Методы NLP*
для обнаружения веб-атак

Александра Мурзина
machine learning engineer
Positive Technologies

*natural language processing

А про что доклад?

- Почему вам нужен WAF
- Мягкое введение в машинное обучение
- Рекуррентные нейронные сети для обнаружения атак
- Автоэнкодер для поиска аномалий
- Демонстрация PoC

https://github.com/PositiveTechnologies/seq2seq-web-attack-detection

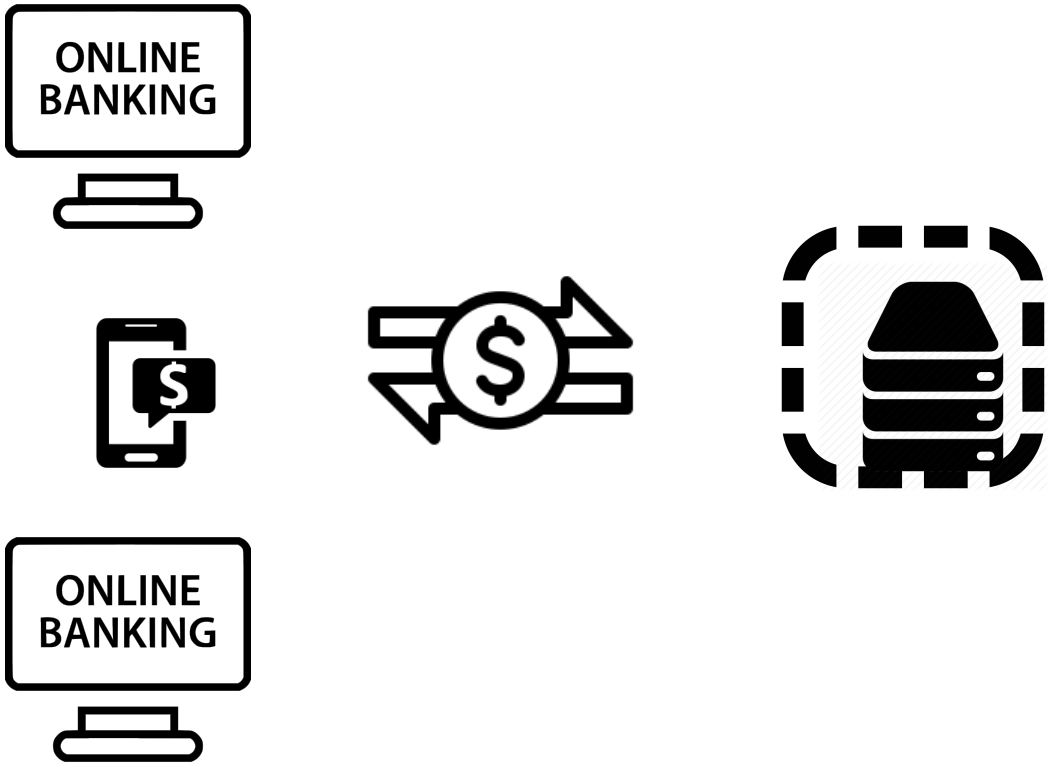
код, данные, докер образ на GitHub

https://aivillage.org/posts/detecting-web-attacks-rnn/

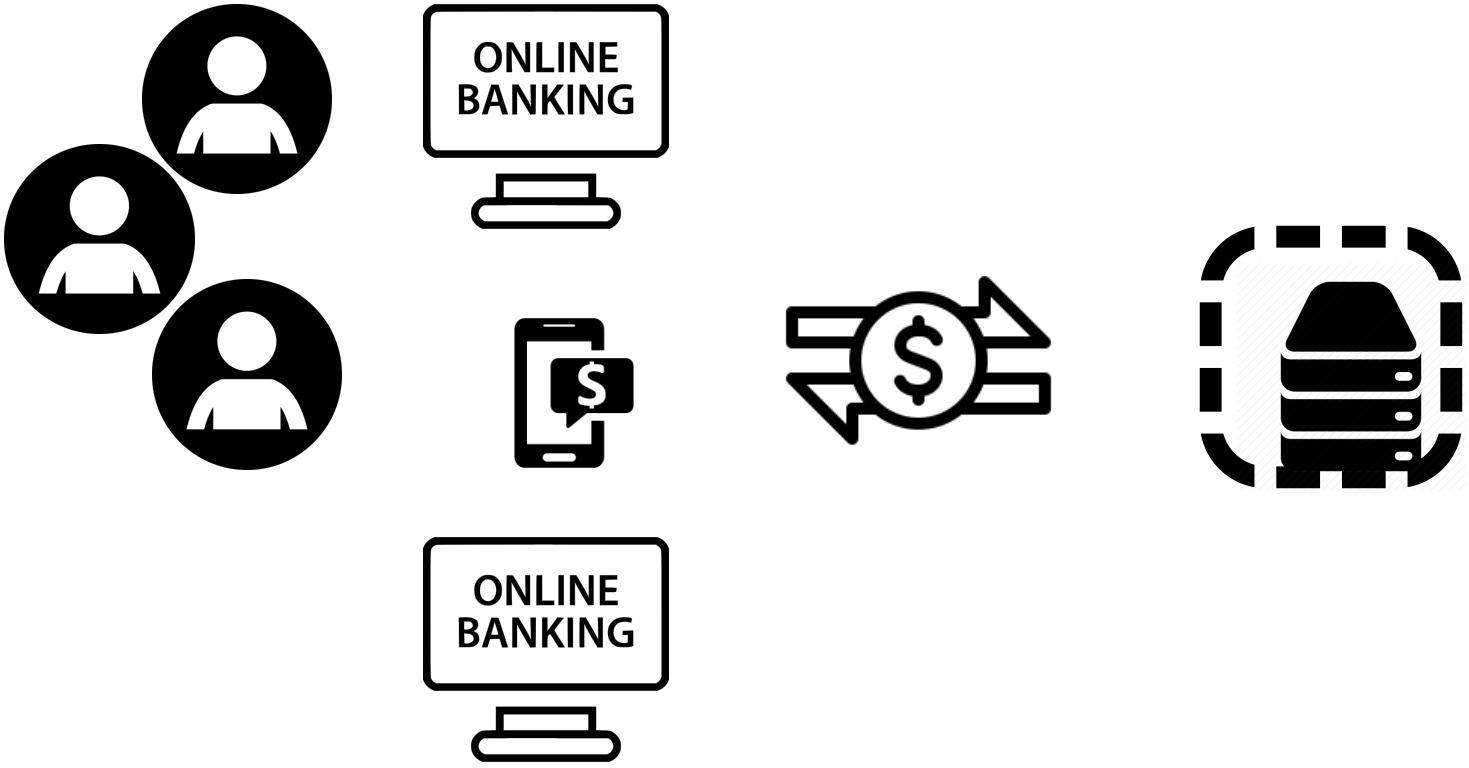
пост с описанием решения для AI Village (DEFCON 26)

Что такое WAF
и
почему он вам нужен

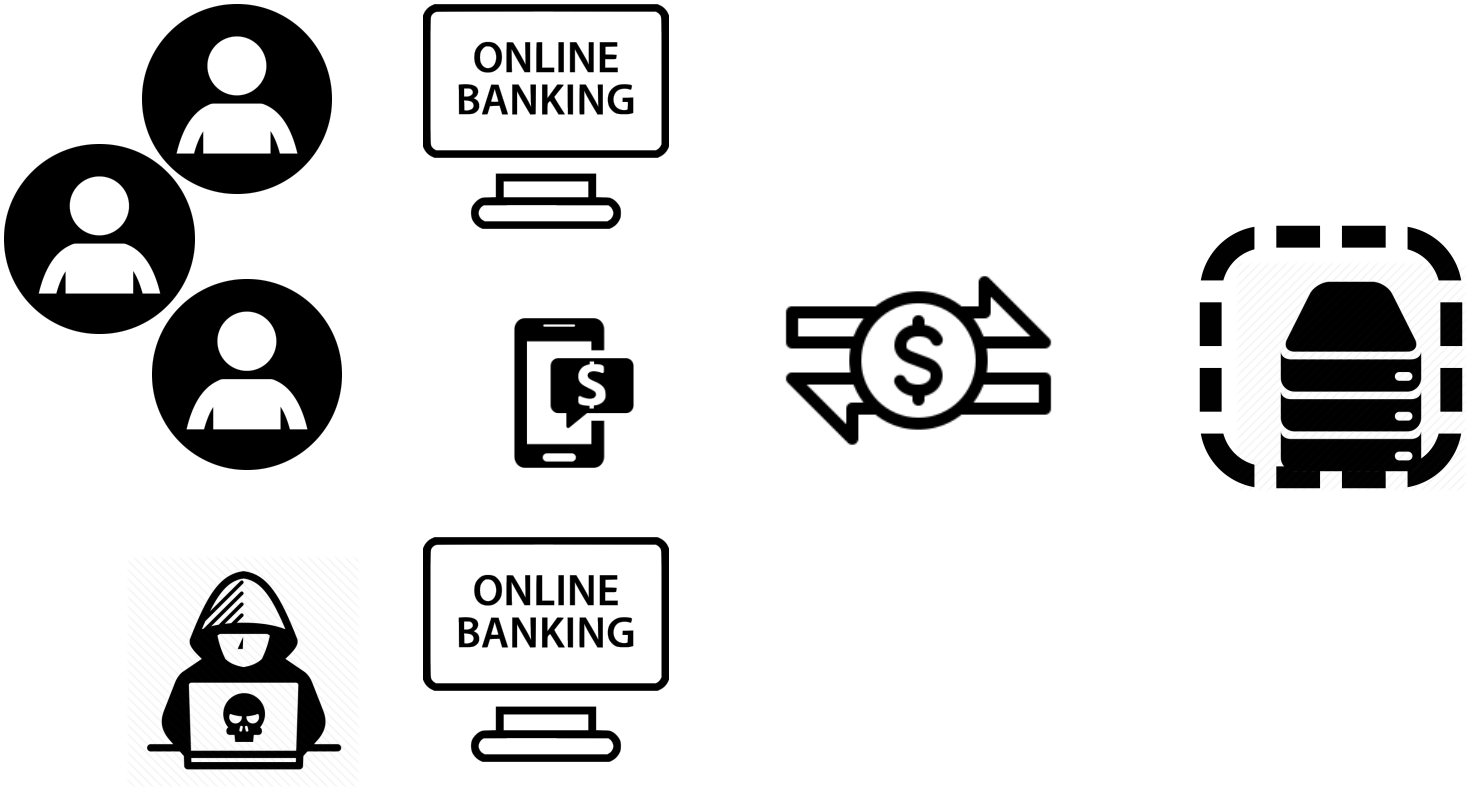
Что такое WAF



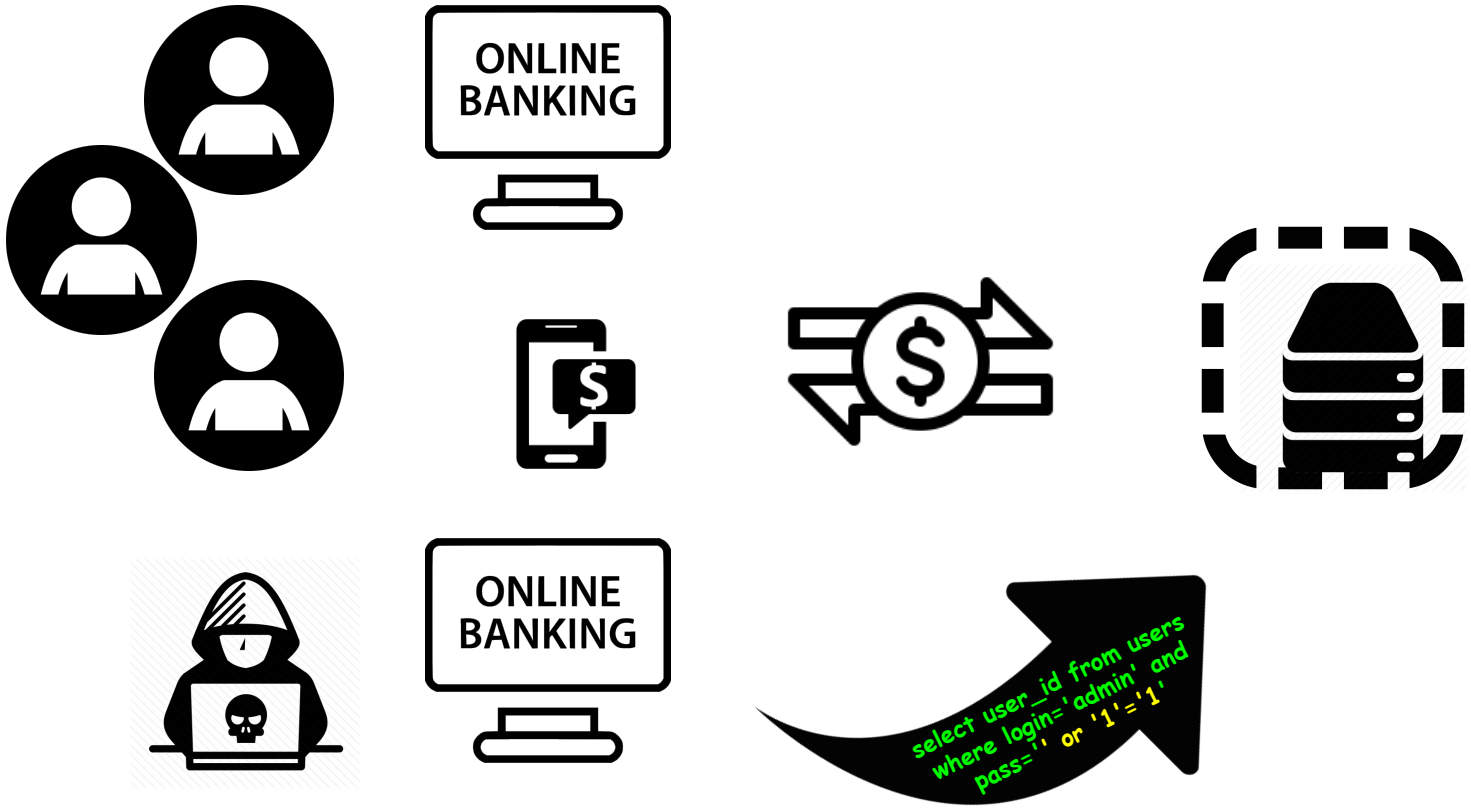
Что такое WAF



Что такое WAF



Что такое WAF



Что такое WAF

Что такое WAF

Но!

Что такое WAF

- OWASP Top 10
- 0-day
- недостаточно компетентных человеческих ресурсов

...

Что такое WAF

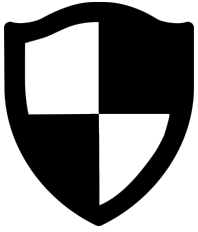
Network Firewall

- L3/L4
- Port
- Ip address
- ACL

Что такое WAF

Network Firewall

- L3/L4
- Port
- Ip address
- ACL

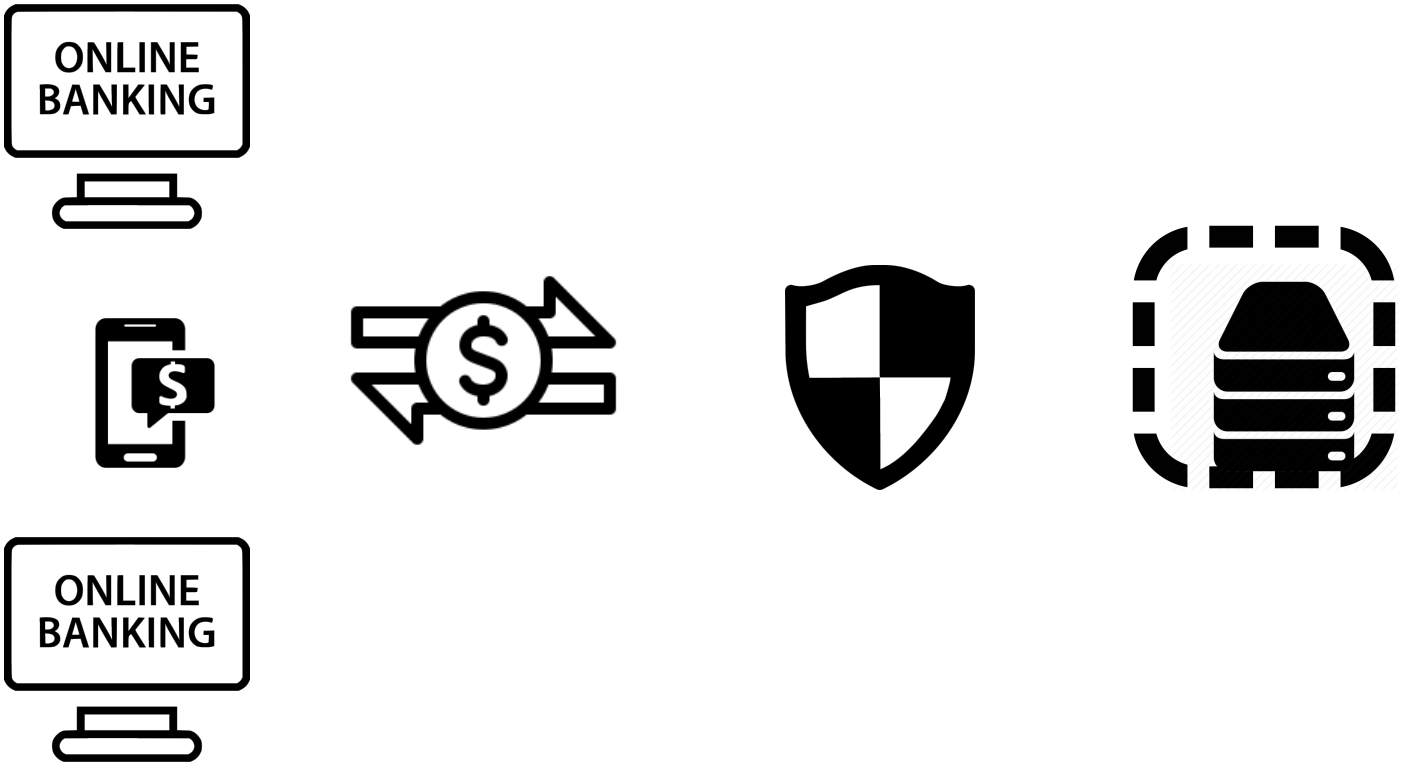


Web Application Firewall

- L7

OWASP Top 10
HTTP Flood
Slowloris
...

Что такое WAF



Типы атак

с точки зрения WAF

Типы атак

с точки зрения WAF

Time series-based:

- Web scraping
- Brute Forcing
- Fingerprinting
- Scanning
- L7 DDoS

HTTP Request/Response-based:

- SQL Injection
- Cross Site Scripting
- XML External Entities Injection
- Path Traversal
- OS Commanding
- Object Injection
- ...

Типы атак

с точки зрения WAF

Time series-based:

- Web scraping
- Brute Forcing
- Fingerprinting
- Scanning
- L7 DDoS

Мы работаем с этими

HTTP Request/Response-based:

- SQL Injection
- Cross Site Scripting
- XML External Entities Injection
- Path Traversal
- OS Commanding
- Object Injection
- ...

SQL Injection?

```
GET
/rest/gadget/1.0/issueTable/jql?num=10&tableContext=jira.table.cols.dashboard&addDefault=true&enableSorting=true&
paging=true&showActions=true&jql=assignee+%3D+currentUser()+AND+resolution+%3D+unresolved+ORDER+BY+priority+DESC%
2C+created+ASC&sortBy=&startIndex=0&_=1533129227137 HTTP/1.1
Host: bugtracking.local
Accept-Encoding: gzip, deflate
Accept: */*
Accept-Language: en
User-Agent: Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; Win64; x64; Trident/5.0)
Connection: close
```

```
assignee = currentUser() AND resolution = unresolved ORDER BY priority DESC, created ASC
```

SQL Injection?

```
GET
/rest/gadget/1.0/issueTable/jql?num=10&tableContext=jira.table.cols.dashboard&addDefault=true&enableSorting=true&
paging=true&showActions=true&jql=assignee+%3D+currentUser()+AND+resolution+%3D+unresolved+ORDER+BY+priority+DESC%
2C+created+ASC&sortBy=&startIndex=0&_=1533129227137 HTTP/1.1
Host: bugtracking.local
Accept-Encoding: gzip, deflate
Accept: */*
Accept-Language: en
User-Agent: Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; Win64; x64; Trident/5.0)
Connection: close
```

```
assignee = currentUser() AND resolution = unresolved ORDER BY priority DESC, created ASC
```

SQL Injection?

```
GET
/rest/gadget/1.0/issueTable/jql?num=10&tableContext=jira.table.cols.dashboard&addDefault=true&enableSorting=true&
paging=true&showActions=true&jql=assignee+%3D+currentUser()+AND+resolution+%3D+unresolved+ORDER+BY+priority+DESC%
2C+created+ASC&sortBy=&startIndex=0&_=1533129227137 HTTP/1.1
Host: bugtracking.local
Accept-Encoding: gzip, deflate
Accept: */*
Accept-Language: en
User-Agent: Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; Win64; x64; Trident/5.0)
Connection: close
```

```
assignee = currentUser() AND resolution = unresolved ORDER BY priority DESC, created ASC
```

Cross Site Scripting?

```
POST /json/topic/?action=save HTTP/1.1
Host: habr.com
Connection: keep-alive
Content-Length: 129
Origin: https://habr.com
X-Requested-With: XMLHttpRequest
User-Agent: Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/66.0.3359.139
Safari/537.36
Content-Type: application/x-www-form-urlencoded; charset=UTF-8
Cookie: PHPSESSID=aasghtnlfls38i1f1n7hb5gn64;

id=&post_type=simple&title=&text=%3Cp%3ECheck+out+my+%3Ca+href%3D%22http%3A%2F%2Fhome.page%22%3Eblog%3C%2Fa%3E!%3
C%2Fp%3E&draft=1
```

Cross Site Scripting?

```
POST /json/topic/?action=save HTTP/1.1
Host: habr.com
Connection: keep-alive
Content-Length: 129
Origin: https://habr.com
X-Requested-With: XMLHttpRequest
User-Agent: Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/66.0.3359.139
Safari/537.36
Content-Type: application/x-www-form-urlencoded; charset=UTF-8
Cookie: PHPSESSID=aasghtnlfls38i1f1n7hb5gn64;

id=&post_type=simple&title=&text=%3Cp%3ECheck+out+my+%3Ca+href%3D%22http%3A%2F%2Fhome.page%22%3Eblog%3C%2Fa%3E!%3
C%2Fp%3E&draft=1
```


```
<p>Check out my <a href="http://home.page">blog</a>!</p>
```

Cross Site Scripting?

```
POST /json/topic/?action=save HTTP/1.1
Host: habr.com
Connection: keep-alive
Content-Length: 129
Origin: https://habr.com
X-Requested-With: XMLHttpRequest
User-Agent: Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/66.0.3359.139
Safari/537.36
Content-Type: application/x-www-form-urlencoded; charset=UTF-8
Cookie: PHPSESSID=aasghtnlfls38i1f1n7hb5gn64;

id=&post_type=simple&title=&text=%3Cp%3ECheck+out+my+%3Ca+href%3D%22http%3A%2F%2Fhome.page%22%3Eblo%3C%2Fa%3E!%3C%2Fp%3E&draft=1
```

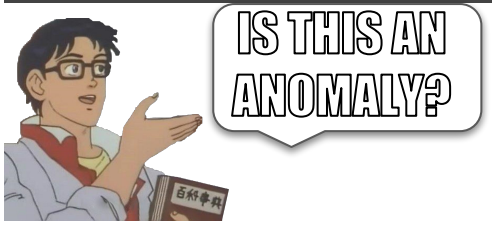

```
<p>Check out my <a href="http://home.page">blog</a>!</p>
```

Normal user registration?

```
POST /index.php/component/users/?task=user.register HTTP/1.1
Host: joomla.local
Connection: close
Accept-Encoding: gzip, deflate
Accept: */*
User-Agent: Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; Win64; x64; Trident/5.0)
Content-Length: 412
Content-Type: application/x-www-form-urlencoded

form[option]=com_users&user[password1]=password&user[username]=hacker&form[email2]=user@example.com&form[password
2]=password&user[email2]=user@example.com&form[task]=user.register&user[password2]=password&user[name]=user&user[
email1]=user@example.com&user[groups][]=7&form[name]=user&user[activation]=0&test=1&form[password1]=password&form
[username]=user&form[email1]=user@example.com&user[block]=0
```

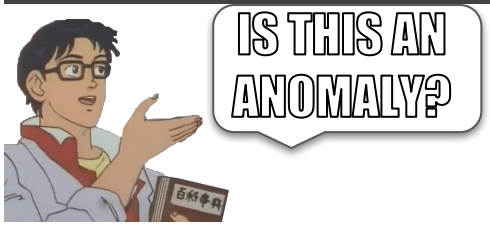
Normal user registration?

```
POST /index.php/component/users/?task=user.register HTTP/1.1
Host: joomla.local
Connection: close
Accept-Encoding: gzip, deflate
Accept: */*
User-Agent: Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; Win64; x64; Trident/5.0)
Content-Length: 412
Content-Type: application/x-www-form-urlencoded

form[option]=com_users&user[password1]=password&user[username]=hacker&form[email2]=user@example.com&form[password2]=password&user[email2]=user@example.com&form[task]=user.register&user[password2]=password&user[name]=user&user[email1]=user@example.com&user[groups][]=7&form[name]=user&user[activation]=0&test=1&form[password1]=password&form[username]=user&form[email1]=user@example.com&user[block]=0
```



Normal user registration?

```
POST /index.php/component/users/?task=user.register HTTP/1.1
Host: joomla.local
Connection: close
Accept-Encoding: gzip, deflate
Accept: */*
User-Agent: Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; Win64; x64; Trident/5.0)
Content-Length: 412
Content-Type: application/x-www-form-urlencoded

form[option]=com_users&user[password1]=password&user[username]=hacker&form[email2]=user@example.com&form[password
2]=password&user[email2]=user@example.com&form[task]=user.register&user[password2]=pass[illegible]d&user[name]=user&user[
email1]=user@example.com&user[groups][]=7&form[name]=user&user[activation]=0&test=1&fo[illegible]ssword1]=password&form
[username]=user&form[email1]=user@example.com&user[block]=0
```



Joomla <3.6.4 Privilege Elevation

Сопоставление шаблонов (pattern matching)

+ Эффективно обнаруживает известные вектора атак
+ Легко поддерживать
+ Вполне быстрые
+ Предсказуемое и интерпретируемое поведение
+ Могут работать из “коробки”

– Тоже можно атаковать (например ReDoS)
– Их возможно легко обойти
– Не эффективны при защите от неизвестных векторов атак (например 0-days)
– Требуют глубоких знаний о веб-безопасности
– Много ложно-положительных срабатываний

Машинное обучение (machine learning)

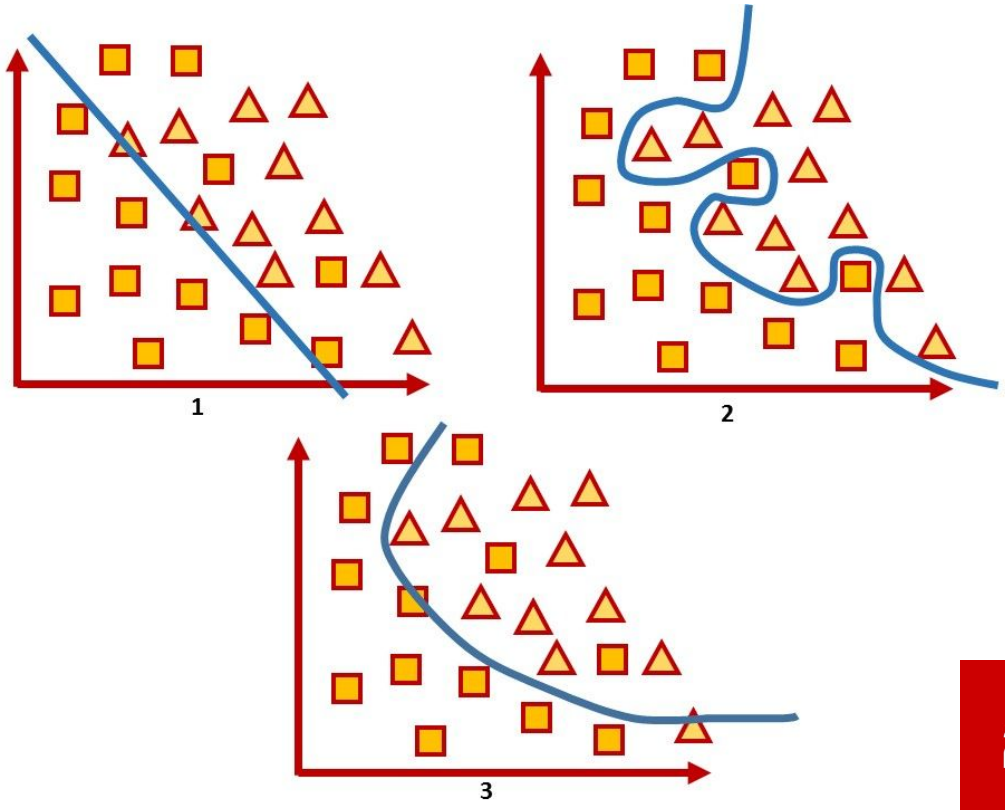
+ Может анализировать данные, которые раньше система не видела
+ Обычно не так просто обойти
+ Обученная модель довольно быстро принимает решение
+ Не требуется глубоких знаний предметной области

- Необходимо время для обучения модели
- Трудно интерпретировать результаты
- Недетерминированное поведение
- Трудно поддерживать

Машинное обучение (machine learning)

Суть машинного обучения:

- все зависимости можно описать какими-то функциями
- любую функцию можно представить в виде полинома
- теперь задача - аппроксимировать полином



Особенности ML в ИБ

- Чаще всего задачи связаны с поиском паттернов в данных и обнаружением аномалий
- Термины outlier и anomaly взаимозаменяемы
- Важное различие между обнаружением выбросов (outlier detection) и обнаружением новизны (novelty detection)
- Нельзя путать их с любыми отсылками на обычные (normal) или стандартные данные (standard)
- Стоимость ошибок I и II рода могут быть значительно разными
- Ошибки I рода заставляют отключать систему
- Модели должны уметь адаптироваться к новым данным

Так, а к вашей задаче как это все относится?

Так, а к вашей задаче как это все относится?

Глобальная цель:

- Создать модель машинного обучения, которая не требует предварительного извлечения признаков
- Модель должна решать задачу обнаружения аномалий в HTTP запросах
- Модель должна отдавать интерпретируемые результаты

Обнаружение аномалий в HTTP запросах

с помощью машинного обучения

Обнаружение аномалий в HTTP запросах

с помощью машинного обучения

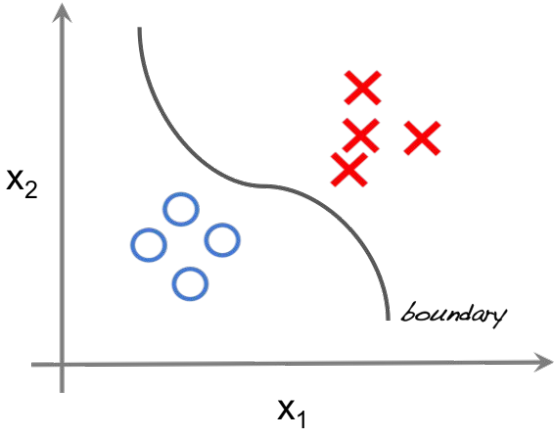
Now is the winter of our discontent
Made glorious summer by this sun of York;
And all the clouds that lour'd upon our house
In the deep bosom of the ocean buried.
Now are our brows bound with victorious wreaths;
Our bruised arms hung up for monuments;
Our stern alarums changed to merry meetings,
Our dreadful marches to delightful measures.

```
POST /vulnbank/online/api.php HTTP/1.1
Host: 10.0.212.25
Connection: keep-alive
Content-Length: 59
Accept: application/json, text/javascript,
*/*; q=0.01
Origin: http://10.0.212.25
X-Requested-With: XMLHttpRequest
User-Agent: Mozilla/5.0 (X11; Linux x86_64)
AppleWeb...
```

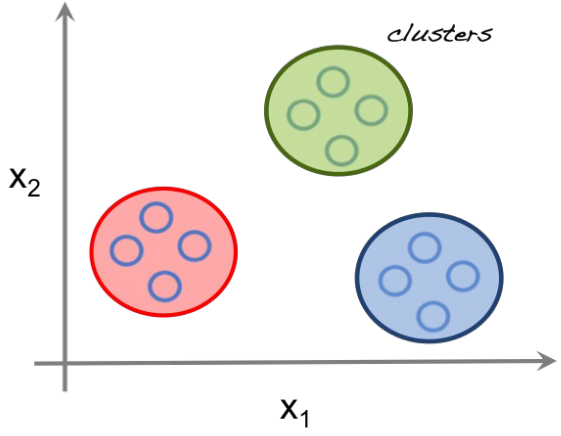
Виды обучения

- Обучение с учителем (supervised learning) - объекты и ответы
- Обучение без учителя (unsupervised learning) - только объекты

Supervised learning

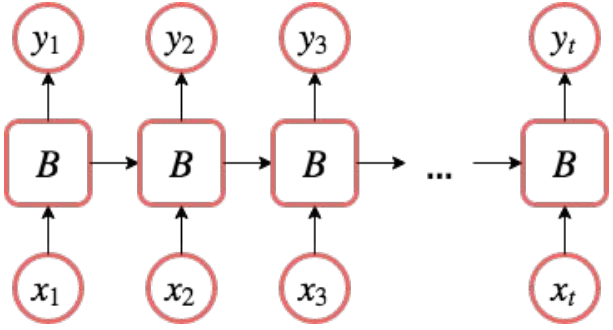


Unsupervised learning



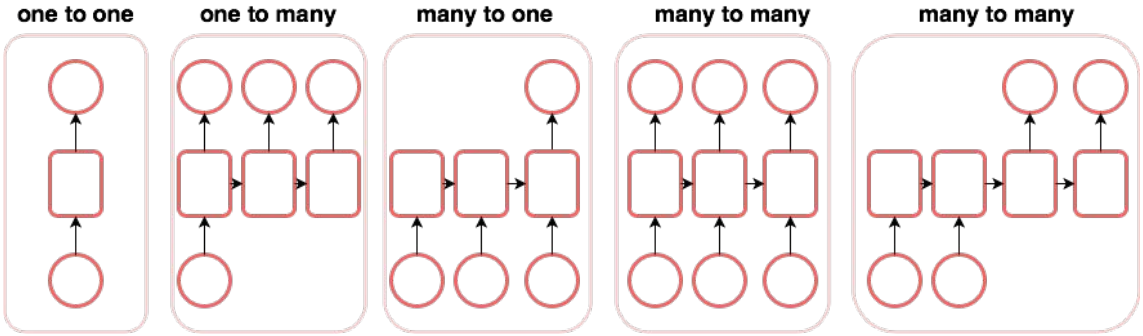
Рекуррентная нейронная сеть (RNN)

- RNN состоит из последовательно связанных блоков, поэтому может обрабатывать последовательности (текст, речь и т.п.)
- Плохо обрабатывает длинные последовательности



X - входная последовательность
B - RNN блок
y - выход сети

Рекуррентная нейронная сеть (RNN)

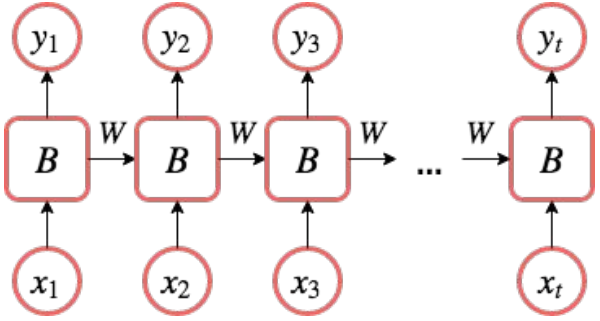


1 - сеть без RNN с входом и выходом фиксированного размера
2 - выход в виде последовательности (вход - картинка, выход - описание картинки)
3 - вход в виде последовательности (сентимент анализ: вход - текст, выход - позитивный или негативный смысл)
4 - вход и выход последовательность (классификация видео, где для каждого кадра - определенный лейбл)
5 - вход и выход последовательность (машинный перевод)

Рекуррентная нейронная сеть (RNN)

RNN плохо обрабатывают длинные последовательности

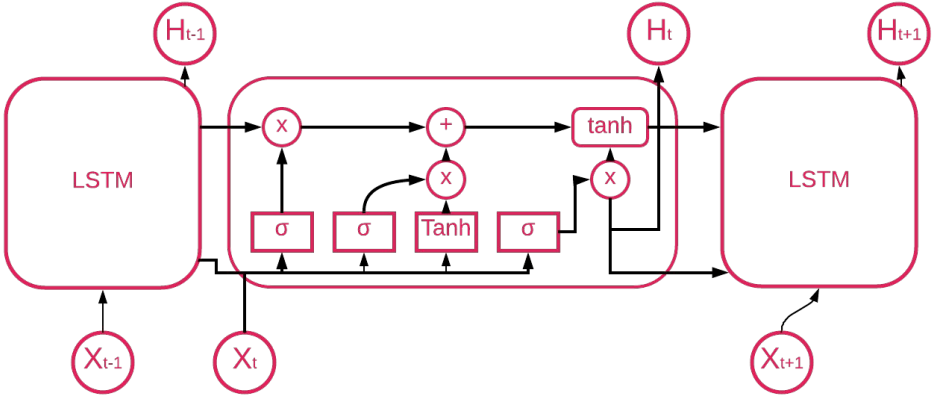
- Использование одной матрицы весов при переходе от одного блока к другому
- Обратное распространение ошибки: градиент ~ матрица весов в большой степени
- Градиент очень большой или очень маленький
- Проблема: взрывающиеся и затухающие градиенты
- Результат: информация не проходит через всю сеть



Рекуррентная нейронная сеть (RNN)

Один из способов решения проблемы затухающих градиентов - использование других RNN архитектур: LSTM или GRU.

В LSTM (Long short-term memory, долгая краткосрочная память) информация проходит через всю сеть.



Классификатор на основе LSTM

- HTTP - это текстовый протокол
- Каждая строка как отдельное предложение
- Каждая строка - последовательность, т.к. значение параметра зависит от имени

```
POST /vulnbank/online/api.php HTTP/1.1
Host: 10.0.212.25
Connection: keep-alive
Content-Length: 59
Accept: application/json, text/javascript,
*/*; q=0.01
Origin: http://10.0.212.25
X-Requested-With: XMLHttpRequest
User-Agent: Mozilla/5.0 (X11; Linux x86_64)
AppleWeb...
```

```
Now is the winter of our discontent
Made glorious summer by this sun of York;
And all the clouds that lour'd upon our house
In the deep bosom of the ocean buried.
Now are our brows bound with victorious wreaths;
Our bruised arms hung up for monuments;
Our stern alarums changed to merry meetings,
Our dreadful marches to delightful measures.
```

Классификатор на основе LSTM

- Собрать нормальные запросы с приложения
- Сгенерировать вредоносные запросы
- Построить классификатор

Классификатор на основе LSTM

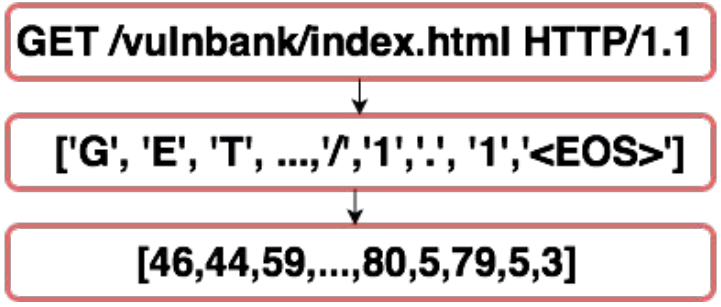
- Собрать нормальные запросы с приложения
- Сгенерировать вредоносные запросы
- Построить классификатор

Объект	Метка
`GET /api/posts?author=mallory&category='%20or%20'1'%20=%20'`	1
`GET /api/posts?author=alice&category=sports`	0

Классификатор на основе LSTM

Преобразование данных:

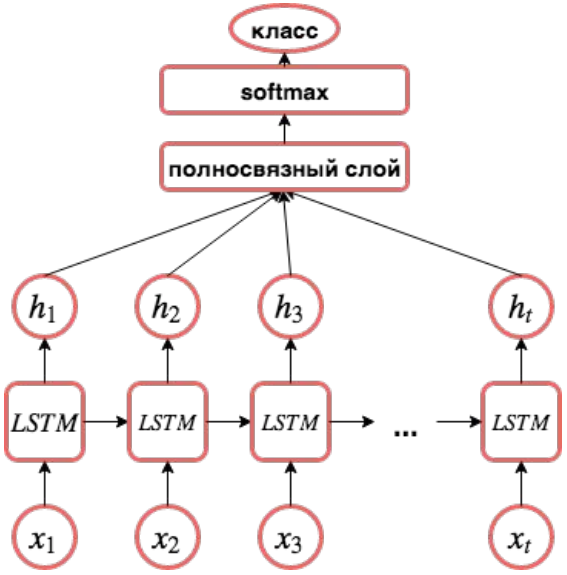
- запрос
- вектор символов
- вектор чисел



Классификатор на основе LSTM

Классификатор состоит из:

- LSTM сети
- Полносвязный линейный слой (fully connected layer)
- Softmax для получения вероятностей классов



Классификатор на основе LSTM

Положительные моменты:

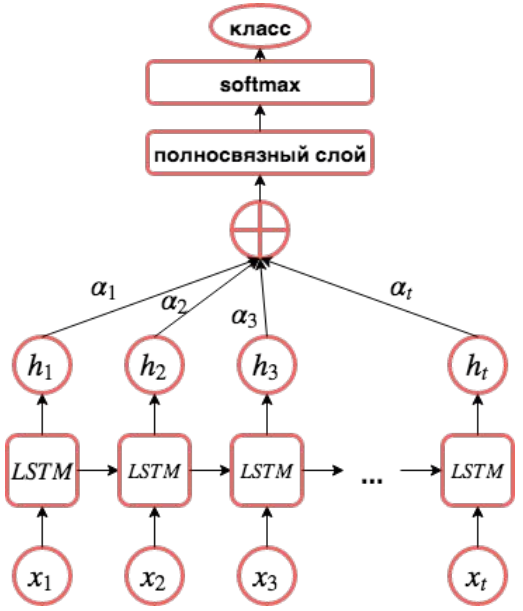
- LSTM используют для обработки последовательностей
- Построили классификатор
- Оценили результатат
- Все хорошо, но ...

Есть некоторые проблемы:

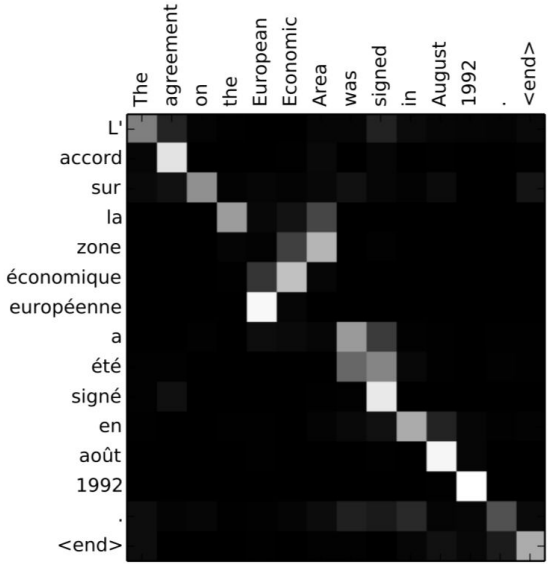
- Результаты не интерпретируемые
- Сложно генерировать вредоносный класс
- Нужна ручная разметка

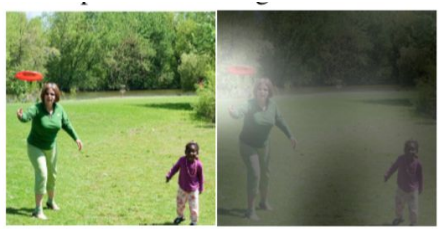
LSTM и механизмы внимания

Механизмы внимания (Attention) позволяют определять элементы (отдельные символы, слова), которые сильнее всего влияют на результат предсказания модели.

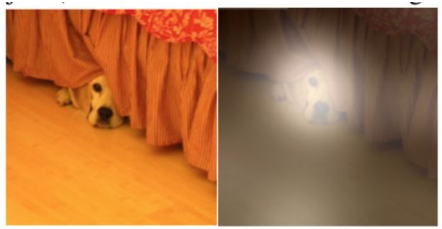


LSTM и механизмы внимания





A woman is throwing a frisbee in a park.



A dog is standing on a hardwood floor.

A little girl sitting on a bed with a teddy bear.

A group of people sitting on a boat in the water.

1. Bahdanau, D., Cho, K., & Bengio, Y. (2014)
2. Xu K. et al. (2015)

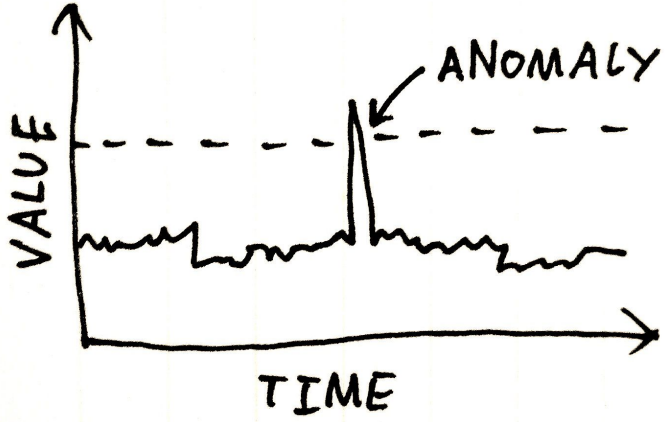
LSTM и механизмы внимания

- Добавили слой Attention
- Улучшили качество классификации
- Улучшили интерпретируемость модели
- Но не решили другие проблемы с классификацией: генерация вредоносного класса, ручная разметка

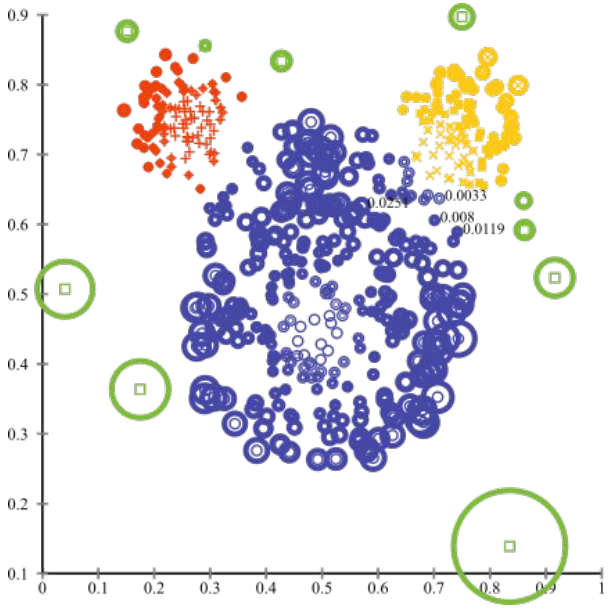
Определение аномалий

Задача определения аномалий (Anomaly Detection) - задача обучения без учителя, т.е. у объектов нет метки.

Обучение осуществляется на нормальных данных. Те данные, что будут сильно отклоняться - аномалии.



Определение аномалий



Определение аномалий

Наша задача хорошо обобщается проблемой определения аномалий

- Не нужно размечать данные
- Не нужно генерировать вредоносные примеры
- Вредоносных примеров сильно меньше, чем нормальных запросов

Определение аномалий

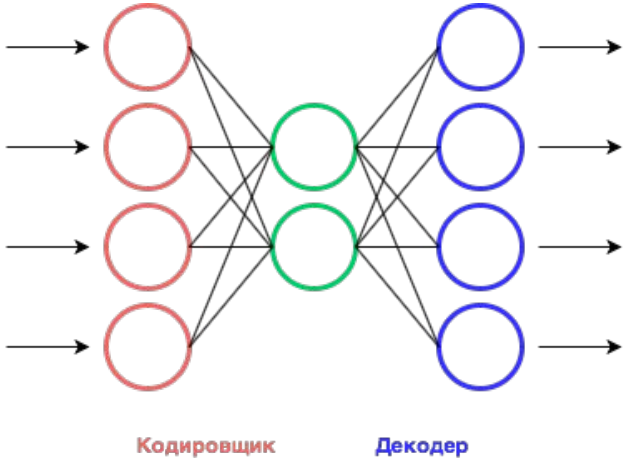
Идея по прежнему в том, чтобы анализировать последовательность символов.

И обнаруживать подпоследовательность символов не типичную для текущей последовательности.

Объект	Метка
`GET /api/posts?author=mallory&category='%20or%20'1'%20=%20'`	1
`GET /api/posts?author=alice&category=sports`	0

Автокодировщик

Задача кодировщика состоит в том, чтобы сжать входные данные в специальное представление.

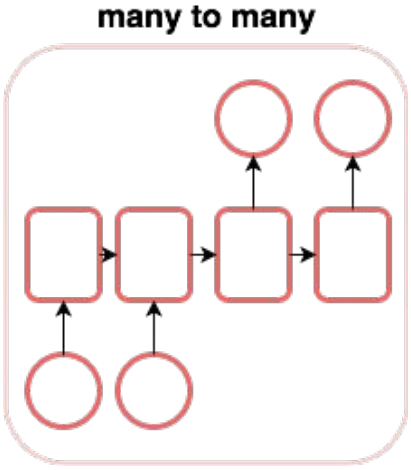


Задача декодера - научиться правильно декодировать из специального представления в исходную форму.

Seq2Seq

Seq2Seq модель для машинного перевода:

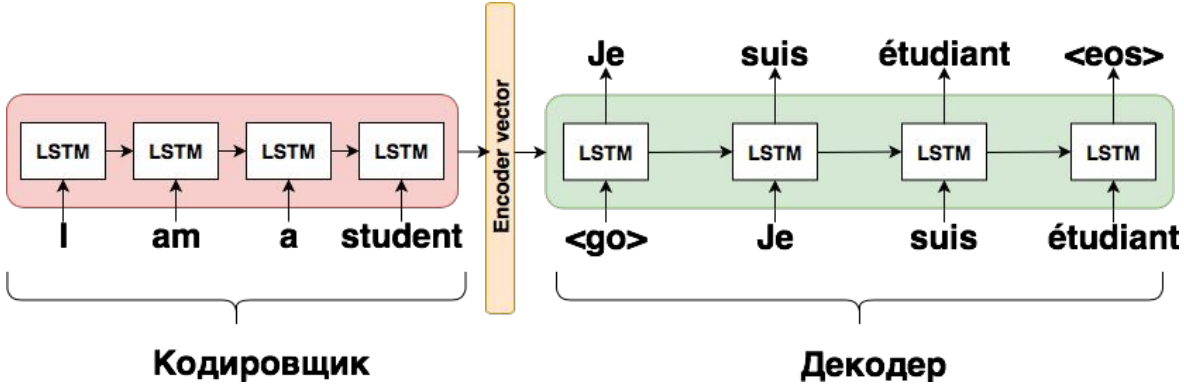
- Состоит из двух многослойных LSTM: кодировщик и декодер



Seq2Seq

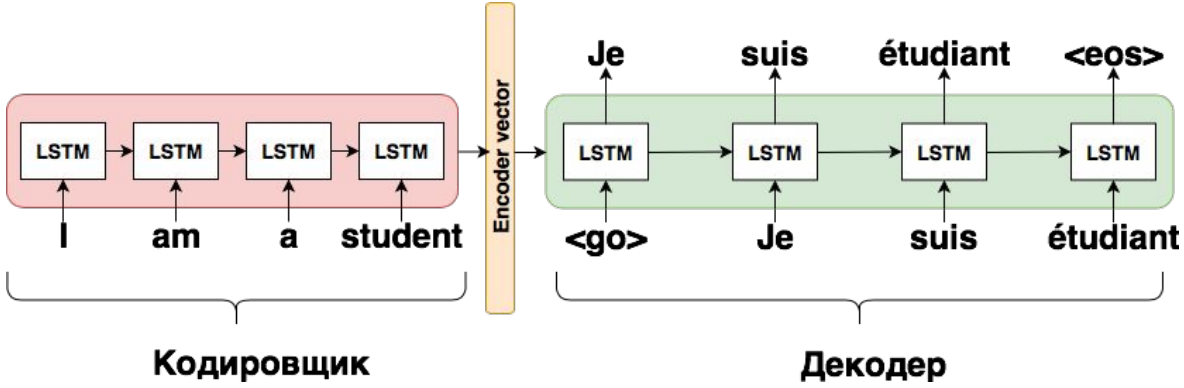
Seq2Seq модель для машинного перевода:

- Состоит из двух многослойных LSTM: кодировщик и декодер



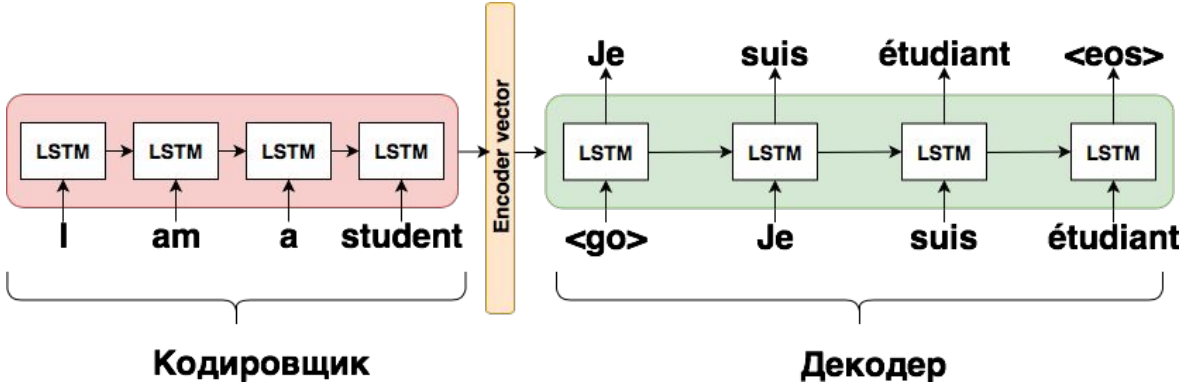
Определение аномалий

- Не слова, а символы
- Целевая переменная - та же последовательность, что и на входе
- Выход - вероятность символа в данном месте последовательности



Определение аномалий

Кроме вероятностей на выходе получаем значение функции потерь (loss). Этот показатель используется при принятии решения. Если значение выше порога, то это аномалия.



Определение аномалий

- Модель выводит вероятности каждого символа и значение функции потерь
- Все запросы с высоким значением функции потерь считаются вредоносными
- Для таких запросов вероятности для аномальных символов будут низкими

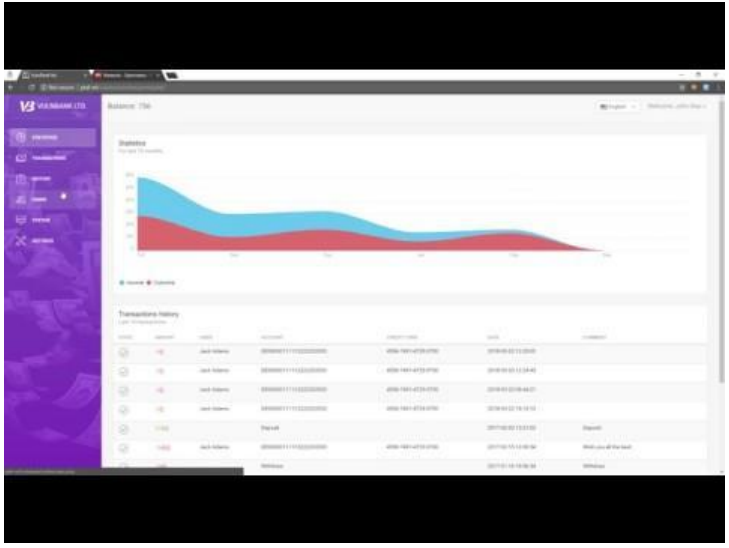
Определение аномалий

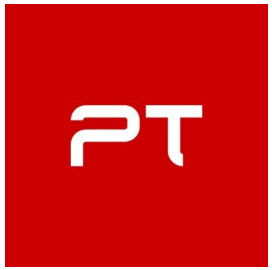
Визуализация аномального запроса

```
POST /vulnbank/online/api.php HTTP/1.1
Host: 10.0.212.25
User-Agent: Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:59.0) Gecko/20100101 Firefox/59.0
Accept: application/json, text/javascript, */*; q=0.01
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
Referer: http://10.0.212.25/vulnbank/online/login.php
Content-Type: application/x-www-form-urlencoded
X-Requested-With: XMLHttpRequest
Content-Length: 76
Cookie: PHPSESSID=mlacs0uiou344i3fa53s7raut6
Connection: keep-alive

type=user&action=login&username=none'+union+select+1,2,login,password,5,6,7,NULL,NULL,10,11,12,13,14,15,16,17+from+users+limit+1+--1
```

Демонстрация PoC

It’s Showtime!





Александра Мурзина

machine learning engineer

AMurzina@ptsecurity.com

https://github.com/PositiveTechnologies/seq2seq-web-attack-detection

код, данные, докер образ на GitHub

https://aivillage.org/posts/detecting-web-attacks-rnn/

пост с описанием решения для AI Village (DEFCON 26)