Machine Learning for Realistic Cyber Deception

David Liebowitz

May 2022

Penten

Canberra-based cyber technology business

Australian Defence and Government focus

penten

Penten

Business units

- Secure Mobility
- Tactical Communications Security
- Applied AI

penten

Secure Mobility

Remote access to classified networks

- AltoCrypt Stik
- AltoCrypt Phone

Secure Mobility

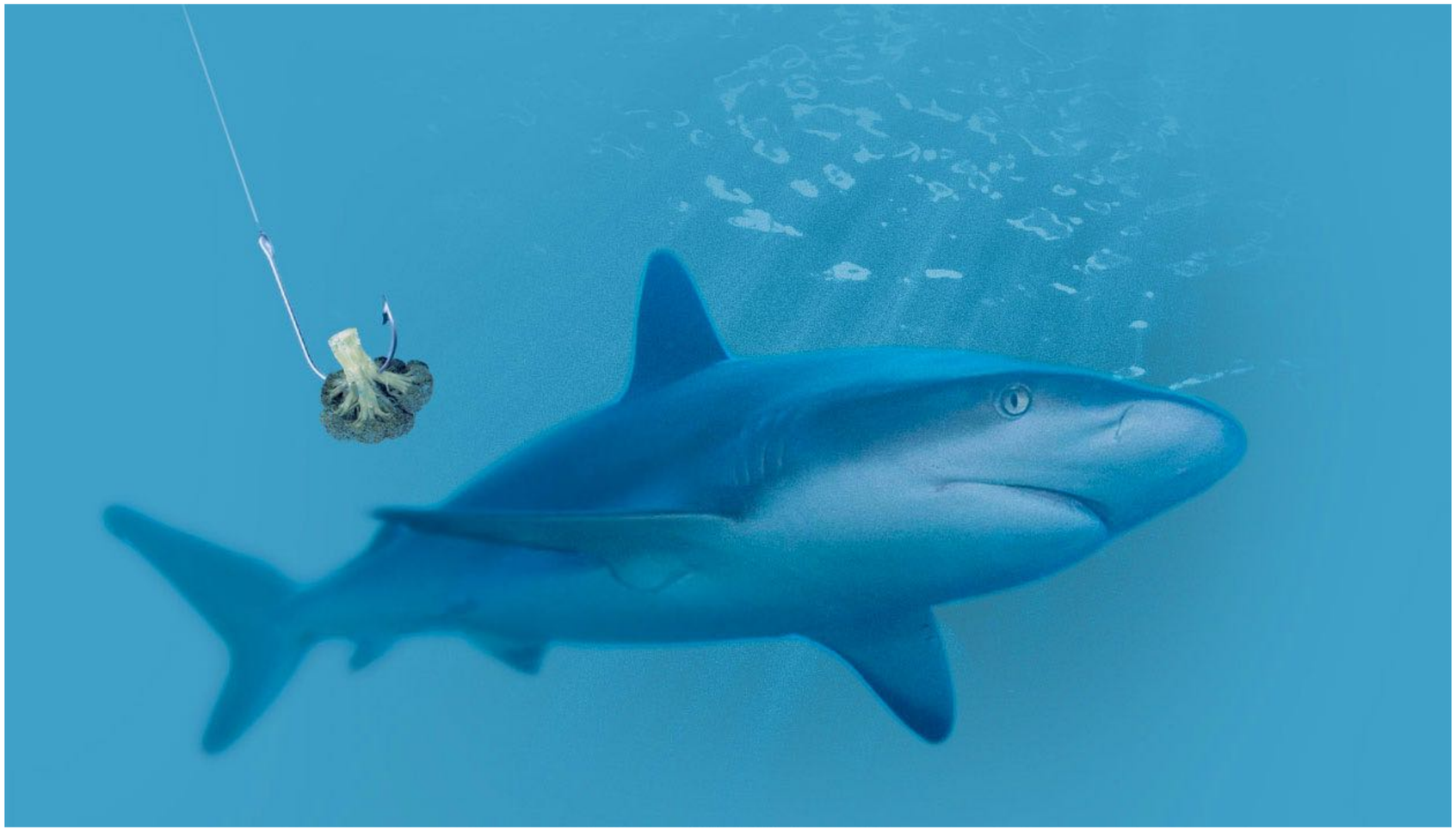
Portable acoustic suppression

- AltoCrypt pBox

Tactical Communications Security

Helix

Applied AI



Cyber deception

penten

Deception Research

CYBER SECURITY
COOPERATIVE
RESEARCH
CENTRE

UNSW
SYDNEY

Cyber Deception Research Lab *http://www.cse.unsw.edu.au/~dliebowitz/*

penten

Deception

My working definition:

Deception is a (deliberate) attempt to manipulate the beliefs of others in order to influence their behaviour.

Deception is everywhere

Deception in Nature

See: Martin Stevens, *Cheats and Deceits: How Animals and Plants Exploit and Mislead,* Oxford University Press, 2016

pentenq

Deception in Nature

Bolas spider

Image: Judy Gallagher
Flickr: https://www.flickr.com/photos/52450054@N04/24843011518/

Deception in Nature

Pitcher plants

and *batch capture*

https://en.wikipedia.org/wiki/Nepenthes_rafflesiana

Evolved to Lie?

We are thoroughgoing liars, even to ourselves. Our most prized possession— language— not only strengthens our ability to lie but greatly extends its range. We can lie about events distant in space and time, the details and meaning of the behavior of others, our innermost thoughts and desires, and so on.

- Robert Trivers

The Trickster

Warfare

All warfare is based on deception - Sun Tzu

Warfare

In wartime, truth is so precious that she should always be attended by a bodyguard of lies. - Winston Churchill

Image: ErrantX (https://commons.wikimedia.org/wiki/File:Map_of_Operation_Bodyguard_subordinate_plans.png), “Map of Operation Bodyguard subordinate plans“, https://creativecommons.org/licenses/by-sa/3.0/legalcode

Sport

Deception for Cyber Security

1. Where there is competition or conflict, deception emerges.

2. Deception is a valuable complement to existing security measures.

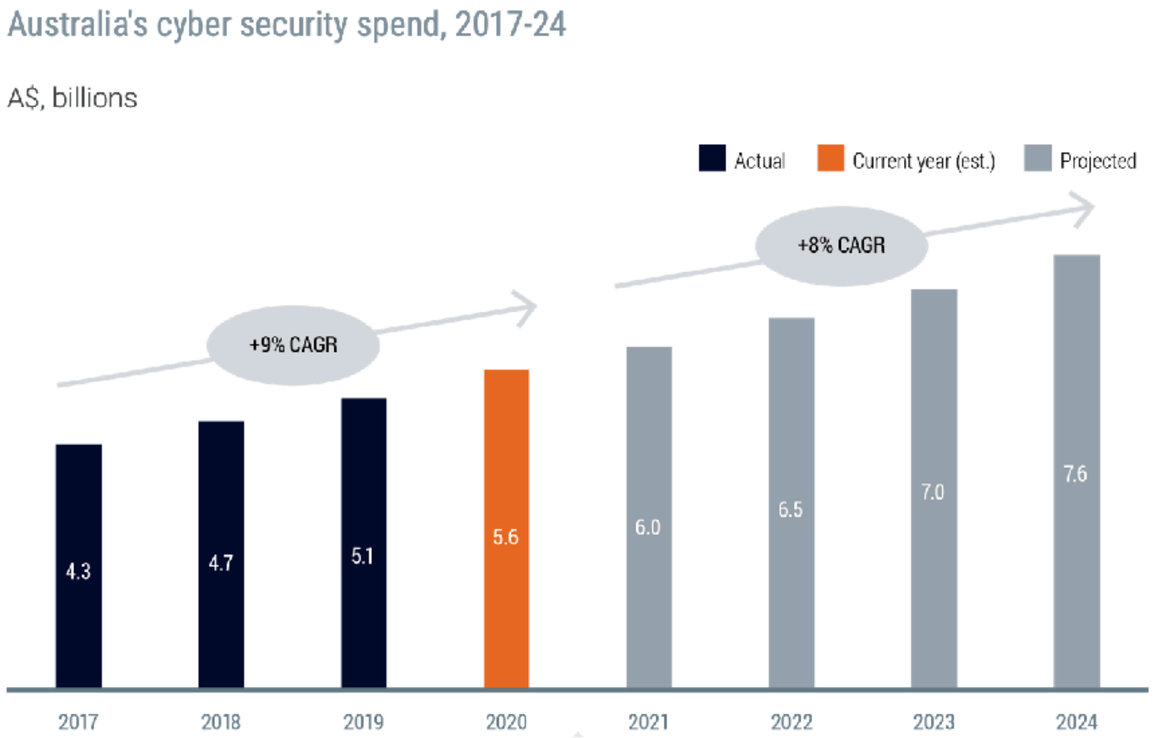
Deception for Cyber Security

Australia spent

AUD 5.6 billion

on cyber security in 2020

Source: *Australia's Cyber Security Sector Competitiveness Plan 2020,* Australian Cyber Security Growth Network



But breaches continue…

pentenq

Deception for Cyber Security

The average time to identify and contain a data breach is

287 days.

Source: *IBM/Ponemon Cost of a Data Breach Report 2021*

Deception Can Help

Honeypots

Cyber Deception usually means honeypots

A honeypot is security resource whose value lies in being probed, attacked or compromised.

- Lance Spitzner

Honeypots

- Put fake stuff on the network
- Legitimate users have no reason to interact with it
- Any interaction is suspicious
  => breach discovery

- Honeypots are selective interaction filters

Types of Honeypots

- Can be mail/database/web servers
- Can also be **honeytokens** - honeypots that aren't computers:
  - Documents and files - honeyfiles
  - Database records - honeydata
  - Credentials - honeycredentials
  - …
  - Honey + Anything =  Deception

pentenq

Tracers

- Designed to be stolen:
  - Create unique tokens
  - Deploy then on the network
  - Track/monitor *or* beacon
  - Can discover and track data theft

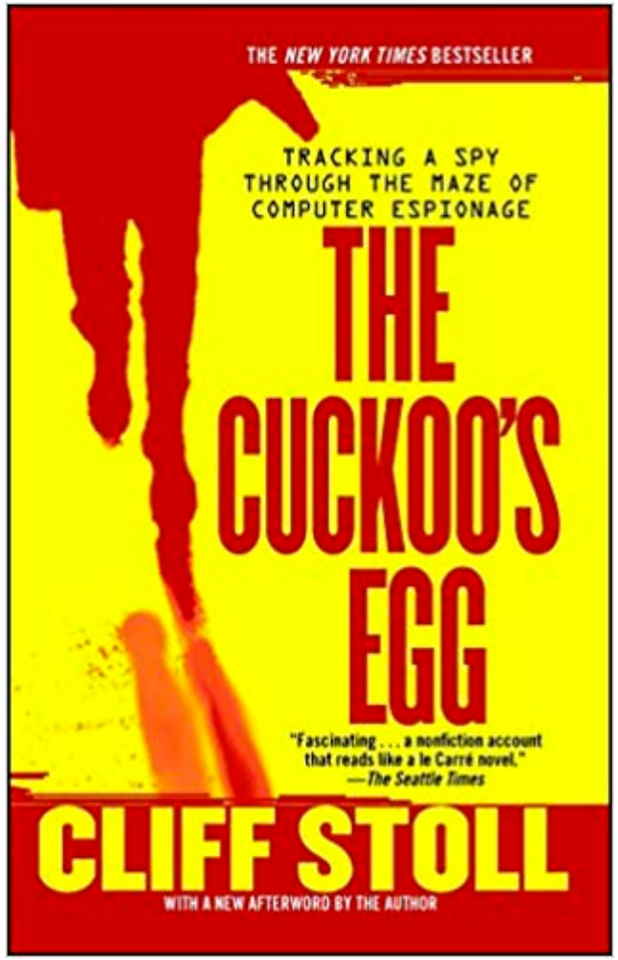
Other Benefits

- Adversary intent
- Tactics, Tools and Procedures (TTPs)
- Delay and frustrate

- Deceptions must be realistic

The Cuckoos' Egg

- Clifford Stoll, Lawrence Berkeley National Laboratory in California
- Used deceptive documents
- Didn't shut the intruder out



Now Add Machine Learning

- Realistic deceptions work better
- But tedious and expensive to create
- If only we could learn from real examples…
- ML/AI

Some ML Tools

- Language models
- Graphs
- Temporal Point Processes

Language Models

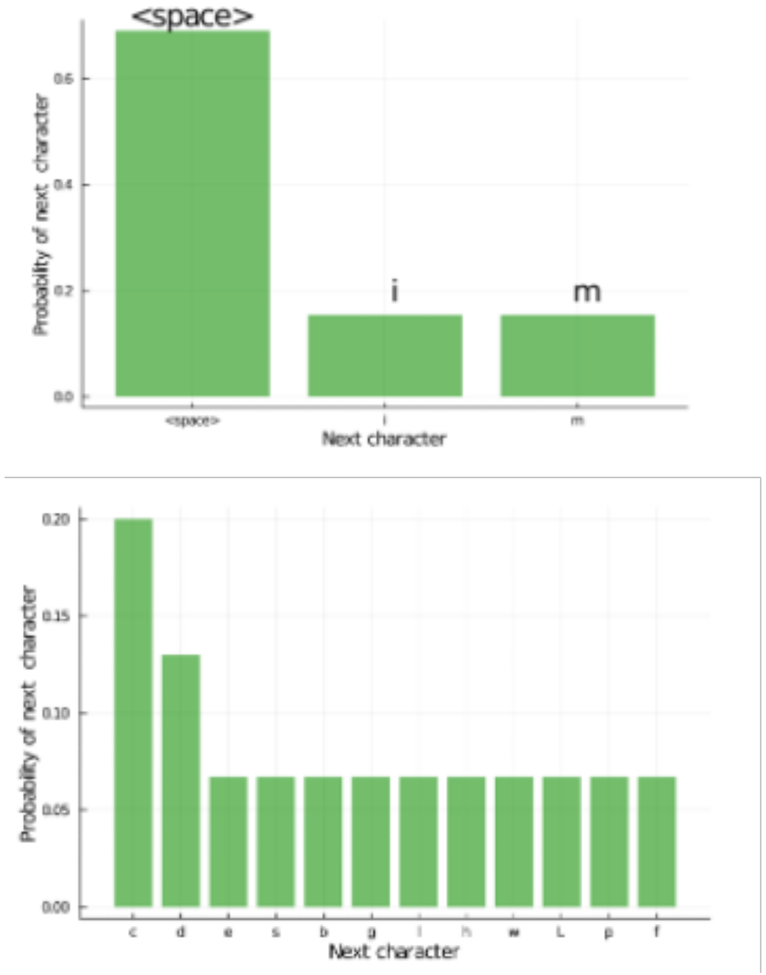
- Models probabilities of word sequences
- Train a model on some text and then generate
- Simple example: character model
  - Pick a *memory length*
  - Go through training data in overlapping windows and
  - For every *memory length* sequence, count instances of the next character

Training

```
In a hole in the ground there lived a hobbit.
Not a nasty, dirty, wet hole, filled with the
ends of worms and an oozy smell, nor yet a dry,
bare, sandy hole with nothing in it to sit down
on or to eat: it was a hobbit-hole, and that
means comfort.
```

Character Based Model

“led with th”	‘e': 1.0
“ed with the”	' ‘: 0.69, ‘i’: 0.154, ‘m': 0.154
“d with the “:	‘c': 0.2, 'd': 0.13, ‘e': 0.067, ’s': 0.067, ‘b': 0.067,‘g': 0.067, ’l’: 0.067, ‘h': 0.067,'w': 0.067, ’L': 0.067,'p': 0.067, 'f': 0.067



Generation Sample

As a matter of fact two nights and the reason. It was lit by a great red fire in their midst and there with thirty or forty armed guards. Gandalf took their leave at last of Beorn, and they were in shining armour, and red light before last. Just think of a really hard one. This he thought, but before long the whole turn of affairs. He had by now had more than enough of the Mountain creaking and cracking his skull on the low arch of the path. “I am the clue-finder, the web-cutter, the stinging!” And he did. He darted backwards and forwards, slashing at spider-threads, hacking at the snapped painter that was made here as a guardroom. There were many stones lying in what appeared. Back swirled the dragon up against the magic it had once obeyed would ever open that door there was nothing to eat as quick as he could, he managed it somehow, he could not stay here, and departed with nothing in it to sit down on or to eat: it was a hobbit-hole.

With Words

In a hole in the ground there lived a hobbit. Not a nasty, dirty, wet hole, filled with the ends of worms and an oozy smell, nor yet a dry, bare, sandy hole with nothing in it to sit down on or to eat: it was a hobbit-hole, and that means comfort.

With Words

“In a hole in the” `‘ground': 0.5, ‘rock’: 5`

Need a lot more text to get variety in generation

This May Look Familiar…

The Bell System Technical Journal

Vol. XXVII *July, 1948* *No. 3*

A Mathematical Theory of Communication

By C. E. SHANNON

Introduction

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and

This May Look Familiar...

3. The Series of Approximations to English

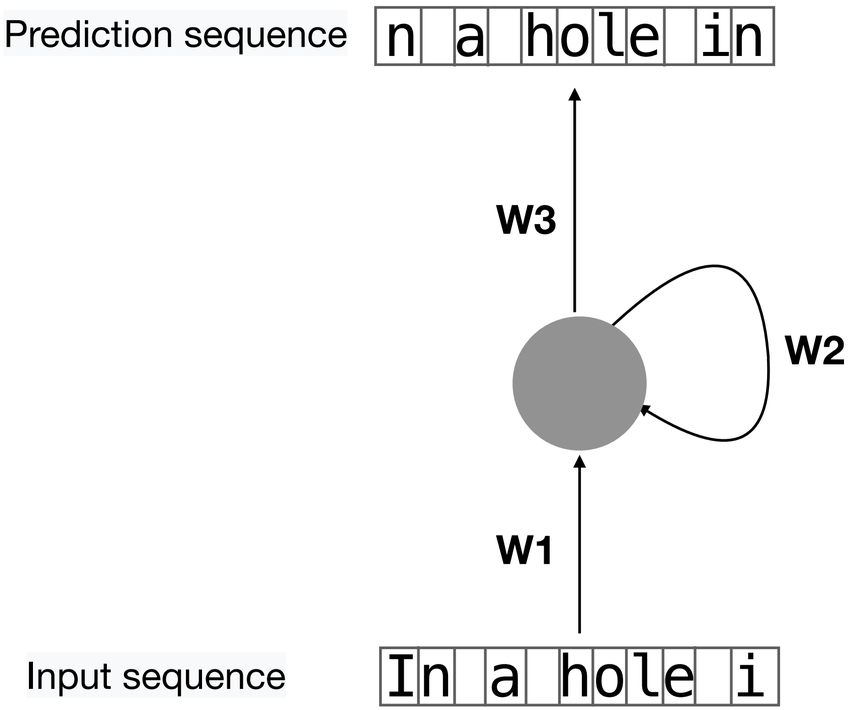
To give a visual idea of how this series of processes approaches a language, typical sequences in the approximations to English have been constructed and are given below. In all cases we have assumed a 27-symbol "alphabet," the 26 letters and a space.

1. Zero-order approximation (symbols independent and equi-probable).
   XFOML RXKHRJFFJUJ ZLPWCFWKCYJ
   FFJEYVKCQSGXYD QPAAMKBZAACIBZLHJQD
2. First-order approximation (symbols independent but with frequencies of English text).
   OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI
   ALHENHTTPA OOBTTVA NAH BRL
3. Second-order approximation (digram structure as in English).
   ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY
   ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO
   TIZIN ANDY TOBE SEACE CTISBE
4. Third-order approximation (trigram structure as in English).
   IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID
   PONDENOME OF DEMONSTURES OF THE REPTAGIN IS
   REGOACTIONA OF CRE
5. First-Order Word Approximation. Rather than continue with tetra-gram, ··· , n-gram structure it is easier and better to jump at this point to word units. Here words are chosen independently but with their appropriate frequencies.
   REPRESENTING AND SPEEDILY IS AN GOOD APT OR
   COME CAN DIFFERENT NATURAL HERE HE THE A IN
   CAME THE TO OF TO EXPERT GRAY COME TO FUR-
   NISHES THE LINE MESSAGE HAD BE THESE.
6. Second-Order Word Approximation. The word transition probabilities are correct but no further structure is included.
   THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH
   WRITER THAT THE CHARACTER OF THIS POINT IS
   THEREFORE ANOTHER METHOD FOR THE LETTERS
   THAT THE TIME OF WHO EVER TOLD THE PROBLEM
   FOR AN UNEXPECTED

The resemblance to ordinary English text increases quite noticeably at each of the above steps. Note that these samples have reasonably good structure out to about twice the range that is taken into account in their construction. Thus in (3) the statistical process insures reasonable text for two-letter sequence, but four-letter sequences from the sample can usually be fitted into good sentences. In (6) sequences of four or more

Recurrent Neural Networks

- Deep Learning language models
- RNNs have long memory



- See: Andrey Karpathy, *The Unreasonable Effectiveness of Recurrent Neural Networks,* http://karpathy.github.io/2015/05/21/rnn-effectiveness/

pentena

Language is Complicated

During the 1998 Test series against England, Cronje scored five consecutive fifties, having failed to score one in the nine previous Tests against them. In his fiftieth Test, at Trent Bridge he scored 126, his sixth and last Test century and his first in 29 matches.

Extracted from the Hansie Cronje Wikipedia page

Language is Complicated

During the 1998 Test series against England, Cronje scored five consecutive fifties, having failed to score one in the nine previous Tests against them. In his fiftieth Test, at Trent Bridge he scored 126, his sixth and last Test century and his first in 29 matches.

104 characters and 16 words between England and them.

Language is Complicated

During the 1998 Test series against England, Cronje scored five consecutive fifties, having failed to score one in the nine previous Tests against them. In his fiftieth Test, at Trent Bridge he scored 126, his sixth and last Test century and his first in 29 matches.

105 characters/17 words between	Cronje and his
139 characters/23 words	Cronje and he
155 characters/26 words	Cronje and his
191 characters/33 words	Cronje and his

Transformers

- Uses attention
- Can be trained in parallel
- GPT-2, Open AI (June 2018), 1.5 B parameters
- Turing NLG, Microsoft (Feb 2020) ,17 B parameters
- GPT-3, Open AI, (May 2020), 175 B parameters

and growing…

Can generate coherent paragraphs with a prompt

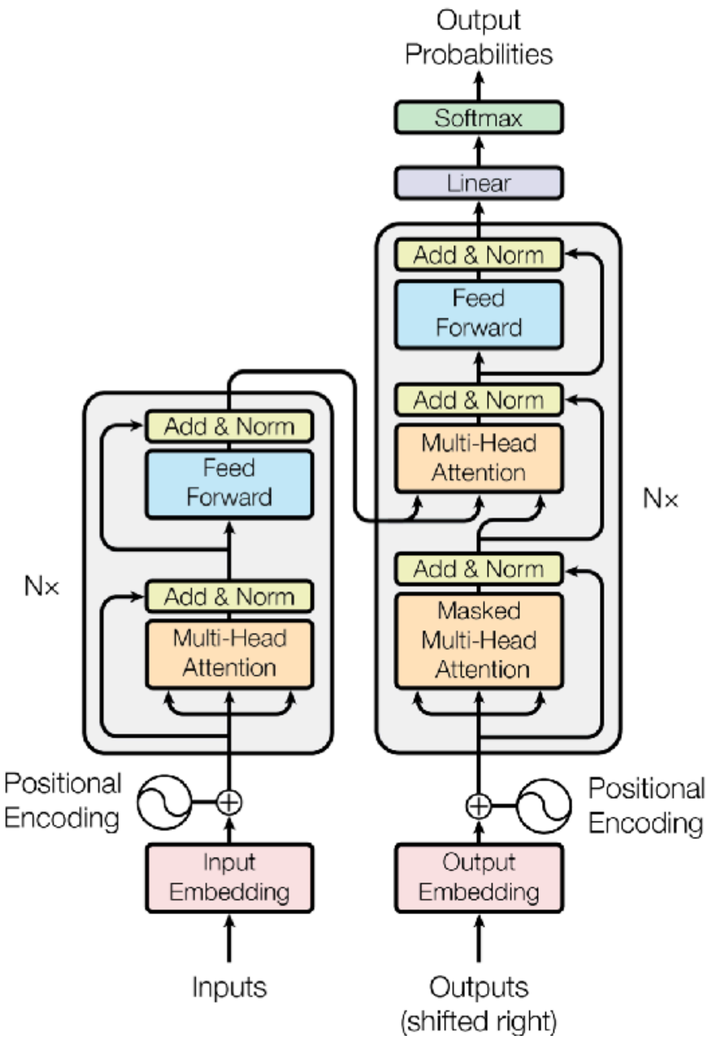
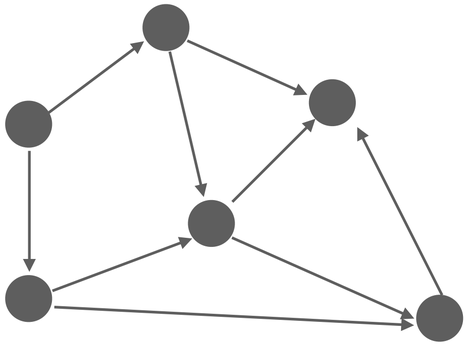


Image from: Vaswani *et al*, *Attention is all you need*. In Advances in Neural Information Processing Systems, 2017

Graphs

- Many systems have graph representations
  - Document structure
  - File system
  - IT networks
  - Communications/social networks
  - …

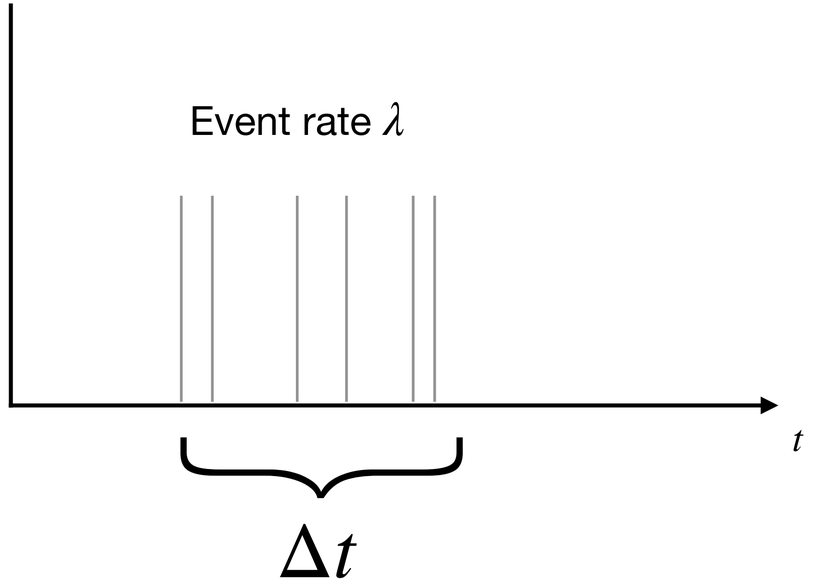


Graphs

- Deep Learning on graphs
  - Graph Recurrent Attention Network (GRAN) Liao *et al*, NeurIPS, 2019

Communications with TPPs

- When do events take place?
- Temporal Point Processes
  - Poisson process with rate λ:
  - p(n events in Δt) = $\frac{(\lambda \Delta t)^n}{n!} \exp^{-\lambda \Delta t}$
  - Memoryless
  - Events independent

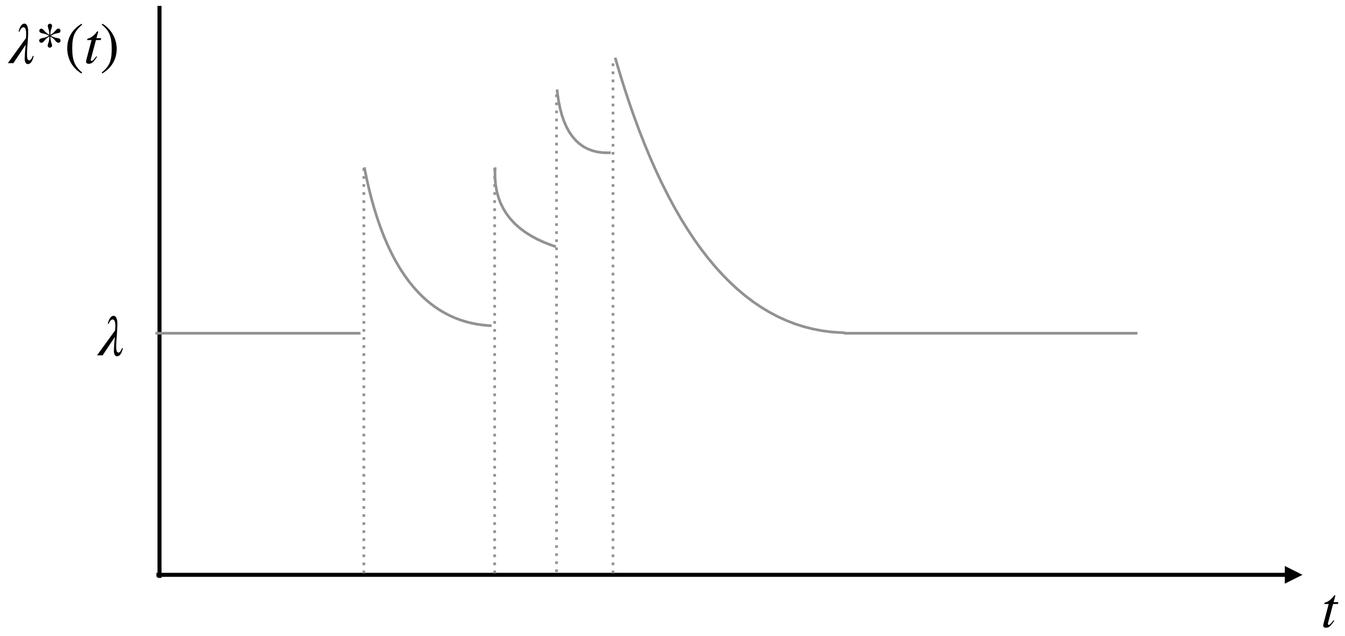


Communications with TPPs

- Hawkes process: $\lambda^*(t) = \lambda + \sum_{t_i<t} \mu(t - t_i)$

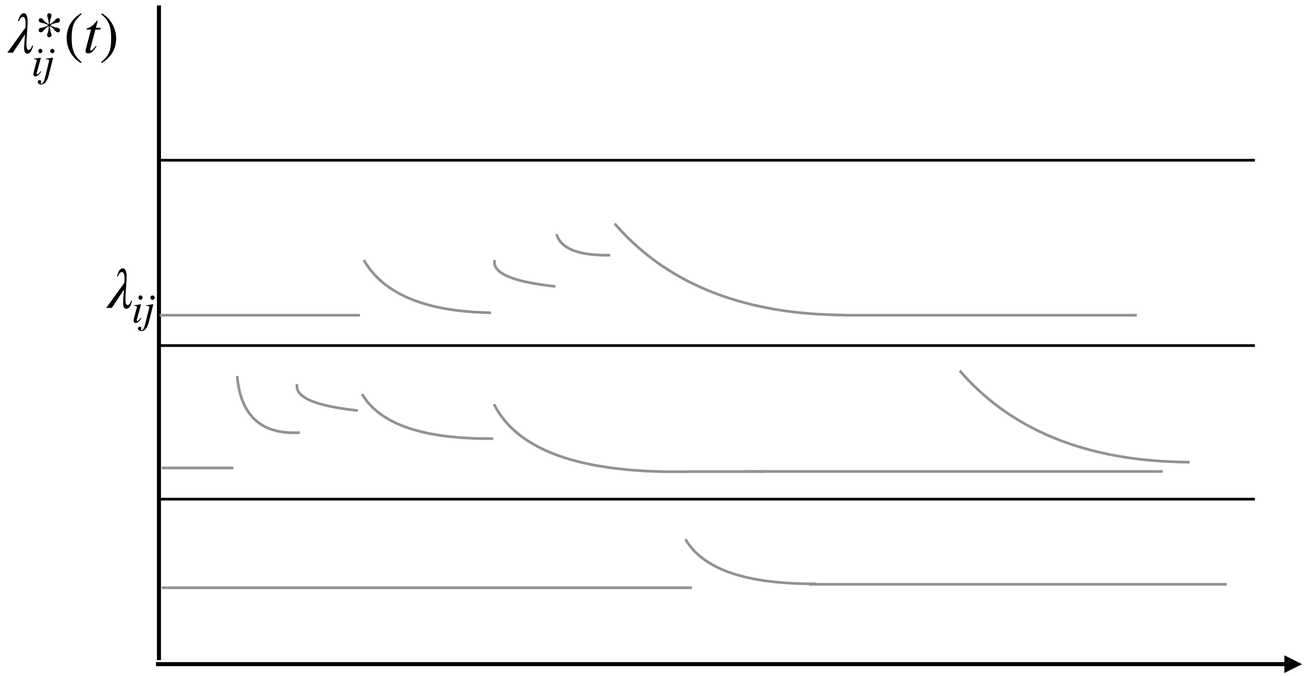
  with $\mu(t) = \alpha e^{-\beta t}$

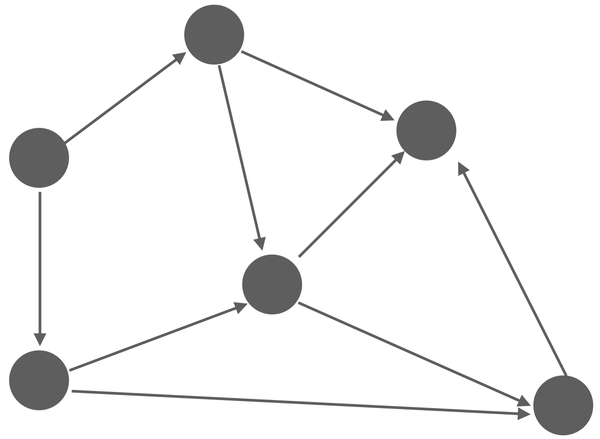
- Base rate λ and *self-exciting* events



Communications with TPPs

- *Marked* TPPs model temporal interactions on network
- α_{ij} and β_{ij} for pairs of communicating nodes





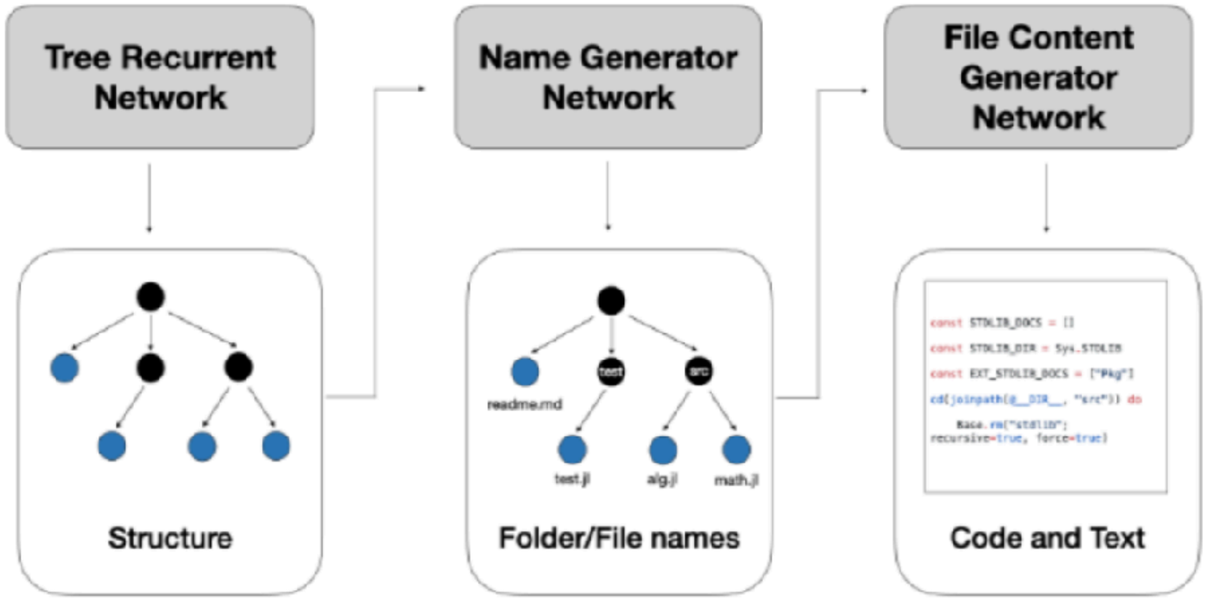
Communications with TPPs

- Neural Point Processes
- RNN to embed event sequence
- Intensity Free TPP:
  - Embedding vector and metadata parametrise inter-event time distribution
  - Mixture of lognormal

See: Intensity Free TPP from O. Shchur, M. Biloš, and S. Günnemann, *Intensity-free learning of temporal point processes*, ICLR, 2019.

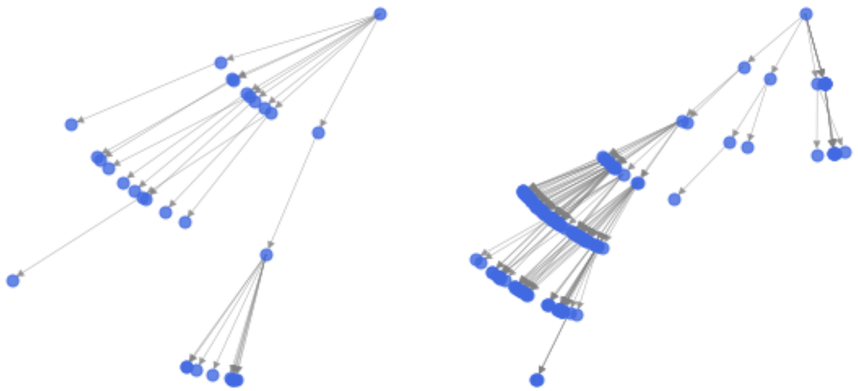
HoneyCode: Fake Repositories

- Fake file trees, file names and code
- Trees from modified GRAN
- Filenames and code from character RNN



D. Nguyen, D. Liebowitz, S. Nepal, and S. Kanhere, *Honeycode: Automating deceptive software repositories with deep generative models*, in Proceedings of the 54th Hawaii International Conference on SystemSciences, 2021

HoneyCode: Fake Repositories



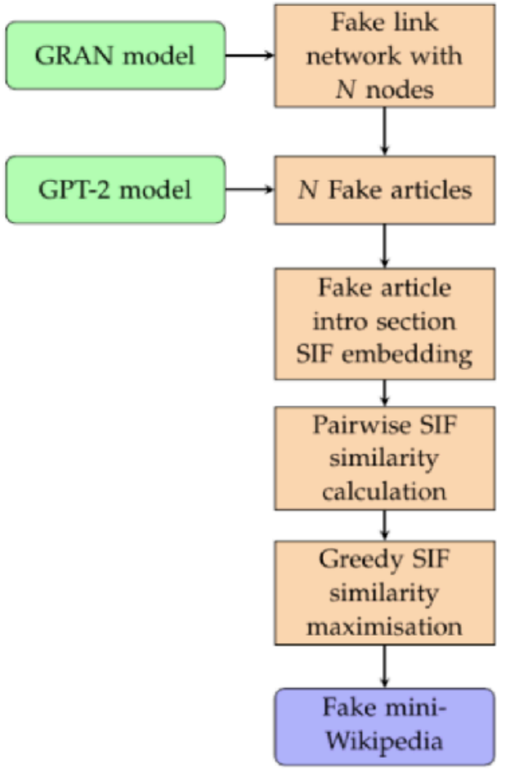
```
function sparse!(d::AbstractMixtureModel, x)
    K = ncomponentwise_logpdf!(Matrix{eltype(x)}(undef, nd, n), one(x))
    logc0 = /x * detach(1)
    logc0 = max(length(I0)*length(J)+1

    for i = 1:(length(Acolptr) <= 1)
        VR = similar(A[1], T, nrows, ncols)
    elseif size(X, 2) == 1
        logX = log(u)
        cm2 /= sqrt(1 - abs2(z)/2) * sqrt(z2)
        alpha3 = max(unsqueezeb, Tuple{Float64})
        $fname(pp, pos, durbin, sort)
            return (false, Int[])
        end
    else
        if !in_single_quotes
            const_prop_profitable(buf) == 0 && return false
            write(buffer, ' ')
        end
    end
    return nothing
    return write_project(env)
end
```

D. Nguyen, D. Liebowitz, S. Nepal, and S. Kanhere, *Honeycode: Automating deceptive software repositories with deep generative models*, in Proceedings of the 54th Hawaii International Conference on SystemSciences, 2021

penten

WikiGen

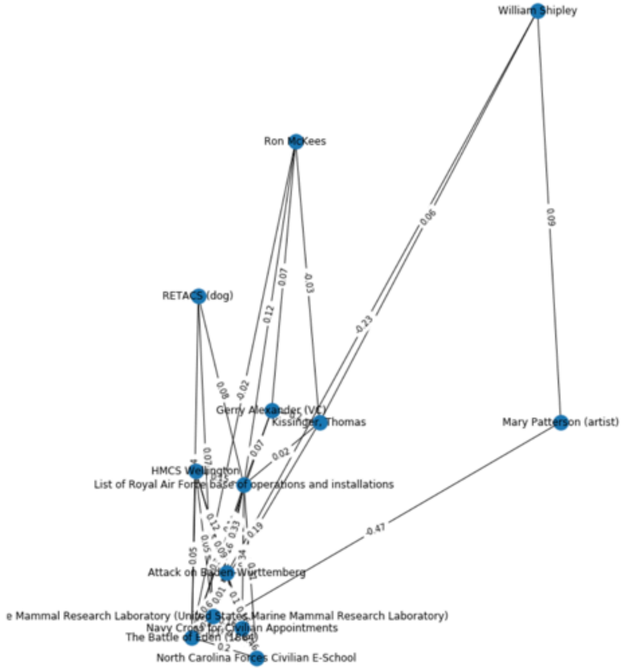
- Fake page content and structure from GPT-2
- Network from GRAN
- SIF embeddings to match linked content



M. Longland, *Generating fake websites: Wikigen*, ANU, Tech. Rep., 2020, https://github.com/longland-m/wikigen

penten

WikiGen



== Article Start ==

Albert Le Roux

Albert Le Roux (August 14, 1879 – July 16, 1961) was a Canadian historian and political activist. He served in the Canadian Infantry during the First World War.

== Early life and education ==

Le Roux was born in 1879 in Ridgeway, Ontario, the son of Alphonse Le Roux. As a young man, Le Roux attended Mount Allison School in Toronto until his father left the family when Le Roux was about twelve. He was subsequently educated at Osgoode Hall and Mount Allison College.

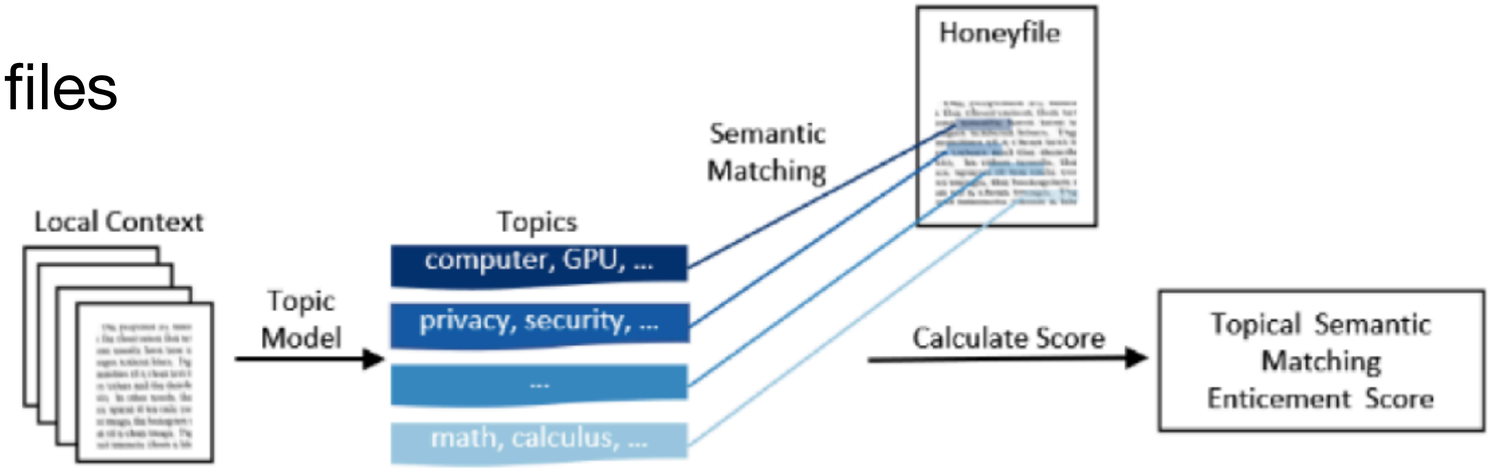
== Military career ==

Le Roux entered the Canadian Infantry in the Second Canadian Expeditionary Force in 1916. Le Roux was commissioned in the men's 1st Battalion in January 1917, and was seconded to the 2nd Battalion. Le Roux was promoted to lieutenant in November 1917 and captain in November 1918. In 1917, Le Roux was in command of a battalion in the Italian Campaign.

M. Longland, *Generating fake websites: Wikigen*, ANU, Tech. Rep., 2020.https://github.com/longland-m/wikigen

Deception Metrics

- Enticement: drawing adversary to the honeypot
- Compare honeyfile text to real files



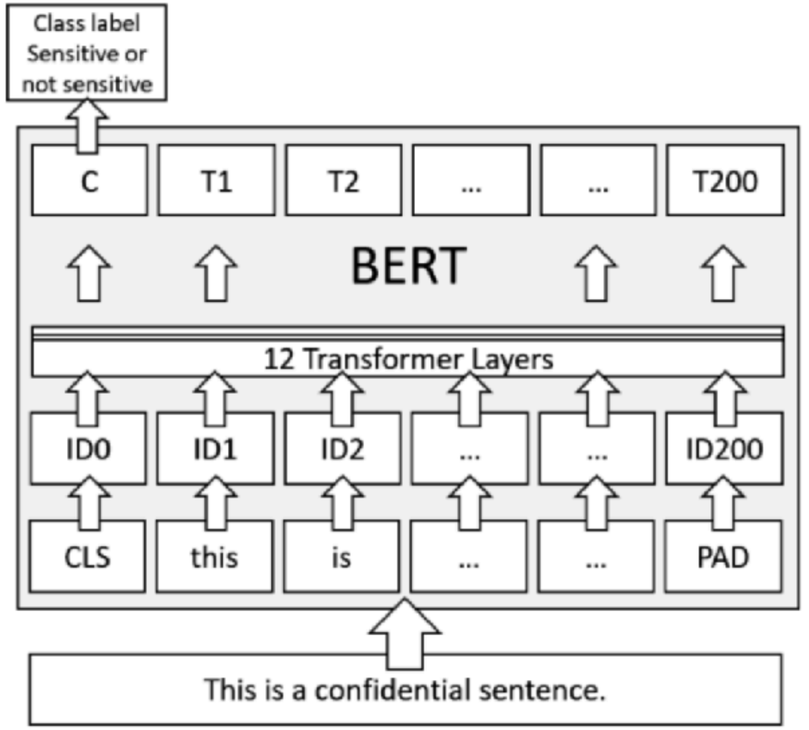
- Topic modelling, semantic matching

R. Timmer, D. Liebowitz, S. Nepal, and S. Kanhere, *TSM: Measuring the Enticement of Honeyfiles with Natural Language Processing*, Proceedings of the 55th Hawaii International Conference on System Sciences, 2022

penten

Sensitive Content Detection

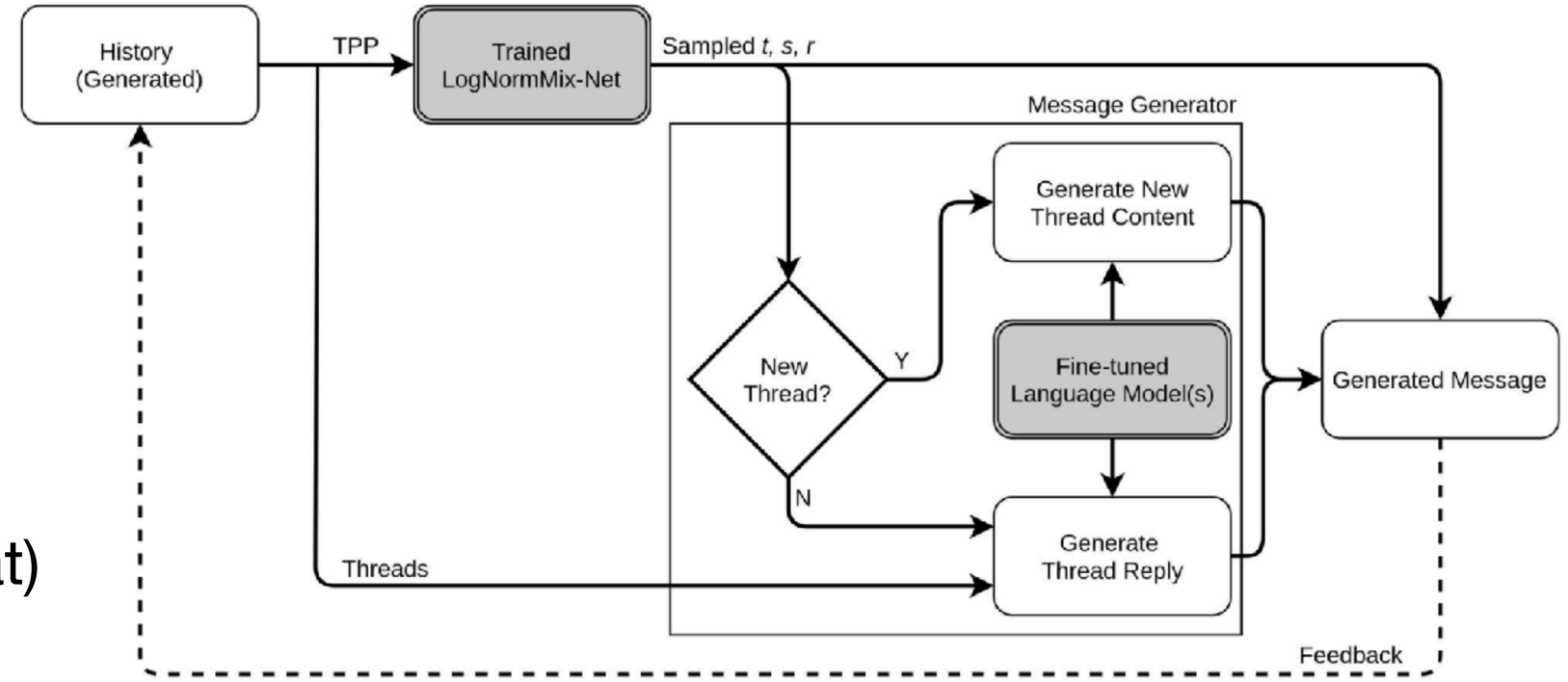
- Automated sensitive content detection
- Monsanto trial data labeled at sentence level
- BERT can do pretty good detection with small training sets



R. Timmer, D. Liebowitz, S. Nepal, and S. Kanhere, *Can pre-trained transformers be used in detecting complex sensitive sentences? - a Monsanto case study*, Proc. the 3rd IEEE International Conference on Trust, Privacy and Security in Intelligent Systems, and Applications, 2021.

Networked Communications

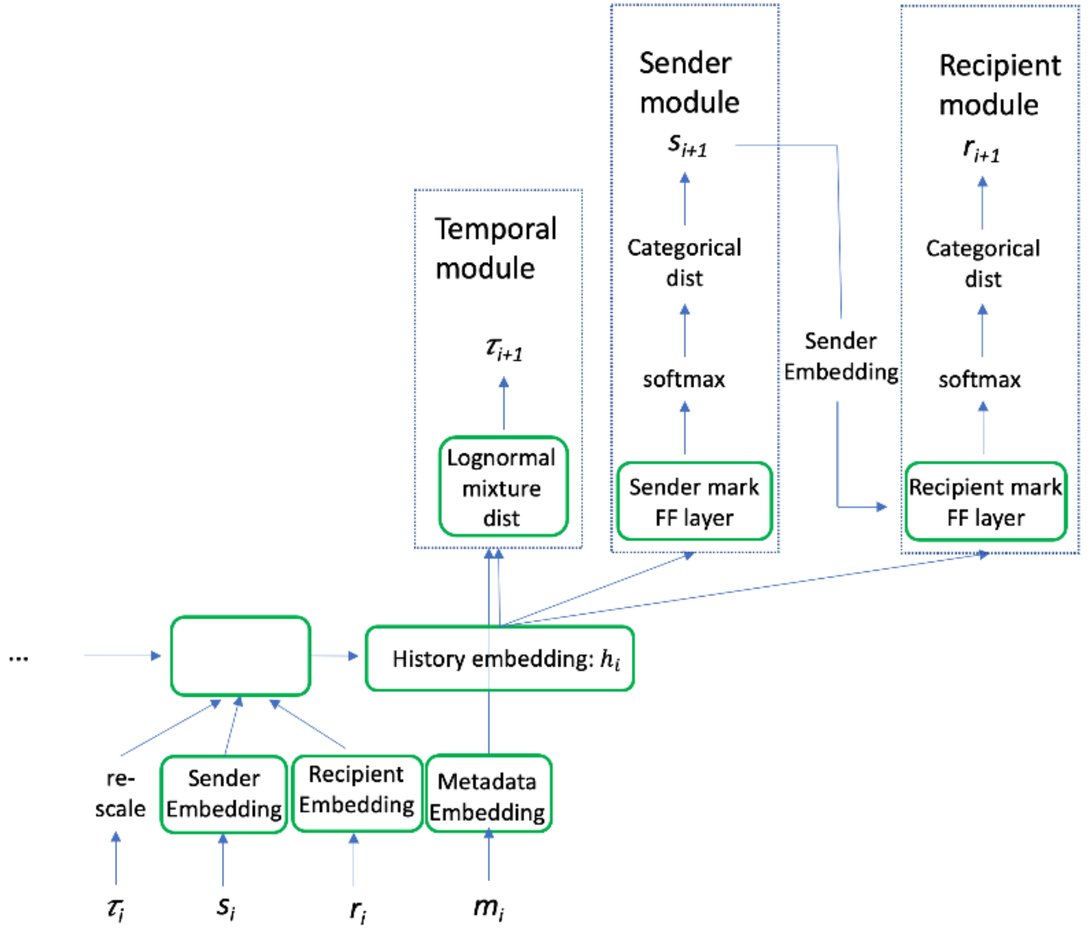
- Marked TPP
- Language model
- Simulate Direct Messaging (email, social networks chat)



K. Moore, C. J. Christopher, D. Liebowitz, S. Nepal, and R. Selvey, *Modelling direct messaging networks with multiple recipients for cyber deception*, To appear in 7th IEEE European Symposium on Security and Privacy, June 2022

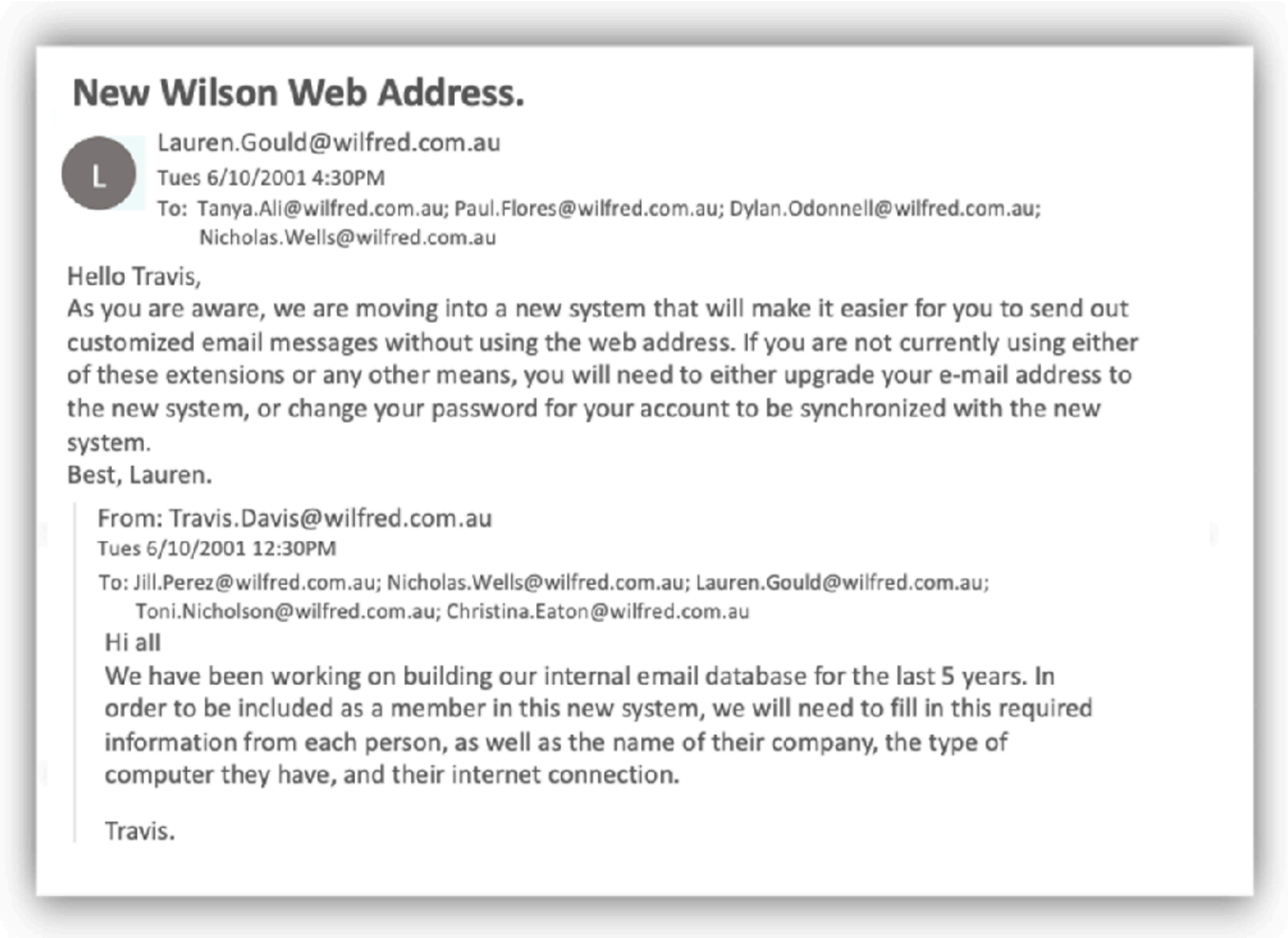
penten

Networked Communications



K. Moore, C. J. Christopher, D. Liebowitz, S. Nepal, and R. Selvey, *Modelling direct messaging networks with multiple recipients for cyber deception*, To appear in 7th IEEE European Symposium on Security and Privacy, June 2022

Networked Communications



New Wilson Web Address.

L Lauren.Gould@wilfred.com.au
Tues 6/10/2001 4:30PM
To: Tanya.Ali@wilfred.com.au; Paul.Flores@wilfred.com.au; Dylan.Odonnell@wilfred.com.au; Nicholas.Wells@wilfred.com.au

Hello Travis,
As you are aware, we are moving into a new system that will make it easier for you to send out customized email messages without using the web address. If you are not currently using either of these extensions or any other means, you will need to either upgrade your e-mail address to the new system, or change your password for your account to be synchronized with the new system.
Best, Lauren.

> From: Travis.Davis@wilfred.com.au
> Tues 6/10/2001 12:30PM
> To: Jill.Perez@wilfred.com.au; Nicholas.Wells@wilfred.com.au; Lauren.Gould@wilfred.com.au; Toni.Nicholson@wilfred.com.au; Christina.Eaton@wilfred.com.au
>
> Hi all
> We have been working on building our internal email database for the last 5 years. In order to be included as a member in this new system, we will need to fill in this required information from each person, as well as the name of their company, the type of computer they have, and their internet connection.
>
> Travis.

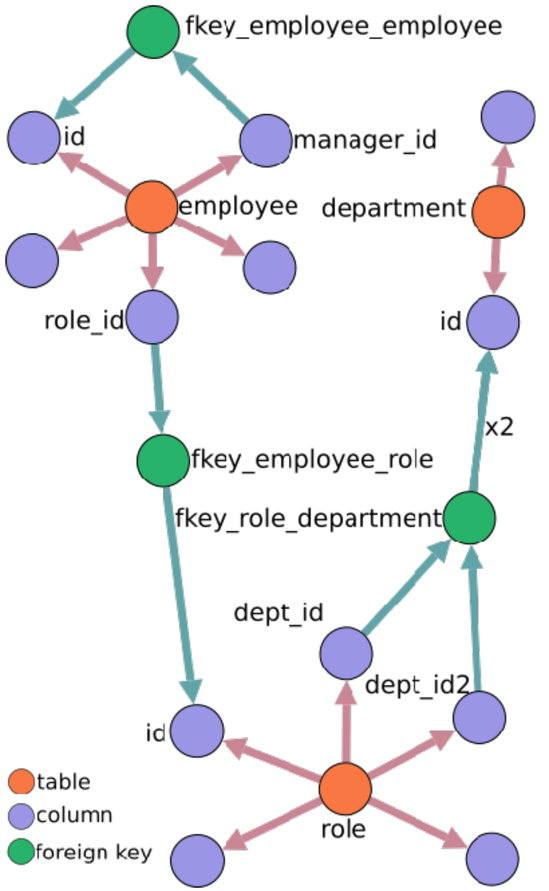
K. Moore, C. J. Christopher, D. Liebowitz, S. Nepal, and R. Selvey, *Modelling direct messaging networks with multiple recipients for cyber deception*, To appear in 7th IEEE European Symposium on Security and Privacy, June 2022

penten

Database Generation

Towards novel database generation:

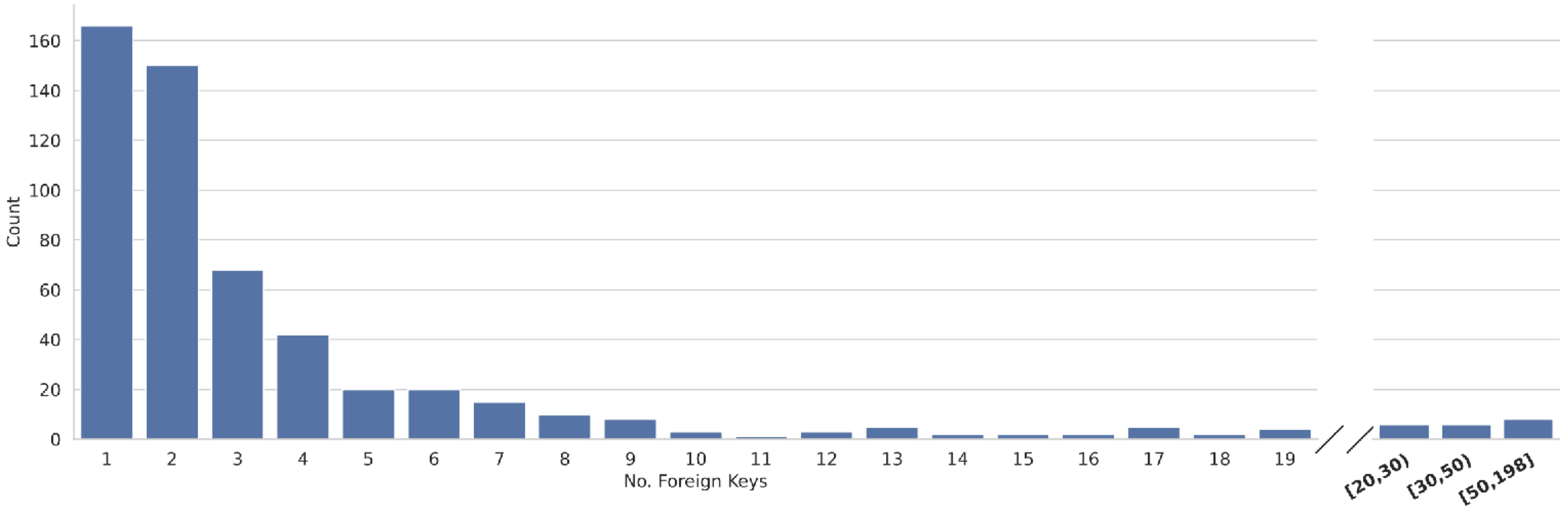
- Automatically scrape and parse schema creation SQL scripts on Github
- Standardised format
- Implied foreign keys extracted
- ~ 2500 schemas and growing



C Christopher, K Moore, D Liebowitz, *SchemaDB: Structures in Relational Datasets,* https://arxiv.org/abs/2111.12835

penten

Database Generation



C Christopher, K Moore, D Liebowitz, *SchemaDB: Structures in Relational Datasets,* https://arxiv.org/abs/2111.12835

pentenq

HoneyTrace

- Data Loss Intelligence
- HoneyTrace.io

pentenq

HoneyTrace

Track information most important to your business

Extend detection beyond your network perimeter

Helps to answer “who, what and when?”

Set up in minutes

No installation of software on your system or network

penten

HoneyTrace

Upload document or create digital twin

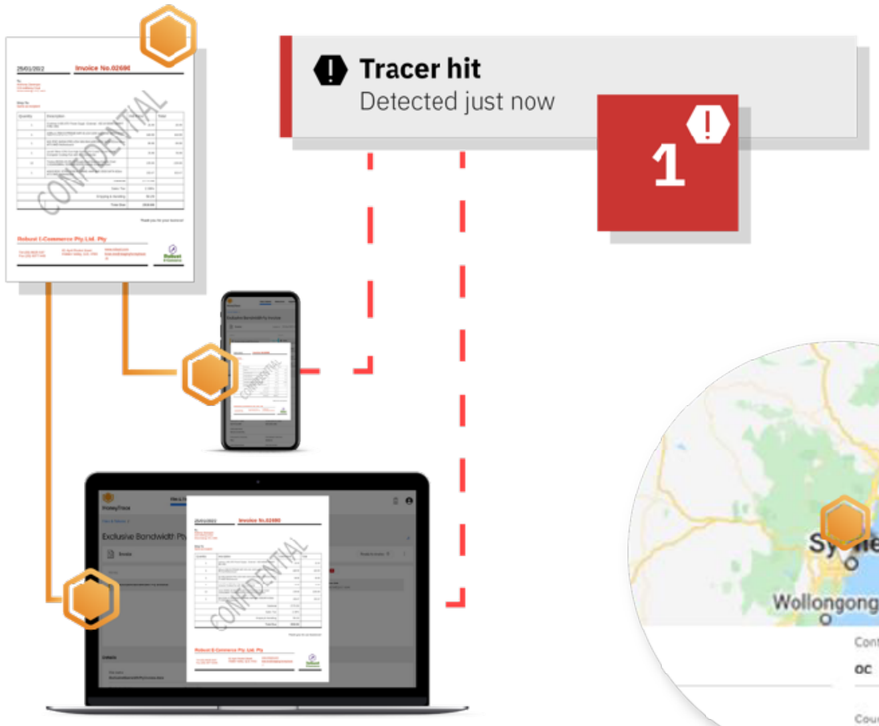
Select one or more tracers

Upload HoneyTrace files into environment

Threat actor interacts with HoneyTrace files

Receive an alert

- Email
- Search
- Beacon
- URL Redirect
- EDR Integration
- Honeypot



penten

Questions?