Improving the Performance of Intrusion Detection using Dialog-based Payload Aggregation

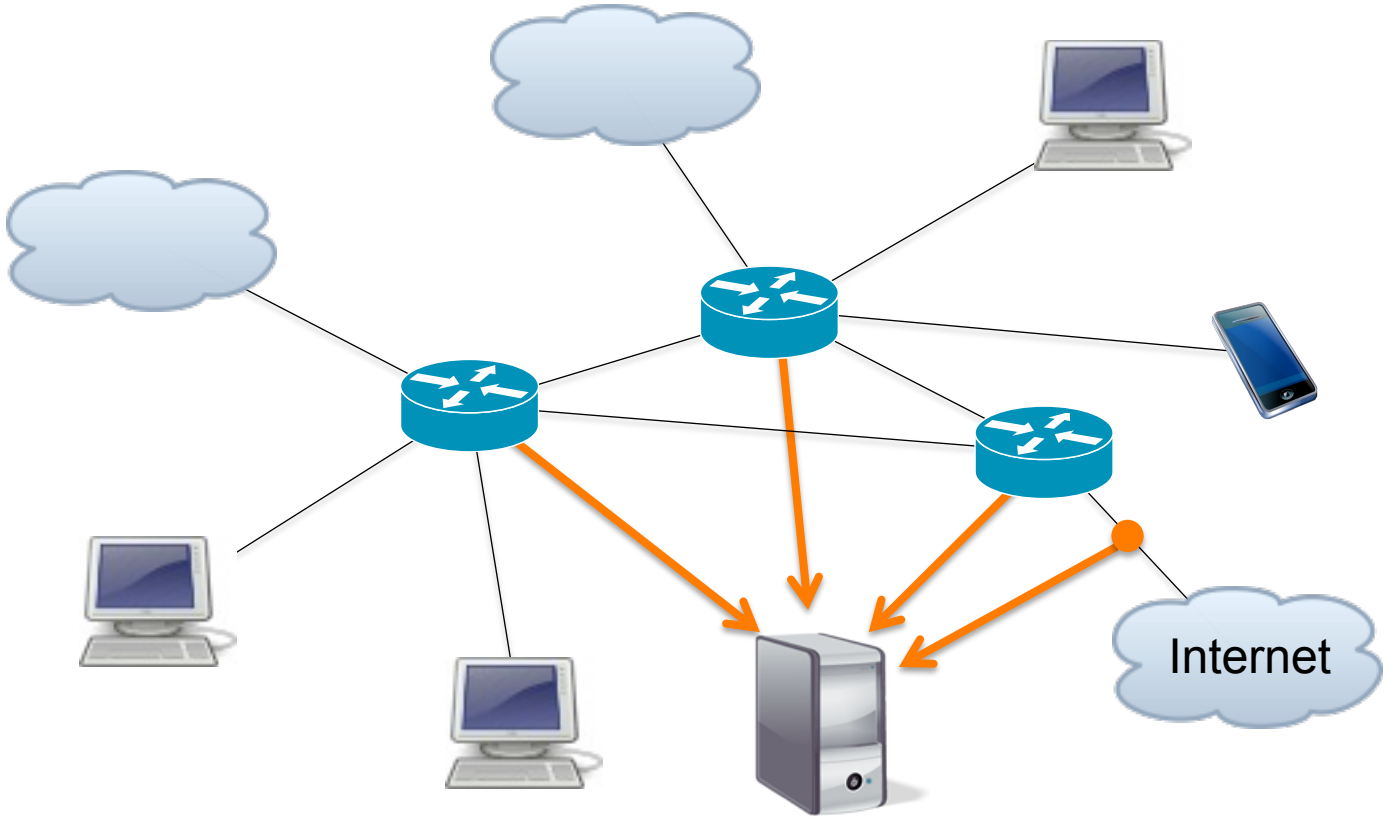
Tobias Limmer, Falko Dressler

Chair for Computer Networks and Communication Systems
University of Erlangen-Nürnberg

limmer at cs.fau.de

15-04-2011

Network-based Intrusion Detection

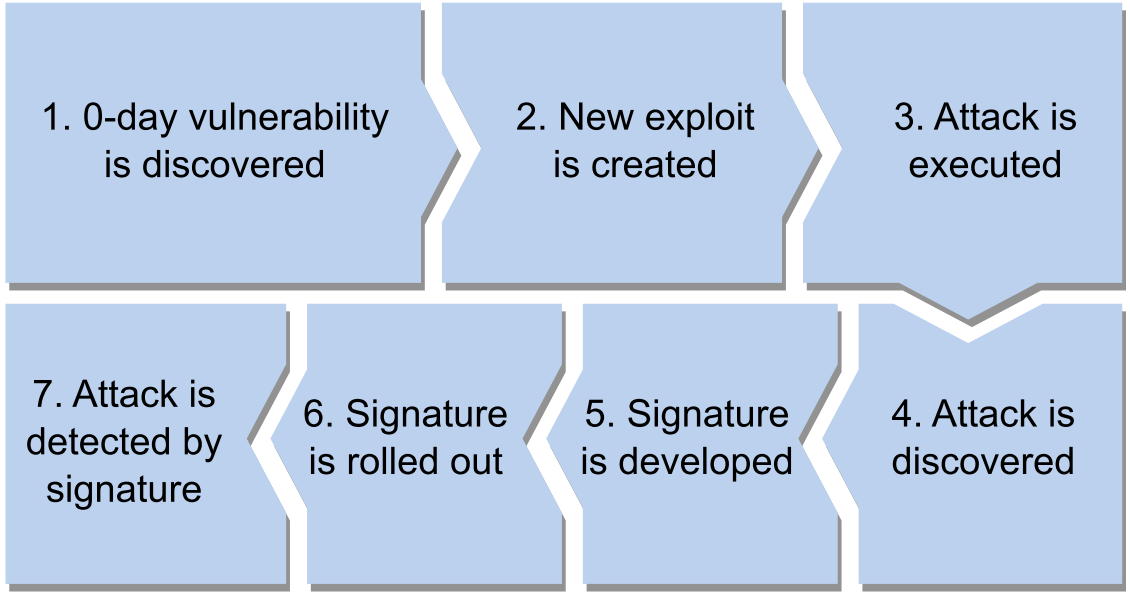


Motivation

- Focus on IDS based on payload analysis using signatures
- Performance problem for these IDSs implemented in software:
  - Processing rate: 200 MBit/s
  - Common data rate of network link: 10 GBit/s
  - ~100 IDS instances needed to analyze fully loaded link (!)
- Multiple suggestions for improvement already available:
  - FPGAs, graphic cards
  - Improved matching algorithms
  - Filtering based on header data (IP addresses, ports)
  - Parallelization
  - . . .

Signature-based Detection

- Typical signature generation process:



- Similar for all signatures!

Payload-based IDS

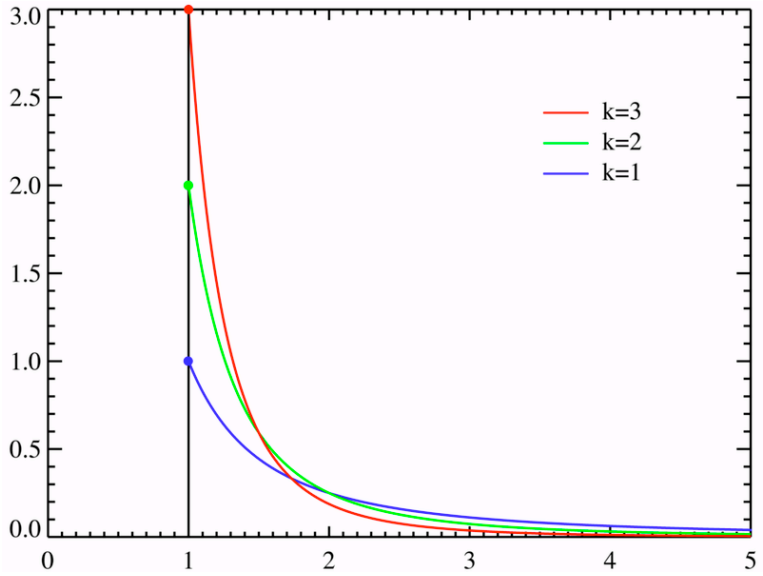
- Common signature features:
  - Header filters: protocol, IPs, ports
  - Payload matches:
    - simple and with regular expressions
    - match restrictions within packets
- Popular implementations: Snort, Bro
- Example signature:

```
alert tcp $EXTERNAL_NET any -> $HOME_NET 21 (msg:"ET EXPLOIT GuildFTPd CWD
and LIST Command Heap Overflow - POC-1"; flow:established; content:"cwd";
depth:4; nocase; dsize:>74; pcre:"/(\/\.){70,}/i"; sid:2008776; rev:3;)
```

- Evasion is possible:
  - Exploitation of protocol ambiguities (→ normalization)
  - Data encryption (→ “SSL-terminators”)
  - Use of unknown attacks / communication protocols (→ anomaly-based IDS?)

Heavy-tailed Network Traffic

- “Heavy-hitters”
- What means heavy-tailed?
  - Pareto-distribution with shape parameter k<2
- Multiple parts within a connection:
  - Dialog between server and client
  - Transfer of bulk data
  - Examples:
    - HTTP: request/response and URI content from server
    - POP3/IMAP: capability handshake, login, request, mail content

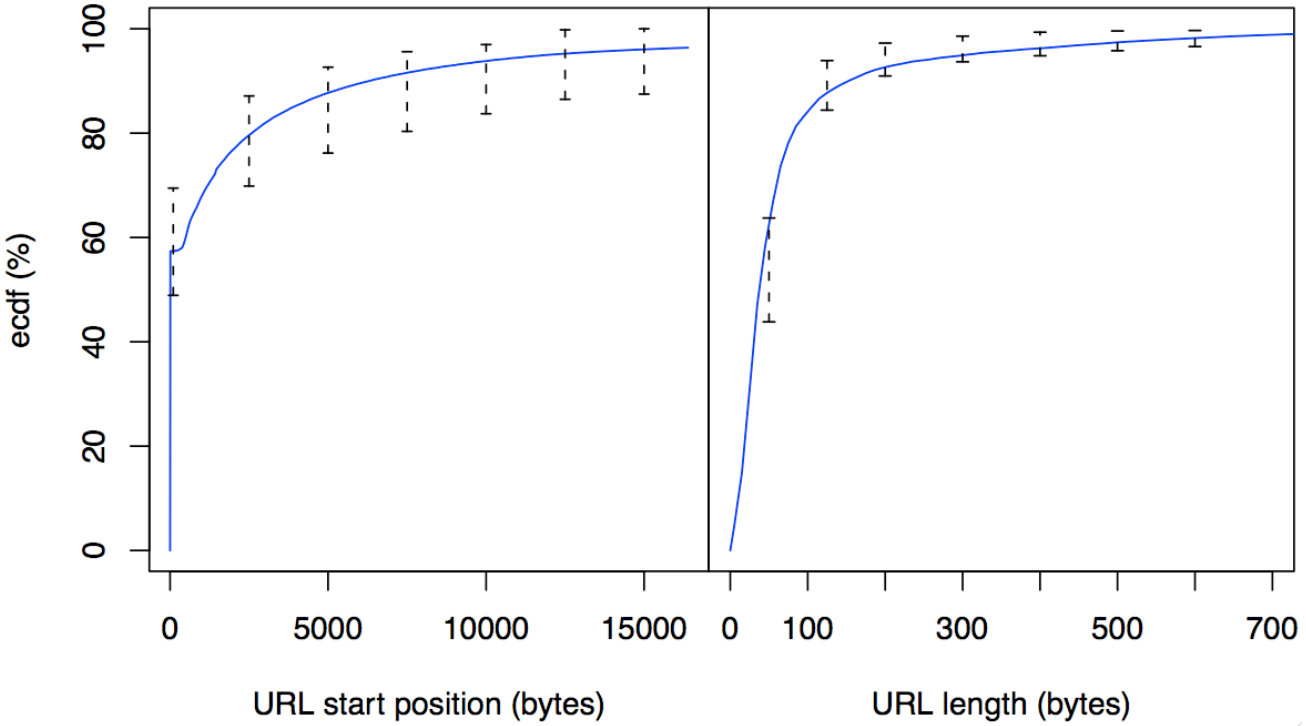


[source: wikipedia.org]

- Hypothesis: Bulk data not interesting for attack detection!
- First approach: Capture payload from beginning of connection
  - Examples: Time-Machine, FPA

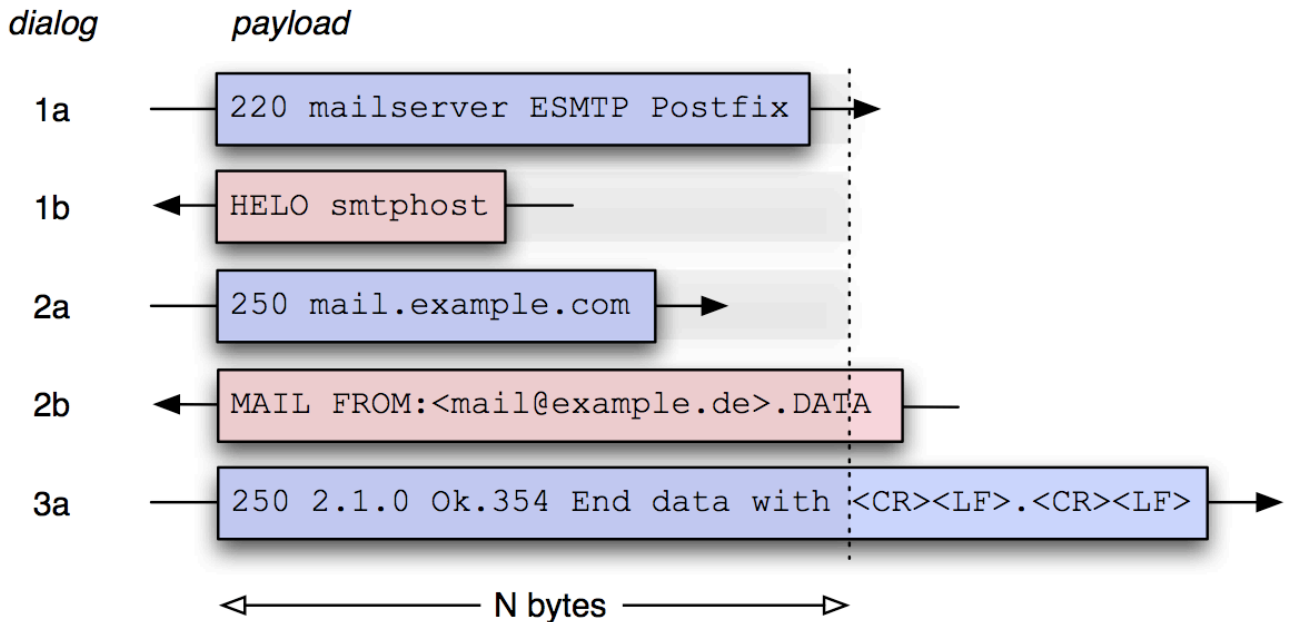
Dialog-based Communication

- Capturing payload from the start of connection is not sufficient
  - Example: HTTP pipelining



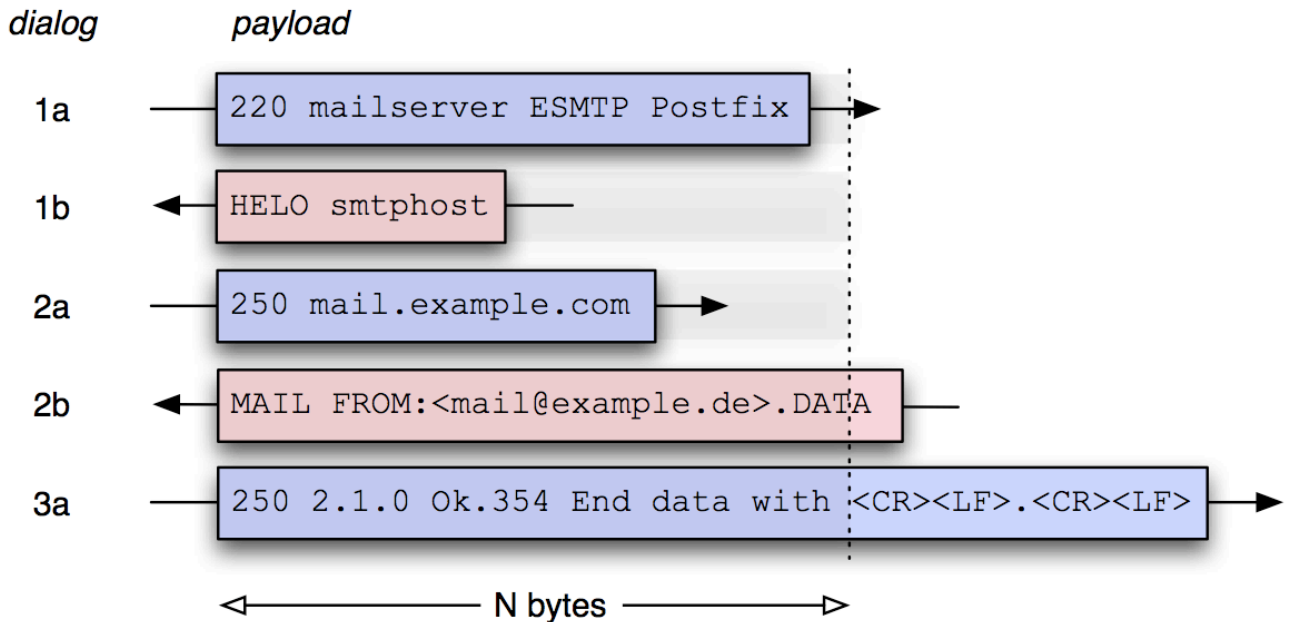
- Make use of typical request-response pattern in protocols!

Dialog-based Payload Aggregation



- Capture “dialogs” between communication endpoints
  - Use communication direction for selecting payload
  - On each direction change, start recording *n* bytes of payload

Dialog-based Payload Aggregation

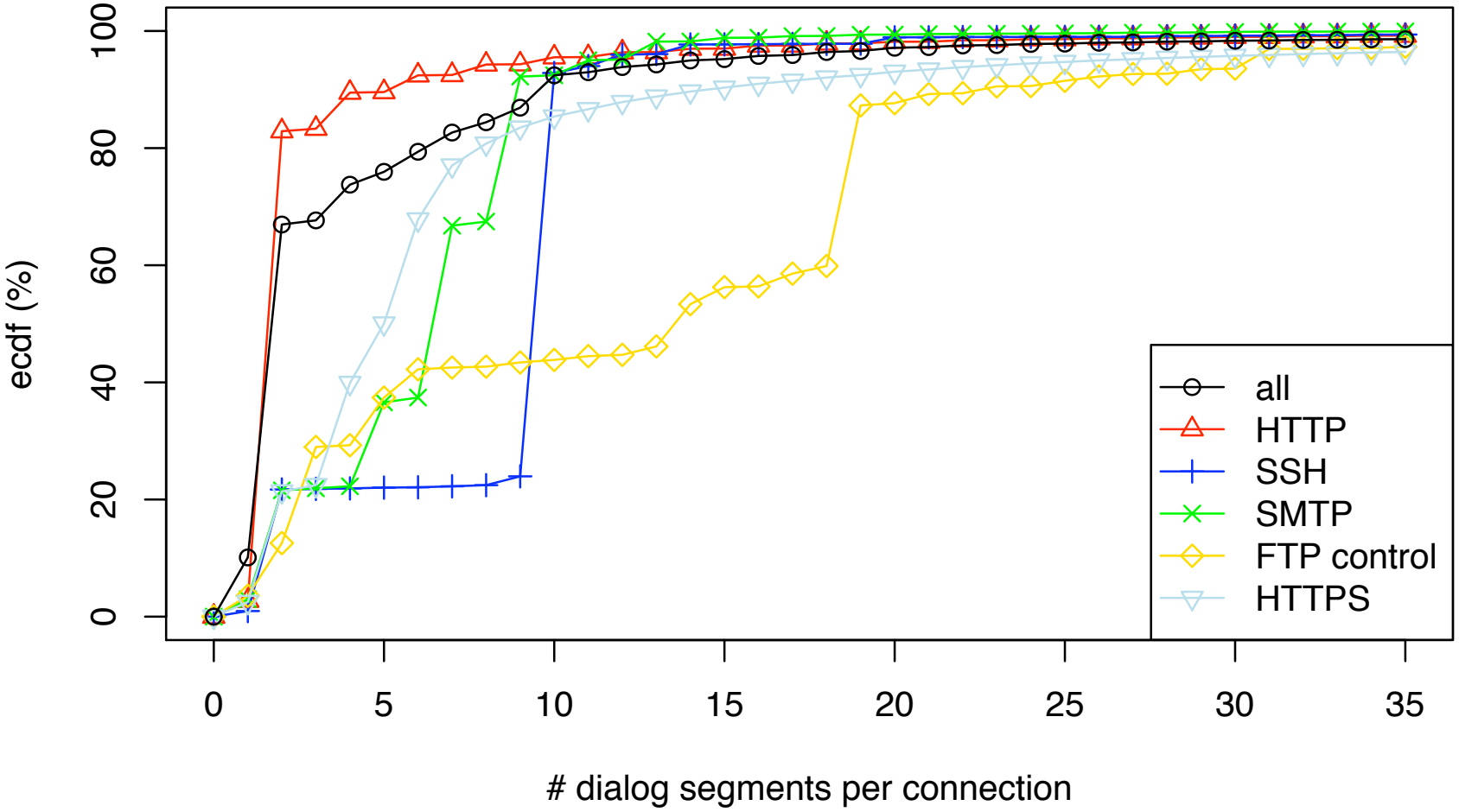


- Application layer analysis is not needed, transport layer contains enough information
  - TCP: sequence numbers
  - UDP: packet order

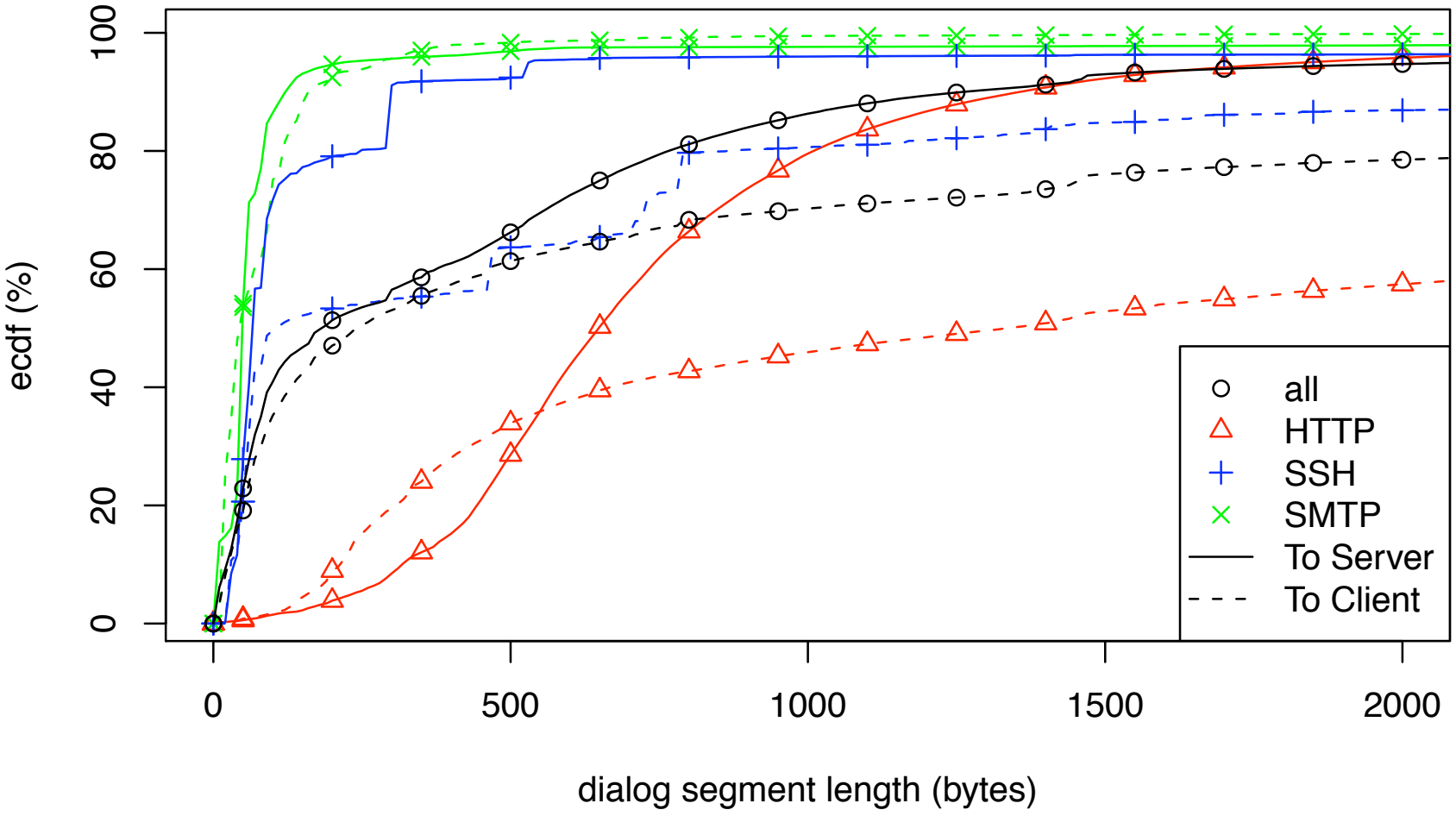
Evaluation

- With live network trace from a University
  - 8 10min packet traces per day over period of 3 months
  - 16.8 TiB of data
  - Anonymized
- Used three rule sets for Snort
  - Excluded rules that did not match payload for patterns
  - Sourcefire (SF), 5600 rules
  - EmergingThreats (ET), 9400 rules
  - BotHunter (BH), 2500 rules
- Collected events from 858 rules
  - Filtered all rules with <10 events
  - Analyzed 526 rules

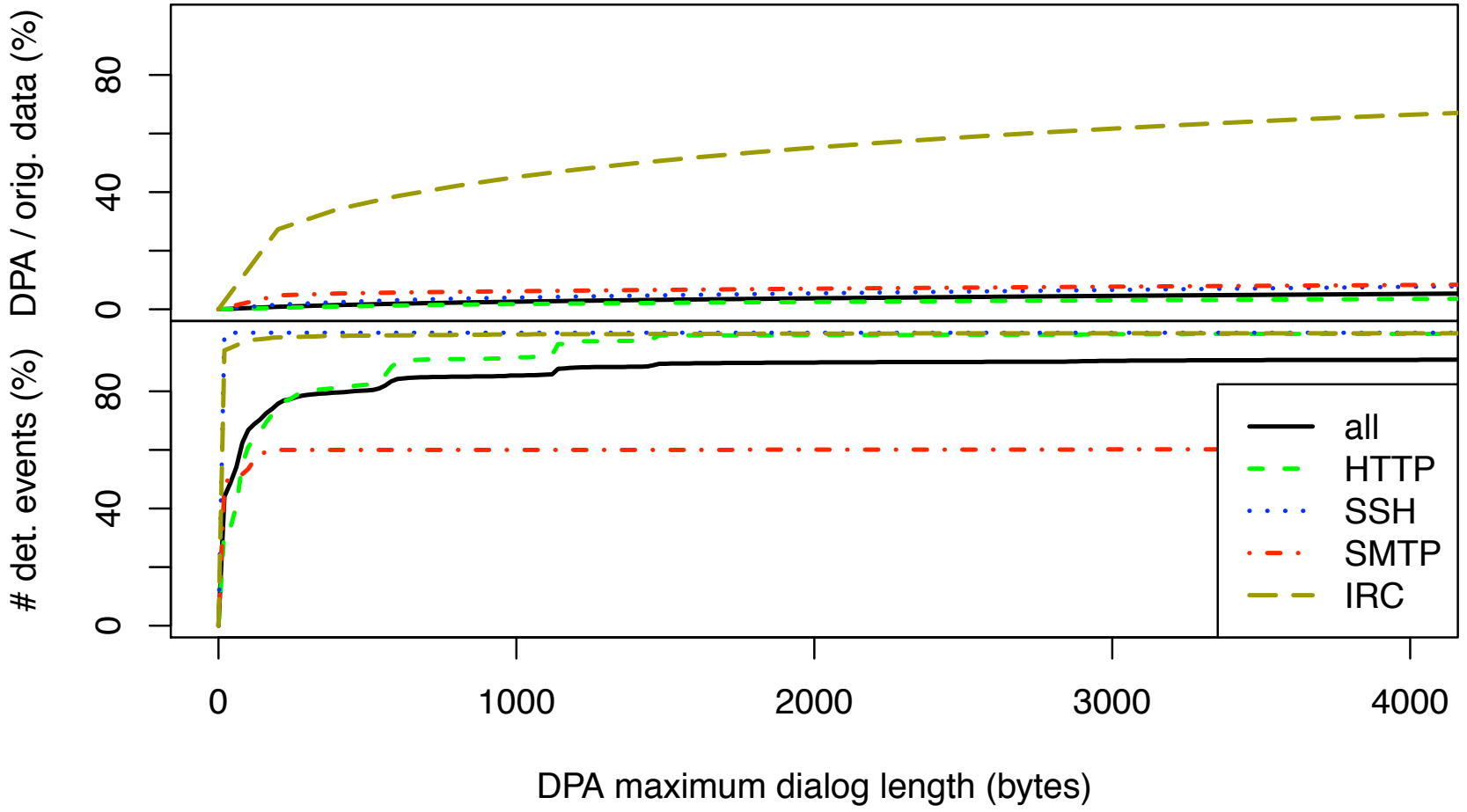
Dialog Segments 1



Dialog Segments 2

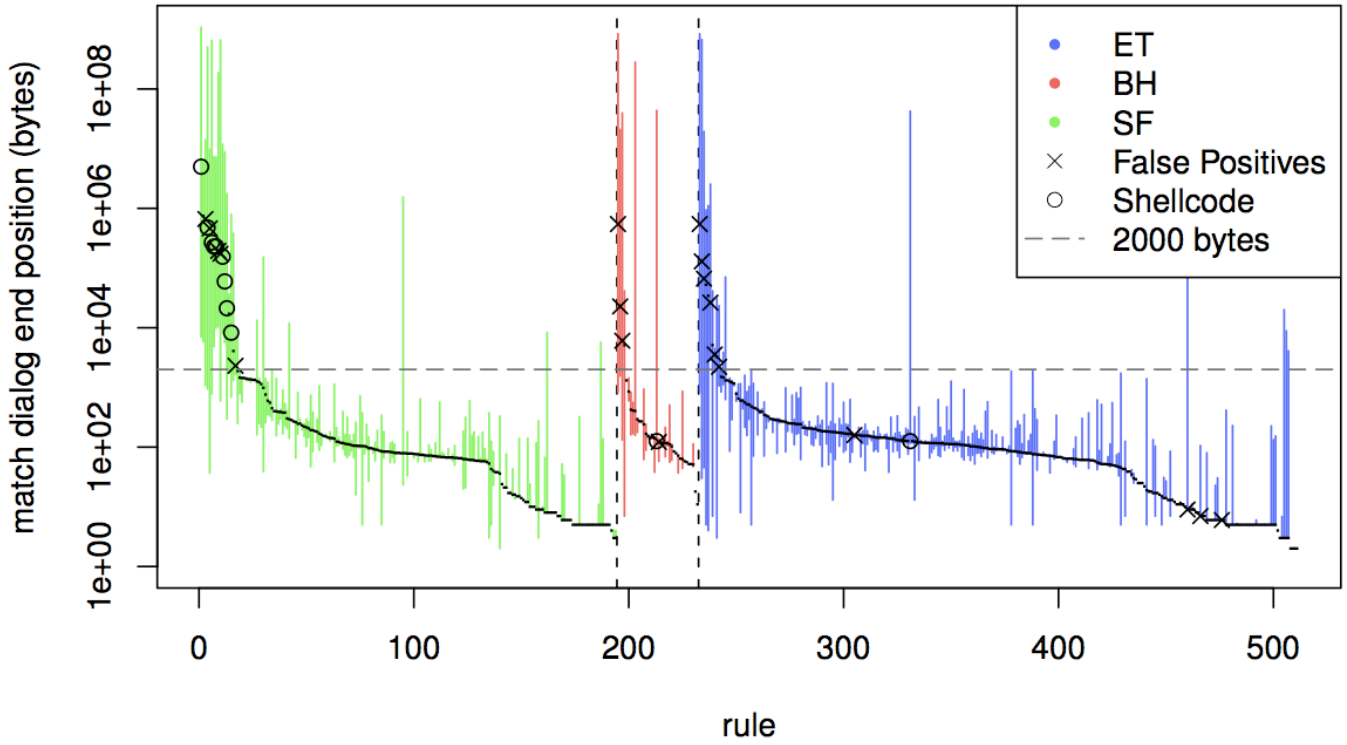


DPA Data Reduction



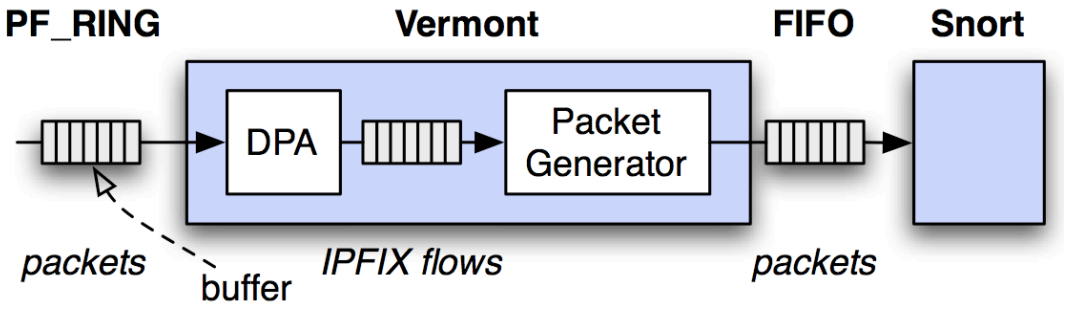
DPA Detection Evaluation

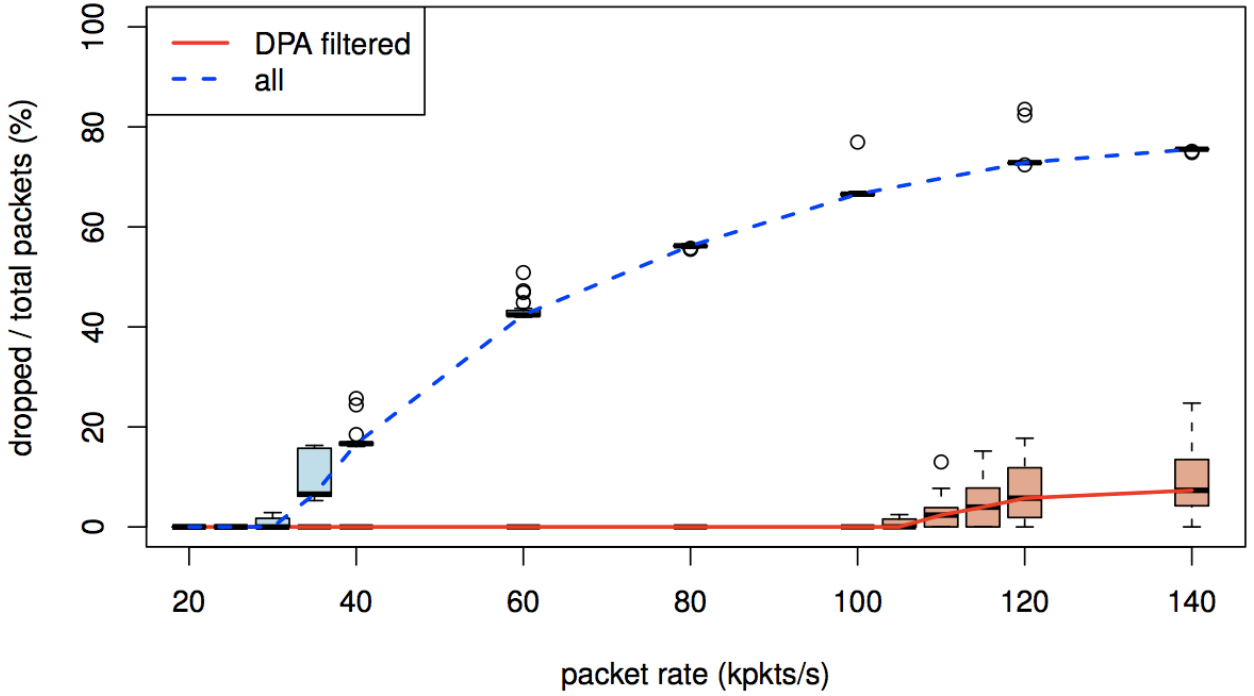
- IDS signature match position relative to start of dialog segment



- Only 1/20 of network data was analyzed by IDS
- 9 out of 10 events were still detected!

DPA Performance





Conclusion

- Introduced Dialog-based Payload Aggregation (DPA)
  - Works out-of-the-box with popular IDSs!
- Results with 2000 byte boundary:
  - 96% of traffic was filtered out
  - 90% of events were detected
  - Problematic events: Shellcode, False-positives
- Future work:
  - Add new match position restriction to signatures which is relative to start of dialog segments
  - Use for forensic analysis
  - Combine DPA with other intrusion detection methodologies

The End

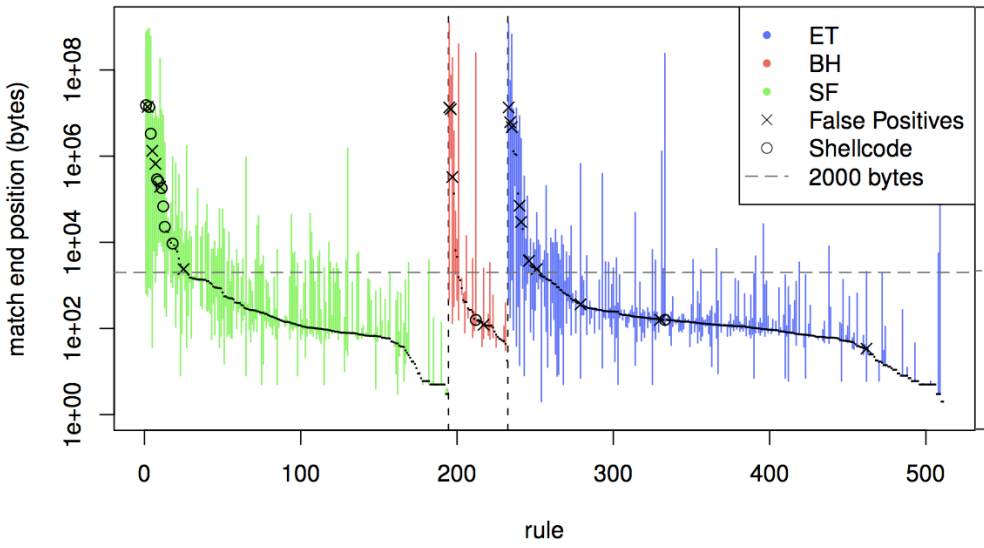
Thanks for your attention!

Questions?

DPA Detection Evaluation 2

- Comparison:

relative to connection start



relative to dialog segment start