Enhancing Deep Learning DGA Detection Models Using Separate Character Embedding

Vikash Yadav
Data Scientist, Royal Bank of Canada

Content

Problem
- Malware Life Cycle
- Domain Generation Algorithm

Existing Solution
- Traditional Approaches
- Machine Learning Approaches

RNN model
- Recurrent Neural Network based detecting technique
- Unified Architecture

Enhancement
- Separate Embedding Model
- Improvement in detecting unknown DGA

Problem

- Malware Life Cycle
- Domain Generation Algorithm

Existing Solution

- Traditional Approaches
- Machine Learning Approaches

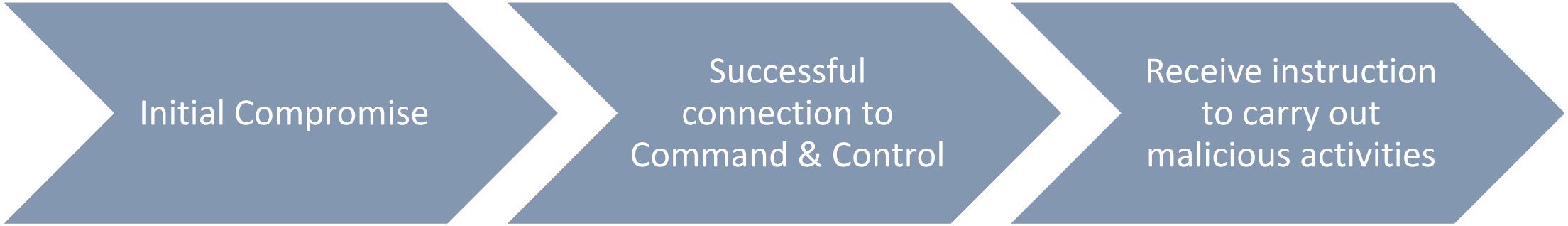
RNN model

- Recurrent Neural Network based detecting technique
- Unified Architecture

Enhancement

- Separate Embedding Model
- Improvement in detecting unknown DGA

Life Cycle of a Malware



Life Cycle of a Malware

Domain Generation Algorithm (DGA)

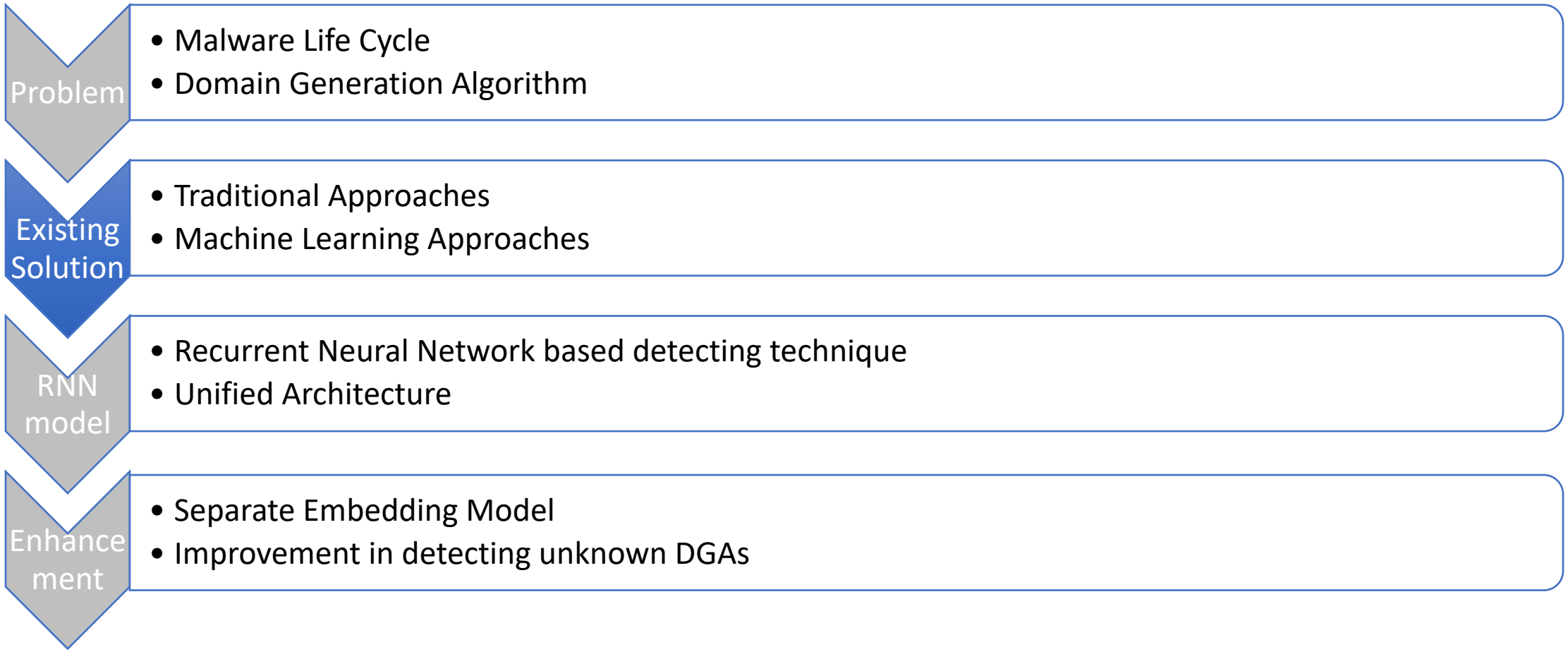
- DGA uses a seed value and/or time-dependent element to avoid command and control domains or IPs being seized or sinkhole

```
def generate_domain(year, month, day):
    """Generates a domain name for the given date."""
    domain = ""

    for i in range(16):
        year = ((year ^ 8 * year) >> 11) ^ ((year & 0xFFFFFFF0) << 17)
        month = ((month ^ 4 * month) >> 25) ^ 16 * (month & 0xFFFFFFF8)
        day = ((day ^ (day << 13)) >> 19) ^ ((day & 0xFFFFFFFE) << 12)
        domain += chr(((year ^ month ^ day) % 25) + 97)

    return domain
```

Credit: Wikipedia


Problem
• Malware Life Cycle
• Domain Generation Algorithm
Existing Solution
• Traditional Approaches
• Machine Learning Approaches
RNN model
• Recurrent Neural Network based detecting technique
• Unified Architecture
Enhance ment
• Separate Embedding Model
• Improvement in detecting unknown DGAs

Stopping DGA Malware – traditional approach

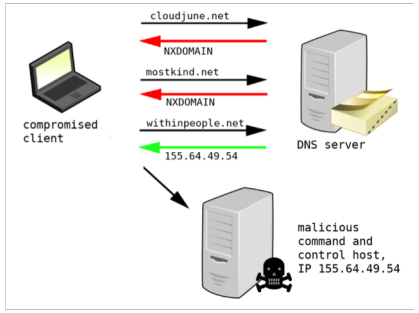
- Reverse engineer the binary to identify the DGA
- Blacklist the domain name and IP address of C2 server
- Sinkhole the C2 communication by registering the domain in advance

Stopping DGA Malware – traditional approach

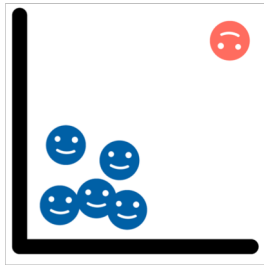
- Reactive
- Time consuming
- Not scalable

Stopping DGA Malware – ML based approach

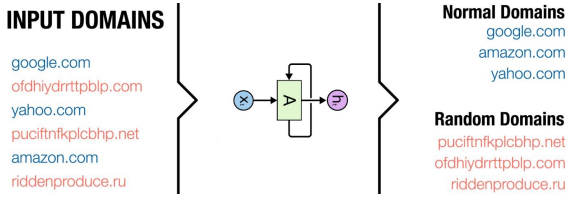
- NXDOMAIN DNS request based detection



- ML approach using handcrafted features

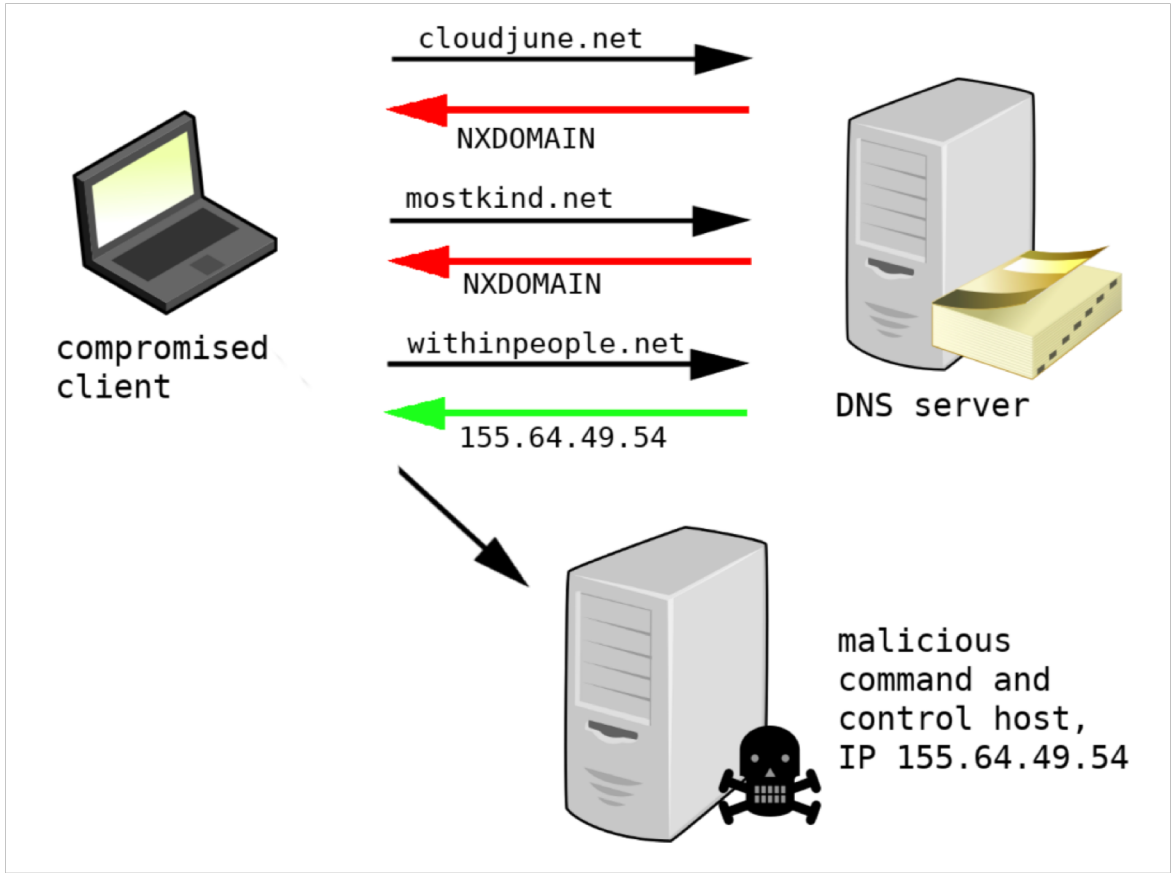


- RNN based detection



NXDOMAIN DNS request based detection

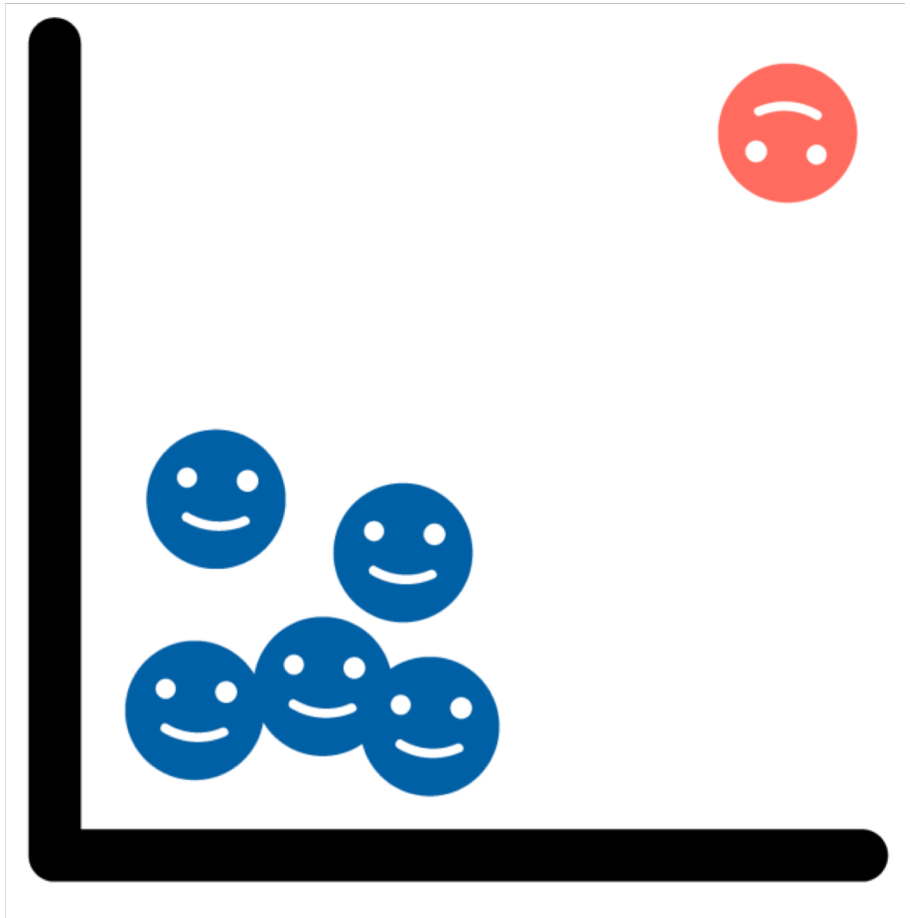
- DGA generates a large number of domains of which only a select few are registered to host a C2 server
- A client making requests to a large number of NXDomains is potentially hosting a DGA malware



Credit: Detecting DGA domains with recurrent neural networks and side information

ML approach using handcrafted features

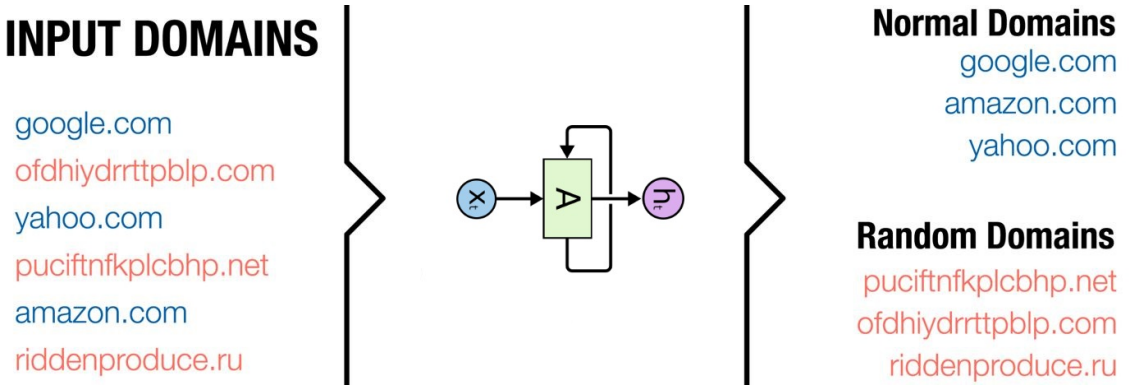
- Entropy, Length of the domain etc.
- Number of vowels vs consonants in the domain
- Periodicity of the request
- Popularity of the domain
- Total byte sent and received

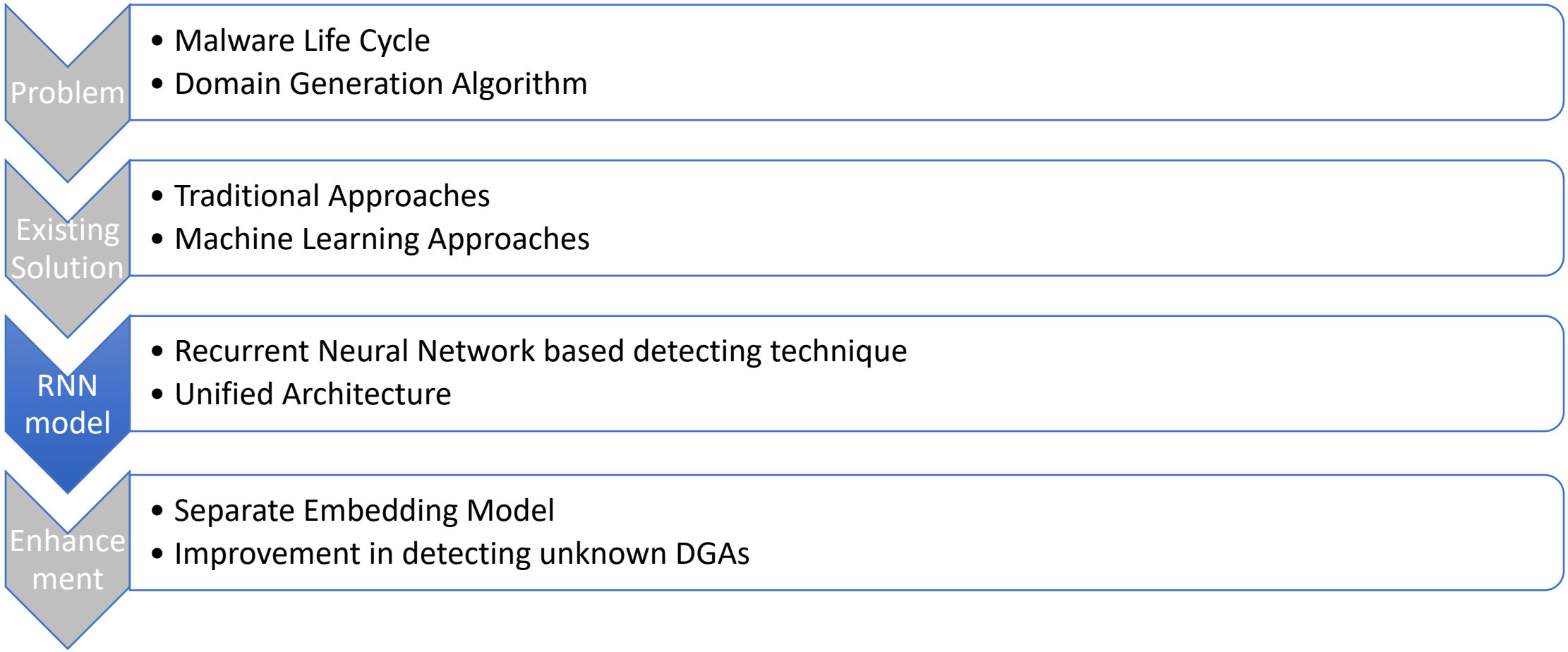


12

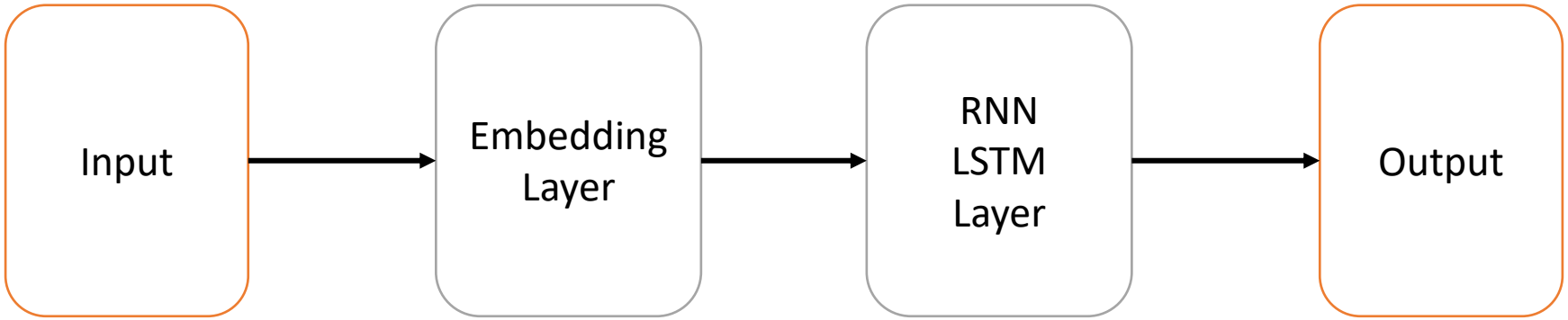
RNN based detection

- No explicit feature engineering required
- Proactive
- Easy to build and deploy
- Easy to retrain outdated models
- Highly scalable
- Highly accurate

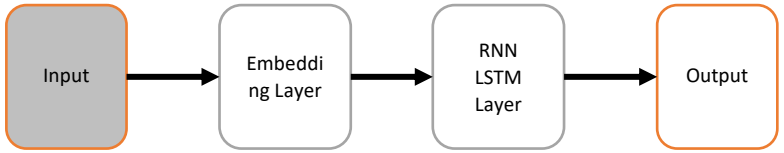



Problem
• Malware Life Cycle
• Domain Generation Algorithm
Existing Solution
• Traditional Approaches
• Machine Learning Approaches
RNN model
• Recurrent Neural Network based detecting technique
• Unified Architecture
Enhance ment
• Separate Embedding Model
• Improvement in detecting unknown DGAs

Unified RNN Model Architecture



Dataset for RNN model



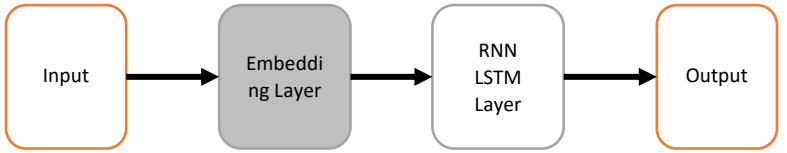
Benign Domains

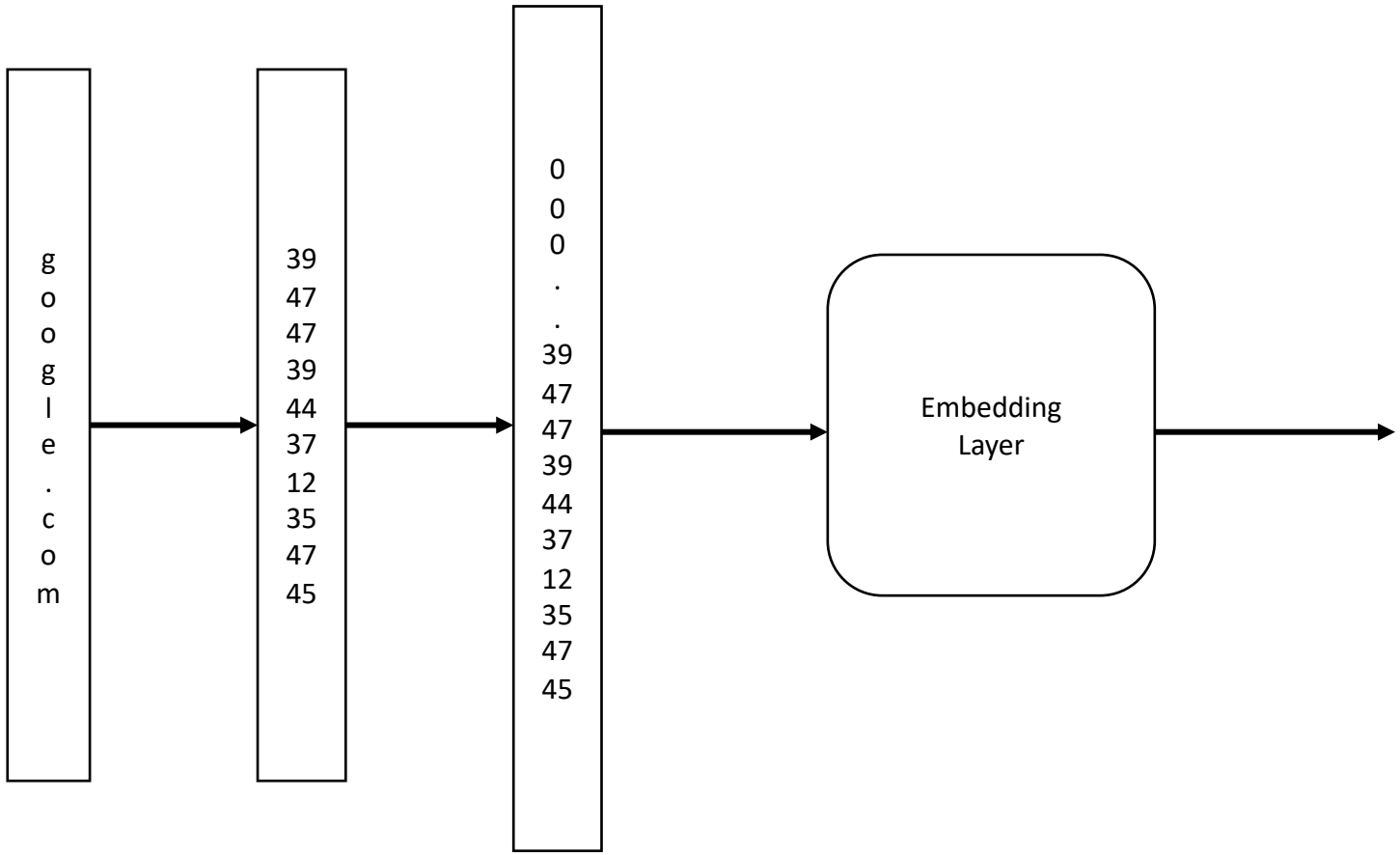
- Alexa top million domains
- Cisco top million domains
- ~1.8 million unique domains
  - google.com
  - youtube.com
  - facebook.com
  - baidu.com
  - wikipedia.org
  - yahoo.com

DGA Domains

- http://data.netlab.360.com/dga/#virut
- ~1.1 million unique domains
  - ydqtkptuwsa.org
  - bnnkqwzmy.biz
  - glrmwqh.net
  - ibymtpyd.info
  - bxyozfikd.ws
  - nvjwoofansjbh.ru

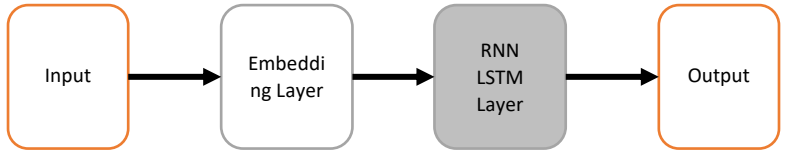
Character Embedding

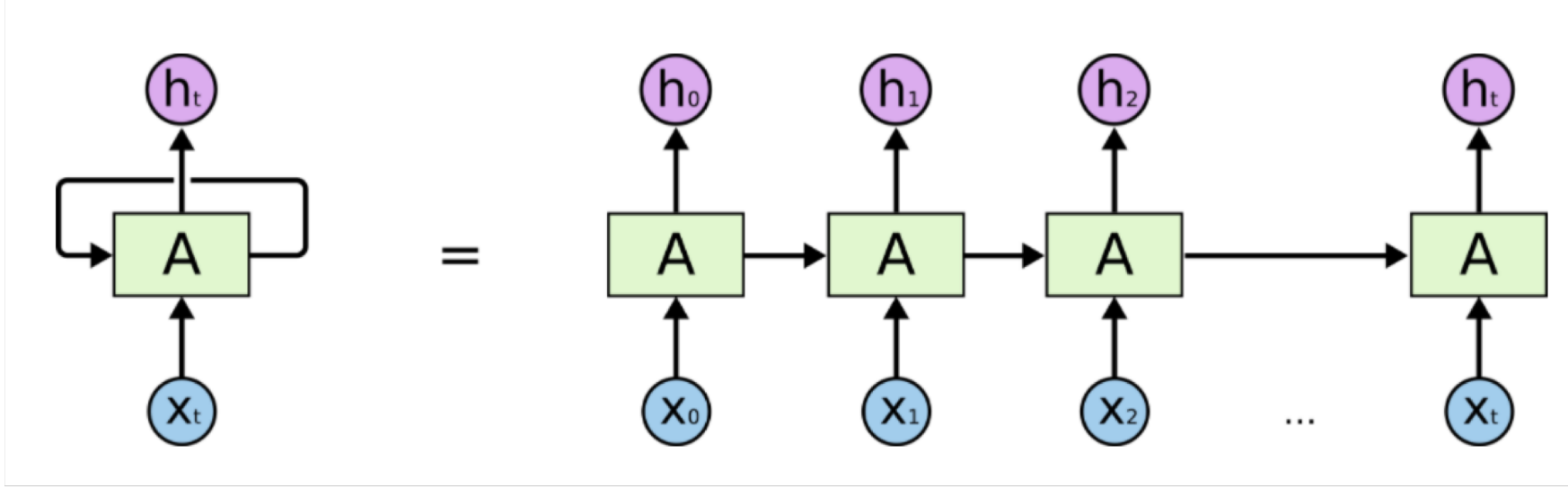




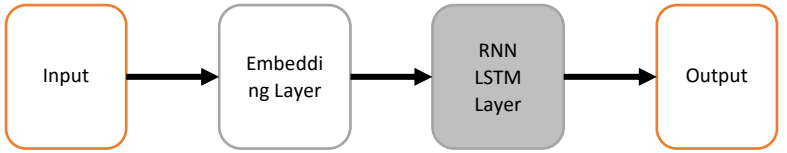
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
.							
.							
.							
-0.57	-0.01	0.51	2.95	1.4	1.03	1.24	0.04
-1.09	0.95	0.29	0.61	-0.72	1.67	1.38	-0.96
-1.09	0.95	0.29	0.61	-0.72	1.67	1.38	-0.96
-0.57	-0.01	0.51	2.95	1.4	1.03	1.24	0.04
0.48	-0.82	0.49	-1.06	0.8	0.83	-2.97	-0.35
-0.51	-0.16	-0.9	-1.65	-0.2	0.6	0.34	0.94
0.27	0.4	-2	0.12	2.09	-2.12	2.38	-1.4
0.3	-0.3	2.02	1.68	-0.98	0.63	0.19	1.36
-1.09	0.95	0.29	0.61	-0.72	1.67	1.38	-0.96
-2.02	-1.69	-0.46	0.26	1.97	0.58	-0.16	-0.45

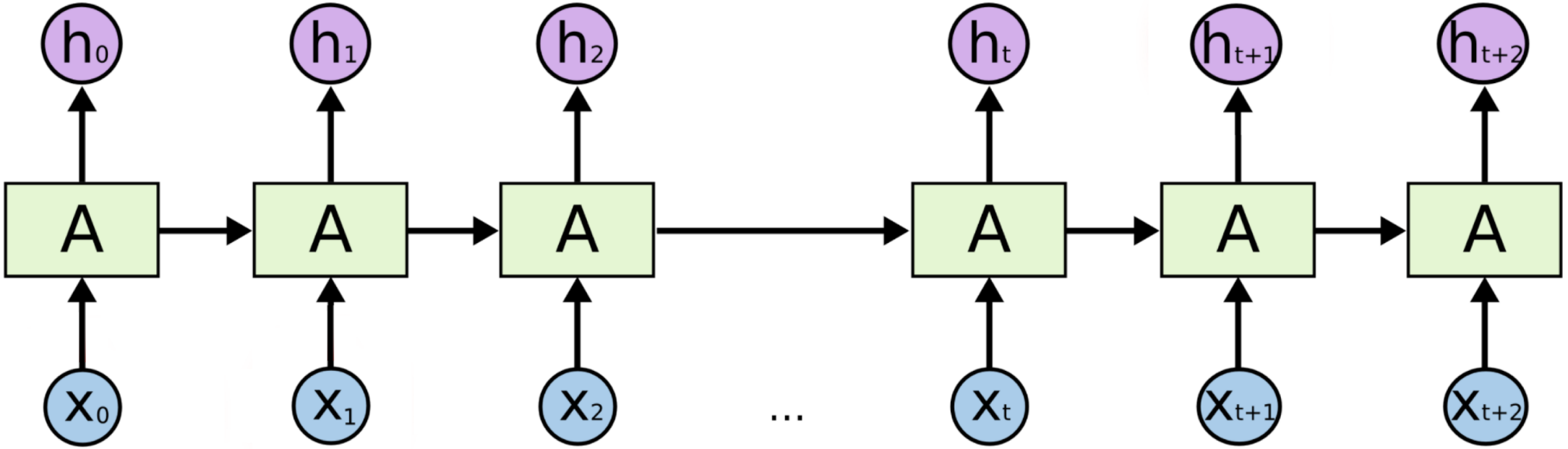
Recurrent Neural Network





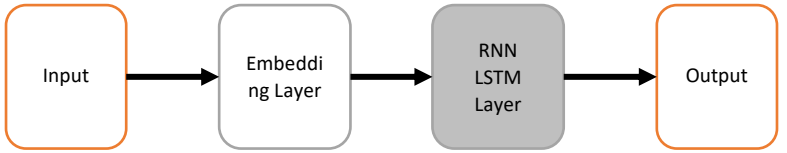
Recurrent Neural Network

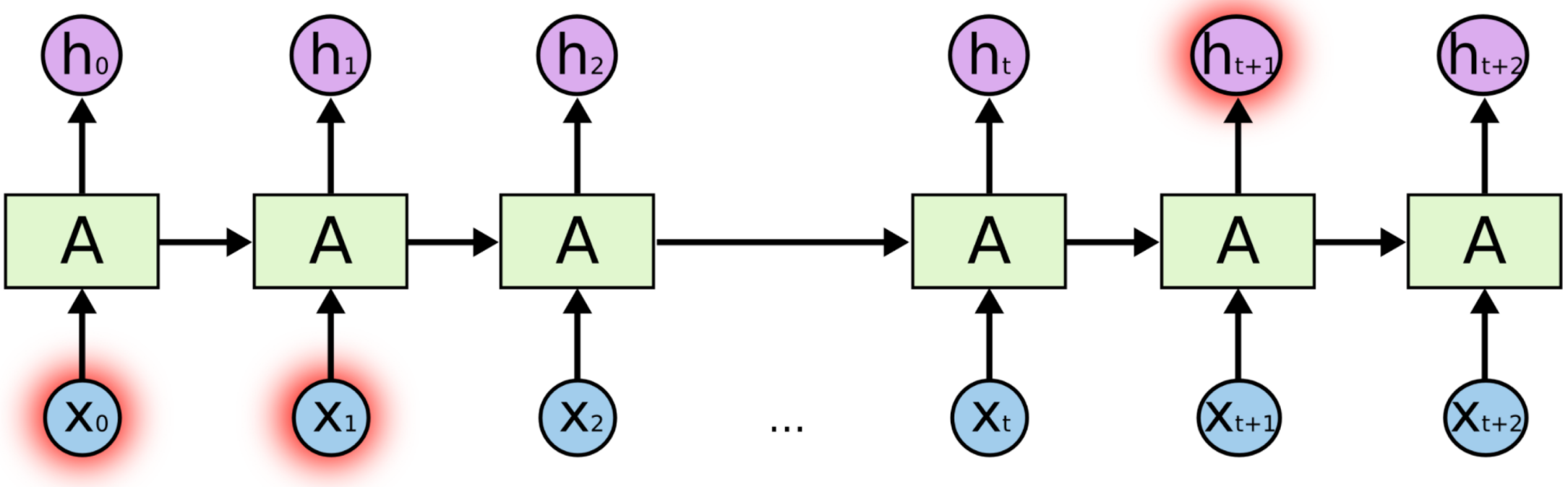




I grew up in France … I speak fluent ___?

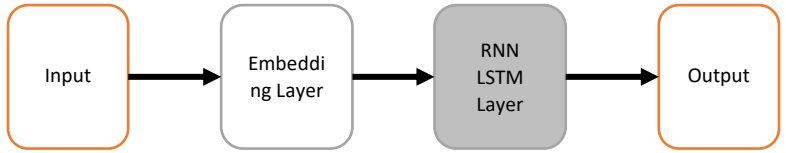
Recurrent Neural Network

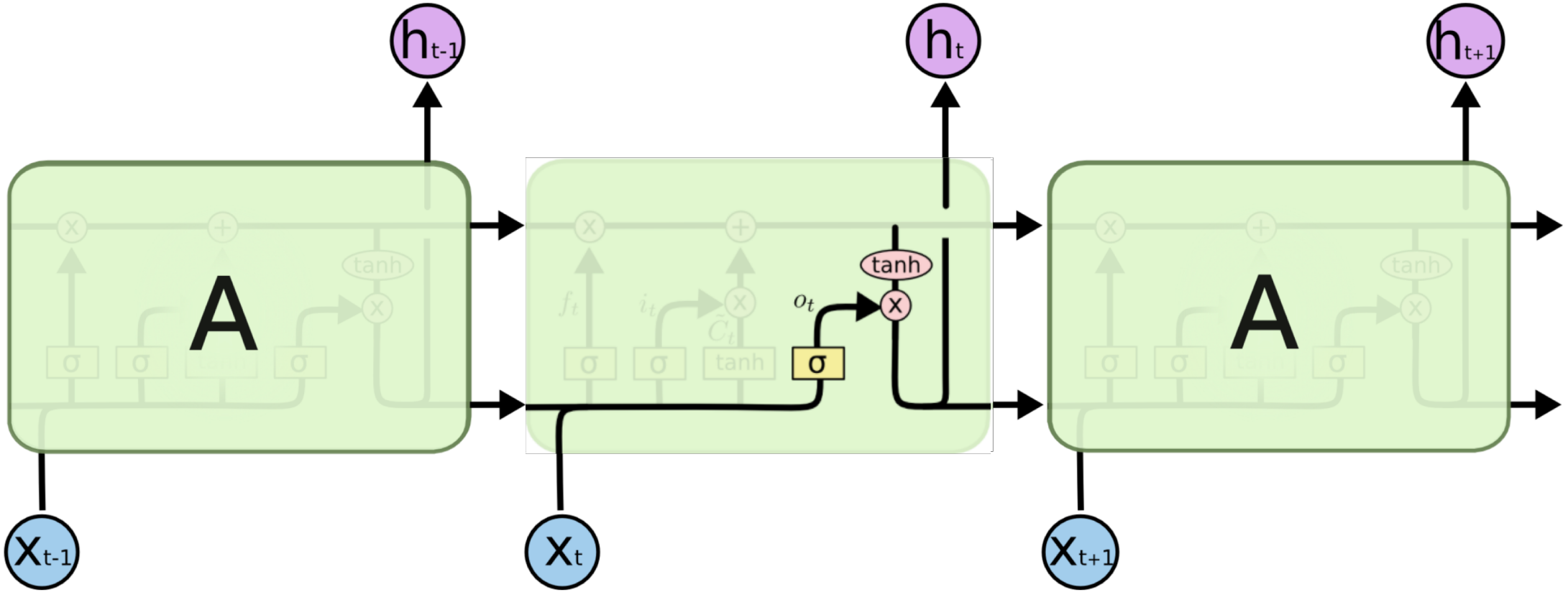




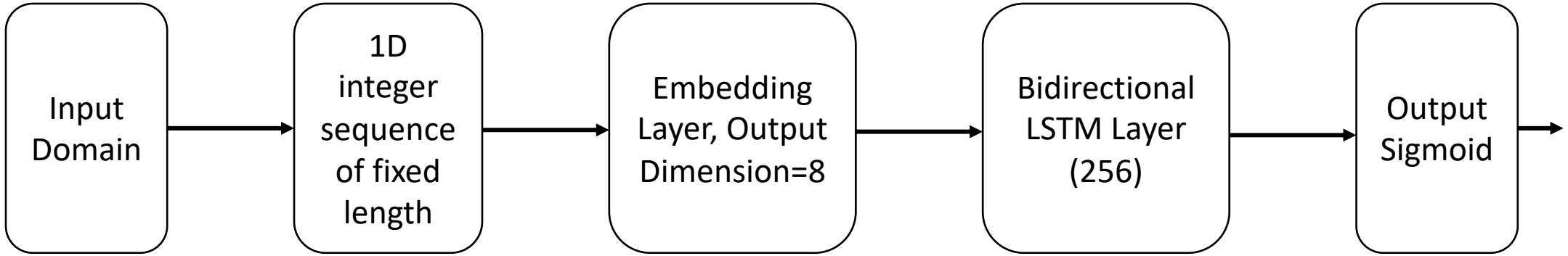
I grew up in France … I speak fluent *French*?

Long Short-Term Memory





Unified RNN Model Architecture



Result

Test Accuracy for known DGA types

Label	Record Count	Unified Model Accuracy	F score
Benign	750153	0.9946	0.9874
Malicious	415976	0.9845	

Very high accuracy for detecting known DGA types

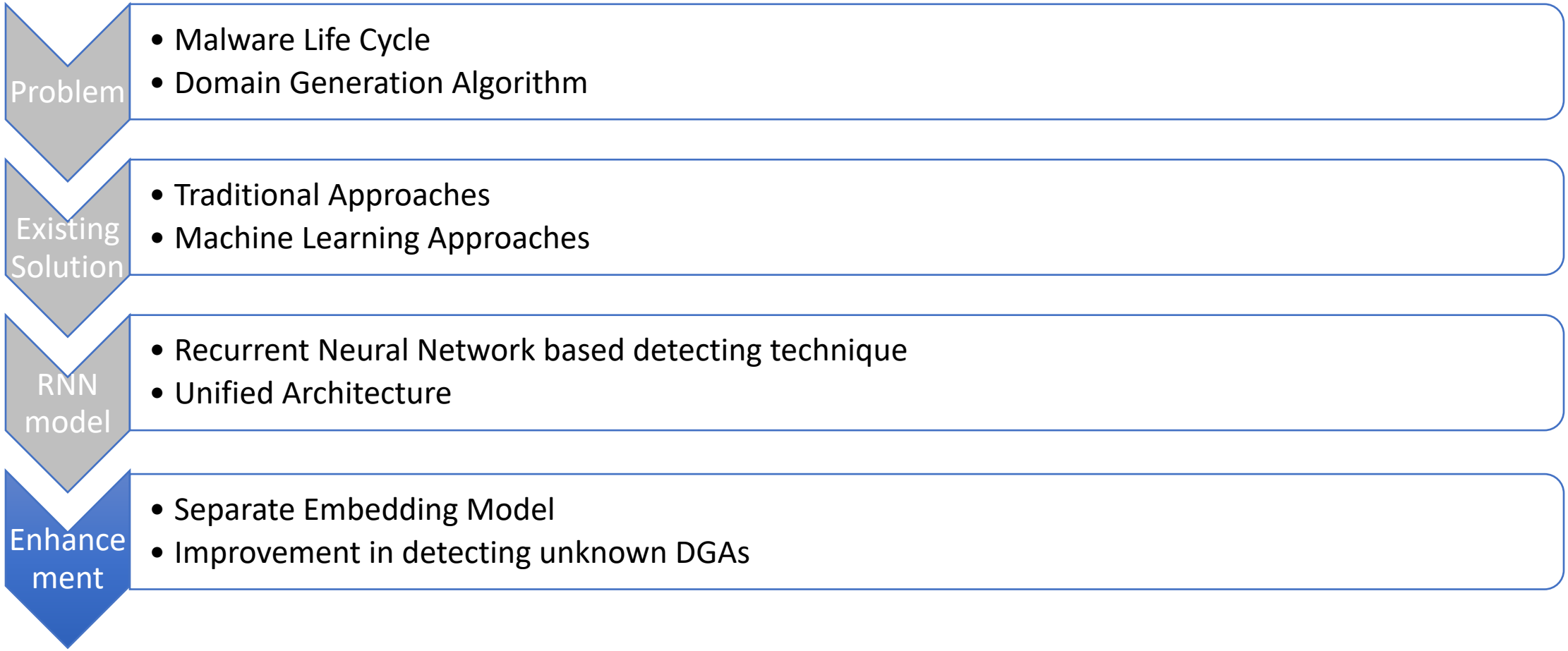
Result

Detection Accuracy for unknown DGA types –

Label	Record Count	Unified Model Accuracy	Sample
chinad	1000	0.996000	qowhi81jvoid4j0m.biz 29cqdf6obnq462yv.com
ramnit	15080	0.718899	jrkaxdlkvhgsiyknhw.com mtsoexdphaqliva.com
shifu	2554	0.438919	urkaelt.info rsymdhk.info

Limitation

- Accuracy suffers for unknown DGA type
- Possible overfitting to training data
- Embedding representation is specific to the training data and is not representative of English language


Problem
• Malware Life Cycle
• Domain Generation Algorithm
Existing Solution
• Traditional Approaches
• Machine Learning Approaches
RNN model
• Recurrent Neural Network based detecting technique
• Unified Architecture
Enhance ment
• Separate Embedding Model
• Improvement in detecting unknown DGAs

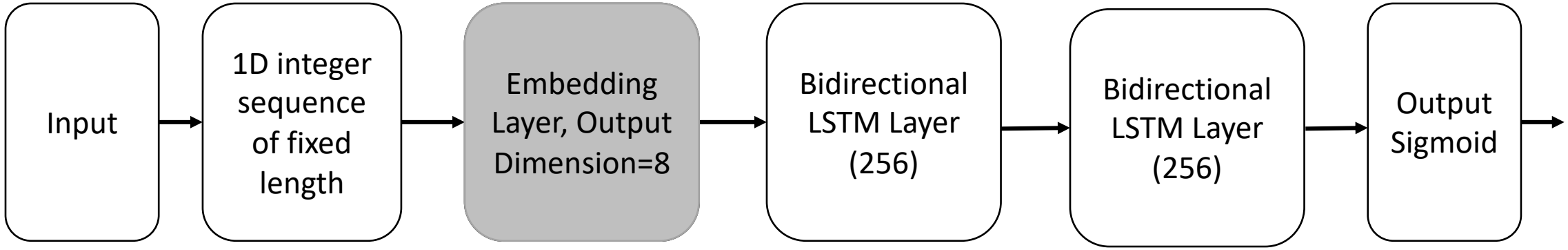
Training Separate Character Embedding Model

- Learn embedding representation to capture the contextual information of the English language by training on articles from popular US newspapers

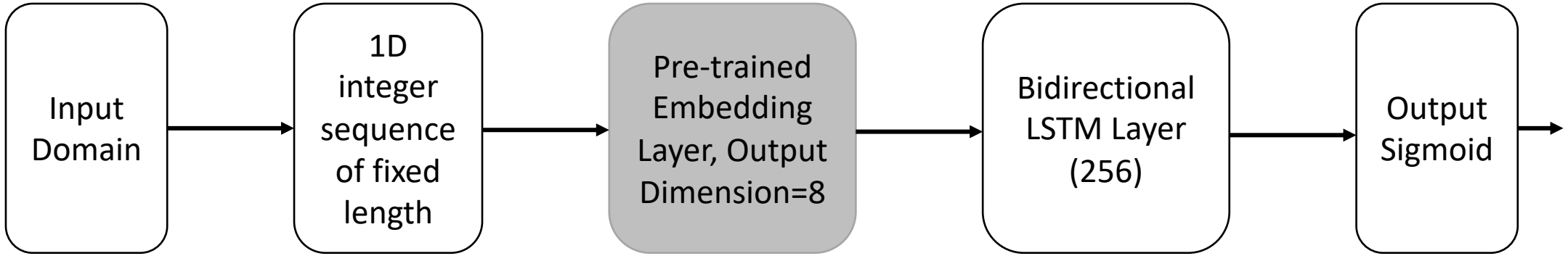
- Use this general representation to transform domain names

- The error is calculated based on the model's ability to predict the next character in the sequence

Learning Character Embedding



Separate Character Embedding based RNN Model Architecture



Result

Test Accuracy for known DGA types

Label	Record Count	Unified Model Accuracy
Benign	750153	0.9946
Malicious	415976	0.9845

Result

Test Accuracy for known DGA types

Label	Record Count	Unified Model Accuracy	Separate Embedding Model Accuracy	F Score Embedding Model
Benign	750153	0.9946	0.9922	0.9875
Malicious	415976	0.9845	0.9889	

Result

Detection Accuracy for unknown DGA types –

Label	Record Count	Unified Model Accuracy
chinad	1000	0.996000
ramnit	15080	0.718899
shifu	2554	0.438919

Result

Detection Accuracy for unknown DGA types –

Label	Record Count	Unified Model Accuracy	Separate Embedding Model Accuracy	% Increase
chinad	1000	0.996000	0.998000	0.2
ramnit	15080	0.718899	0.768833	5.0
shifu	2554	0.438919	0.831245	39.23

Wrapping Up

- LSTM based RNNs are highly effective in detecting DGA
- Our proposed changes can improve detection accuracy for unknown DGA malware
- RNN based detection is proactive rather than reactive