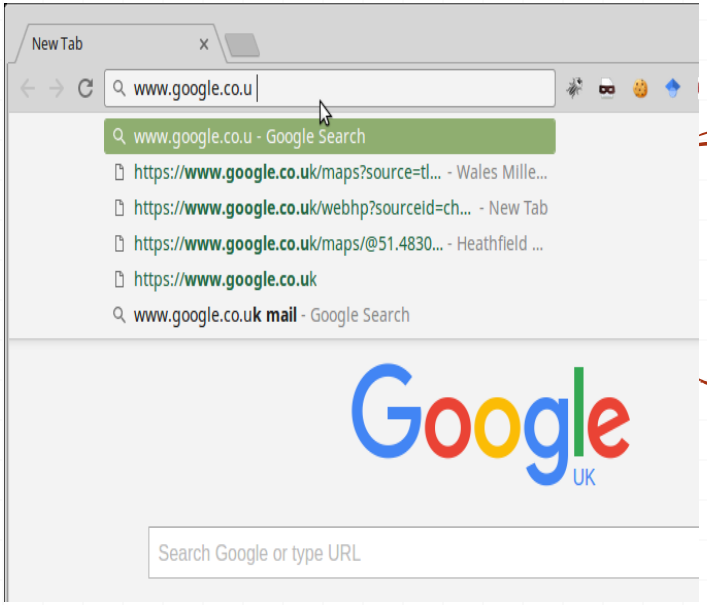
Unsupervised Approach for Detecting Low-Rate Attacks on Network Traffic with Autoencoder

Baskoro Adi Pratomo
Pete Burnap
George Theodorakopoulos

What network traffic looks like

What you see on your monitor



What your computer actually sends and receives

```
GET / HTTP/1.1
Host: www.google.co.uk
Connection: keep-alive
Upgrade-Insecure-Requests: 1
User-Agent: Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/
54.0.2840.71 Safari/537.36
X-Chrome-UMA-Enabled: 1
X-Client-Data: CIq2yQEIpbbJAQipncoB
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Accept-Encoding: gzip, deflate, sdch
Accept-Language: en-US,en;q=0.8,id;q=0.6
Cookie:
```

Censored

```
HTTP/1.1 302 Found
Location: https://www.google.co.uk/?gws_rd=ssl
Cache-Control: private
Content-Type: text/html; charset=UTF-8
Date: Fri, 12 May 2017 13:05:43 GMT
Server: gws
Content-Length: 233
X-XSS-Protection: 1; mode=block
X-Frame-Options: SAMEORIGIN

<HTML><HEAD><meta http-equiv="content-type" content="text/html;charset=utf-8">
<TITLE>302 Moved</TITLE></HEAD><BODY>
<H1>302 Moved</H1>
The document has moved
<A HREF="https://www.google.co.uk/?gws_rd=ssl">here</A>.
</BODY></HTML>
```

Low-Rate Attacks

- Exploit
  - Piece of code that takes advantage of bugs/vulnerabilites in order to cause unintended behaviour
  - i.e. Heartbleed, SQL Injection, etc
- Backdoor
  - Software which is planted by attacker to have another entrance to the system

Dst	Host name	HTTP request
108.61.210.244	geospotrima.com	GET / HTTP/1.1
108.61.210.244	geospotrima.com	GET /19-0.gif HTTP/1.1
108.61.165.65	tracetoolbox.ga	GET /BFIOCUsDSwRFCA.html HTTP/1.1
108.61.165.65	tracetoolbox.ga	GET /VktHTABQDgxFC0oCTwYEBQQCDQkMAANPAQQZAwQAFwgDBh0PA0tXWwc HTTP/1.1
108.61.165.65	tracetoolbox.ga	GET /VVpbVUsGVgwARQRPAksBAwMCAAEJAwUDSwYDHwUGDBcHBgYZCARNAklvUXR1cnt4T0hA HTTP/1.1
108.61.165.65	tracetoolbox.ga	GET /VVpbVUsGVgwARQRPAksBAwMCAAEJAwUDSwYDHwUGDBcHBgYZCARNAkl7UmxDYG55dXhdSQs HTTP/1.1
108.61.210.244	geospotrima.com	GET /favicon.ico HTTP/1.1
74.125.129.103	www.google.com	GET / HTTP/1.1
74.125.129.103	www.google.com	GET / HTTP/1.1

Nuclear Exploit Kit

```
POST /shellme.aspx HTTP/1.1
Cache-Control: no-cache
X-Forwarded-For: 81.47.81.45
Referer: http://192.168.33.138
Content-Type: application/x-www-form-urlencoded
User-Agent: Mozilla/4.0 (compatible; MSIE 6.0; windows NT 5.1)
Host: 192.168.33.138
Content-Length: 1107
Connection: Close

Password=Response.Write("->|");var err:Exception;try{eval
(System.Text.Encoding.GetEncoding(65001).GetString(System.Convert.FromBase64String
("dmFyIGM9bmV3IFN5c3Rlbs5EaWFnbm9zdGljcy5Qcm9jZXNzU3RhcnRJbmZvKFN5c3Rlbs5UZXhOLkVuY29kaW5
nLkdldEVuY29kaW5nKDY1MDAxKS5HZXRTdHJpbmcoU3lzdGvtLkNvbnZlcnQuRnJvbUJhc2U2NFN0cmluZyhSZXF1
ZxN0Lkl0Zw1bInoxIlOpKSk7dmFyIGU9bmV3IFN5c3Rlbs5EaWFnbm9zdGljcy5Qcm9jZXNzKCk7dmFyIG91dDpTe
XN0ZW0uSU8uU3RyZWFtUmVhZGVyLEVJolN5c3Rlbs5JTy5TdHJlYW1SZWFkZXI7Yy5Vc2VTaGvsbEv4ZWn1dGU9Zm
Fsc2U7Yy5SZWRpcmVjdFN0YW5kYXJkT3V0CHV0PXRydWU7Yy5SZWRpcmVjdFN0YW5kYXJkRXJyb3I9dHJ1ZTtlLlN
0YXJOSw5mbz1jo2MuQXJndW1lbnRzPSIvYyAiK1N5c3Rlbs5UZXhOLkvuY29kaW5nLkdldEvuY29kaW5nKDY1MDAx
KS5HZXRTdHJpbmcoU3lzdGvtLkNvbnZlcnQuRnJvbUJhc2U2NFN0cmluZyhSZXF1ZXN0Lkl0Zw1bInoyIlOpKTtlL
lN0YXJOKCk7b3vOPWUuU3RhbmRhcmRPdXRwdXQ7RUk9Z5TdGFuZGFyZEVycm9yO2UuQ2xvc2UoKTtSZXNwb25zZS
5Xcml0ZShvdXQuUmVhZFRvRW5kKCkrRUkuUmVhZFRvRW5kKCkpow%3D%3D")),"unsafe");}catch(err)
{Response.Write("ERROR:// "%2Berr.message);}Response.Write("|<-");Response.End
();&z1=Y21k&z2=Y2QgL2QgImM6XGluZXRwdWJcd3d3cm9vdFwiJndob2FtaSZly2hvIFtTXSZjZCZly2hvIFtFXQ
%3D%3D HTTP/1.1 200 OK
Connection: close
Date: Thu, 06 Jun 2013 18:48:51 GMT
Server: Microsoft-IIS/6.0
X-AspNet-Version: 1.1.4322
Cache-Control: private
Content-Type: text/html; charset=utf-8
Content-Length: 66

->|nt authority\network service
[S]
C:\Inetpub\wwwroot
[E]
|<-
```

Traffic from attacker

Response from the victim

Why doing all of these?

- Statistical properties of traffic obtained from packet header may not be enough
- Classification is most likely to perform badly on imbalanced dataset
- Difficulty of having a labelled dataset for training purposes in real world

Previous works

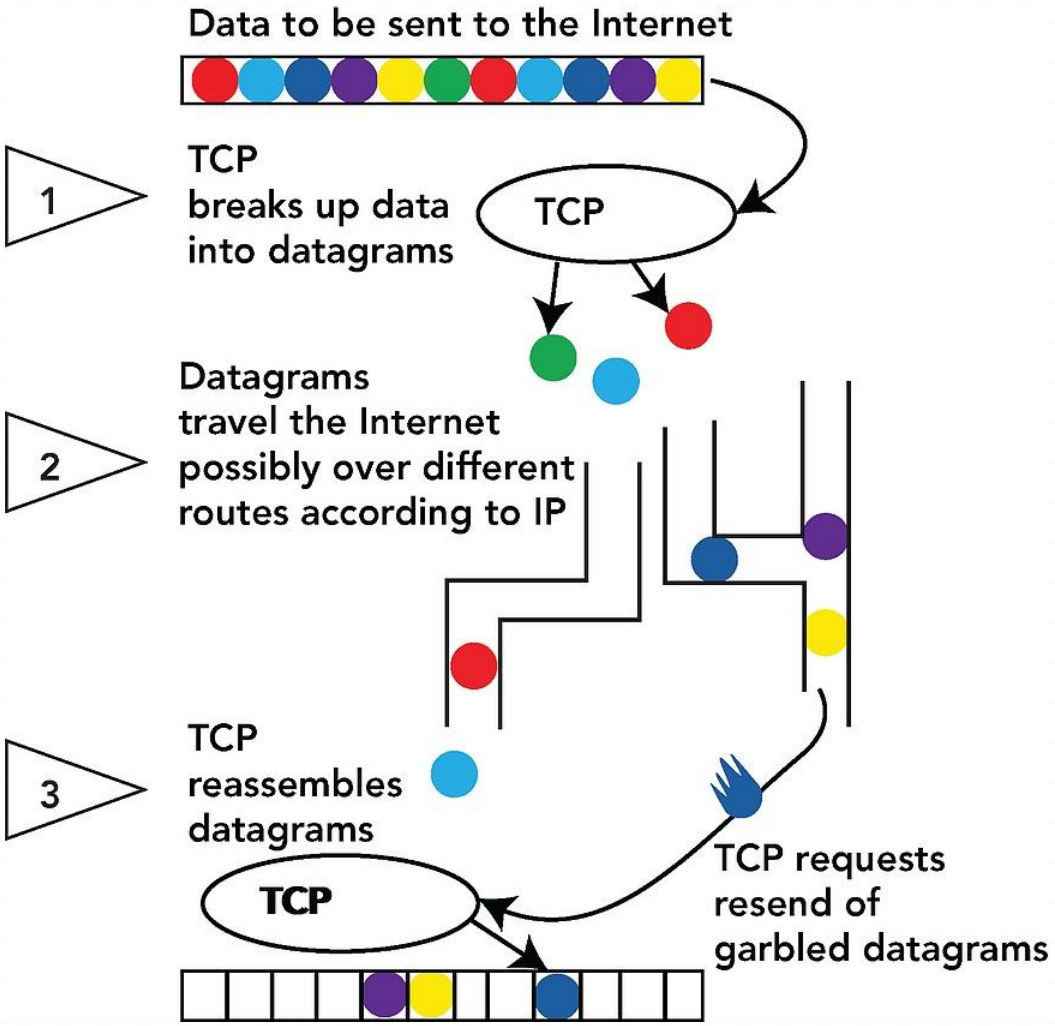
- PAYL
- McPAD
- OCPAD

- Anagram
- Poseidon
- HMM PAYL
- RePIDS

Methodology

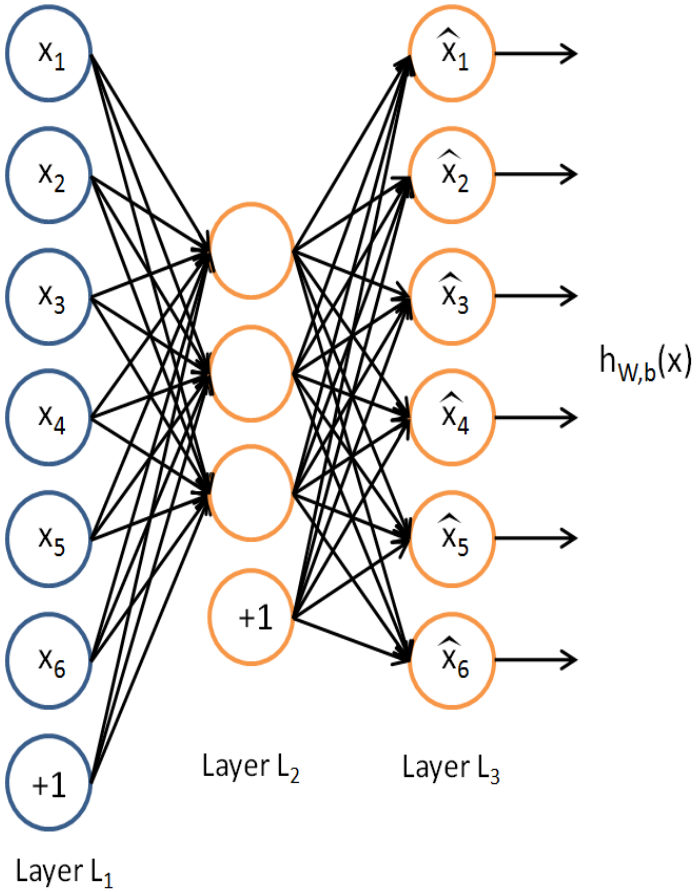
The complicated part

TCP Stream Reassembly

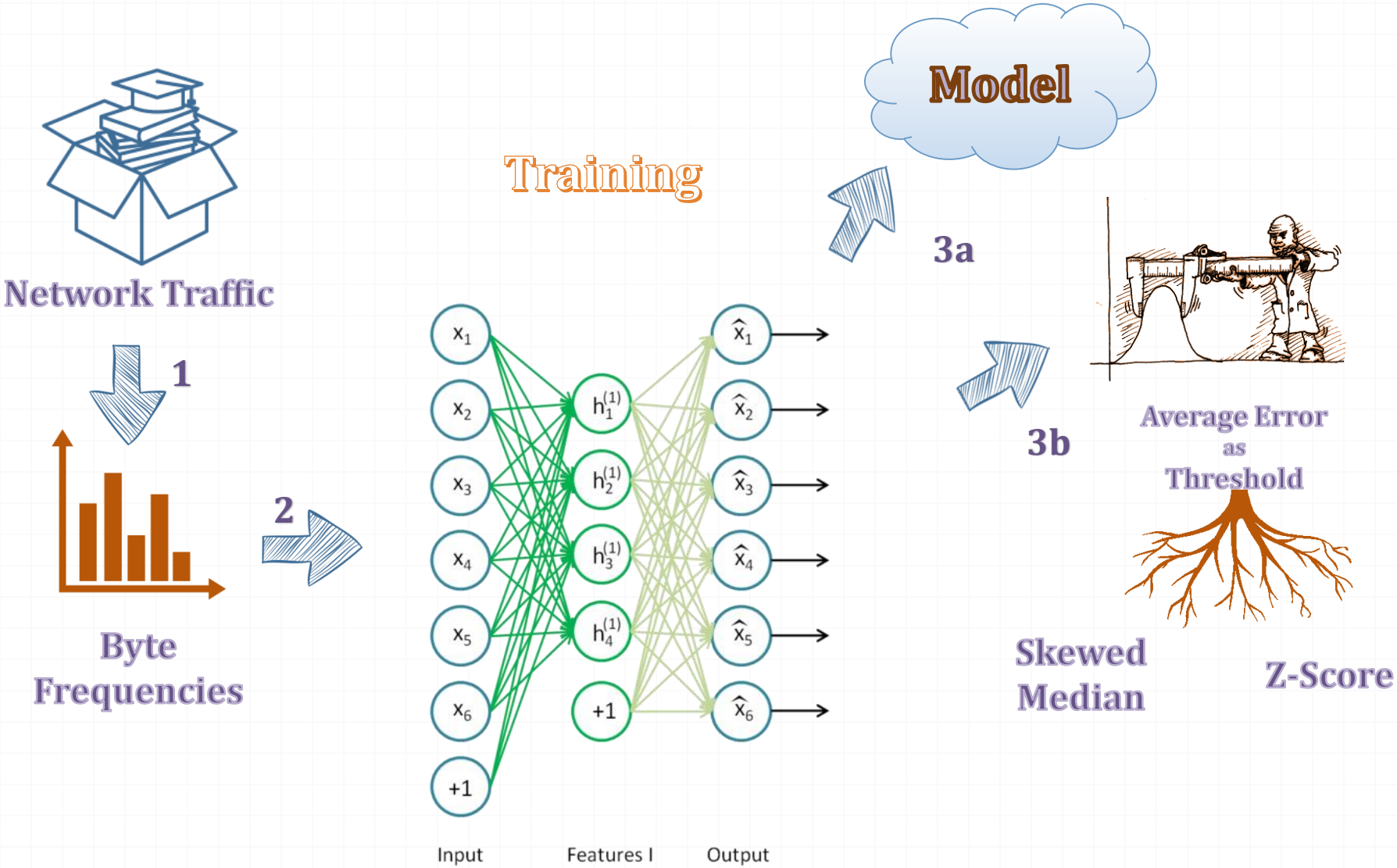


What is an Autoencoder?

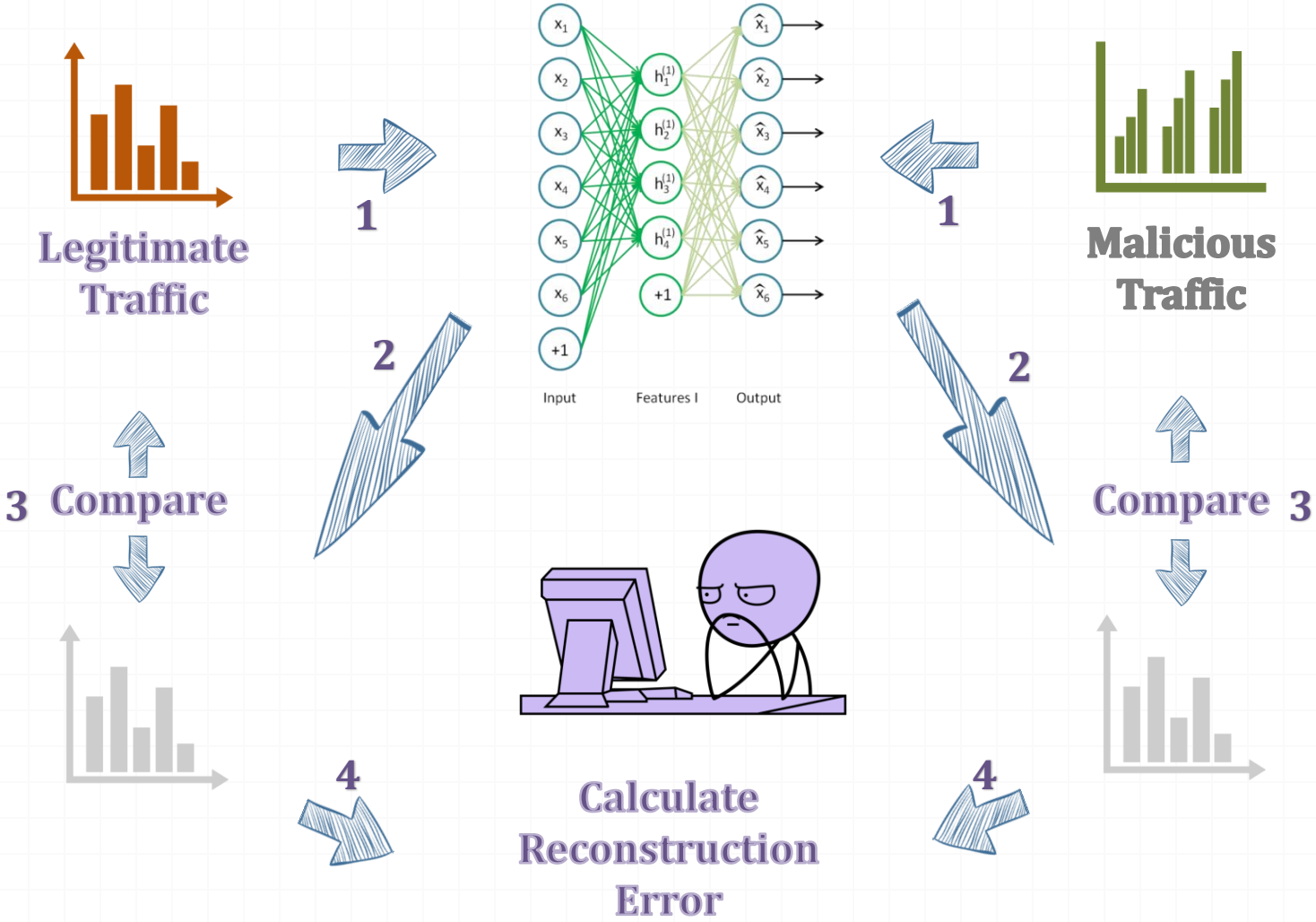
- Autoencoder neural network
  - An unsupervised learning algorithm that applies backpropagation, setting the target values to be equal to the inputs
- May be used as outlier detection
  - By calculating the total of reconstruction errors



How it learns

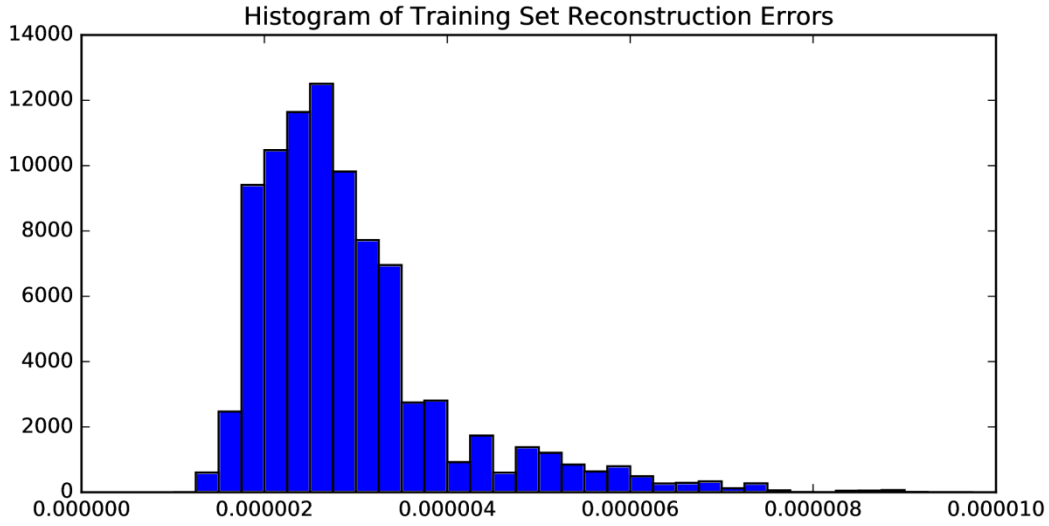


How it detects malicious traffic



Defining Threshold

- Usual approach:
  - Mean + (2 or 3) * Standard Deviation
- The distribution of Reconstruction Errors of training set is skewed



Defining Threshold (2)

- Skewed Median
  - MedCouple (MC): $$MC(F) = \underset{x_i < m < x_i}{\text{median}}\, h(x_i, x_j)$$

$$h(x_i, x_j) = \frac{(x_j - m) - (m - x_i)}{x_j - x_i}$$

  - Threshold: $$Q_3 + e^{3*MC} * 1.5 * IQR, \text{ if } MC \geq 0$$

$$Q_3 + e^{4*MC} * 1.5 * IQR, \text{ if } MC < 0$$

- Median Absolute Deviation (Z-Score method)

$$MAD = \text{median}(\{(|e_i - \text{median}(E)|), e_i \in E, 0 < i < n\})$$

$$z = \frac{0.6745 * (|e - \text{median}(R)|)}{MAD}$$

Experimental Results (Dataset)

- UNSW NB15 Dataset
  - Training set:
    - 93406 HTTP connections
    - 41395 SMTP connections
  - Testing set:
    - 131842 legitimate traffic (HTTP & SMTP)
    - 17041 malicious HTTP connections
    - 4631 malicious SMTP connections

Experimental Results (Comparison)

Method	DR (%)	FPR (%)
Autoencoder (T_{IQR} Threshold)	68.91	0.99
Autoencoder (Z-Score Threshold)	100	8.01
OCPAD (1-gram)	19.88	0.00
OCPAD (3-gram) (HTTP only)	29.08	8.85
PAYL	36.1	0.05

Experimental Results (The Effect of Hidden Layers)

Number of Neurons in Each Hidden Layer	Threshold Method	DR (%)	FPR (%)
200	T_{IQR}	68.91	0.99
200-100-200	T_{IQR}	68.21	0.63
200-100-50-100-200	T_{IQR}	69.21	3.16
200	Z-Score	100	8.11
200-100-200	Z-Score	100	8.01
200-100-50-100-200	Z-Score	100	8.1

Conclusion

- The proposed method can detect outliers (malicious traffic)
- Skewed median is more cautious, resulting in lower false positive rate
- Z-score method has got better detection rate with the cost of increasing false positive

Future Works

- Combining 1D Convolutional Network and Autoencoder to capture sequence attributes of network traffic
- Trying Recurrent Neural Network as an outlier detector

Thank you

Any question?