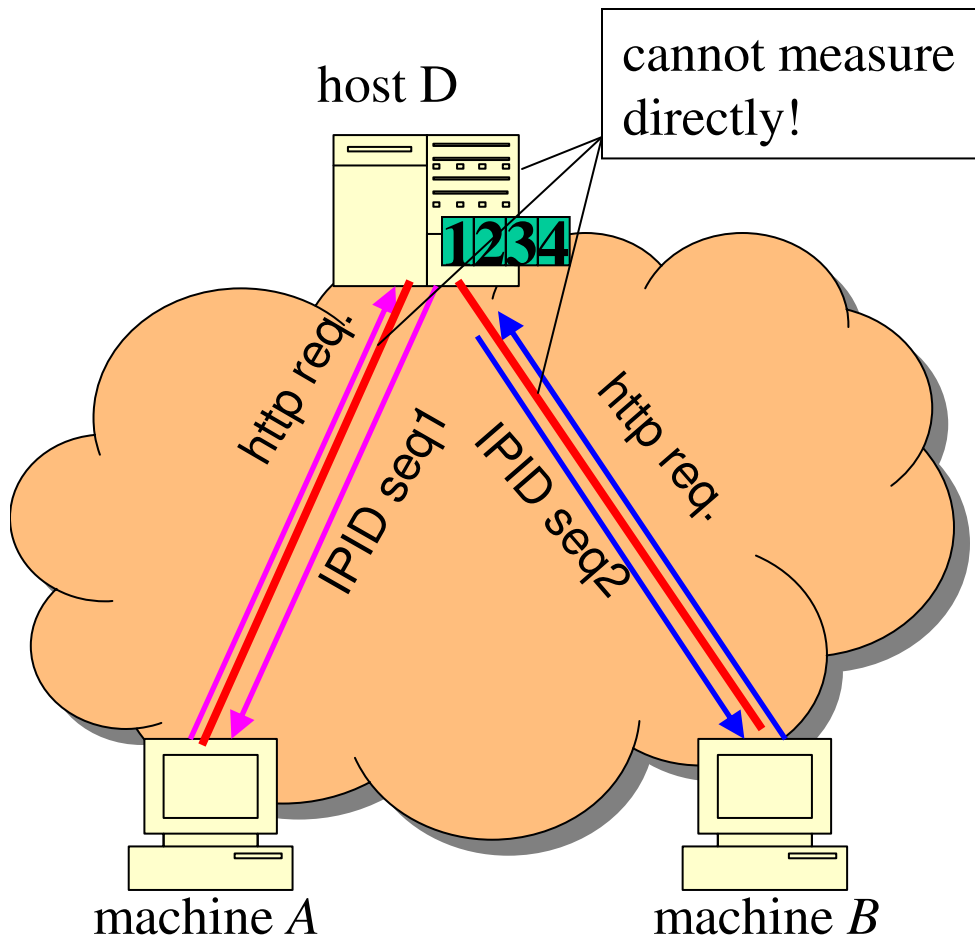
- ❑ What is the IPID field?
- ❑ Motivation for this work
- ❑ Previous work using IPID
- ❑ 3 new uses of IPID
 - ❑ inferring amount of network-internal traffic generated by server
 - ❑ inferring num. of load-balancing servers
 - ❑ inferring one-way delay differences
- ❑ Summary, future work

What is the IPID field?

- ❑ **IPID**
 - ❑ 16-bit field in IP header
 - ❑ contains current value of counter in IP stack
- ❑ **Global IPID: one global counter for host**
 - ❑ Windows 95/98/ME/XP/2000, Linux 2.2-, FreeBSD, Mac OS [insecure.org]
- ❑ **Per-session IPID: separate per-session counter**
 - ❑ Linux 2.3+ w/o path MTU discovery
- ❑ **Random/constant IPID**
 - ❑ recent OpenBSD, Solaris [insecure.org], Linux 2.3+ with path MTU enabled



Motivation



- ❑ Inferring network path and end system characteristics
 - ❑ Hard: limited information
- ❑ Exploit IPID field in IP header to infer characteristics
- ❑ No extensive study of IPID use, inference techniques₄

Classification:

Past use of the IPID field

- ❑ Measuring traffic activity
 - ❑ $\Delta\text{IPID}(i)$: packet counts between $T(i)$ and $T(i-1)$
 - ❑ Idle host scanning [insecure.org]
- ❑ Clustering of sources
 - ❑ Small ΔIPID between packets close in time
 - ❑ Router alias detection [Mahajan], NATed host counting [Bellovin], and load-balanced server counting [insecure.org]
- ❑ Identifying packet loss, duplication, and arrival order
 - ❑ Gap, repetition, order in IPID sequence
 - ❑ Passive monitor [Jaiswal], routers [Mahajan]

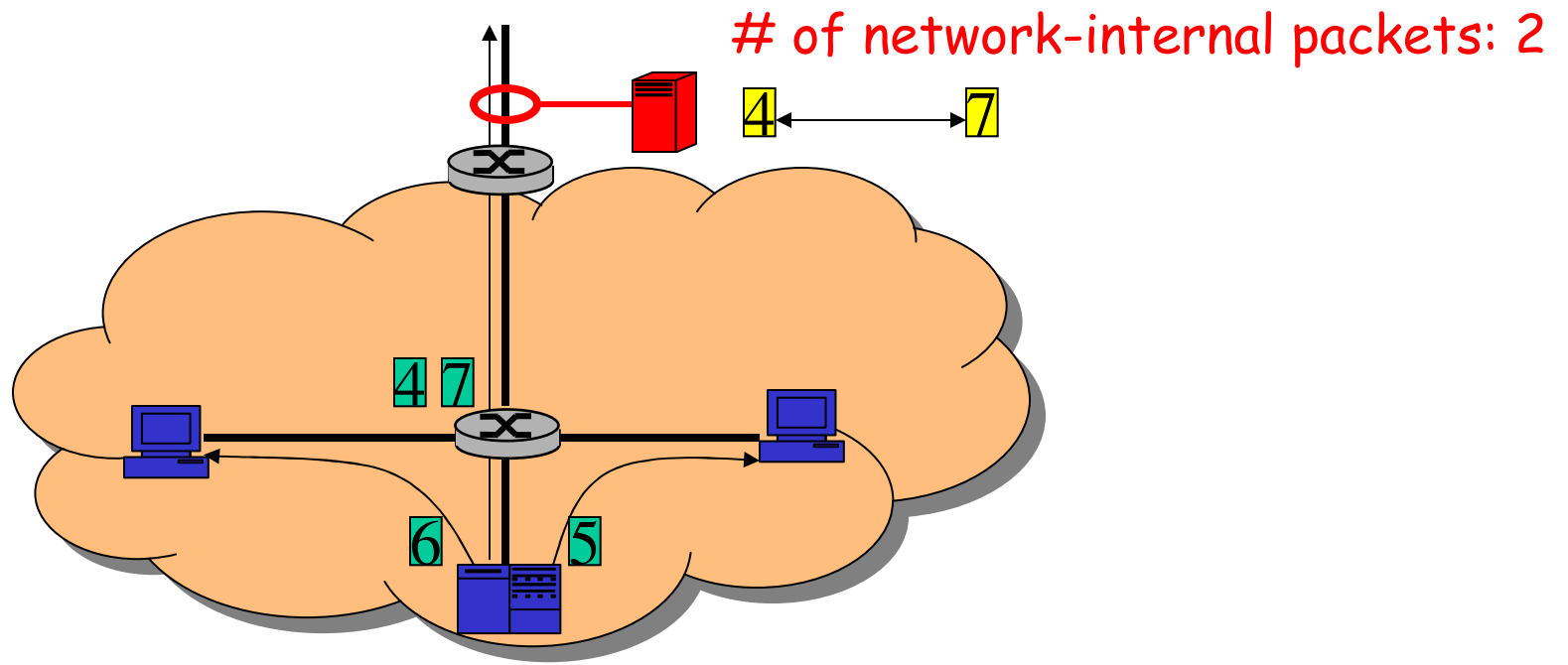
Contribution:

3 new uses of the IPID

- ❑ Measuring traffic activity
 - ❑ Amount of network-internal traffic generated by a server
- ❑ Clustering of sources
 - ❑ Load-balanced server counting
- ❑ Identifying packet loss, duplication, and arrival order
 - ❑ One-way delay differences

1: Inferring amount of internal traffic from server

- ❑ **Goal:** infer amount of network-internal traffic generated by server
- ❑ **Approach:** Passively monitor IPID at gateway router



1: Inferring amount of internal traffic from server - Algorithm

□ Main Idea

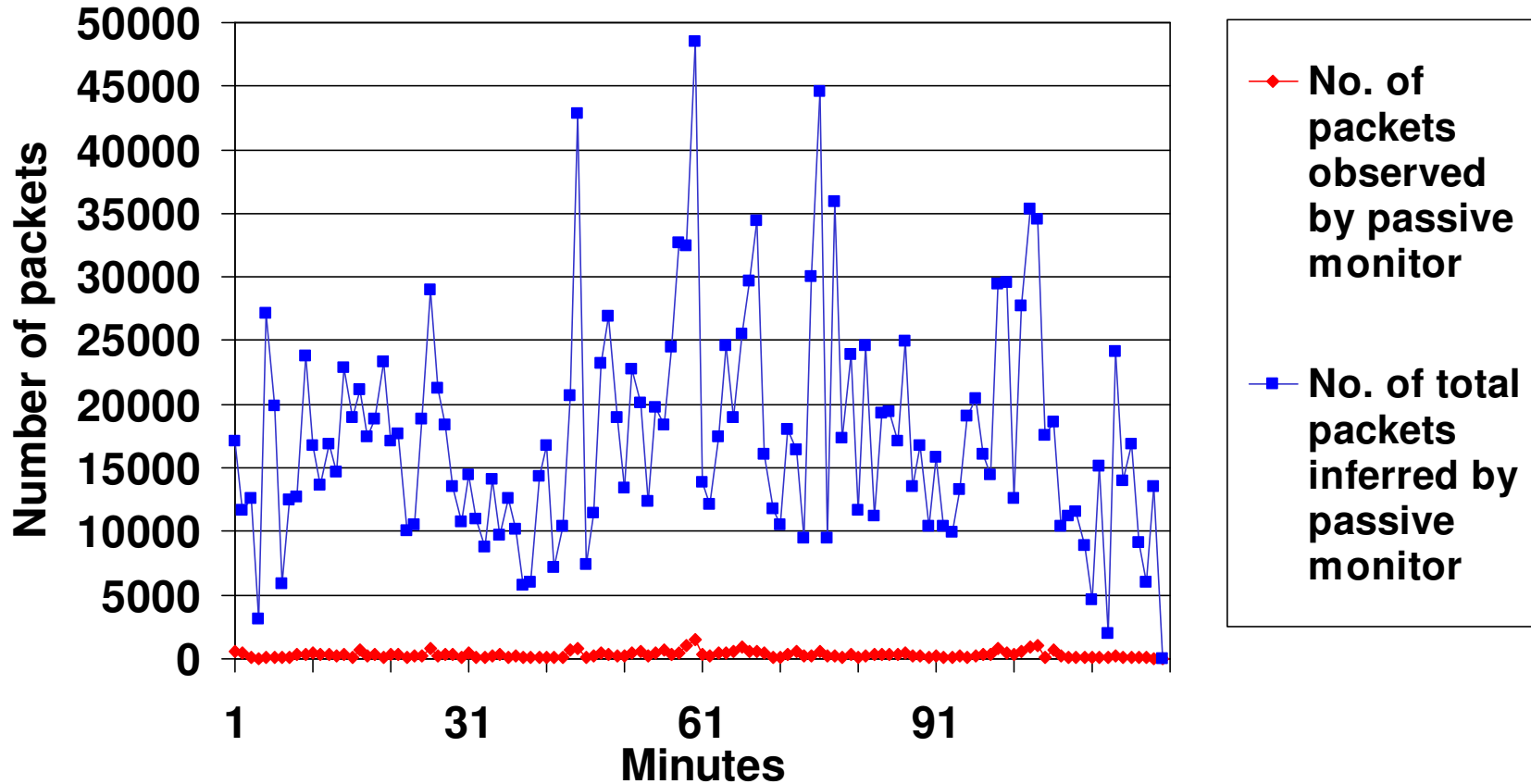
- (Δ IPID of two adjacent outbound packets at gateway) - 1 :
 - # of packets sent to hosts in internal network

□ How to handle multiple wraparounds?

- Augment with adaptive active measurement
- Exponentially weighted moving average (EWMA) for timeout

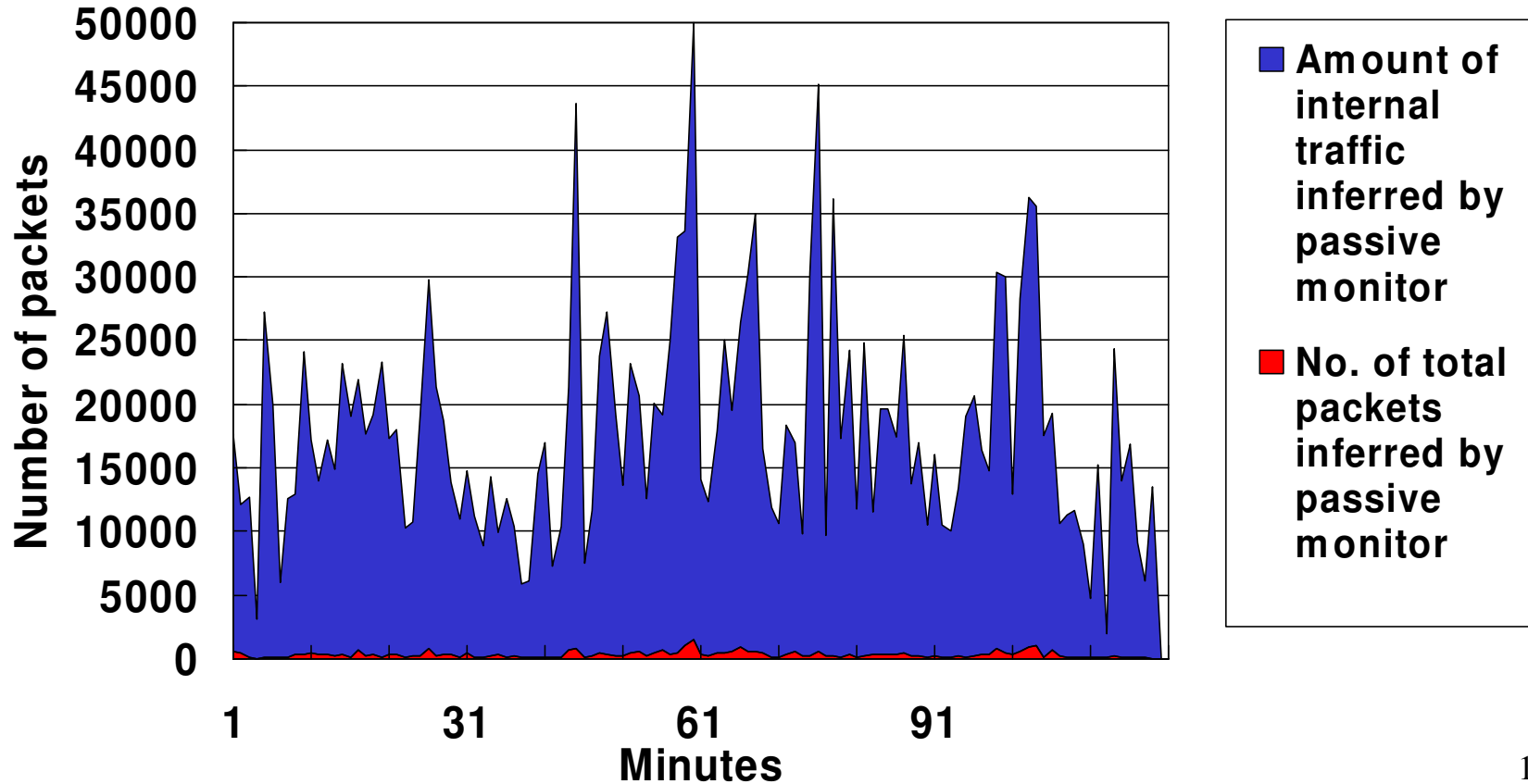
1: Inferring amount of internal traffic from server - Experimental result

No. of packets generated by server in each minute



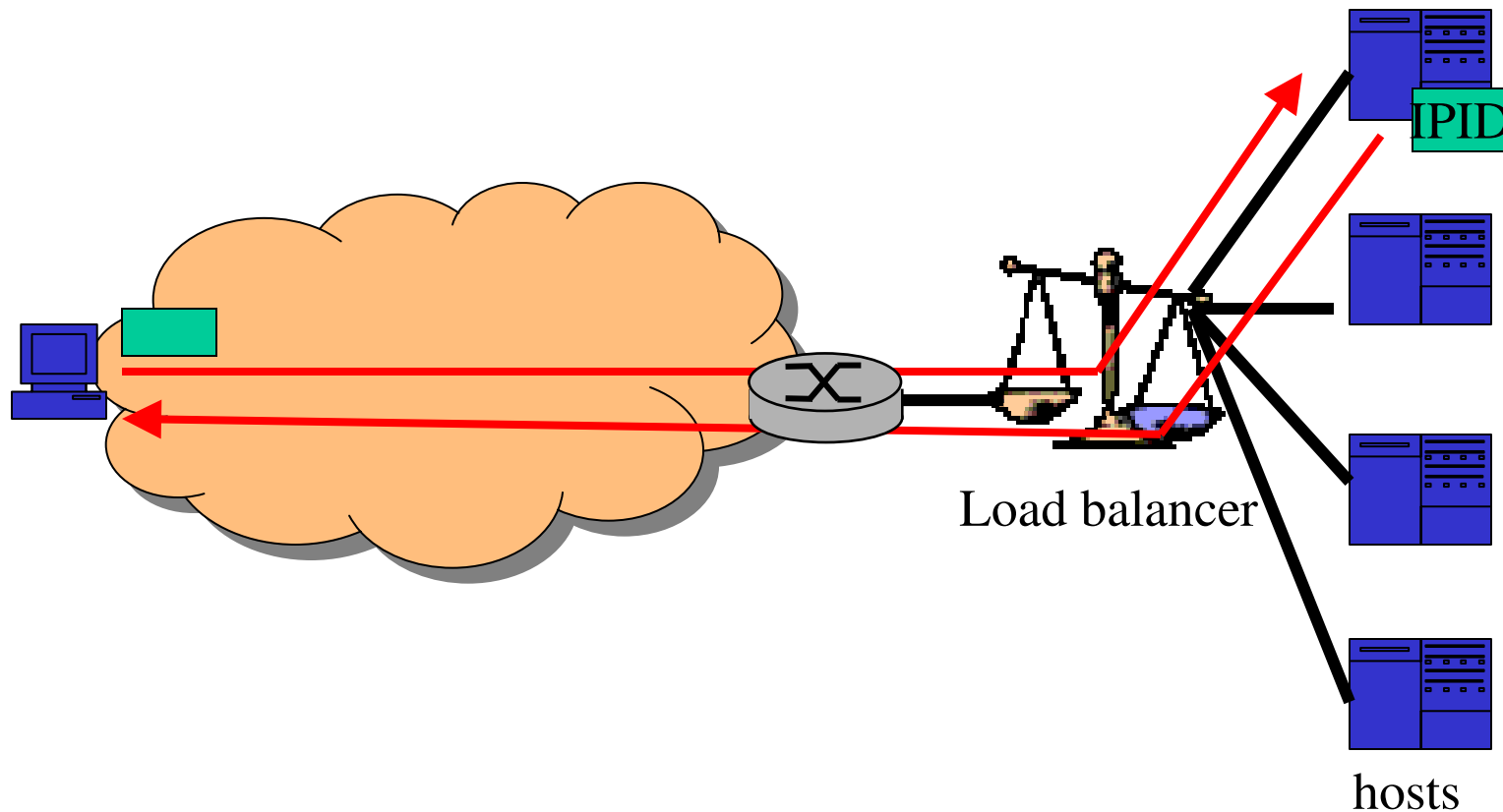
1: Inferring amount of internal traffic from server - Experimental result

No. of packets generated by server in each minute



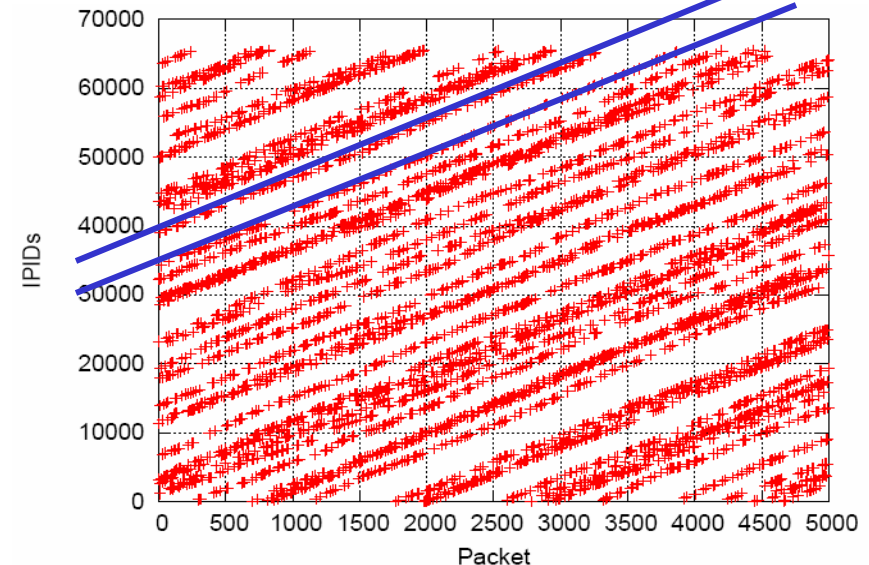
2: Inferring # of load-balancing servers

- IPIDs returned from a load-balancing server



2: Inferring # of load-balancing servers

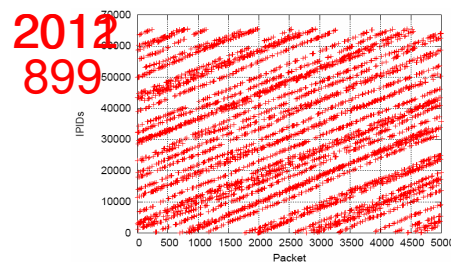
- ❑ IPIDs returned from load-balancing servers
- ❑ **goal**: infer # of load-balancing servers based on IPID values
- ❑ **approach**: classify IPIDs into distinct sequences, with # of sequences being the estimate



Commercial web server

2: Inferring # of load-balancing servers - Algorithm

- define threshold T
- rule 1: if new IPID has distance larger than T to all existing sequences, create new sequence;
- rule 2: attach new IPID to closest sequence;

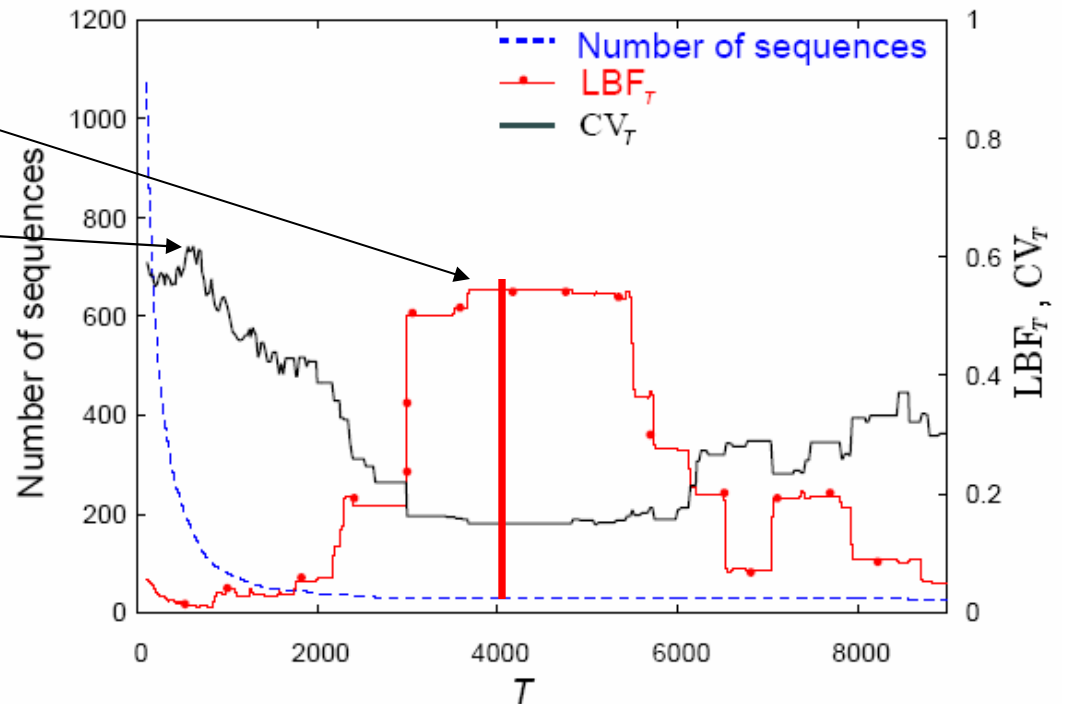


seq1
2011-899 > T
seq2
[rule 1] [rule 2]

2: Inferring # of load-balancing servers - Experimental result

$$LBF_T = \frac{\text{min sequence size}}{\text{max sequence size}}$$

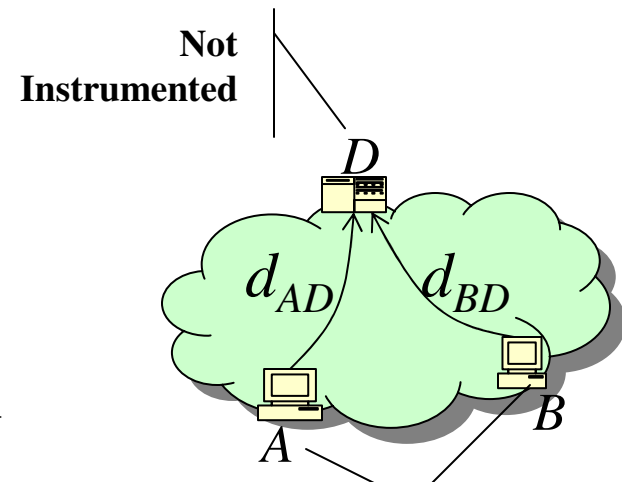
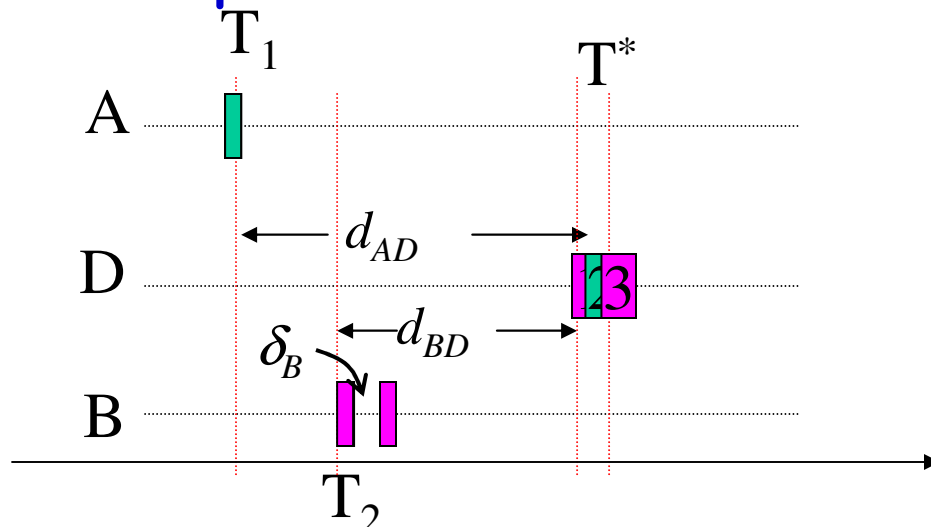
$$CV_T = \frac{\text{std deviation of sequence size}}{\text{mean of sequence size}}$$



An appropriate $T(=4000)$ has a maximum LBF_T and minimum CV_T

3: Inferring one-way delay difference

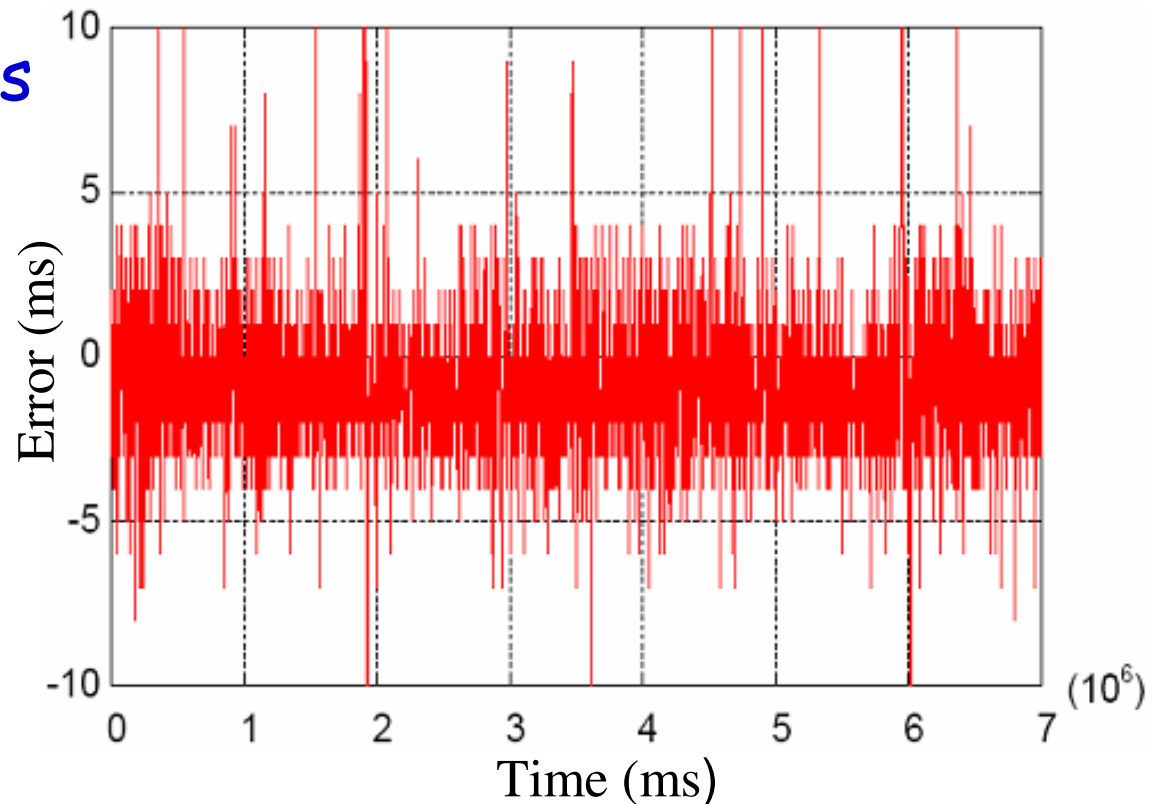
- ❑ **goal:** infer diff of one-way delay d_{AD} and d_{BD}
- ❑ **approach:** instrumented (synchronized) A and B send packets to non-instrumented host D



$$\delta_B \ll 1 \implies \begin{cases} T^* - d_{AD} \approx T_1 \\ T^* - d_{BD} \approx T_2 \end{cases} \implies T_2 - T_1 \approx d_{AD} - d_{BD}$$

3: Inferring one-way delay difference - Experimental result

- A : Unifacs (Brazil), B : UMN (MN), D : UMASS (MA)
- Actual diff: 230ms
- $\delta_A = 1000\text{ms}$
- $\delta_B = 3\text{ms}$



Most of difference are within δ_B 16

Summary and future work

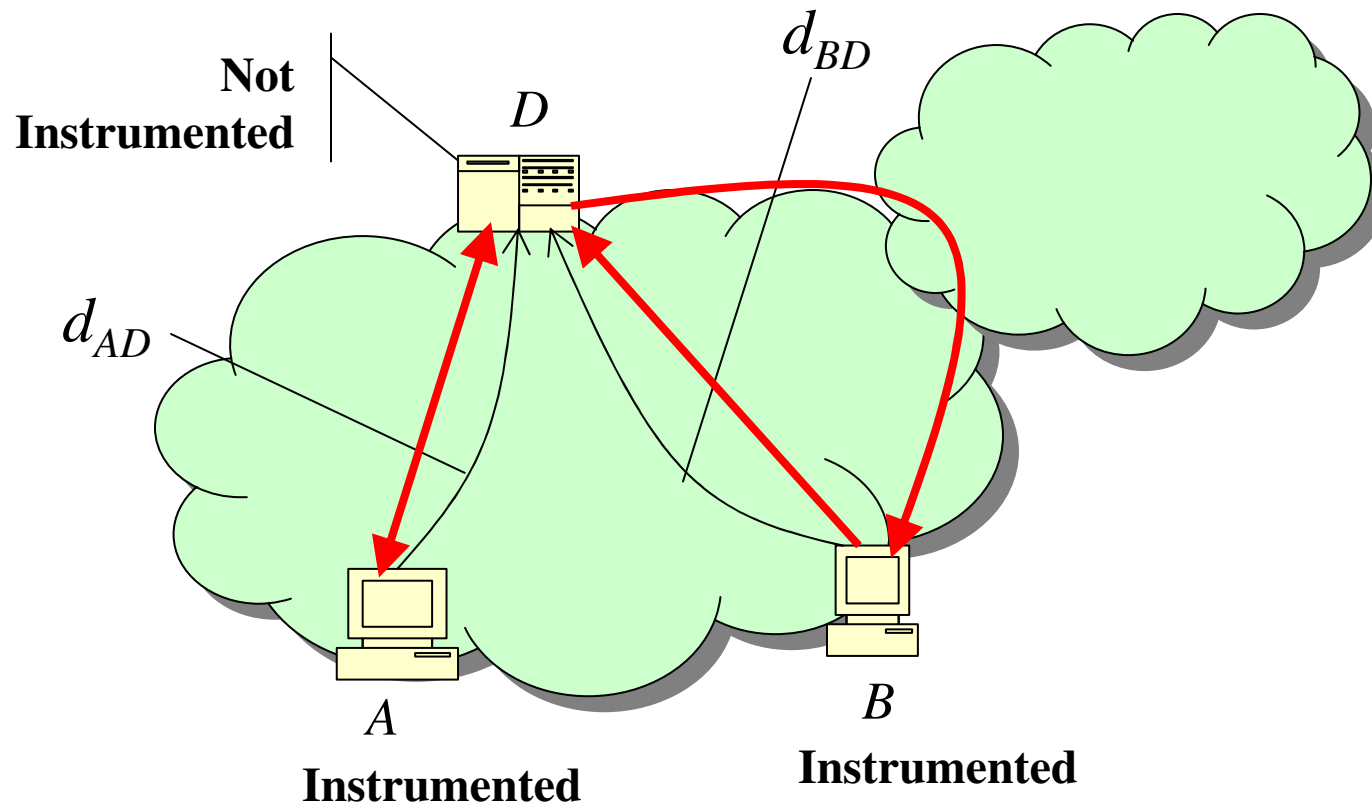
- ❑ Classified IPID field usages
- ❑ Proposed 3 new uses for IPID and techniques
- ❑ Empirical validation of proposed techniques
- ❑ Future work
 - ❑ Evaluation of EWMA-based algorithm
 - ❑ Deal with high jitter for one-way delay diff

The End

Backup Slides
follow....

Why 3 new uses of IPID interesting?

- ❑ **Inferring amount of network-internal traffic**
 - ❑ Characterizing spatial distribution of traffic: from a "single" measurement point!
 - ❑ Useful for traffic modeling
- ❑ **Load-balancing server counting**
 - ❑ Security hazard (DoS attack)
 - ❑ Sales information
- ❑ **One-way delay diff measurement?**
 - ❑ Estimation of one-way delay of asymmetric path, based on known RTT of symmetric path



$$d_{AD} \approx \text{RTT}_{AD} / 2$$

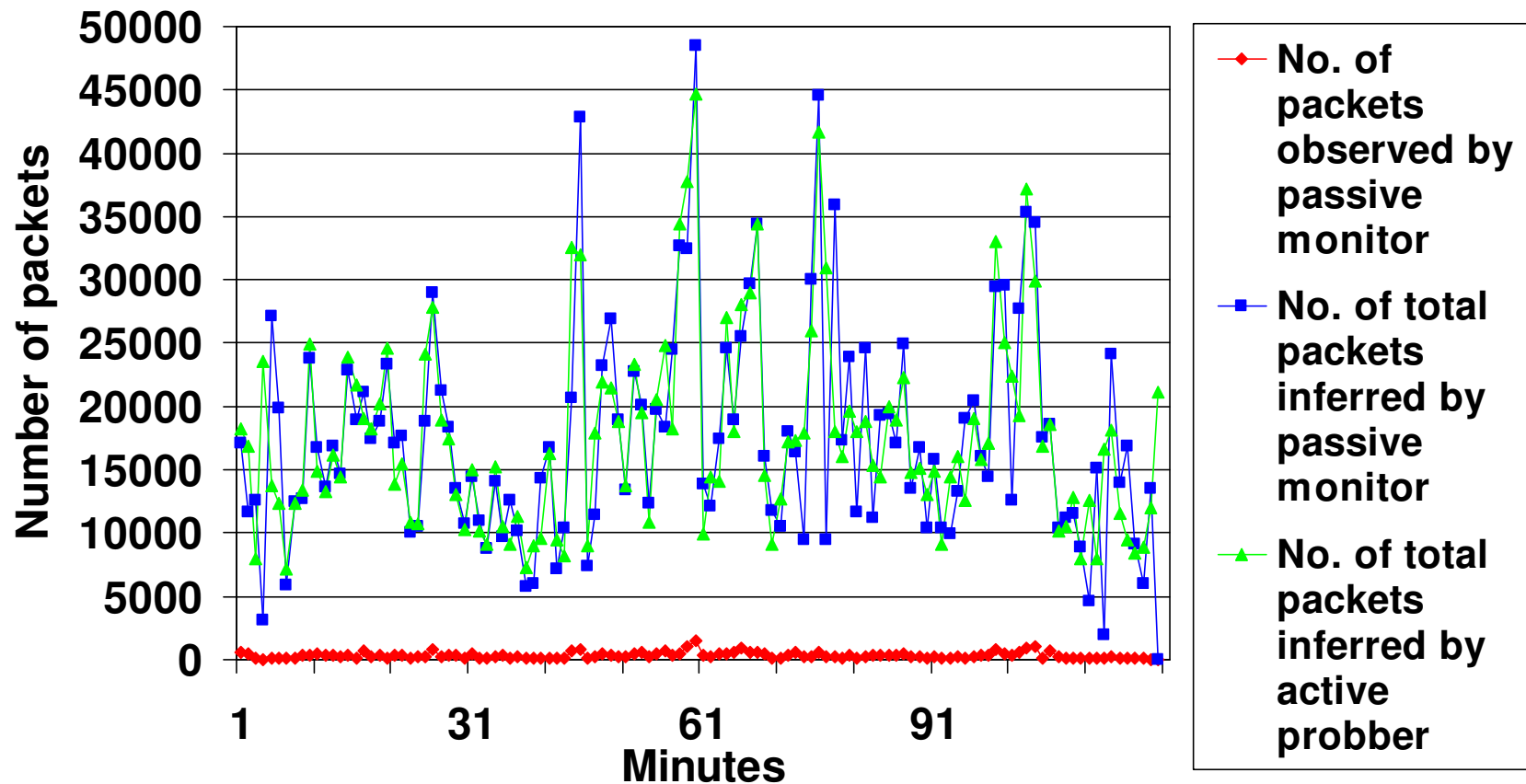
$$d_{BD} \approx d_{AD} - (\text{one-way delay diff})$$

Questions?

- ❑ Some traffic may stay on a LAN
 - ❑ indeed, some are sent to the "loopback" address, and stay within the host
- ❑ Validation of load-balancing server counting?
 - ❑?
- ❑ Validation of internal traffic inference?
 - ❑ Lab test
 - ❑ Active measurement test
 - ❑ See next slide!

Inferring amount of internal traffic from a server : Experimental result

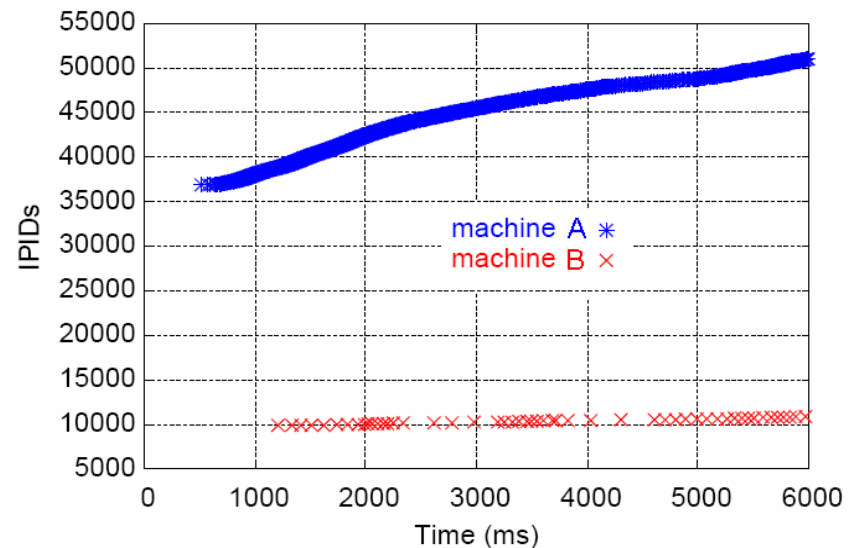
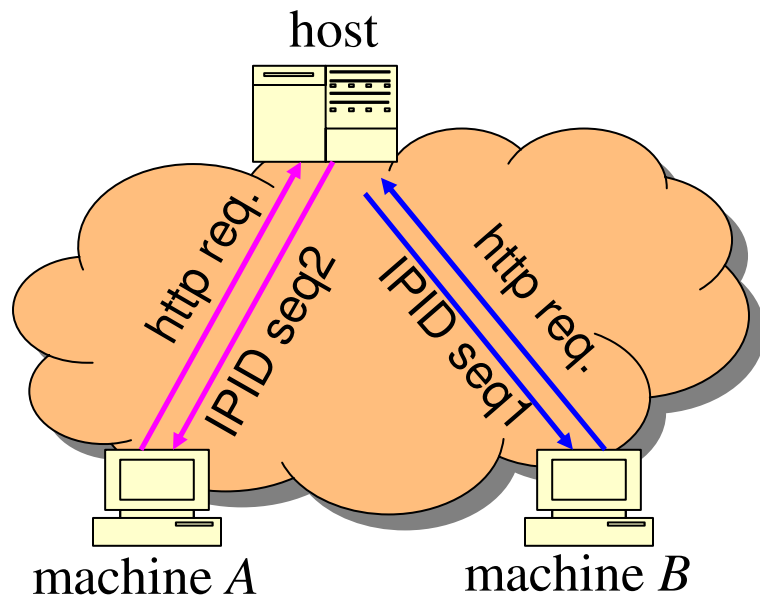
No. of packets generated by server in each minute



Cross correlation coefficient = 0.96

Verifying global IPIDs

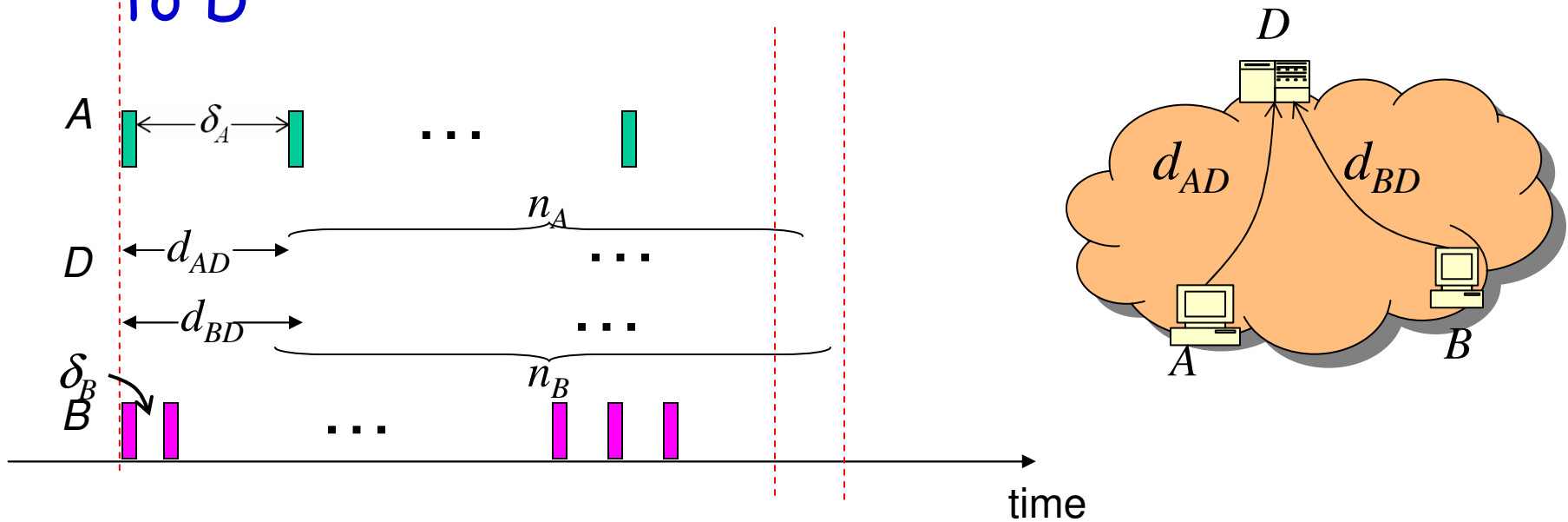
- Simultaneously sending requests from different machines with different rates



The host has non-global IPID

Inferring one-way delay difference (More complex animation)

- **goal:** infer diff of one-way delay d_{AD} and d_{BD}
- **approach:** synchronized A and B send packets to D

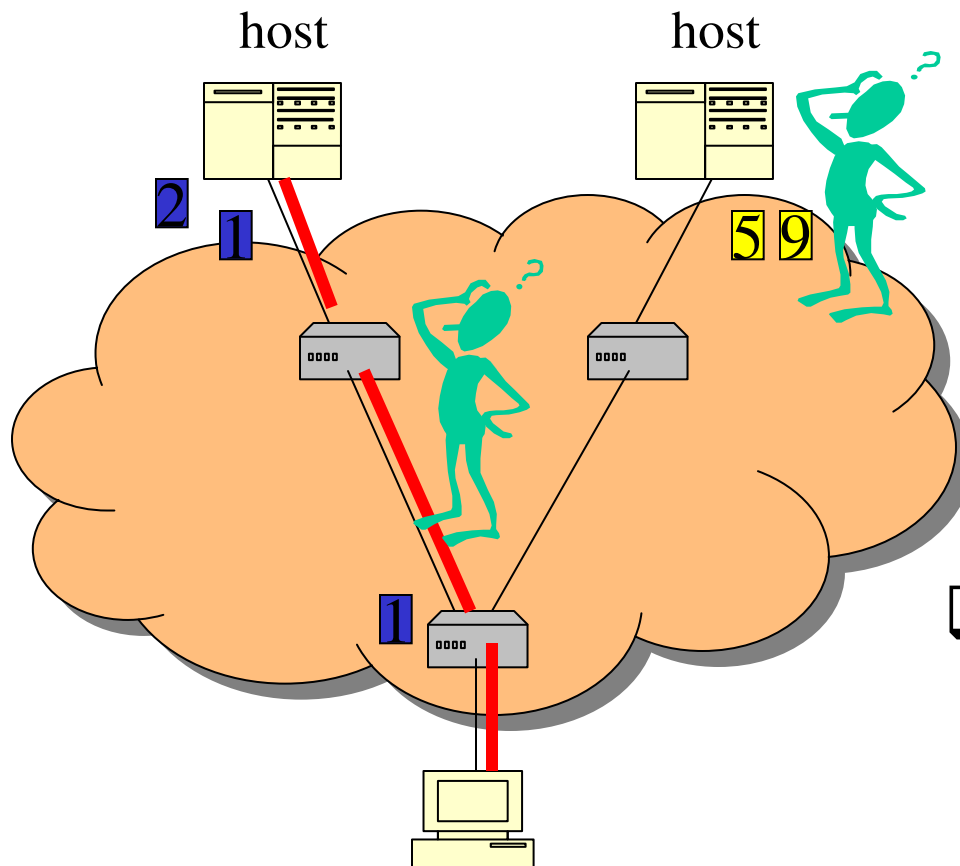


$$d_{BD} + (n_B - 1)\delta_B \leq d_{AD} + n_A\delta_A \leq d_{BD} + n_B\delta_B$$

$$\Rightarrow (n_B - 1)\delta_B - n_A\delta_A \leq d_{AD} - d_{BD} \leq n_B\delta_B - n_A\delta_A$$

The following slides
are for reference only

Motivation (Full version)

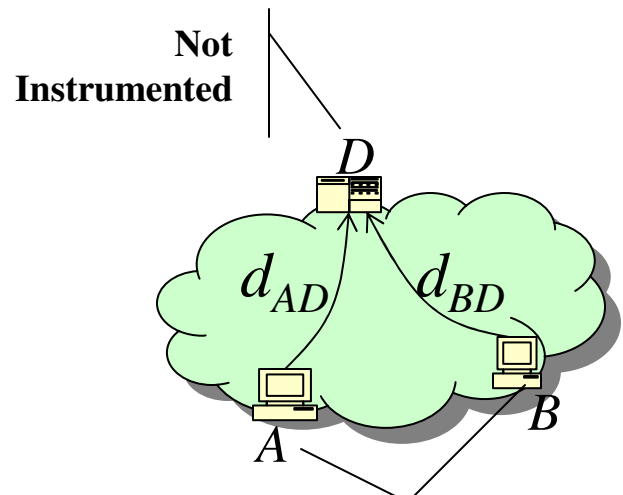
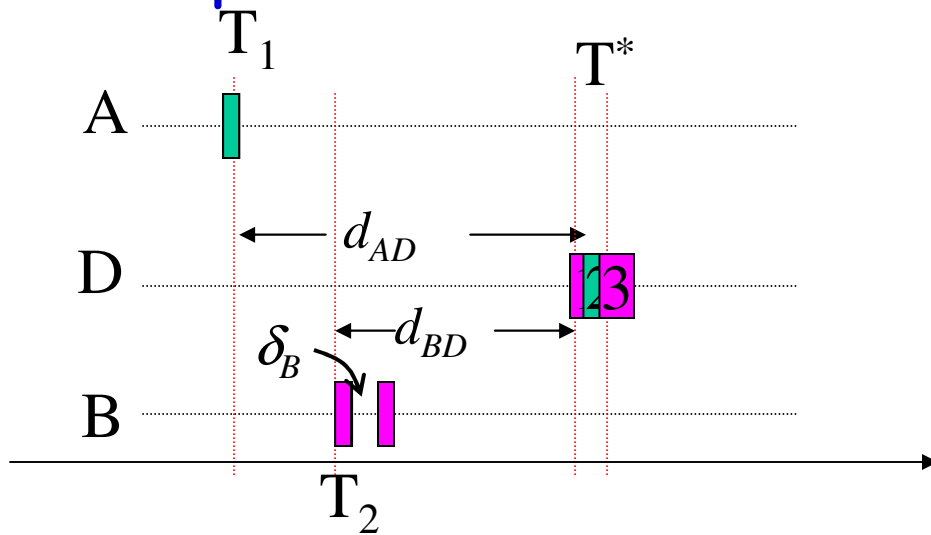


- ❑ Inferring network path and end system characteristics
 - ❑ Important and hard
- ❑ Exploitation of global IPID implementation
- ❑ No extensive study on IPID usages and inference techniques

Our contribution: classification of IPID usages and new techniques

Inferring one-way delay difference

- ❑ **goal:** infer diff of one-way delay d_{AD} and d_{BD}
- ❑ **approach:** instrumented (synchronized) A and B send packets to non-instrumented host D



$$\delta_B \ll 1 \implies \begin{cases} T_1 + d_{AD} \approx T^* \\ T_2 + d_{BD} \approx T^* \end{cases} \implies T_2 - T_1 \approx d_{AD} - d_{BD}$$