Selective capping of packet payloads for network analysis and management

Víctor Uceda, Miguel Rodríguez, **Javier Ramos**, José Luis García-Dorado and Javier Aracil

Universidad Autónoma de Madrid
7th International Workshop on Traffic Monitoring and Analysis
Barcelona, Spain

April 23,2015

High Performance Computing and Networking

Outline

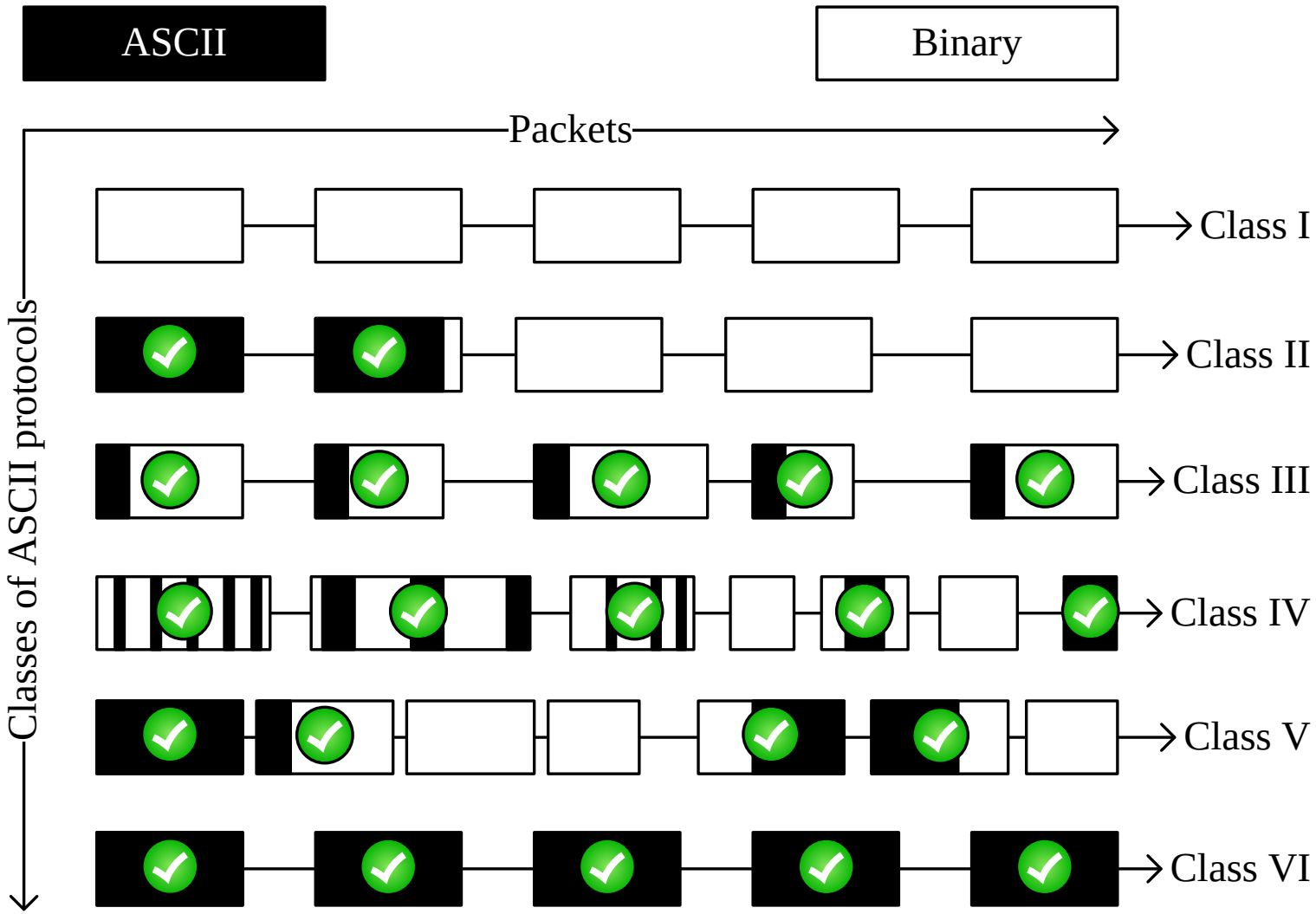
Outline

Introduction

- Traffic capture and storage at multi-Gb/s rates is a demanding task:
  - Capture throughput
  - Write throughput
  - Storage capacity
- Huge amount of useless packet payloads (encrypted traffic)
- A packet is only interesting for storage if it can be interpreted by network analysts:
  - Unknown binary protocols cannot be interpreted
  - Text-based or well-known binary protocols can be interpreted

Introduction
Objectives

1. Cap packets to reduce the storage capacity needed:
   - Cap unknown binary protocols
   - Keep text-based and well-known binary protocols
2. Detect packets containing ASCII bytes as representative of human readable characters
3. Identify and categorize protocols based on ASCII distribution
4. Implement capping method in a high-performance driver


ASCII
Binary
Packets
Classes of ASCII protocols
Class I
Class II
Class III
Class IV
Class V
Class VI

Outline

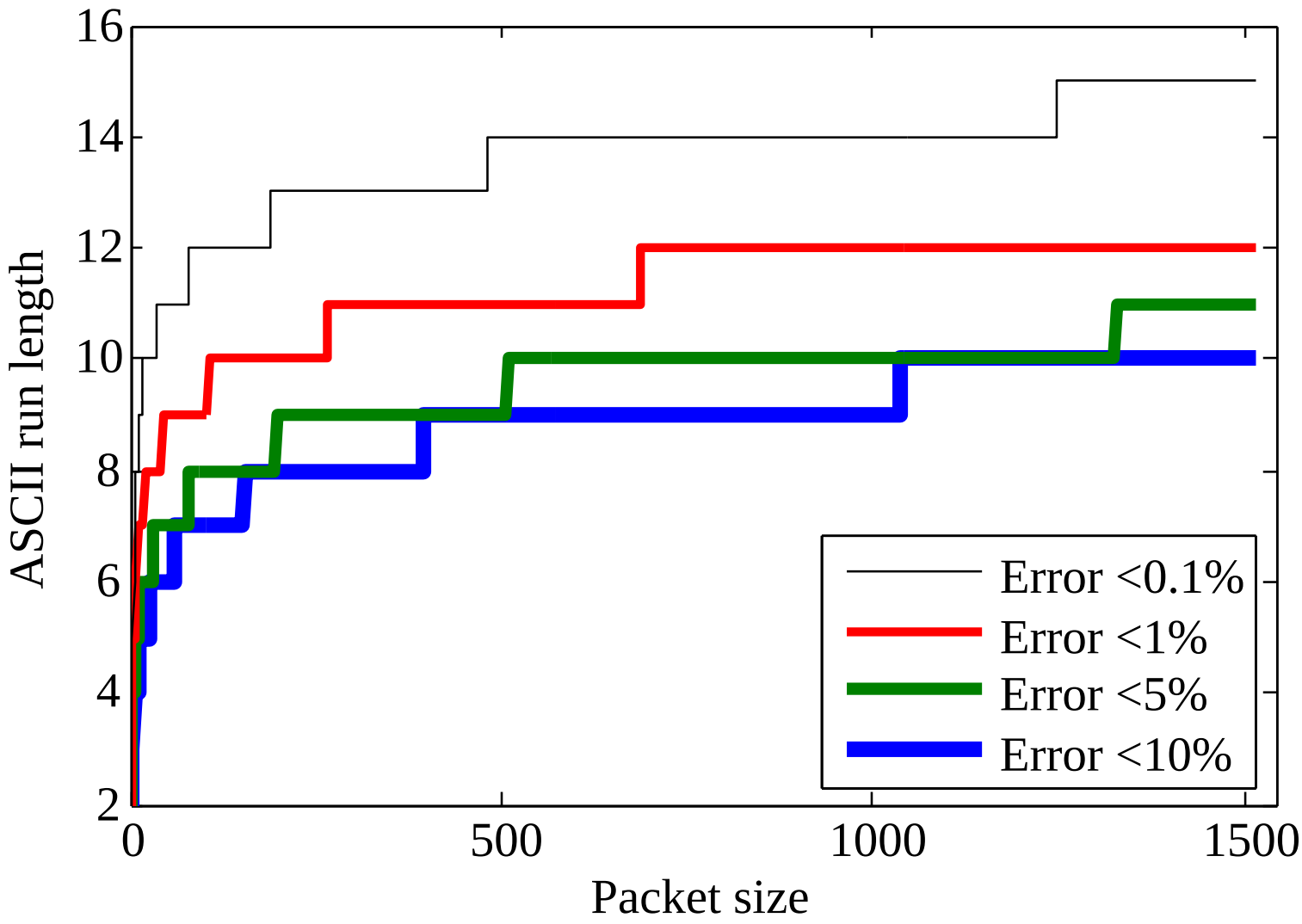
Detection of ASCII traffic

Runs threshold

- Detect ASCII character runs of specific length
- Run-length must be large enough to reduce the error (FP) with respect to a random uniform payload distribution
- The probability of generating a random byte-value contained in the ASCII range is 0.4
- Error can be modeled using a Markov chain with L+1 states, each of it corresponding to have read an ASCII run of L consecutive characters

Detection of ASCII traffic

Runs threshold



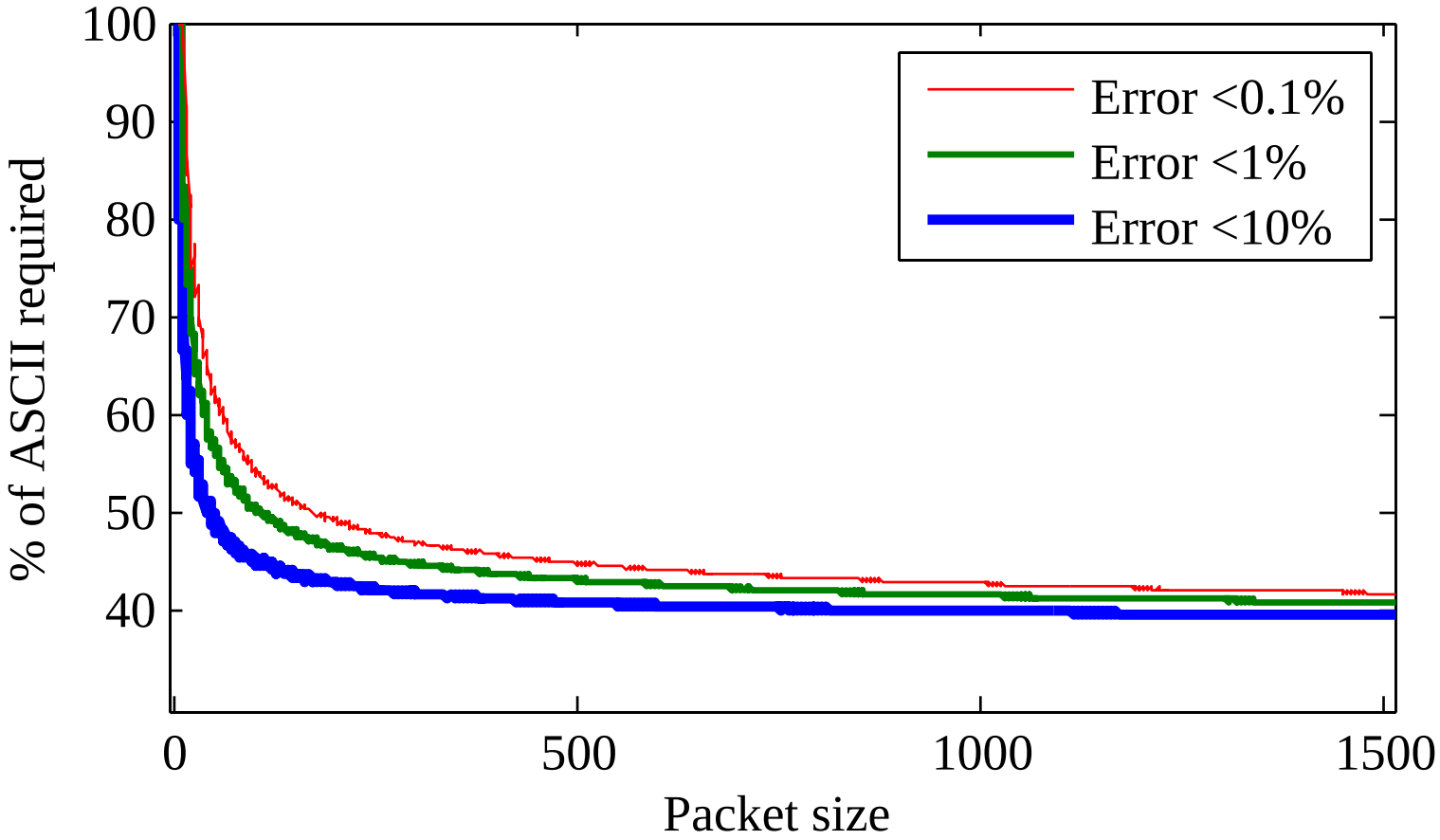
Detection of ASCII traffic

Percentage threshold

- TLV-based protocols contain ASCII characters but is not likely to find long ASCII runs
- A packet is a sequence of N bytes each with probability p of being ASCII and probability 1-p of being a non-ASCII value
- Use binomial distribution to calculate the minimum ASCII percentage required for an error threshold

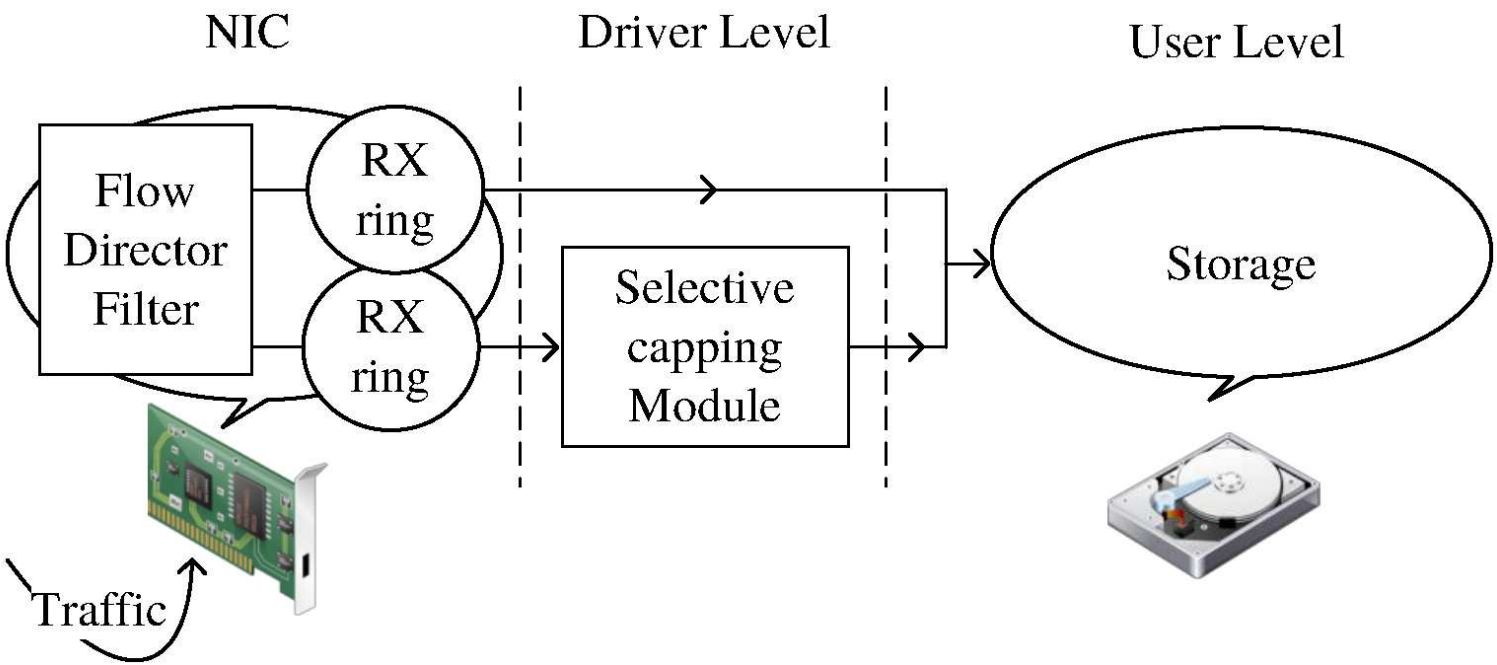
Detection of ASCII traffic

Percentage threshold



Outline

Selective capping sniffer architecture



Outline

Results and discussion
Testbed

- The ASCII-capping module has been implemented over a high-performance network driver (HPCAP)
- Use of commodity server equipped with Intel 10 Gb/s card and RAID-0 storage
- Traffic traces replayed with FPGA
- Performance and compression ratio are evaluated when lossless capture and storage is achieved

Trace	**Capture Point**	**Info.**
1	University	DNS , HTTP , SSH Dropbox , Others
2	HTTP proxy	HTTP
3	Bank network	HTTP, HTTPS, SSH Banking protocols, DNS

Results and discussion
Compression Ratio

Trace	Compression ratio	% of capped packets
1	4.28	45
2	3.33	74
3	3.24	81

Results and discussion
Performance Evaluation

Trace	Avg. throughput (Gb/s) ± Std. Dev.	Avg. packet rate (Kpps) ± Std. Dev.
1	3.1 ± 0.13	2221 ± 100
2	2.5 ± 0.08	856 ± 29
3	3.2 ± 0.15	428 ± 21

Outline

Conclusions

- A solution for on-the-fly packet capping in multi-Gb/s networks has been presented
- Capping is performed identifying ASCII packets based on two different methods
- A taxonomy of protocols based on the amount and distribution of ASCII characters has been presented
- Implementation over a high-performance capture engine (HPCAP):
  - Performance between 2.5 and 3.2 Gb/s
  - Compression ratio between 3 and 4

Outline

Future Work

1. Increase performance to cope with fully-loaded 10 Gb/s links:
   - Use of parallelism paradigms (multi-core programming and GPUs)
   - Use of sampling strategies for ASCII inspecting
   - Use of hardware solutions (NetFPGA)
2. Identify useless ASCII payload using dispersion measurements
3. Detect other codification schemes (UTF-8 and base64)

Thank you for your attention!

Questions?

Selective capping of packet payloads for network analysis and management

Víctor Uceda, Miguel Rodríguez, **Javier Ramos**, José Luis García-Dorado and Javier Aracil

Universidad Autónoma de Madrid
7th International Workshop on Traffic Monitoring and Analysis
Barcelona, Spain

April 23,2015

High Performance Computing and Networking