Using Machine Learning and Word Embedding to Characterise the DDoS Landscape with **DDoS2Vec**

Ravjot Singh Samra
Marinho Barcellos

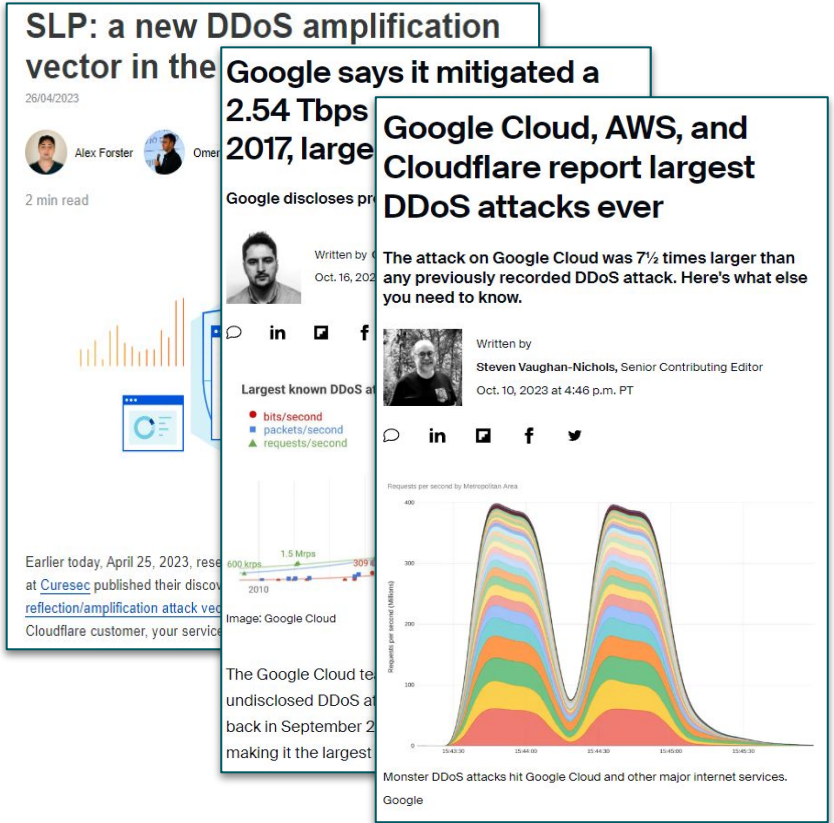
THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Volumetric Distributed Denial of Service (DDoS) Attacks

DDoS has been a plague on the Internet since the beginning

Attacks seem to be ever growing in size and impact

Attackers continuously improve their strategies to cause more damage using less resources

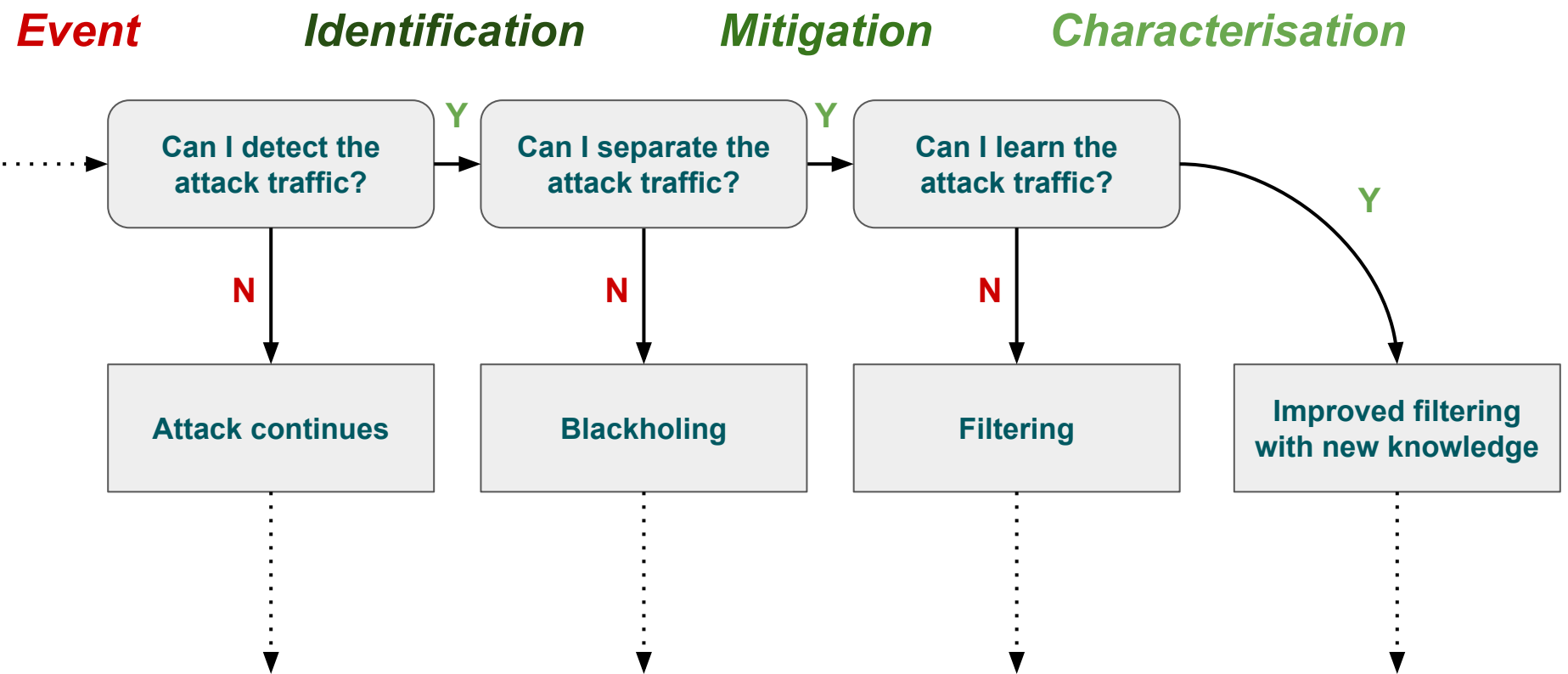


Outline

1. The meaning of DDoS attack *characterisation*
2. Handling data and labels: tiny lab networks versus the Internet
3. Leveraging natural language processing: **DDoS2Vec**
4. Longitudinal analysis on a year's worth of IXP traffic

So, what does "characterisation" mean here?

DDoS Attack Characterisation



Data Requirements

Before we start characterising DDoS attacks, *we need the following for evaluation*:

A realistic **network traffic dataset** with serious scale

and

A set of **ground truth** or **labels** describing **characteristics**

Publicly Available Datasets

Popular ones you may have come across:

- KDD Cup 1999
- DARPA Intrusion Detection Evaluation Dataset (1998, 1999)
- CAIDA UCSD DDoS Dataset (2007)
- UNSW-NB15 (2015)
- CIC-DDoS2019
- NF-UQ-NIDS (2021, combination of older datasets like UNSW-NB15)

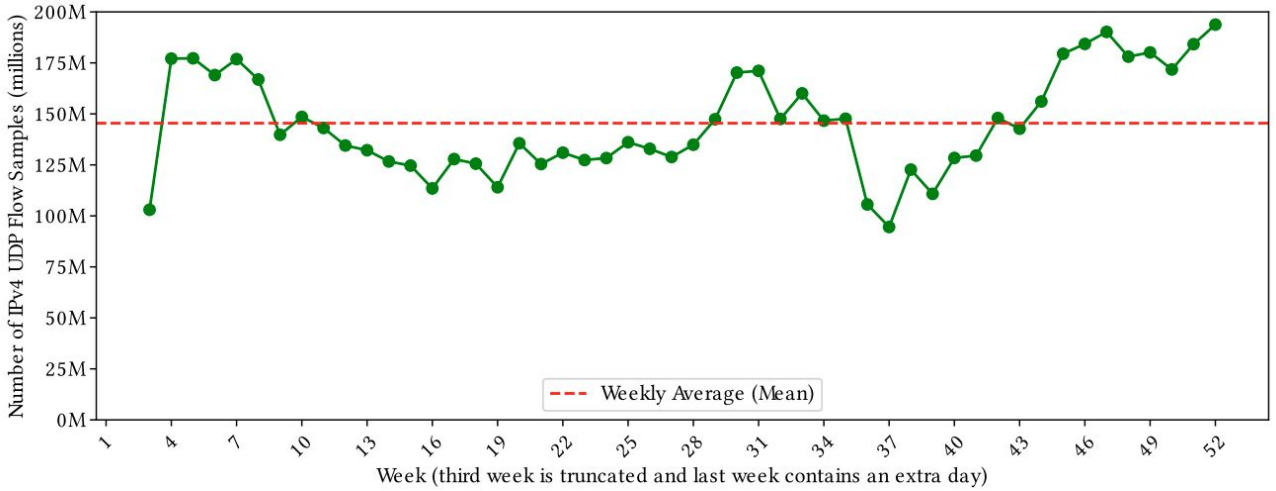
They almost always contain *two general flaws and shortcomings*:

- **Unrealistic and/or unknown attack configurations**
- **Unrealistic network environment scale**

Let's move on to a real-world **alternative**…

IXP Flow Samples

- Private **IXP** flow sample dataset from **2019** (**1:4096** sample rate)
- Medium-sized IXP with *over* **200 member networks**
- Represents *real-world traffic at Internet infrastructure scale*



Obtaining Ground Truth

- **Issue**: our IXP dataset is *unlabelled*
- The recent **IXP Scrubber** work can help us with their filtering rules artefact
- *Vast majority* of the rules are for *UDP only*, which limits our evaluation
- A filtering rule match on a flow can be considered the defining characteristic of the flow

Example Filtering Rule


```
"20d10ae9":{
 "protocol":17,
 "port_src":53,
 "port_dst":2701,
 "packet_size":"(1400,1500]",
 "confidence":1.0,
 "antecedent population":410966
},
```

Natural Language Processing (NLP)

- **NLP** has seen a recent increase in both *interest* and *research*
- Network security research has taken *advantage* of that:

  See **IP2Vec**, **DANTE**, **DarkVec**, *etc.*
- *What about applying such techniques to DDoS attack characterisation?*
- **Untested** on a realistic network traffic dataset
- **Problem**: NLP approaches require *natural language corpora*, not flows

Example Document Corpus: Visualised

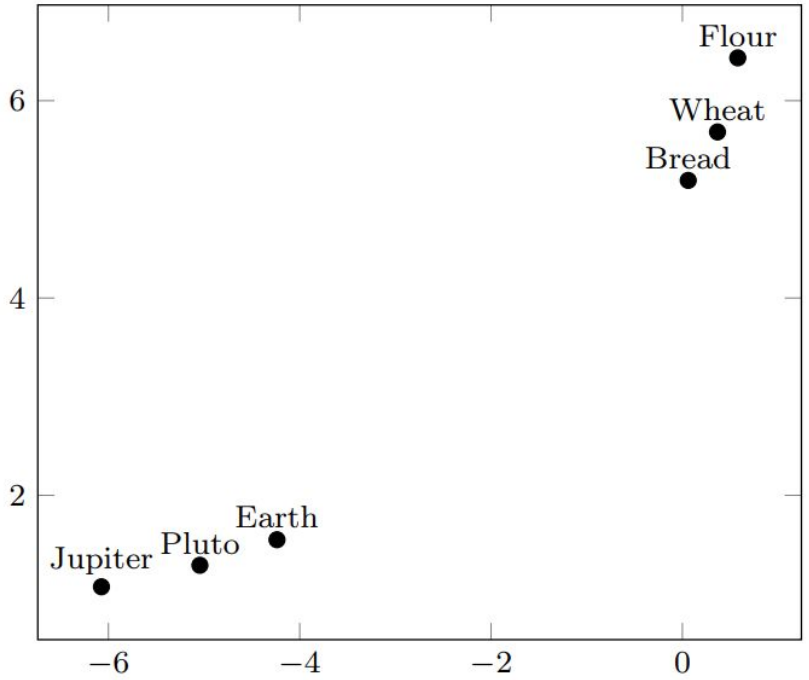
- A **corpus** is a **collection** of sentences, paragraphs, documents, *etc.*
- Previously mentioned work uses sentences; we use documents
- **Example**: a tiny document corpus with Wikipedia articles

Document Tag	**Words**					
Bread	bread	is	a	staple	food	...
Pluto	pluto	minor	planet	designation	pluto	...
Flour	flour	is	a	powder	made	...
Jupiter	jupiter	is	the	fifth	planet	...
Earth	earth	is	the	third	planet	...
Wheat	wheat	is	a	grass	widely	...

Example Document Corpus: "2Vec"

- How can we find similar articles?
- Turn **documents** *into an* **embedding: a unified vector space**
- We can use *Doc2Vec* for this

Document Tag	Vector	
Bread	0.061	5.192
Pluto	−5.044	1.291
Flour	0.579	6.434
Jupiter	−6.073	1.073
Earth	−4.238	1.550
Wheat	0.367	5.683



Flow Corpus Generation

Overall Goal

Flow Records  *Vector Space*

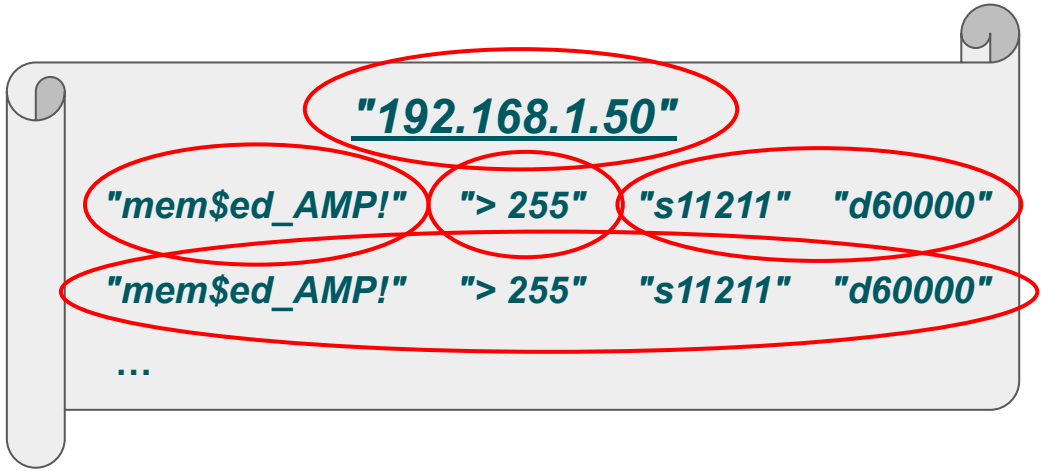
- We can **convert flow records into a document corpus** first
- The words will need to describe *flow-level behaviour and patterns*
- There is **no** standard "correct" way to do this: trial and error

Flow Corpus Generation: Example

Field	**1st Flow**
Timestamp (initial packet, UTC)	1648468800
Source IP Address	192.168.1.40
Destination IP Address	192.168.1.50
Source Port	11211
Destination Port	60000
Packets	2
Bytes	2230
Protocol	UDP

Flow Corpus Generation: Example…

Field	1st Flow
Timestamp (initial packet, UTC)	1648468800
Source IP Address	192.168.1.40
Destination IP Address	192.168.1.50
Source Port	11211
Destination Port	60000
Packets	2
Bytes	2230
Protocol	UDP



***Ready** for **input** into an NLP technique…*

NLP Techniques & Approaches

- Many NLP approaches are *compatible*

- **Word2Vec**: most relevant to prior work…

  but requires a document-to-sentence conversion

- **Doc2Vec**: essentially a document-based modification of Word2Vec

  No changes to the corpus required

- **Latent Semantic Analysis (LSA)**: a much **older** approach

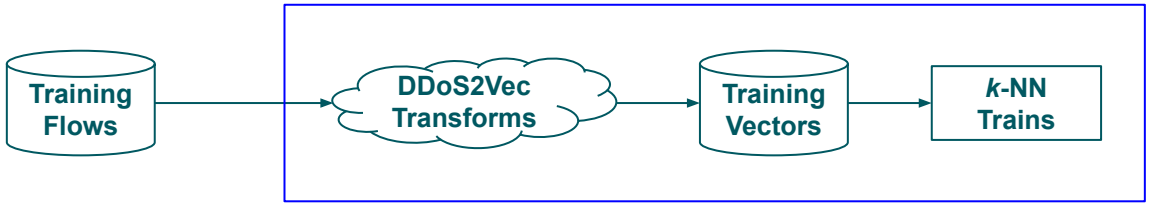
  Performs the best, *despite* its **simplicity** — a **key** part of **DDoS2Vec**

Longitudinal Analysis: Multi-Label Classification

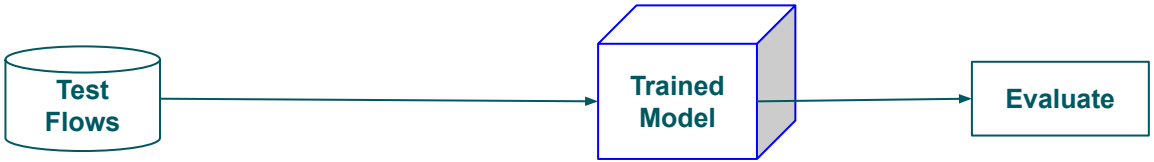
Challenge: *predict the IXP Scrubber filtering rules* that apply to traffic destined for a potential *unseen* victim IP address in each month of 2019

Classifier: Distance weighted *k*-NN (*k* = 10)

For training month (June 2019):



For every other (testing) month in 2019:



Longitudinal Analysis: Classification Performance

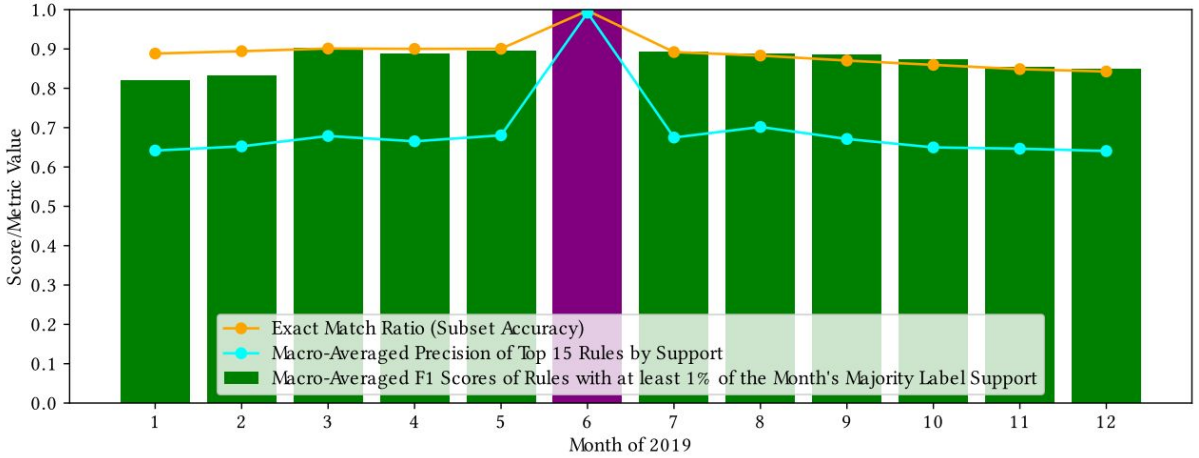


Fig. 3. Classification performance over 2019 of a DDoS2Vec embedding trained on 2019-06-01 — 2019-07-01.

- One training month **does not contain all** attack characteristics
- For classification performance: *sharp* initial drop-off, *subtle* decline

Longitudinal Analysis: Time Performance

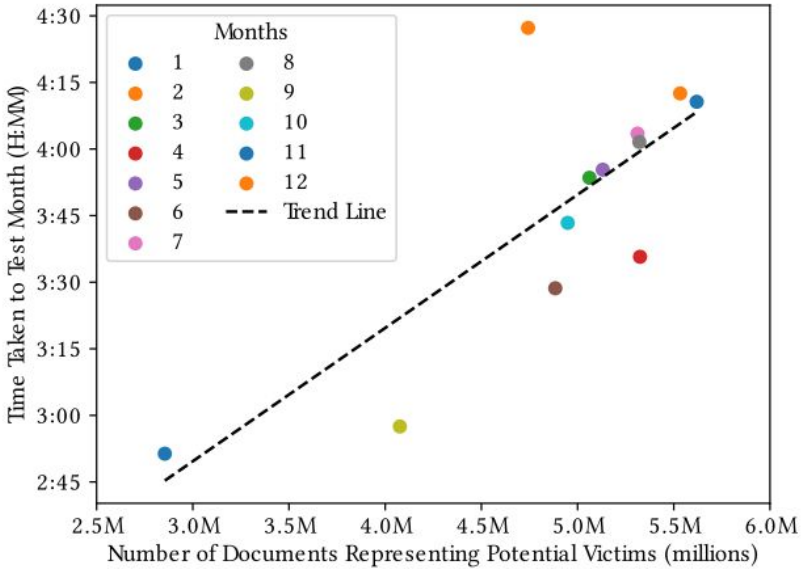


Fig. 5. Time taken for months based on corpus size.

Limitations & Future Work

- **Evaluation was held back** to UDP-based volumetric DDoS attacks
  → We require a real dataset with more labelled characteristics in general
- **Limited comparison** to other approaches
  → We are unaware of other possible multi-label classification baselines
- **Behind** the state-of-the-art in NLP — **no** deep learning or LLMs here
  → Doc2Vec, Word2Vec, LSA, *etc.* are *at minimum roughly a decade old*

Key Takeaways

- **Publicly available datasets** created in lab environments are **inadequate**
- **DDoS2Vec** *can characterise volumetric DDoS attacks in a highly novel way*
  It can do so *across time* with a reasonable performance drift
- **NLP** is a **promising** concept to leverage for DDoS attack characterisation
  We recommend experimenting with flow corpus generation, NLP techniques, *etc.*

Thanks for listening