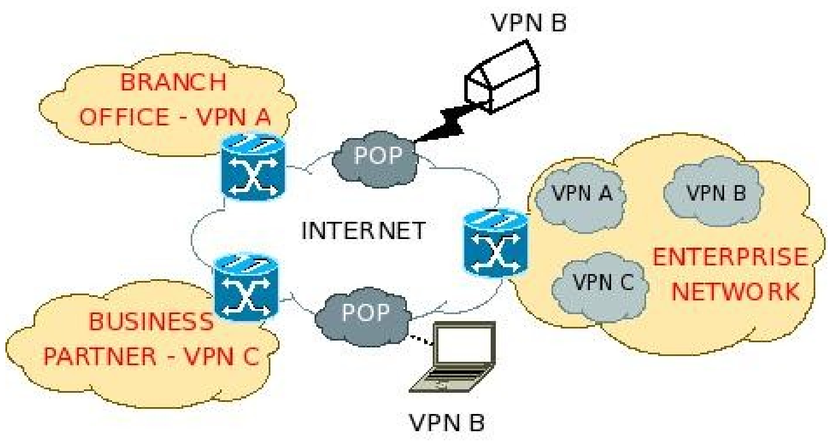
Optimal design of performance measurement experiments for complex, large-scale networks

Charalampos Rotsos, Andrew Moore, Computer Laboratory, University of Cambridge
In collaboration with John Schormans, Steven Gilmour and Ben Parker Queen Mary, University of London.

A BRAZIL PROJECT

Scope



MOTIVATION

I want to perform optimal resource allocation
• How can I define the performance of a VPN within an enterprise network?

I need to be able to be able to verify SLA
• How can I measure accurately network performance for an application?

OBJECTIVE

Every observation made in a computer network is subject to measurement error due to causes such as implementation choices, hardware imperfections and timer precision.
• Can we parametrise this error?

• What are the the traffic characteristics that describe performance of a network application?

Focus

PROBLEM

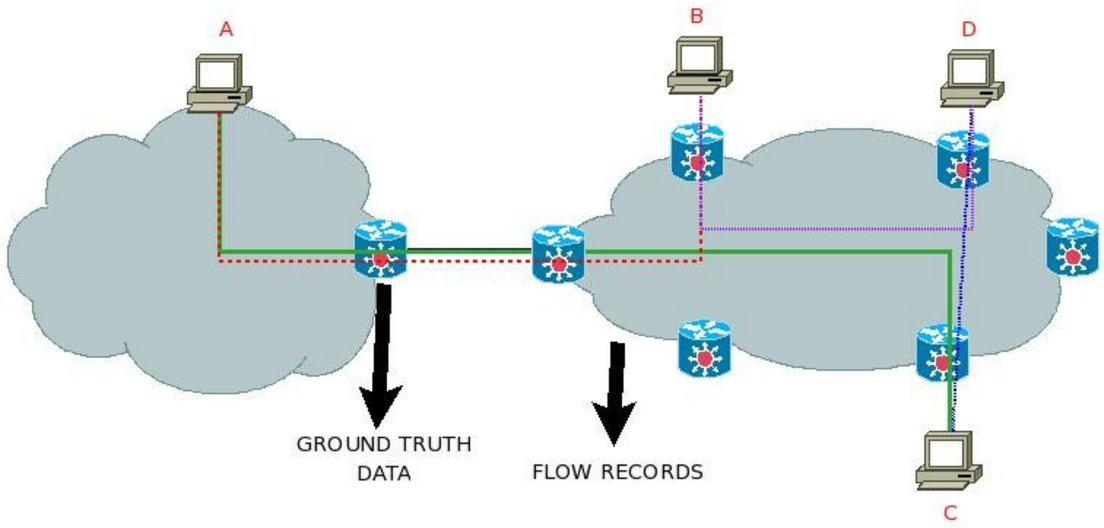
As a first step in this performance analysis we believe that it is important to understand what are the applications in a network .

• Can we formulate an automatic method with low resource demands to measure what applications exist in a network?

EXISTING APPROACHES

IDS / Anomaly detection	fast implementations	depends on protocol specifications, task specific
Deep packet inspection	accurate results	Full payload, fails on encryption & protocol changes
Statistical analysis[1][3]	Flow granularity, can run online on fast link	Requires diverse ground truth data for training,
Connection Pattern / BLINC[2]	low information requirement	host granularity, fails to adjust on small protocol changes, complex design

Can we fuse these different approaches to reduce the effects of the disadvantages and keep only the advantages?



METHOD

By using ground truth data and flow records we can formulate identifiers based on:
- Connection patterns
- Flow statistics
- Host behaviour

If we observe the behaviour of the hosts within a network we can match them to existing identification and derive the type of the application.

• What information do I need to identify the application of A-B and A-C flows?

I know what type of application runs on the A-B and A-C flows and I have flow records from all intermediate routers.

• Can I infer what applications run on the B-D and C-D lines?

CHALLENGES

• What is the minimum information required to characterise a flow?

• How accurate is application identification in the presence of sampling?

• What is the effect on the accuracy of an application identification algorithm when the data covers only a subset of the network?

• Can we develop techniques that adapt to new protocols and recognise them as unknown or match them to applications with similar behaviour?

• How can we adapt to changes in the application mix over time?

References

1. Andrew W. Moore , Denis Zuev, Internet traffic classification using bayesian analysis techniques, SIGMETRICS ' 05
2. Karagiannis, T., Papagiannaki, K., and Faloutsos, M. BLINC: multilevel traffic classification in the dark. SIGCOMM '05
3. Jiang, H., Moore, A. W., Ge, Z., Jin, S., and Wang, J. Lightweight application classification for network management. INM '07

EPSRC
Engineering and Physical Sciences Research Council

UNIVERSITY OF CAMBRIDGE
800 YEARS 1209 ~ 2009