An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djibrilla.amadoukountche@gmail.com

15/03/2018

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djib-

rilla.amadoukountch

ata, Data, D

Clients HTTP protocol

Message: Request/Respoi

Parameters

Intrusion detection in Web

Web intrusion: Origin the problem

Web intrusion: How solve the problem

/hat is an IDPS ?

IDPS Detection Method: Behavioral

Agenda

Data, Data, Data

Intrusion detection in Web applications

RoCaWeb

Conclusion

Reading list

IDPS Detection Method: Behavioral Intrusion detection

An application of

Machine Learning to Intrusion Detection and Prevention in Web **Applications**

Event:Ruhrgebiet Al & Data Science Meetup Author: Djibrilla Amadou Kountche

Email: djibrilla.amadoukountch



An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djib-

Event:Ruhrgebiet

rilla.amadoukountch

Data, Data

Clients

HTTP protoco

Message: Request/Respon

Parameters

Intrusion detection in Web

application

Web intrusion: Origin of the problem

Web intrusion: How t solve the problem

What is an IDPS 1

Knowledge based

Method:
Behavioral

Data, Data, Data

Suppose, as a Data scientist you have been given:

- ► Logs from web servers
- Logs from developpers (functional tests logs)
- Logs from applications
- Packet captures from your Network admin (or maybe NSA and BND)

And you are asked to get usefull information from these data for intrusion detection and prevention in web applications.

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche Email: diib-

rilla.amadoukountch

Data, Data, Data

Web application Clients

HTTP protocol

Message:

Request/Resp Parameters

Intrusion detection in Web

pplications

Web intrusion: Origin of the problem Web intrusion: How to solve the problem

hat is an IDPS ?

Method:
Behavioral
Intrusion detection
Algorithmic approaches

Web application

Our first problem will be a definition problem. What is a Web application ?

An application of Machine Learning to Intrusion Detection and Prevention in Web

Applications

Event:Ruhrgebiet

Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djib-

rilla.amadoukountch

Data, Data,

Web application Clients

HTTP protocol
Message:

Message: Request/Response arameters

Intrusion detection in Web

applications

Web intrusion: Origin of the problem

Web intrusion: Origin of the problem

Web intrusion: How to

what is an IDPS ?

IDPS Detection Method:

Behavioral Intrusion detection Algorithmic approaches

Web application

Our first problem will be a definition problem. What is a Web application ?

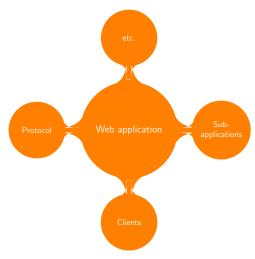


Figure: An attempt to represent a web application

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djibrilla.amadoukountch

Data, Data, D

Web application

Clients
HTTP protocol

Message: Request/Respon

i arameter:

Intrusion detection in Web

Web intrusion: Origin of the problem

What is an IDPS

Knowledge base

IDPS Detection Method: Behavioral Intrusion detection

Clients

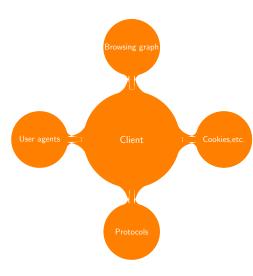


Figure: An example of client model

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djib-

rilla.amadoukountch

Data, Data,

Web applicat

HTTP protor

Message: Request/Respons

Parameters

Intrusion detection in Web

Web intrusion: Origin of the problem

Web intrusion: How t solve the problem

What is an IDPS Knowledge based

IDPS Detection
Method:
Behavioral
Intrusion detection
Algorithmic approaches

HTTP protocol

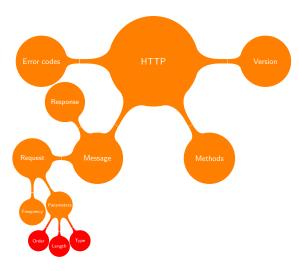


Figure: Data obtained from the HTTP protocol

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djibrilla.amadoukountch

ata, Data, D

Veb application

HTTP protocol

Message: Request/Respon

Request/Respor

Intrusion detection in Web

Web intrusion: Origin or the problem

solve the problem
What is an IDPS ?

Knowledge bases

IDPS Detection
Method:
Behavioral
Intrusion detection
Algorithmic approaches

Message

start-line

```
generic-message = start-line
          *message-header
          CRLF
           [ message-body ]
```

= Request-Line | Status-Line

An application of Machine Learning to Intrusion Detection and Prevention in Web

Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche

Email: djibrilla.amadoukountch

Data, Data,

Web application Clients

HTTP protocol

Message:

Request/Response

Parameters

Intrusion detection in Web

applications

Web intrusion: Origin of the problem

the problem

Web intrusion: Origin of the problem

Web intrusion: How to solve the problem

nat is an IDPS ?

Method:
Behavioral
Intrusion detection
Algorithmic approaches

Request

POST / HTTP/1.1

Host: meinesite.de

/path/code?param1=value1¶m2=value2

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djib-

Event:Ruhrgebiet

rilla.amadoukountch

Message:

Request/Response

Response

HTTP/1.1 200 OK

(trocated)

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djib-

rilla.amadoukountch

Message:

Request/Response

Parameters

Consider captured values from a POST Request

```
/ path/code? param_1 = value_{11} \& param_2 = value_{21} / path/code? param_1 = value_{12}
```

. . .

```
/path/code?param_1 = value_{1n}\&param_2 = value_{2n}
Example
```

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

AI & Data Science Meetup Author: Djibrilla Amadou Kountche

Event:Ruhrgebiet

Email: djibrilla.amadoukountch

Data, Data, I

Web application

Clients
HTTP protocol

Message:

Request/Response

Parameters

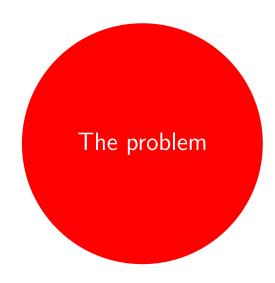
Intrusion detection in Web

Web intrusion: Orig the problem

Web intrusion: How to solve the problem

hat is an IDPS ?

IDPS Detection
Method:
Behavioral
Intrusion detection
Algorithmic approaches



An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djib-

rilla.amadoukountch

Data, Data,

Web application

HTTP protocol Message:

Request/Response

Parameters

Intrusion detection in Web

Web intrusion: Origin of the problem

solve the problem

What is an IDPS

IDPS Detection Method: Behavioral

How many of you follows Secure Development Life cycles?

1. TRAINING	2. REQUIREMENTS	3. DESIGN	\$. IMPLEMENTATION	5. VERIFICATION	6. RELEASE	7. RESPONSE
Core Security Training	Establish Security Requirements	Establish Design Requirements	8. Use Approved Tools	11. Perform Dynamic Analysis	14. Create an Incident Response Plan	Execute Incident Response Plan
	Create Quality Gates/Bug Bars	Perform Attack Surface Analysis/ Reduction	9. Deprecate Unsafe Functions	12. Perform Fuzz Testing	15. Conduct Final Security Review	
	Perform Security and Privacy Risk Assessments	7. Use Threat Modeling	10. Perform Static Analysis	13. Conduct Attack Surface Review	16. Certify Release and Archive	

Figure: Microsoft SDL. Source

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche

Email: djibrilla.amadoukountch

Data, Data,

Clients

Message:

Parameters

Intrusion detection in Web

lications

Web intrusion: Origin of

the problem
Web intrusion: How to

What is an IDPS

IDPS Detection
Method:
Behavioral
Intrusion detection
Algorithmic approach 5

Complex Web applications?

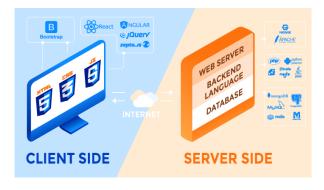


Figure: Example of the diversity of web technologies. Source

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djib-

rilla.amadoukountch

Data, Data, [

vveb applicatio Clients

HTTP protocol Message:

Parameters

Intrusion detection in Web

Web intrusion: Origin of the problem

solve the problem
What is an IDPS ?

IDPS Detection Method:

An attacker has many opportunities...

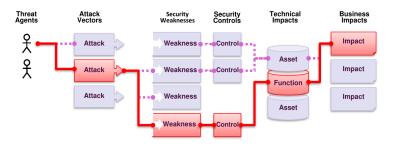


Figure: An attacker only needs the weakest link. Source

An application of Machine Learning to Intrusion Detection and Prevention in Web **Applications**

Event:Ruhrgebiet Al & Data Science Meetup

Author: Djibrilla Amadou Kountche Email: diibrilla.amadoukountch

Web intrusion: Origin of

the problem

An attacker only needs the weakest point.



Figure: An attacker only needs the weakest link.

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet
AI & Data Science
Meetup
Author: Djibrilla
Amadou Kountche
Email: djib-

rilla.amadoukountch

Data, Data, I

Web application

HTTP protocol Message:

Message: Request/Response Parameters

Intrusion detection in Web

Web intrusion: Origin of the problem

Web intrusion: How to solve the problem
What is an IDPS ?

IDPS Detection Method: Behavioral

How to solve the problem

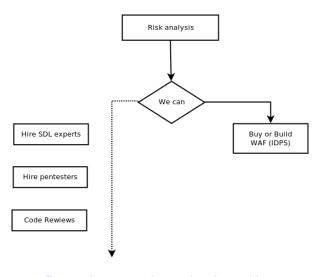


Figure: An approach to solve the problem

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

AI & Data Science Meetup Author: Djibrilla Amadou Kountche

Event:Ruhrgebiet

Email: djibrilla.amadoukountch

Data, Data

Clients

Message:

Request/Respo Parameters

Intrusion detection in Web

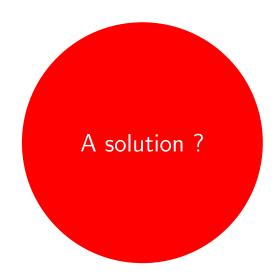
Web intrusion: Origin o

Web intrusion: How to solve the problem

What is an IDPS ?

IDPS Dete

Method:
Behavioral
Intrusion detection
Algorithmic approaches



An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet

AI & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djib-

rilla.amadoukountch

Data, Data,

Web application Clients

HTTP protocol Message:

Request/Response

Parameters

Intrusion detection in Web

Web intrusion: Origin of the problem

the problem

Web intrusion: How to

solve the problem
What is an IDPS ?

IDPS Detecti

Method: Behavioral

What is an IDPS?

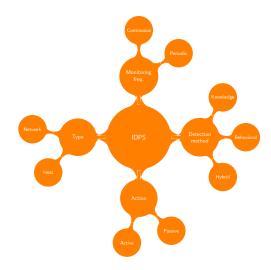


Figure: Components and different types of IDPS

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djib-

rilla.amadoukountch

Data, Data,

Web application

Message: Request/Respon

Parameters

Intrusion detection in Web

Web intrusion: Origin of the problem

solve the problem

What is an IDPS ?

IDPS Detection
Method:
Behavioral
Intrusion detection
Algorithmic approaches

IDPS Detection Method: Knowledge-based



Figure: Knownledge-based intrusion detection. Source

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet

Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djibrilla.amadoukountch

Knowledge based

IDPS Detection Method: Knowledge-based



Issues with the oracle:

- Frequency of updates of the databases
- Polymorphic malicious programs
- Zeroday attacks

An application of Machine Learning to Intrusion Detection and Prevention in Web **Applications**

Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djibrilla.amadoukountch

Event:Ruhrgebiet

Knowledge based

Behavioral Intrusion detection



Figure: Is something abnormal with Mr. Zuckerberg? Source

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djibrilla.amadoukountch

IDPS Detection Method: Behavioral Intrusion detection

Behavioral Intrusion detection



Issues with the normal behavior:

- Is it really possible to model (cover) the behavior?
- What about new behaviors (new trends) ?
- False positive rate

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djib-

rilla.amadoukountch

Data Data

Web application

Clients

Message: Request/Respon

Parameters

Internal and a

Intrusion detection
in Web
applications

Web intrusion: Origin of the problem

olve the problem What is an IDPS

What is an IDPS

IDPS Detection Method: Behavioral Intrusion detection Algorithmic approach 50

Algorithmic approaches used for Behavioral Intrusion detection



Figure: Algorithmic approaches

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djibrilla amadoukountch

Data, Data,

Web application Clients

HTTP protocol

Message:
Request/Respons

Parameters

Intrusion detection in Web

Web intrusion: Ori the problem

Web intrusion:

What is an IDPS

Knowledge based

Method: Behavioral Intrusion detection

What we consider as Learning

Imporant remark

Learning is the process where we build a model representing the normal behavior of the application.

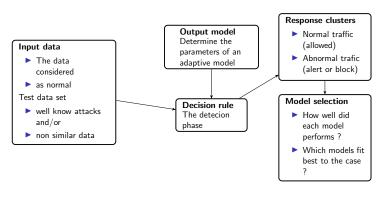


Figure: The overall view of the learning phase

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djibrilla amadoukountch

ata Data D

Web application Clients HTTP protocol

Message: Request/Resp Parameters

Intrusion detecti in Web applications

applications
Web intrusion: O

the problem

Web intrusion: How t solve the problem

What is an IDPS Knowledge bases

IDPS Detection Method: Behavioral Intrusion detection Algorithmic approach 20 / 50

How well are we learning

Performance measures

- Classification error rate
- ► False positive rate
- ► False negative rate /
- Execution time

	True class		
	Normal	Abnormal	
Normal	TP	FP	
Abnormal	FN	TN	

An application of Machine Learning to Intrusion Detection and Prevention in Web **Applications**

Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djib-

rilla.amadoukountch

Event:Ruhrgebiet

Librairies





The most interesting libraries for anomaly detection

- Implemented in Java
- ► License GPLv3
- Anomaly detection algorithms:
 - Distance-based
 - Local factor family
 - Angle-based
 - clustering based
 - etc.

An application of Machine Learning to Intrusion Detection and Prevention in Web **Applications**

Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djibrilla.amadoukountch

Event:Ruhrgebiet

Librairies

Lingpipe

- Implemented in Java
- Free and Paid License
- Anomaly detection algorithms:
 - General machine learning algorithms
 - Sentiment analysis algorithms (LDA)

An application of Machine Learning to Intrusion Detection and Prevention in Web **Applications**

Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djib-

Event:Ruhrgebiet

rilla.amadoukountch

Librairies



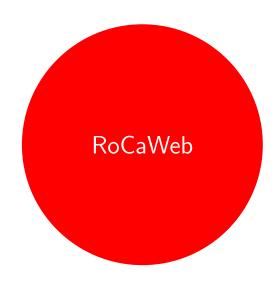
- ▶ Implemented in LuaJIT C, C++
- BSD license
- GPU support
- Anomaly detection algorithms:
 - Statistical libraries such as cephes
 - General purpose machine learning algorithms
- Integration in Modsecurity

An application of Machine Learning to Intrusion Detection and Prevention in Web **Applications**

Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: diib-

Event:Ruhrgebiet

rilla.amadoukountch



An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djib-

Event:Ruhrgebiet

rilla.amadoukountch

Data, Data,

Web application Clients

HTTP protoco

Request/Response

Parameters

Intrusion detection in Web

Web intrusion: Origin o the problem

Web intrusion: How to solve the problem

What is an IDPS ?

Knowledge based

Method:
Behavioral

RoCaWeb

About

- A research project funded by a RAPID grant
- Aimed building a WAF using Behavioral mehtods
- Implemented in Java, Lua, Scala
- Open source GPLv3 licence
- https://github.com/dakountche/RoCaWeb

An application of Machine Learning to Intrusion Detection and Prevention in Web **Applications**

Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: diib-

Event:Ruhrgebiet

rilla.amadoukountch

RoCaWeb: Architecture

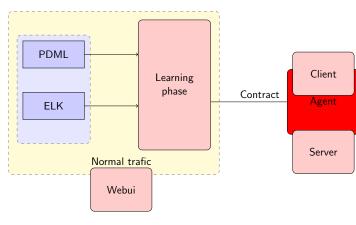


Figure: The RoCaWeb Architecture

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djib-

rilla.amadoukountch

Event:Ruhrgebiet

Data, Data, D

Web application
Clients

Message: Request/Resp Parameters

Intrusion detection in Web

Web intrusion: (

the problem Web intrusion: Ho

Vhat is an IDPS ?

Knowledge based

IDPS Detection
Method:
Behavioral
Intrusion detection
Algorithmic approaches

RoCaWeb: Learning algorithms

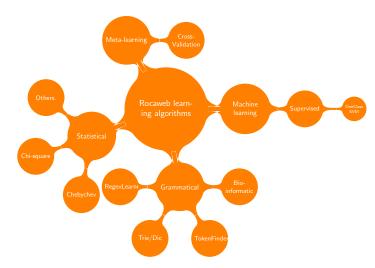


Figure: Learning algorithms considered for RoCaWeb

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djibrilla.amadoukountch

Data, Data,

Clients

Message:

Parameters

.

Intrusion detection in Web

Web intrusion: Origin of the problem

Web intrusion: How solve the problem

What is an IDP:

IDPS Detection Method:

Behavioral Intrusion detection

Grammatical inference

Definition

Let:

- $ightharpoonup \varSigma$ be a finite alphabet of symbols
- lacksquare Σ^* be the set of all finite strings of symbols from Σ
- ▶ langage be any sub-set of Σ^*

The learning problem

Given:

- ightharpoonup a langage L define over Σ
- a sequential or structured data (strings, words) drown form L

Determine the automata or a grammar of L.

Output model

A grammar explaning the data (See the Chomsky Hierarchy)

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche

Email: djibrilla.amadoukountch

D . D . D

Clients
HTTP protocol
Message:

'arameters

Intrusion detection in Web applications

Veb intrusion: Origin he problem

ve the problem nat is an IDPS ?

Knowledge based

IDPS Detection

Method: Behavioral Intrusion detection

Grammatical inference: Example

Example

- A = [a zA Z0 9]
- L is a langage define over A
- ► Observed set: Ich, esse, gern, in, Essen

Infer the grammar of this langage Is "J'aime manger à Essen" normal given this langage ? An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

AI & Data Science Meetup Author: Djibrilla Amadou Kountche Email: diib-

Event:Ruhrgebiet

rilla.amadoukountch

Data, Data,

Clients

HTTP protocol

Message: Request/Response

Parameters

Intrusion detection in Web

plication

Web intrusion: O the problem

lve the problem hat is an IDPS ?

Knowledge based

Method:
Behavioral
Intrusion detection
Algorithmic apprachas

Grammatical inference: RegexLearner

Learning phase

Given:

- ► S: a set of strings
- R set of Regular expressions
- ightharpoonup Find r, the most appropriate regex describing S

Detection phase

Given a string s:

- s is normal if it validates r
- abnormal otherwire

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet

AI & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djibrilla amadoukountch

Data Data

Data, Data,

Clients

Message: Request/Response

rarameters

Intrusion detection in Web

pplication

Web intrusion:

Web intrusion the problem

olve the problem /hat is an IDPS ?

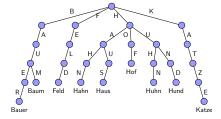
/hat is an IDPS ?

IDPS Detection
Method:
Behavioral
Intrusion detection
Algorithmic app. 29 / 50

Grammatical inference: Trie structure

Learning phase

- ► Given S a set of strings
- ▶ Build a trie



Detection phase

Given a string s:

- **s** is normal if it is in the Trie
- abnormal otherwire

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djibrilla.amadoukountch

Event:Ruhrgebiet

Data Data I

Data, Data, L

Clients
HTTP protocol

Message: Request/Respo

Parameters

Intrusion detection in Web

Web intrusion: Original the problem

solve the problem
What is an IDPS 7

What is an IDPS ?

IDPS Detection Method: Behavioral

Grammatical inference: Sequence alignment

Origin

Bio-informatics: computational approach to molecular biology.

Usage

Where given:

- ▶ An alphabet: $\alpha = \{A, G, T, C\}$
- ightharpoonup And a langage S define on α
- ightharpoonup The goal is to compare the sequences from S:
- ► Aligning these sequences can allow to:
 - Compare species

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche

Email: diib-

rilla.amadoukountch

HTTP protocol
Message:

Parameters

Intrusion detection in Web

Web intrusion: Origin of the problem

hat is an IDPS ?

PS Detection ethod: ehavioral

Grammatical inference: Sequence alignment

Origin

 Bio-informatics: computational approach to molecular biology.



Figure: Example of sequece alginment (From Wikipedia)

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djibrilla amadoukountch

Web application Clients HTTP protocol Message:

Parameters

Intrusion detection in Web

Web intrusion: Origin the problem Web intrusion: How to solve the problem

What is an IDPS Knowledge based

IDPS Detection Method: Behavioral Intrusion detection

Grammatical inference: Sequence Alignment

Definition: Pair alignment

Given:

- $ightharpoonup s=s_1\dots s_n$ and $t=t_1\dots t_m$ two strings defined on an alphabet \varSigma
- $ightharpoonup \beta$ the gap character, $\beta \ni \Sigma$

Summarizing alignment

- Insertion : the gap characters in the first string;
- Deletion : the gap character in the second string;
- ▶ Match or correspondence: the 2 characters are identical
- Mismatch : the two characters are different from each other and gap

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla

Amadou Kountche Email: djibrilla.amadoukountch

Data, Data, [

Clients
HTTP protocol
Message:

Request/Respons

Intrusion detection in Web

Web intrusion: Origin of the problem Web intrusion: How to

What is an IDPS ? Knowledge based

IDPS Detection Method: Behavioral

Intrusion detection 32 / 50 Algorithmic approaches

Grammatical inference:Example of Pair Alignment

Andreas Wulfes andreas wulfes An--ja----an--ja-----Andreas Wulfes andreas wulfes Annika- ---S. annika- ---s. Andreas Wulfes andreas wulfes ----a---rnea--r---ne-Andreas Wulfes andreas wulfes -Bernd Bäumler -bernd bäumler Andreas Wulfes andreas wulfes -Chr----is -chr----is -Andreas Wulfes -andreas wulfes Christine Gangl christine gangl Andreas Wulfes andreas wulfes ----Con ---Con ----con ---con

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djibrilla.amadoukountch

Event:Ruhrgebiet

Data, Data,

Clients
HTTP protocol

Message: Request/Response Parameters

Intrusion detection in Web

Web intrusion: Origin of the problem Web intrusion: How to solve the problem

What is an IDPS

IDPS Detection Method: Behavioral Intrusion detection Algorithmic approaches

Pair alignment: Approaches

Dynamic programming

- Exact solution: Dynamic programming approach
- BLAST

Dynamic programming: Score matrix

- The first column and the first row are filled with $(j \times penalty)$ and $(i \times penalty)$
- Any other cell is filled with :

$$M(i,j) = \max \left\{ egin{array}{ll} M(i-1,j) + \textit{penalty} & (i) \ M(i,j-1) + \textit{penalty} & (d)(1) \ M(i-1,j-1) + d(s_i,t_j) & (m) \end{array}
ight.$$

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche

Email: diib-

rilla.amadoukountch

Data, Data, Data

Clients
HTTP protocol
Message:
Request/Response

ntrusion detect

in Web applications

the problem

Web intrusion: How to solve the problem

What is an IDPS?

IDPS Detection Method:

Behavioral Intrusion detection Algorithmic approaches

Grammatical inference: Sequence alignment

Definition: Score

$$\delta(s',t') = \sum_{i=1}^{I} d(s'_i,t'_i)$$

$$d(s'_i,t'_i) = \begin{cases} \text{match} & \text{if } s'_i,t'_i \in \Sigma \text{ and } s'_i = t'_j \\ \text{mismatch} & \text{if } s'_i,t'_j \in \Sigma \text{ and } s'_i \neq t'_j \end{cases}$$

$$\text{penalty} & \text{if } s'_i \text{ or } t'_j \text{ equals } \beta$$

$$(2)$$

to Intrusion Detection and Prevention in Web **Applications** Event:Ruhrgebiet

An application of

Machine Learning

Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djibrilla.amadoukountch

Grammatical inference: Multiple sequence alignment

Multiple sequence alignment

Given:

- ▶ *n* sequences $s_1, ..., s_n$ defined on an alphabet Σ having different lengths
- $\triangleright \beta$ is the gap character

A multiple alignment of s_1, \ldots, s_n is an $n-uplet(s'_1, \ldots, s'_n)$ of length $l \geq |s_i|, i \in [1 \ldots n]$ on the alphabet Σ' where the following conditions hold :

- 1. $|s_i| = |s_j|, \forall i, j \in [1 \dots n]$
- 2. $h(s_i') = s_i, \bigvee, i \in [1 \dots n]$
- 3. there is no row in the $(n \times I)$ -matrix where there is only gap characters.

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet
AI & Data Science
Meetup
Author: Djibrilla

Amadou Kountche Email: djibrilla.amadoukountch

Web application

Clients
HTTP protocol
Message:
Request/Response

ntrusion detection n Web

applications
Web intrusion: Origin of the problem

What is an IDPS ?

IDPS Detection Method: Behavioral Intrusion detection Algorithmic approach 50

Grammatical inference: Multiple sequence alignment

Solution

- Exact computation (Difficult)
- Combining with pair alignment (Heuristic)

Algorithms

- UPGMA
- ► T-Coffee
- etc

An application of Machine Learning to Intrusion Detection and Prevention in Web **Applications**

Event:Ruhrgebiet Al & Data Science

Meetup Author: Djibrilla Amadou Kountche Email: diibrilla.amadoukountch

Grammatical inference: Example Multiple sequence alignment

Parameters

- ▶ Pair alignment algorithm : Needleman Wunsch
- ► Gap character : +, match = 1.0, mismatch = 0.0, penalty = -1.0

Learning Data Set

GARFIELD THE LAST FAT CAT GARFIELD THE FAST CAT GARFIELD THE VERY FAST CAT THE FAST CAT

Table: The GARFIELD dataset.

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet

AI & Data Science Meetup Author: Djibrilla Amadou Kountche Email: diib-

rilla.amadoukountch

Data, Data,

VVeb application Clients

Message: Request/Respons

Parameters

in Web

Web intrusion

solve the problem
What is an IDPS ?

Knowledge based

Method: Behavioral Intrusion detection Algorithmic approaches

Grammatical inference: Multiple sequence alignment

GARFIELD THE VERY FAST CAT +++++++++THE++++ FAST CAT GARFIELD THE+++++ FAST CAT GARFIELD THE LAST FA+T CAT

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djibrilla.amadoukountch

Event:Ruhrgebiet

ata, Data, D

HTTP protocol

Message:
Request/Response

Parameters

Intrusion detection in Web

Web intrusion: the problem

the problem Web intrusion: How solve the problem

/hat is an IDPS ?

IDPS Detection
Method:
Behavioral
Intrusion detection
Algorithmic approaches

Learning the regular expression

Learning the regular expression

- Alignment does not generate regular expression;
- Sequences of same size;
- Learn the corresponding regular expression.

	Result of AMAA																								
G	Α	R	F	П	Ε	L	D		Т	Н	E		V	E	R	Υ	П	F	Α	S	Т	П	C	Α	Т
+	+	+	+	+	+	+	+	+	Т	Н	Ε	+	+	+	+	+	П	F	Α	S	Т	H	C	Α	Т
G	Α	R	F	1	Е	L	D	ĺ	Т	н	E	+	+	+	+	+	П	F	Α	S	Т	H	C	Α	Т
G	Α	R	F	1	Е	L	D	İ	Т	Н	Ε		L	Α	S	Т	П	F	Α	+	Т	H	C	Α	Т
	[GARFIELD]{0,1}THE[LAST]{0,1} FA[S]{0,1}T CAT																								

Table: Generation of the regular expression for the GARFIELD dataset.

An application of Machine Learning to Intrusion Detection and Prevention in Web **Applications**

Event:Ruhrgebiet Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: diib-

rilla.amadoukountch

The Chebychev inequality applied to the length

Learning phase

Given:

- a parameter A of a request;
- $ightharpoonup A = \{a_i, i = 1...n\}$ values of this parameters are collected

Determine:

- $\triangleright \mu$: the sample mean of the lengths of a_i ;
- $\triangleright \sigma$: the variance:

An application of Machine Learning to Intrusion Detection and Prevention in Web **Applications**

Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: diib-

Event:Ruhrgebiet

rilla.amadoukountch

The Chebychev inequality applied to the length

The detection phase

Soient:

- \triangleright a_k : the current value to test
- $\triangleright \mu, \sigma$: the mean and variance.

Variante :

$$p(|X - \mu| > |I - \mu|) < p(I) = \begin{cases} \frac{\sigma^2}{(I - \mu)^2} & \text{if } 1 \ge \mu \\ 1 & \text{otherwise} \end{cases}$$
 (4)

I the current length. Return p(I)

to Intrusion Detection and Prevention in Web **Applications** Event:Ruhrgebiet

An application of

Machine Learning

Meetup Author: Djibrilla Amadou Kountche Email: diibrilla.amadoukountch

Al & Data Science

Character distribution

Learning phase

Determine the reference character distribution

Given:

- \triangleright $A = \{a_1, a_2, ..., a_n\};$
- $\triangleright \Sigma$ an alphabet;
- $ightharpoonup a_i$ defined on Σ^* .

A character distribution is défined by :

$$CD = \{n_i, i = 1..k\}$$
 (5)

An application of Machine Learning to Intrusion Detection and Prevention in Web **Applications**

Event:Ruhrgebiet Al & Data Science Meetup Author: Djibrilla

Amadou Kountche Email: diibrilla.amadoukountch

Character distribution

Reference character distribution

- For each value of the parameter determine its CD
- $ightharpoonup RCD = \{f_i = n_i/k, i = 1...k\}$ is the mean CD
- ▶ Where k is the size of the alphabet

Detection phase

Given:

- the CD a value and its length
- the RCD computed earlier

Determine the value of the χ^2 probability

An application of Machine Learning to Intrusion Detection and Prevention in Web **Applications**

Event:Ruhrgebiet Al & Data Science Meetup

Author: Djibrilla Amadou Kountche Email: diibrilla.amadoukountch

RoCaWeb: Agent

Reserve proxy

Apache configure as Reverse proxy for each website

Modsecurity

- Contract are formated in the Modsecurity format
- Validation for each method is implemented in Lua
- Modsecurity phases:
 - Request header
 - 2. Request body
 - Response Header
 - Response body
 - Logging

An application of Machine Learning to Intrusion Detection and Prevention in Web **Applications**

Event:Ruhrgebiet Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: diibrilla.amadoukountch

RoCaWeb: User interface

b ×				Data - Mozilla Firefox (Private Browsing)	∨ ♦ ×
Data		× +			©
€)⇒	C &		① localhost:9001	🛡 🏚 🔍 Search	W • □ ≡
Ham	epage 📊 Forum 🛙	WK 4	Most Visited Em Machine Learning Em Information security Em Scoop@F Em C-ITS E	⊞ WSN	Becklaands 🗎 Toda 🛅 Working in Germany 🔉 Mazilla News
			Data		
			Sources	Visualisation	
			8 a v	View 🖸 🔻 🔻	Editor
			2018-03-12-08.14.54.947_fichier_output_week35.pdmi. ^		
	USER		2		
	HEP		Preprocessed	b	
			nolane nolane keresal com		
			Home Team Blog		© 2015 RoCalifeb Team

Figure: RoCaWeb User Interface

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet AI & Data Science Meetup Author: Djibrilla Amadou Kountche

Email: djibrilla.amadoukountch

Data, Data,

Clients
HTTP protocol

Message: Request/Resp Parameters

Intrusion detection in Web applications

Web intrusion: (

olve the problem

What is an IDPS

IDPS Detection Method: Behavioral

RoCaWeb: The docker platform

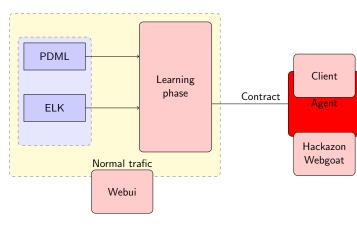


Figure: The RoCaWeb docker platform

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: djibrilla.amadoukountch

Event:Ruhrgebiet

Data, Data, I

Clients
HTTP protocol

Message: Request/Res Parameters

Intrusion detection in Web

application

Web intrusion: Orig the problem

What is an IDPS

Knowledge based

IDPS Detection
Method:
Behavioral
Intrusion detection
Algorithmic approaches

Similar projects

- Apache metron
- Vulture

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Al & Data Science Meetup Author: Djibrilla Amadou Kountche

Event:Ruhrgebiet

Email: djibrilla.amadoukountch

Conclusion and perspectives

► Fondamental limitation of Behavioral based

An application of Machine Learning to Intrusion Detection and Prevention in Web Applications

Event:Ruhrgebiet
AI & Data Science
Meetup
Author: Djibrilla
Amadou Kountche

Email: djibrilla.amadoukountch

Data, Data, I

Web application

Clients HTTP protoc

HTTP protoco

lessage: equest/Response rameters

trusion detecti

in Web applications

Web intrusion: Origin the problem

solve the problem

What is an IDPS ?

hat is an IDPS ?

Knowledge based

Method: Behavioral Intrusion detection 49 / 50 Algorithmic approaches

Reading list I

Anamaly detection

Anomaly detection: A survey, Varun Chandola

Grammatical inference

A bibliographical study of grammatical inference, Colin de la Higuera

Fondament Limitation of Anomaly detection:

http://all.net

An application of Machine Learning to Intrusion Detection and Prevention in Web **Applications**

Al & Data Science Meetup Author: Djibrilla Amadou Kountche Email: diib-

rilla.amadoukountch

Event:Ruhrgebiet