



Regular expressions Denial of Service -ReDoS

Michael Hidalgo

OWASP Costa Rica

michael.hidalgo@owasp.org



OWASP

The Open Web Application Security Project



OWASP

The Open Web Application Security Project

- Líder del capítulo de OWASP Costa Rica.
- Software engineer Security Innovation.
- Experiencia en el área financiera.
- Investigador seguridad de aplicaciones.
- Participo en proyectos de OWASP (Plataforma O2, SafeNuGet).



OWASP

The Open Web Application Security Project

- Breve introducción a las expresiones regulares.
- Ejemplo de expresiones regulares usadas frecuentemente
- Autómata determinístico no Finito (NFA)
- Problemas de denegación de servicio
- Demo
- Conclusiones, recomendaciones oportunidades de mejora.



OWASP

The Open Web Application Security Project

- Expresiones regulares son muy usadas

Page 1 of 1

Google Form with Regex

Form Description

Question Title: Year of birth

Help Text: Enter a value between 1900 and 2014

Question Type: Text

Their answer

RegEx for data validation goes here

✓

Data validation:

Regular expression Contains

(19\d{2}|20\d{1-4})

Enter a valid year

Done Required question

A screenshot of the Google Form editor. It shows a question titled "Year of birth" with the help text "Enter a value between 1900 and 2014". The question type is set to "Text". Below the question, there is a "Their answer" field and a "Data validation" section. In the "Data validation" section, the "Regular expression" checkbox is checked, and the expression "(19\d{2}|20\d{1-4})" is entered into the input field. A yellow box highlights this regular expression. A blue annotation "RegEx for data validation goes here" with a checkmark points to the validation input field. At the bottom, there are "Done" and "Required question" buttons.



OWASP

The Open Web Application Security Project

- Expresiones regulares facilitan validación de campos de entrada.

A screenshot of a web application interface. At the top, there is a form with two input fields and a button:

- Name:** Michael Hidalgo
- Email:** email inválido
- Send invitation** (button)

Below the form, the HTML code for the form is displayed:

```
<form class="form-inline">
  <div class="form-group">
    <label for="exampleInputName2">Name</
```

A tooltip is shown over the invalid email field, containing the following message:

! Please include an '@' in the email address. 'email inválido' is missing an '@'.



OWASP

The Open Web Application Security Project

Algunos usos comunes de las expresiones regulares

- Validar formatos como número de teléfono, correos electrónico, direcciones Web, etc.
- Servir como filtro de datos.
- Encontrar cierto texto dentro de un archivo.



OWASP

The Open Web Application Security Project



Pero ¿qué es una expresión regular?

- Cadena que contiene una combinación de caracteres normales y caracteres especiales o meta secuencias.
- Caracteres normales tienen su propio significado.
- Meta caracteres o meta secuencias son caracteres o secuencias que tienen significado (cantidades, ubicaciones etc.).



OWASP

The Open Web Application Security Project

Elementos de una expresión regular

- Coincidencia de Patrones (Pattern matching) : Encontrar una sección de texto que es descrita por la expresión regular (matching).
- Motor de expresiones regulares: Código o programa encargado de buscar ese texto



OWASP

The Open Web Application Security Project



- Expresiones regulares usadas frecuentemente
- Contraseñas : `/^([a-zA-Z_-]{6,18})$/`
- Correo electrónico : `/^([a-zA-Z_\.-]+)@([\da-zA-Z_\.-]+\.)\.([a-zA-Z]{2,6})$/`
- Solo dígitos: `^[0-9]*$`
- Solo letras: `/[a-zA-Z]+/`



OWASP

The Open Web Application Security Project

- Existen muchos lugares donde aprender y probar expresiones regulares.

A screenshot of a Google search results page. The search bar at the top contains the query "regex online". Below the search bar, there are several navigation links: "Web" (which is highlighted in red), "Books", "News", "Videos", "Shopping", "More ▾", and "Search tools". A horizontal line separates this from the search results. The first result is a link to "https://regex101.com/". The snippet for this result reads: "Online regex tester and debugger: JavaScript, Python, PHP ... https://regex101.com/ Online regex tester for PHP, PCRE, JavaScript and Python that highlights pattern and matches on the fly." The snippet is displayed in purple text.

About 50,400,000 results (0.23 seconds)

Online regex tester and debugger: JavaScript, Python, PHP ...
<https://regex101.com/> ▾
Online regex tester for PHP, PCRE, JavaScript and Python that highlights pattern and matches on the fly.

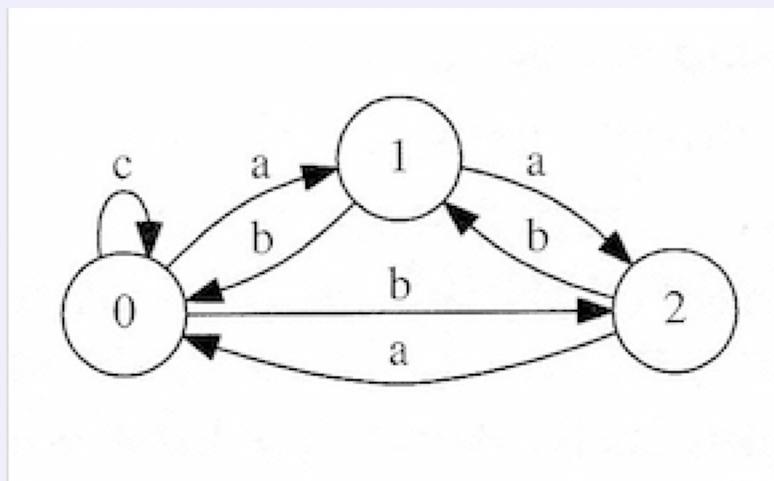


OWASP

The Open Web Application Security Project

Autómata finito no determinístico (NFA)

- Autómata: simple computadora, con estados y la transición de un estado a otro de acuerdo a una secuencia de entradas.





OWASP

The Open Web Application Security Project

Autómata finito no determinístico (NFA)

- Retroceso (Backtracking)
- El motor encuentra fácilmente una coincidencia positiva (positive matching), no obstante confirmar una coincidencia negativa (negative matching) toma más tiempo.



OWASP

The Open Web Application Security Project

Denegación de Servicio a través de Expresiones regulares (ReDoS)

- Implementaciones de expresiones regulares pueden alcanzar situaciones extremas.
- Ralentización crece exponencialmente relacionado al tamaño del dato de entrada.
- Atacante puede ocasionar que el sistema se enfrente a estas situaciones extremas y se comporte de forma lenta por un periodo de tiempo prolongado.



OWASP

The Open Web Application Security Project

El algoritmo problemático de naïve

- Algoritmo de expresiones regulares que implementa un NFA.
- Máquina de estado finito.
- Por cada par de estados y símbolos de entrada puede haber múltiples estado.
- La mayoría de los motores de expresiones regulares usan el algoritmo de naïve.

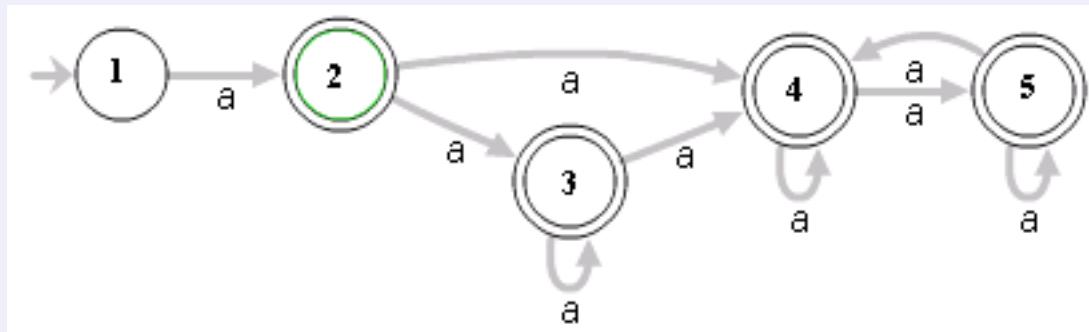


OWASP

The Open Web Application Security Project

Ilustrando el problema

Dada la expresión regular: $^*(a+)^*\$$



Dada una entrada como **aaaaaX** existen 16 posibles rutas (2^4).

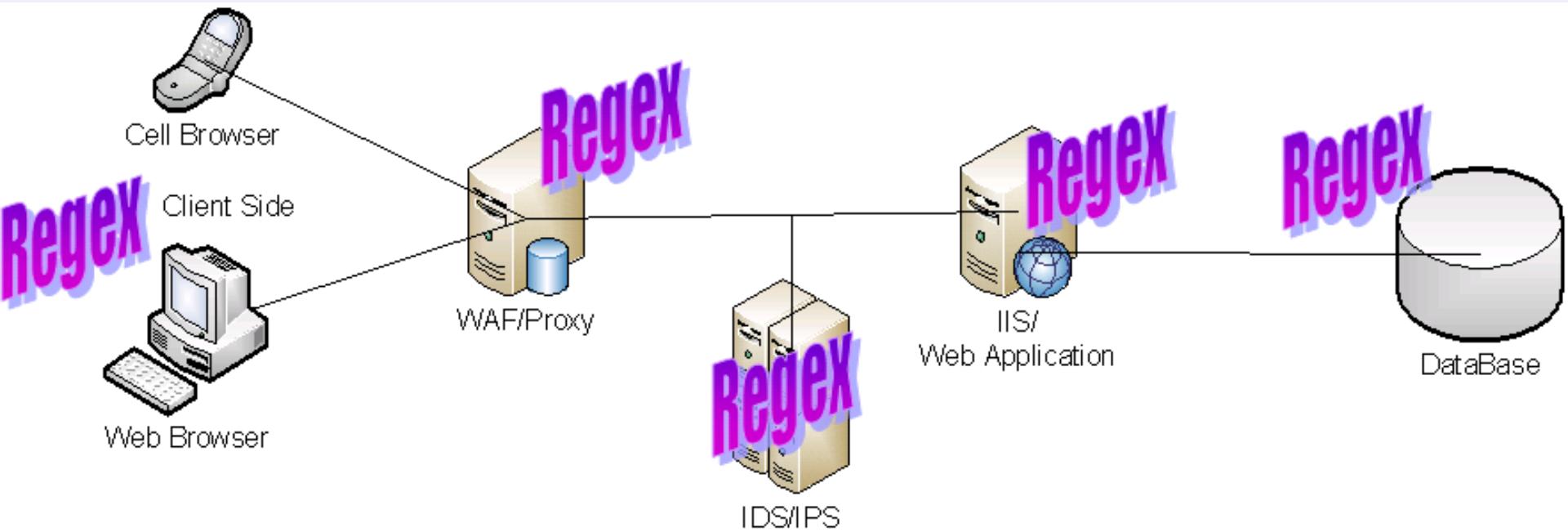
No obstante dada la entradaaaaaaaaaaaaaaaX, hay 65,536 posibles rutas (2^{16}).



OWASP

The Open Web Application Security Project

- La Web se basa en expresiones regulares





OWASP

The Open Web Application Security Project





OWASP

The Open Web Application Security Project

Algunos patrones de expresiones regulares problemáticas

- Agrupamiento con repetición
- $(a^+)^+$
- $([a-zA-Z]^+)^*$
- $(a|aa)^+$
- $(a|a?)^+$
- $(.*a)\{x\} \mid \text{for } x > 10$



OWASP

The Open Web Application Security Project

Conclusiones, recomendaciones, oportunidades de mejora

- Es muy fácil escribir código inseguro.
- Siempre se debe probar las expresiones regulares a través de pruebas unitarias.
- No utilizar expresiones regulares que implementan complejos agrupamientos.
- Siempre se debe validar la longitud del dato de entrada antes de evaluar dicho dato con la expresión regular.



OWASP

The Open Web Application Security Project

Muchas Gracias!





OWASP

The Open Web Application Security Project



• Referencias

1. Hollos, F. Hollos R : Finite Automata and Regular Expressions Problems and Solutions
2. Stubblebine, T. Regular Expressions Pocket Reference.
3. Regular Expressions Denial of the Service Attacks and Defenses
<https://msdn.microsoft.com/en-us/magazine/ff646973.aspx>
4. OWASP https://www.owasp.org/index.php/Regular_expression_Denial_of_Service_-_ReDos
5. <http://www.cs.bham.ac.uk/~hxt/research/reg-exp-sec.pdf>