Detecting Unknown Network Attacks using Language Models

Konrad Rieck and Pavel Laskov
DIMVA 2006, July 13/14
Berlin, Germany

Fraunhofer Institut
Rechnerarchitektur
und Softwaretechnik

The zero-day problem

- How to distinguish *normal* from *unknown*?

```
GET /dimva06/john/martin.html
Accept: */*
Accept-Language: en
Host: www
Connection: keep-alive
```

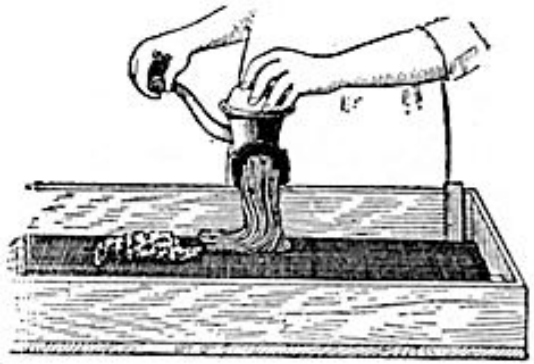
```
GET /scripts/..%%35c../..%%35c../..%%35c../..%%35c
    %%35c../winnt/system32/cmd.exe?/c+dir+c:\ HTTP/1.0
Host: www
Connection: close
```

- Cast intrusion detection into linguistic problem
  - Utilization of machine learning instruments

N-gram models

Connection payload

```
get□/index.html
```



```
g   e
t   □   /
i   n
…
```

Bytes

```
ge  et
t□  □/  /i
in  nd
…
```

2-grams

```
get  et□
t□/  □/i  /in
ind  nde
…
```

3-grams

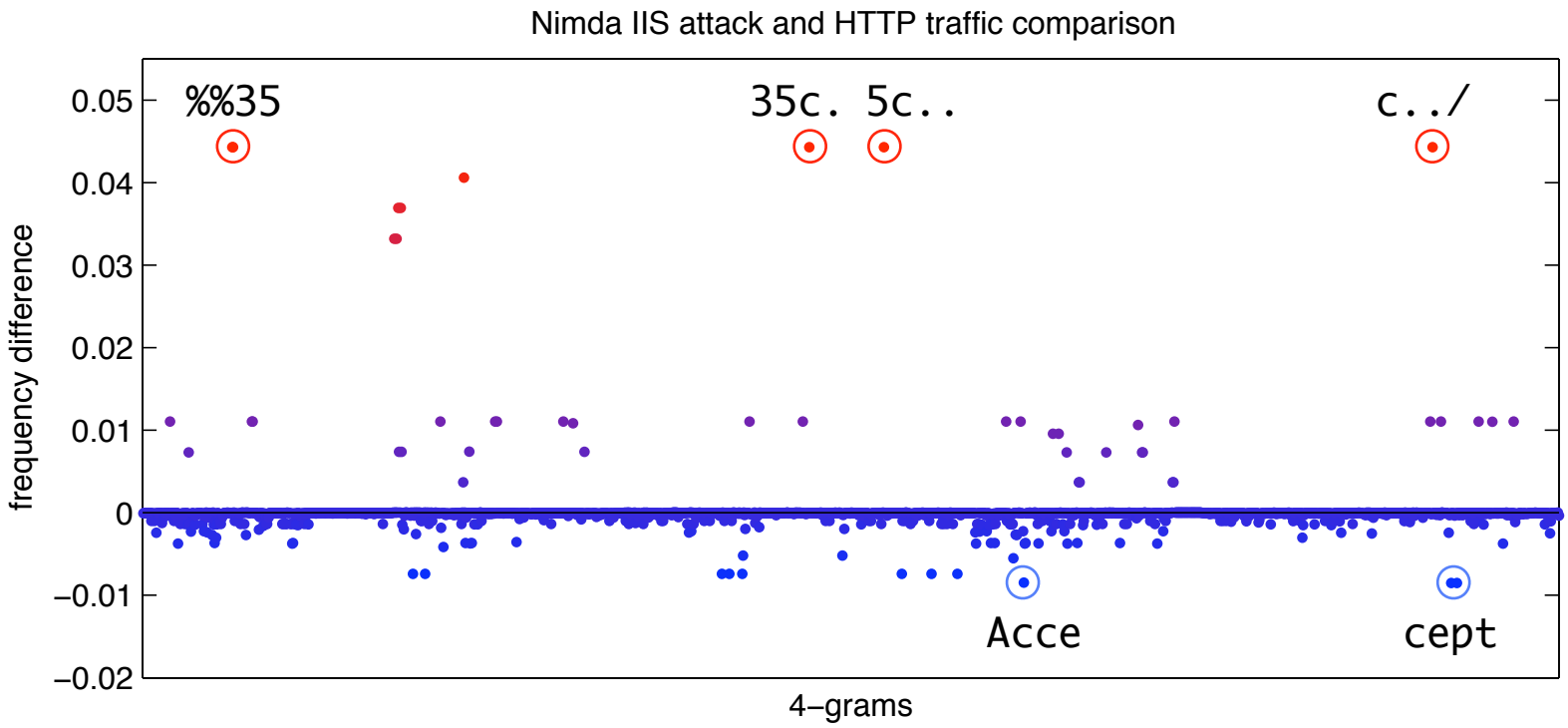
…

n-grams

N-grams in attacks

```
GET /scripts/..%%35c../..%%35c../..%%35c../..%%35c
    %%35c../winnt/system32/cmd.exe?/c+dir+c:\ HTTP/1.0
```

Frequency differences to 4-grams in normal HTTP



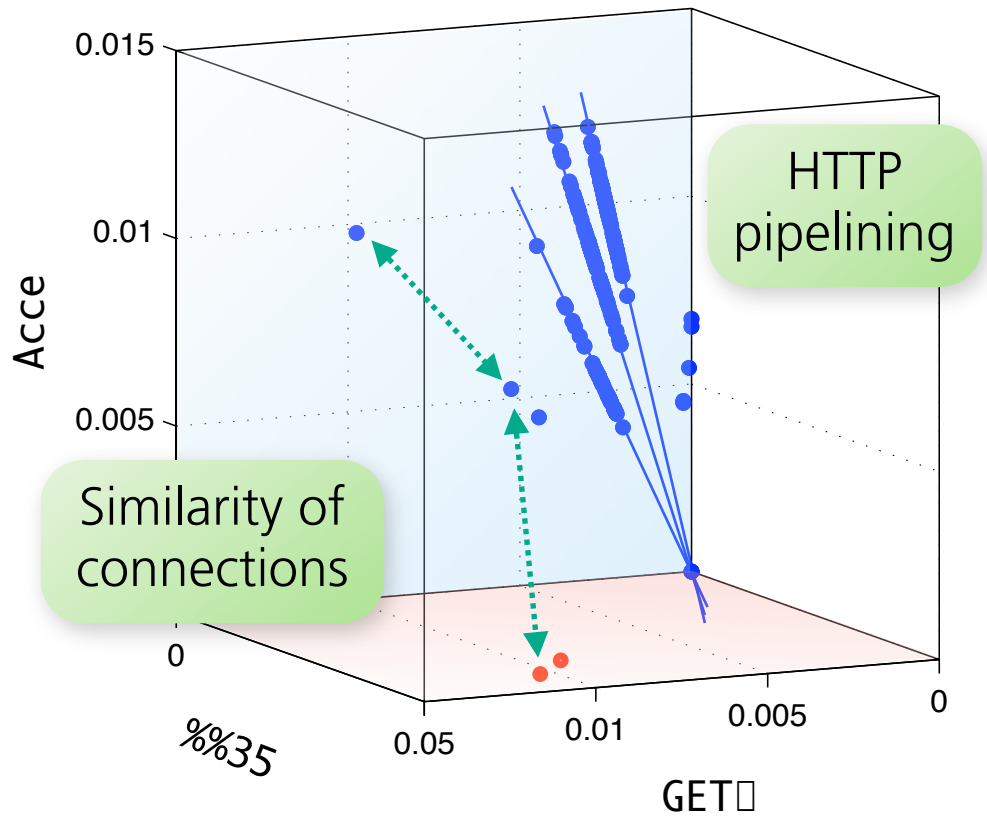
Geometric representation

- A simple example

GET□

Acce

%%35



- Huge feature space
  - 256^n dimensions
- Geometric representation of connections

Similarity measures

- Distances, kernel functions, ... e.g.
  - Manhattan $\sum_{w \in L} |\phi_w(x) - \phi_w(y)|$
  - Minkowski $\sqrt[k]{\sum_{w \in L} |\phi_w(x) - \phi_w(y)|^k}$

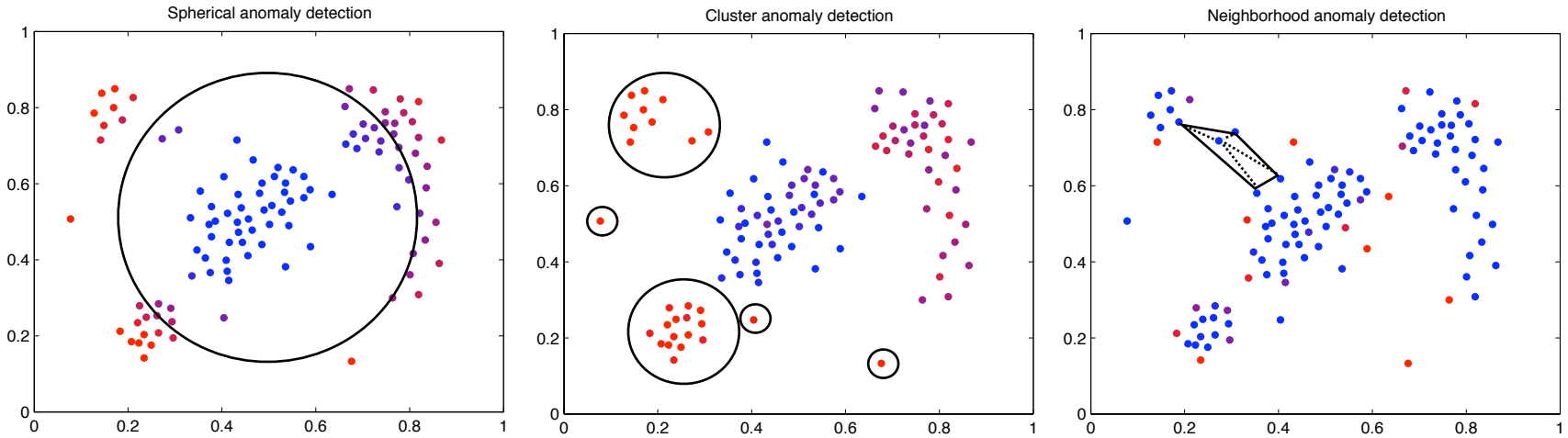
$x, y \in \{0, \ldots, 255\}^*, L = \{0, \ldots, 255\}^n$

$\phi_w(x) =$ frequency of w in sequence x

- Efficient computation not trivial
  - Sparse representation of n-gram frequencies
  - Linear-time algorithms (cf. DIMVA 2006 paper)

Anomaly detection

- Detection of outliers in feature space
  - Exploration of geometry between connections
  - No training phase - no labels required
- Anomaly detection (AD) methods
  - e.g. Spherical AD, Cluster AD, Neighborhood AD



Experiments

- Open questions
  - *Do n-gram models capture semantics sufficient for detection of unknown attacks?*
  - *Can anomaly detection reliably operate at low false-positive rates?*
  - *How does this approach compare to classical signature-based intrusion detection?*

Evaluation data

- **PESIM 2005 data set**
  - Real network traffic to servers at our laboratory
    - *HTTP* Reverse proxies of web sites
    - *FTP* Local file sharing, e.g. photos, media
    - *SMTP* Retransmission flavored with spam
  - Attacks injected by pentest expert (e.g. metasploit)
- **DARPA 1999 data set as reference**
- **Statistical preprocessing**
  - Extraction of 30 independent samples comprising 1000 incoming connection payloads per protocol

Method comparison

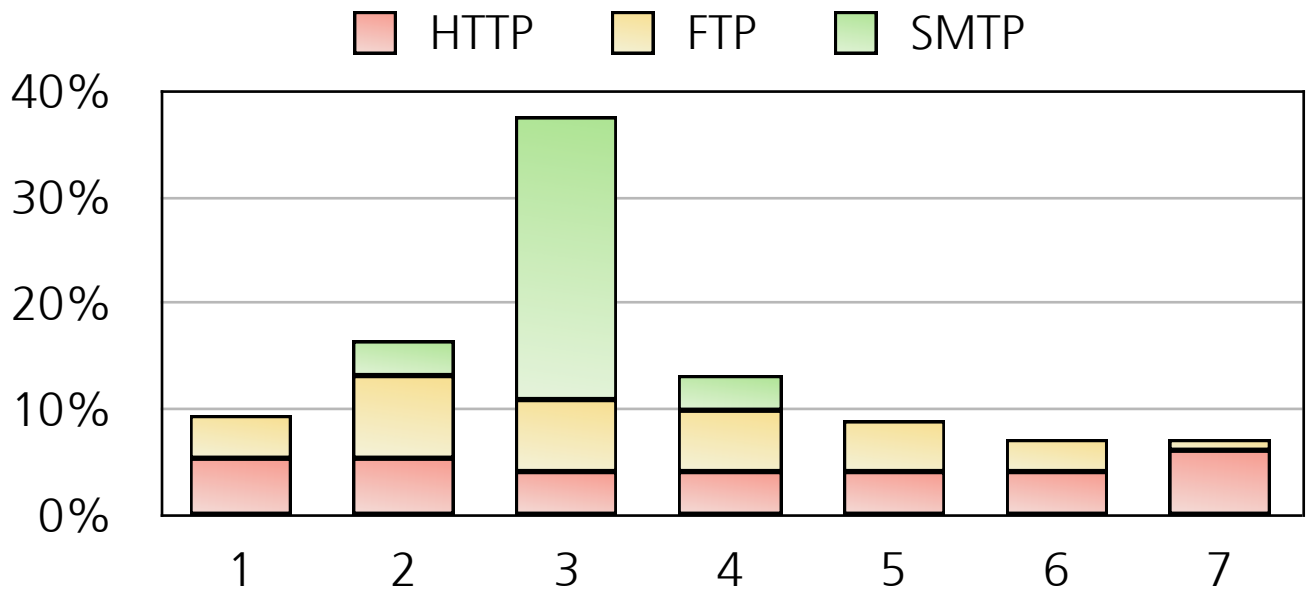
- Comparison of anomaly detection methods
  - Criteria: $AUC_{0.01}$ - Area under ROC within [0, 0.01]
  - Results averaged over n-gram lengths [1,7]

Protocol	Best method	$AUC_{0.01}$
HTTP	Spherical *(qsSVM)*	0.781
FTP	Neighborhood *(Zeta)*	0.746
SMTP	Cluster *(Single-linkage)*	0.756

Bottom line: Different protocols require different anomaly detection methods

N-gram lengths

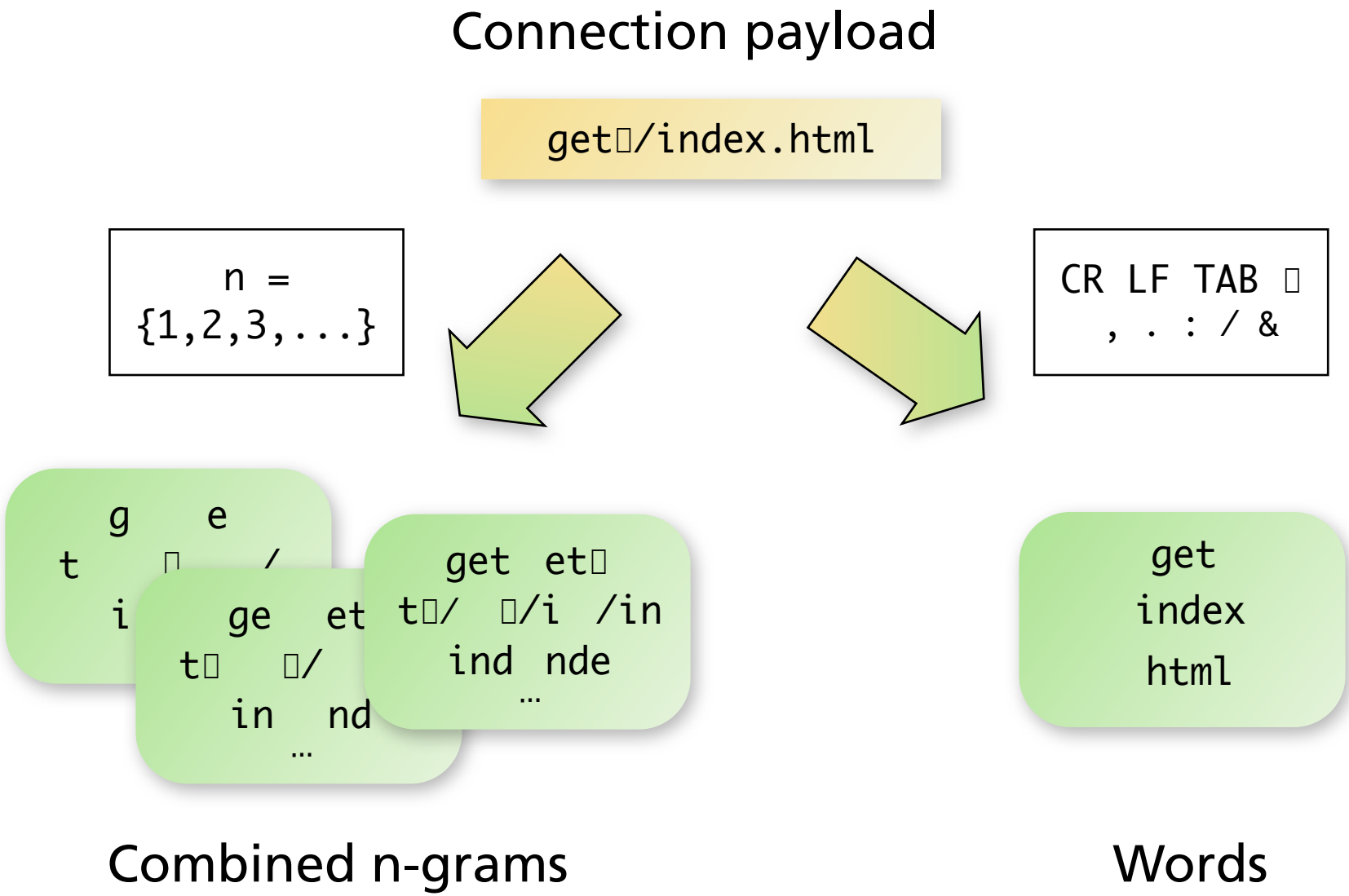
- How does one choose the optimal n-gram length?



Optimal n-gram length per attack

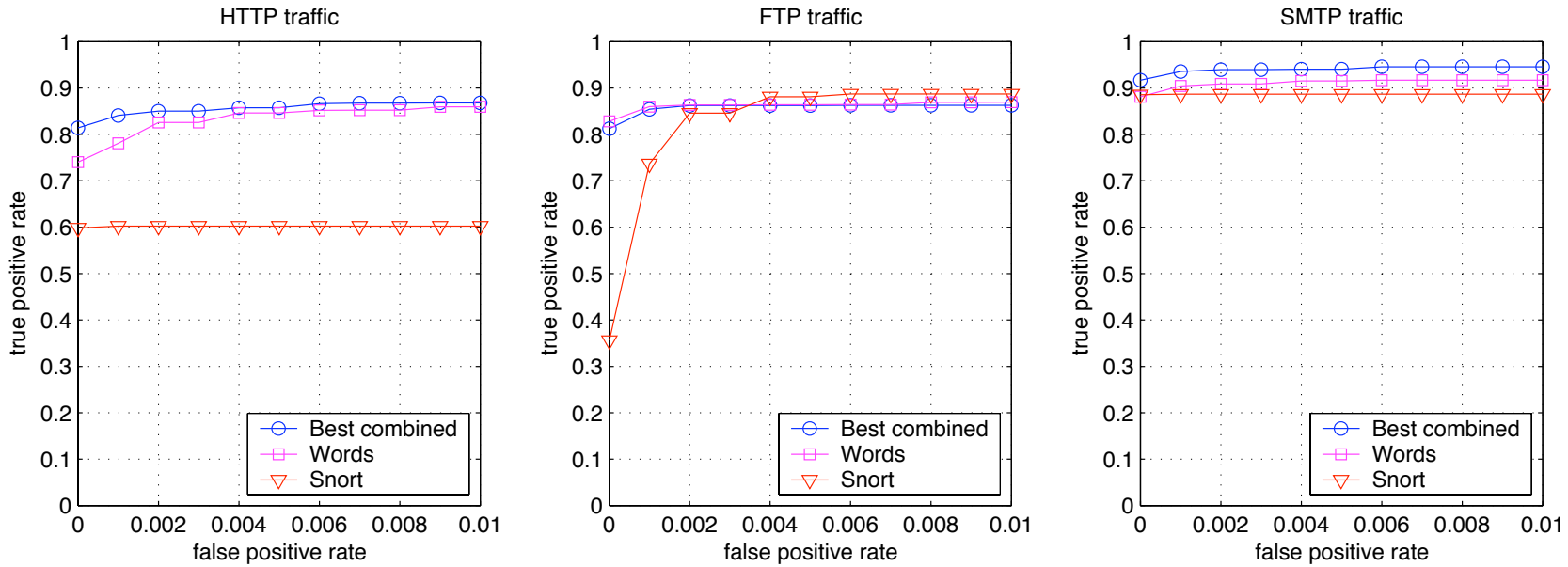
- **No single *n* fits all:** variable-length models required

Variable-length models



Comparison with Snort

- Language models vs. Snort
  - *Combined n-gram (1-7)* and *word models*
  - Snort: Version 2.4.2 with default rules



Conclusions and outlook

- **Language models for intrusion detection**
  - Characteristic patterns in normal traffic and attacks
  - Unsupervised nomaly detection with high accuracy
  - Detection of ~80% unknown network attacks

- **Future perspective**
  - From in vitro to in vivo: *real-time application*
  - Language models as prototypes for signatures?

Outwit language models

- Approaches
  - *Red herring*
    Denial-of-service with random traffic patterns
  - *Creeping poisoning*
    Careful subversion of normal traffic model
  - *Mimicry attacks*
    Adaption of attacks to mimicry normal traffic
- Conclusions
  - (1) Worse for signature-based intrusion detection
  - (2,3) Requires profound insider knowlegde

Questions?