Requerimientos del sistema

Proyecto: Sistema *open source* para la detección páginas web maliciosas

Revisión 1.0

Ficha del documento

|  |  |  |
| --- | --- | --- |
| Fecha | Revisión | Autor |
| 28 de abril de 2017 | 1.0 | Autores del proyecto |

Documento validado en la fecha:

|  |  |
| --- | --- |
| Por el tutor | Por los autores del proyecto |
| Christian Urcuqui | Melisa García P. y Jose Luis Osorio Q. |

Índice de contenido

[Introducción 6](#_Toc481617141)

[Propósito 6](#_Toc481617142)

[Visión general del documento 6](#_Toc481617143)

[Objetivos 7](#_Toc481617144)

[Objetivo general 7](#_Toc481617145)

[Objetivos específicos 7](#_Toc481617146)

[Alcance del sistema 7](#_Toc481617147)

[Contextos de alcance 7](#_Toc481617148)

[Aspectos fuera del alcance 7](#_Toc481617149)

[Referencias 7](#_Toc481617150)

[Descripción General 8](#_Toc481617151)

[Perspectiva del proyecto 8](#_Toc481617152)

[Funciones del proyecto 8](#_Toc481617153)

[Características de los usuarios del sistema 9](#_Toc481617154)

[Suposiciones y Dependencias 9](#_Toc481617155)

[Suposiciones 9](#_Toc481617156)

[Dependencias 9](#_Toc481617157)

[Requisitos específicos 9](#_Toc481617158)

[Funciones 9](#_Toc481617159)

[Particionamiento del sistema 10](#_Toc481617160)

[Requerimientos globales 10](#_Toc481617161)

[Requerimientos finales 12](#_Toc481617162)

[Análisis de Dorfman Completo 13](#_Toc481617163)

[Atributos de calidad del sistema 13](#_Toc481617164)

# Introducción

El proyecto de grado titulado “Sistema *open source* para la detección de ataques en páginas web” es un proyecto conformado por 2 estudiantes: Melisa García Peña y Jose Luis Osorio Quintero, de noveno semestre de la Universidad Icesi. Además de contar con Christian Urcuqui y Andrés Navarro, tutor y cotutor del proyecto respectivamente. El objetivo principal del proyecto es desarrollar y validar un sistema de detección de ataques en páginas web. Para garantizar esto, se trabaja de acuerdo a lo leído en los estudios realizados en artículos investigados.

El documento presente se denomina Especificación de Requisitos de Software (ERS) para el proyecto, cuyo desarrollo está a cargo de los autores del mismo. Esta especificación se ha estructurado basándose en el estándar ANSI/IEEE 830, 1998. El documento se escribe para dar a conocer las necesidades a satisfacer por el proyecto, así como su alcance y funcionamiento.

## Propósito

El propósito de éste documento es mostrar de una forma clara todas las características y especificaciones funcionales y no funcionales que va a tener el sistema a desarrollar, las cuales servirán de soporte y de guía para las fases del ciclo de vida del mismo. En el momento en que los tutores del proyecto acepten el documento, no se podrán adicionar requerimientos que no estén comprendidos bajo el alcance definido. Así mismo, el cumplimiento de todo lo expuesto en este documento, va a significar la finalización completa y correcta del sistema.

## Visión general del documento

El documento tiene la siguiente estructura:

* Sección 1: Introducción  
  Se presentan los objetivos y la delimitación del alcance del proyecto.
* Sección 2: Descripción global del sistema  
  Se presenta de manera general las funcionalidades y restricciones del sistema.
* Sección 3: Requisitos específicos  
  Se realiza un análisis más detallado de los requerimientos para su total entendimiento por parte de los desarrolladores del sistema.

## Objetivos

### Objetivo general

Desarrollar y validar un sistema de detección de ataques en páginas web.

### Objetivos específicos

* Identificar y caracterizar los ataques que se encuentran en las páginas web.
* Diseñar el diagrama de arquitectura del sistema.
* Validar el sistema de monitoreo.

## Alcance del sistema

El sistema de detección va a comprender las siguientes funcionalidades:

* Generación del tráfico web de URL’s ingresadas.
* Generación de la matriz de características de aplicación y red.
* Generación de la matriz binaria.
* Resultados de los algoritmos de *Machine Learning.*
* Diseño de la arquitectura del sistema.

### Contextos de alcance

### Aspectos fuera del alcance

A continuación, se mencionas todos los aspectos que quedan fuera del alcance del proyecto:

* El sistema se reduce a la identificación porcentual de la probabilidad que existe de haber tráfico maligno en el tráfico web ingresado.
* No se implementará el sistema, es decir, solo se deja hasta el diseño de arquitectura del mismo.

### Referencias

# Descripción General

## Perspectiva del proyecto

El proyecto final, debe permitir en la satisfacción del objetivo general del mismo. En éste, los usuarios van a poder verificar el tráfico web que ingresen al sistema, verificando que porcentaje de maldad hay en éste y entregando un informe del mismo, el resultado es manejado dependiendo de los algoritmos utilizados.

## Funciones del proyecto

Las funcionalidades externas del sistema, se pueden visualizar en el siguiente diagrama de casos de uso contextual.

## Características de los usuarios del sistema

|  |  |
| --- | --- |
| **Tipo de Usuario** | Usuario |
| **Responsabilidad** | Genera nuevos datos en el sistema y descarga el informe de análisis |
| **Actividades** | * Ingresar el tráfico web * Descargar el informe del análisis |
| **Interacción con el sistema** | Es el encargado de suministrar la información, los *datasets* con las URL’s. |

|  |  |
| --- | --- |
| **Tipo de usuario** | Administrador |
| **Responsabilidad** | Su función es configurar la plataforma e ir recolectando datos ingresados |
| **Actividades** | * Modificar los algoritmos * Descargar o no el tráfico ingresado por el usuario. |
| **Interacción con el sistema** | Es el encargado de actualizar los algoritmos de entrenamiento, las gráficas del análisis y la información de los reportes generados. |

## Suposiciones y Dependencias

A continuación, se describen los factores de los cuales dependen los requerimientos, esto es, si algunos de estos cambian, los requerimientos también pueden verse alterados. Estos factores se dividen en supuestos y dependencias que afectan el funcionamiento del sistema de detección.

### Suposiciones

En esta sección, se identifican las restricciones que se relacionan con el funcionamiento del sistema, éstas hacen parte de las siguientes suposiciones:

* El usuario externo, ingresa solamente las URL’s que desea analizar.

### Dependencias

En esta sección, se identifican las restricciones de las que depende el buen funcionamiento del sistema:

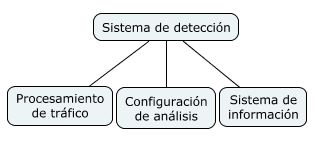
* Buena conexión a internet.

# Requisitos específicos

## Funciones

Ahora, se identifican todas las funcionalidades del sistema, partiendo de los requerimientos globales del mismo, hasta llegar a las divisiones de éstos. El proceso de análisis se realiza mediante el análisis de requerimientos Dorfman.

### Particionamiento del sistema



A continuación, se describen los subsistemas planteados.

|  |  |
| --- | --- |
| **Subsistema** | **Descripción** |
| Procesamiento de tráfico | El procesamiento del tráfico es la generación de las matrices de la capa de aplicación y de red. |
| Configuración de análisis | Es el análisis de las matrices generadas. |
| Sistema de información | Toda la generación de información y gráficas que el sistema debe generar, para mostrar los datos resultantes. |

### Requerimientos globales

|  |  |
| --- | --- |
| RF01 | El sistema debe simular una captura real de cada página web |
| RF02 | El sistema debe generar un análisis por aplicación y por capa de red |
| RF03 | El sistema debe abstraer características de la capa de aplicación y de red |
| RF04 | El sistema debe estar en la capacidad de generar una matriz con valores entre 0 y 1, incluyéndolos, de las características de la capa de aplicación y de red. |
| RF05 | El sistema debe analizar las matrices "binarias" con algoritmos de Machine Learning |
| RF06 | El sistema debe estar en la capacidad de mostrar una matriz final comparando los Falsos Negativos y los Falsos Positivos de cada algoritmo de Machine Learning. |
| RF07 | El sistema debe mostrar el porcentaje de URL's malignas y benignas clasificadas |
| RF08 | El sistema debe generar un diagrama que represente el porcentaje de URL's benignos y malignos que se encuentren en el *dataset* |
| RF09 | El sistema debe generar un diagrama que represente las características de las URL's de un *dataset* |
| RF10 | El sistema debe generar un diagrama que represente el contenido de cada URL |
| RF11 | El sistema debe estar en la capacidad de generar un reporte con el análisis de las URL's |
| RF12 | El sistema debe permitir que el usuario descargue tanto la información gráfica como los informes generados a partir del tráfico de red ingresado |
| RF13 | El sistema debe permitir hacer comparaciones de un mismo trafico aplicando distintos algoritmos de análisis |
| RF14 | El sistema debe permitir al administrador cambiar el algoritmo de entrenamiento |

### Requerimientos finales

|  |  |  |  |
| --- | --- | --- | --- |
| **Requerimiento** | **Subsistema nivel 1** | **Requerimiento subsistema nivel 1** | **Descripción** |
| **RF01** | PT | RF01\_PT01 | El sistema debe estar en la capacidad de simular el tráfico de cada página web ingresada |
| **RF02** | PT | RF02\_PT01 | El sistema debe permitir el análisis de las URL's por la aplicación de la misma |
| RF02\_PT02 | El sistema debe permitir el análisis de las URL's por la capa de red de la misma |
| **RF03** | PT | RF03\_PT01 | El sistema debe abstraer las características de las URL's para la capa de aplicación |
| RF03\_PT02 | El sistema debe abstraer las características de las URL's para la capa de red |
| **RF04** | PT | RF04\_PT01 | El sistema debe estar en la capacidad de generar una matriz con valores entre 0 y 1, incluyéndolos, de las características de la capa de aplicación y de red |
| **RF05** | PT | RF05\_PT01 | El sistema debe analizar las matrices "binarias" con algoritmos de *Machine Learning* |
| **RF06** | SI | RF06\_SI01 | El sistema debe estar en la capacidad de mostrar una matriz final comparando los Falsos Negativos y los Falsos Positivos de cada algoritmo de *Machine Learning* |
| **RF07** | SI | RF07\_SI01 | El sistema debe mostrar el porcentaje de URL's malignas y benignas clasificadas |
| **RF08** | SI | RF08\_SI01 | El sistema debe generar un diagrama que represente el porcentaje de URL's benignos y malignos que se encuentren en el *dataset* |
| **RF09** | SI | RF09\_SI01 | El sistema debe generar un diagrama que represente las características de las URL's de un *dataset* |
| **RF10** | SI | RF10\_SI01 | El sistema debe generar un diagrama que represente el contenido de cada URL |
| **RF11** | SI | RF11\_SI01 | El sistema debe estar en la capacidad de generar un reporte con el análisis de las URL's |
| **RF12** | CA | RF12\_CA01 | El sistema debe permitir que el usuario descargue tanto la información gráfica como los informes generados a partir del tráfico de red ingresado |
| **RF13** | CA | RF13\_CA01 | El sistema debe permitir hacer comparaciones de un mismo trafico aplicando distintos algoritmos de análisis |
| **RF14** | CA | RF14\_CA01 | El sistema debe permitir al administrador cambiar el algoritmo de entrenamiento |

### Análisis de Dorfman Completo

Ver anexo “Análisis de Dorfman. xlsx”

### Atributos de calidad del sistema

Los atributos de calidad que el sistema contiene, son:

* Funcionalidad, debe cumplir con todos los requerimientos planteados.
* Mantenibilidad, se deja todo *open source* y en un lenguaje portable, con el fin que sea fácil la modificación del mismo.
* Portabilidad, se desarrolla en un lenguaje que funciona en la mayoría de sistemas operativos.