# Técnicas para Incrementar la Seguridad en Web Services basados en WSDL

Edgardo Bernardis[(1)], Hernán Bernardis[(1)], Mario M. Berón[(1)], Daniel E. Riesco[(1)], Pedro Rangel Henriques[(2)], Maria Joao V. Pereira[(3)]

[(1)] Departamento de Informática / Facultad Ciencias Físico Matemáticas y Naturales/ Universidad Nacional de San Luis
Ejército de los Andes 950 – San Luis – Argentina
{hbernardis, ebernardis, mberon, driesco}@unsl.edu.ar

[(2)] Departamento de Informática/Universidade do Minho
Braga – Portugal
pedrorangelhenriques@gmail.com

[(3)] Departamento de Informática e Comunicações/ Instituto Politécnico de Bragança
Bragança - Portugal
mjoao@ipb.pt

## RESUMEN

En los tiempos que corren ha cambiado de manera considerable la forma en que las personas y empresas interactúan e intercambian información. Este escenario lo vuelve un blanco importante para todo tipo de actores que desean obtener información útil y valiosa para sí mismos o de terceros. Ante esta situación, es sumamente imprescindible implementar todo tipo de medidas y acciones tendientes a evitar estos ataques. Toda acción, herramienta o metodología enfocada a evitar, contrarrestar o retrasar ataques contra activos sensibles hace referencia a lo que se denomina Seguridad Informática [1].

Actualmente muchos sistemas de software son en realidad una agrupación de servicios web en la nube que se invocan cuando es necesario obtener la información que los mismos proveen. Es por ello, que este artículo, se enfoca en comprender y mejorar la seguridad de los servicios web que utilizan un WSDL (Web Services Description Language) [2] como medio para proveer servicios a través de la red.

**Palabras Clave:** Web Services, Métricas, Comprensión, WSDL, Ofuscación, Seguridad.

## CONTEXTO

La línea de investigación descripta en este artículo se desarrolla en el Laboratorio de Calidad e Ingeniería de Software (LaCIS) de la Universidad Nacional de San Luis; y se encuentra enmarcada dentro del proyecto: "Ingeniería de Software: Estrategias de Desarrollo, Mantenimiento y Migración de Sistemas en la Nube", perteneciente a la misma. Dicho proyecto, es reconocido por el programa de incentivos, y es la continuación de diferentes proyectos de investigación de gran éxito a nivel nacional e internacional.

## 1. INTRODUCCIÓN

El incremento en el uso de los servicios web dentro del desarrollo de sistemas es cada vez mayor. Muchas organizaciones construyen sus sistemas basándose en una arquitectura orientada a servicios web, algunos de ellos se publican al resto del mundo de manera

pública y libre, mientras que otros son utilizados internamente de manera privada por sus equipos de desarrollo y/o funcionalidades propias de la entidad.

Todo Web Service (WS) posee una especificación que provee la información necesaria para invocarlo y utilizarlo adecuadamente. Uno de los estándares de descripción más conocidos es WSDL (Web Service Definition Language) [3]. Las especificaciones WSDL son un dialecto XML, con reglas bien definidas para especificar cada componente del WS. Así como el archivo WSDL sirve para que un agente de software o persona pueda interpretar para luego usar el servicio web que describe, también puede dar información a personas no deseadas o incluso exponer vulnerabilidades de la organización que lo creó y que lo utiliza. Esto se torna sumamente importante para aquellos casos en donde los servicios web pertenecen a bancos, tarjetas de créditos, servicios de compra/venta online, entre otros. Incluso también para los servicios web que no se publican, son privados y necesitan mayor control y seguridad como los que pertenecen a empresas privadas y redes militares. Esto no solo se trata de competencia, los ataques de seguridad como espionaje de información, suplantación de clientes, inyección de comandos y denegación de servicio también son posibles ya que los atacantes pueden aprender sobre los datos intercambiados y los patrones de invocación de los documentos WSDL. Si bien la legibilidad de las descripciones de los servicios hace que los servicios web sean reconocibles, también contribuye a la vulnerabilidad del servicio [4]. Todos contienen información formal (código fuente) e informal (identificadores, comentarios, documentación, etc.) y es en este tipo de información en donde los atacantes hacen foco para obtener información beneficiosa a sus propósitos. Suena lógico entonces incrementar la seguridad que posee un determinado WSDL para evitar e impedir los ataques.

Si se supone la situación particular de un banco que posee múltiples servicios web en su sistema con arquitectura orientada a servicios y desea determinar qué tan vulnerables son a posibles ataques informáticos. Será importante responder a las siguientes preguntas ¿Cómo puede analizar las vulnerabilidades del mismo? ¿Cómo determinar qué tan comprensibles son a personas externas que deseen encontrar vulnerabilidades en sus descripciones para atacar? Ambos escenarios desnudan la necesidad de técnicas de comprensión de servicios web y mediante las mismas, aplicar técnicas para mejorar la seguridad de los WSDLs.

## 2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Este equipo de investigación trabaja actualmente en dos líneas principales de estudio. Una de ellas dedicada a la Extracción y Análisis de la Información y la otra enfocada en la Seguridad de la Información. Estas dos líneas se combinan en este trabajo de regulación en la seguridad de servicios web con especificaciones WSDL. A continuación, se describen las líneas de investigación antes mencionadas de manera resumida destacando las principales actividades llevadas a cabo en esta investigación.

### Línea de Extracción y Análisis

Esta línea se enfoca en obtener información de la especificación o el código fuente de los sistemas, procesarla y analizarla para generar resultados a partir de la misma.
Dentro del contexto de especificaciones WSDL, el análisis y reducción de vulnerabilidades de las especificaciones WSDL posee múltiples etapas. Cada una con su función particular dentro del proceso global. En las subsecciones siguientes se describen brevemente dichas etapas.

**Extracción de Información**

Debido a que las especificaciones WSDL son un dialecto XML, se pueden usar técnicas de compilación sobre las mismas basadas en el parser DOM (Domain Object Model) [5]. Un parser DOM construye una representación interna del WSDL basada en árboles. A partir de funciones específicamente diseñadas para recorrer la representación construida (funciones transversales) se extrae la información deseada. Estas funciones transversales logran extraer los identificadores de cada componente (*name, type element*, *etc.*), la documentación y los comentarios presentes en el WSDL. Adicionalmente, también se extrae información complementaria como cantidad de líneas del archivo, cantidad de espacios de nombres, versión del WSDL, entre otras.

**Preprocesamiento de la Información**

Este proceso prepara los datos extraídos para luego realizar el cálculo de las métricas. Esta etapa implica, por ejemplo, quitar las stop-words, separar las palabras de los nombres de los identificadores y calcular valores de base.

**Cálculo de Métricas**

De la información extraída del WSDL, se calculan múltiples métricas, que pueden ser lógicamente diferenciadas en los siguientes grupos:

- **Métricas de tamaño:** miden la complejidad del WSDL en base a las cantidades de componentes de cada etiqueta dentro del WSDL, como por ejemplo cantidad de tipos complejos, de parámetros, de operaciones, de mensajes, entre otras. Esto permite tener una idea del tamaño de cada sección particular del WSDL y determinar qué tan complejo es, a primera vista, su comprensión.
- **Métricas de calidad:** permiten medir la calidad semántica de la especificación WSDL. Esto es, que tanta información semántica brinda la especificación WSDL respecto del WS que representa y que tan entendible y comprensible en sí es dicha especificación. Alguno de los parámetros que se miden son calidad de palabras, cantidad de correcciones por palabras y palabras en diccionario.
- **Métrica de entendimiento global:** usando LSP (Logic Scoring of Preference) se calcula el grado de entendimiento que posee la especificación WSDL de un WS [6, 7]. Esto se realiza usando como base todas las métricas calculadas relacionadas mediante estructuras de agregación específicamente creadas para este fin.

Toda aplicación web está conformada por distintos tipos de información, tanto formal como informal. El análisis detallado de la misma y el cálculo de métricas permite detectar, dado su grado de entendimiento, que partes son más susceptibles a los ataques. En este punto, es posible definir estrategias que permitan subsanar las vulnerabilidades y proteger las partes que sean susceptibles de ataques [8]; en este caso particular, ataques dirigidos a WSDLs.

## Línea de Seguridad de la Información

Esta línea se enfoca en trabajar la seguridad de la información diseñando y aplicando diferentes técnicas y herramientas diseñadas para tal fin.

En el caso de este trabajo, se aplica ofuscación, transformación de código y otras modificaciones al WSDL para incrementar su seguridad.

**Ofuscación de Código**

Según el Diccionario de la Real Academia, ofuscar significa deslumbrar, turbar la vista, oscurecer, trastornar o confundir las ideas [9]. Es decir, se refiere a encubrir deliberadamente el significado de alguna cosa haciéndola más confusa y complicada

de interpretar, evitando la comprensión de la misma. La palabra ofuscación fue elegida para esta actividad porque connota oscuridad, ininteligibilidad y desconcierto, y porque ayuda a distinguir este enfoque de otros métodos [10]. Especificando más el concepto, Ofuscación de Código hace referencia a un conjunto de transformaciones de código que convierten un programa en uno funcionalmente equivalente, pero ininteligible haciendo difícil su entendimiento y realizar ingeniería inversa sobre el mismo. La ofuscación de código aplica una o más transformaciones de código que hacen que el código sea más resistente al análisis y la manipulación, pero preservan su funcionalidad [11].

**Transformaciones de Código**

La compilación de código se ha convertido en mucho más que solo traducir un programa de computadora en uno ejecutable. Por lo general, los programas se escriben en un lenguaje de alto nivel, dadas las ventajas que estos ofrecen. Sin embargo, tienen ciertas desventajas que hacen que la compilación de código incluya implícitamente numerosas técnicas de optimización. Estas, incluyen eliminar código muerto, asignación de registros óptimas y asignaciones eficientes al objetivo conjunto de instrucciones de la arquitectura [11]. En estas etapas de compilación y/o desarrollo de software se pueden realizar y aplicar distintas transformaciones de código. Una buena ofuscación se compone de una o más transformaciones de código que alteran un programa de tal forma que resulte difícil aplicar técnicas de ingeniería inversa. La única restricción para estas transformaciones, sean manuales o automatizadas, es preservar la funcionalidad original del programa. Las transformaciones de código para ofuscar un programa se pueden dividir en cuatro clases principales: i) Léxicas o de Diseño ii) de Flujo de Control, iii) de Flujo de Datos y iv) Preventivas [12].

**Incrementar la Seguridad**

Utilizando la información extraída del WSDL se pueden manipular diferentes partes del mismo para mejorar su seguridad disminuyendo su nivel de entendimiento. Esto se puede lograr mediante la utilización de funciones de ofuscación y/o encriptación al realizar las modificaciones y/o transformaciones necesarias que aumentaran el nivel de seguridad. Estas transformaciones pueden ser sobre partes específicas del WSDL (identificadores, operaciones, etc.) o en la totalidad del mismo. Dichas modificaciones dependen del nivel de seguridad deseado, partiendo de un nivel básico en donde se ofuscan y/o encriptan partes específicas del WSDL, como por ejemplo el nombre de los identificadores, hasta llegar a un nivel máximo en donde se realiza una transformación completa del WSDL.

## 3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados más destacados obtenidos de esta investigación se mencionan a continuación:

- Se diseñaron estrategias para la extracción de información desde las especificaciones WSDL de los WS.
- Se definieron y crearon múltiples métricas de tamaño y semánticas para obtener información de los WSDLs.
- Se construyó la estructura de agregación LSP que permite vincular todas las métricas las métricas y calcular el grado de entendimiento global del WSDL.
- Se construyó WSDLUDTool, una herramienta que realiza la extracción de información, cálculo de métricas, cálculo del grado de entendimiento del WS usando LSP y la visualización de la información.
- Se modificaron representaciones WSDLs originales y ofuscadas y se le calculó el grado de entendimiento

a ambas para comprobar el cambio de comprensión.

Los objetivos planteados a corto y largo plazo son:

- Crear un módulo que permita recomendar automáticamente cambios en el WSDL para incrementar o decrementar el grado de entendimiento según sea necesario.
- Definir otras estrategias para modificar la comprensión de los WSDLs para hacerlos más seguros.
- Definir, analizar y combinar distintas técnicas y estrategias de ofuscación en WSDLs.

## 4. FORMACIÓN DE RECURSOS HUMANOS

Las tareas realizadas en el contexto de la presente línea de investigación están siendo desarrolladas como parte de trabajos para optar al grado de Magister en Ingeniería de Software. En el futuro se piensa generar diferentes tesis de maestría y doctorado usando como base parte de esta investigación.

## 5. BIBLIOGRAFÍA


[1] ISO/IEC. Iso/iec 27032:2012 information technology - security techniques - guidelines for cybersecurity.

[2] WSDL Specification for W3C https://www.w3.org/TR/wsdl

[3] WSDL Specification for W3C. https://www.w3.org/TR/wsdl.

[4] Pananya Sripairojthikoon, Twittie Senivongse. Concept-Based Readability Measurement and Adjustment for Web Services Descriptions. ICACT Transactions on Advanced Communications Technology (TACT) Vol. 3, Issue 1, January 2014.

[5] Parser DOM specification for W3C. https://www.w3.org/DOM.

[6] Mario M. Berón, Hernán Bernardis, Enrique A. Miranda, Daniel E. Riesco, Maria João Pereira, Pedro Rangel Henriques. "WSDLUD: a Metric to Measure the Understanding Degree of WSDL Descriptions". Proceedings of the 2015 Symposium on Languages, Applications and Technologies, SLATE'15. Madrid, España 2015.

[7] Bernardis, Hernán; Berón Mario; Bernardis, Edgardo; Riesco, Daniel; Henriques, Pedro. "Extracción de información y cálculo de métricas en WSDL 1.1 y 2.0". II Congreso Nacional de Ingeniería Informática / Sistemas de información (CoNaIISI). Argentina. 2014.

[8] Edgardo Bernardis, Hernán Bernardis, Mario Berón, Germán Montejano. "Seguridad en Servicios Web". XIX Workshop de Informática y Ciencias de la Computación (WICC). Buenos Aires, Argentina. Abril de 2017.

[9] C. Collberg, C. Thomborson, and D. Low. A taxonomy of obfuscating transformations. Technical Report #148, Department of Computer Science, The University of Auckland, 1997.

[10] Brunton, Finn and Nissenbaum, Helen. Obfuscation: A user's guide for privacy and protest. MIT Press. 2015.

[11] Cappaert, Jan. Code obfuscation techniques for software protection. Katholieke Universiteit Leuven. 2012.

[12] Collberg, Christian and Thomborson, Clark and Low, Douglas. A taxonomy of obfuscating transformations. Computer Science Technical Reports. The University of Auckland. 1997.