Aplicación del nivel 1 estándar ASVS de OWASP: un caso de estudio

Enrique Brenes[1] and Alexandra Martínez[2]

[1] Posgrado en Computación e Informática, Universidad de Costa Rica
enrique.brenes@ecci.ucr.ac.cr
[2] Escuela de Ciencias de la Computación e Informática, Universidad de Costa Rica
alexandra.martinez@ecci.ucr.ac.cr

Resumen Este caso de estudio explora la aplicabilidad del nivel 1 del estándar para verificación de la seguridad de aplicaciones de OWASP en el contexto de una aplicación web de la industria financiera. Dos analistas de calidad no expertos en seguridad se encargaron de ejecutar el nivel 1 del estándar, reportando en una bitácora varias métricas relativas a las pruebas realizadas, el esfuerzo requerido y las vulnerabilidades encontradas. Luego, el equipo de desarrollo corrigió las vulnerabilidades reportadas, registrando el esfuerzo de corregir cada vulnerabilidad. Finalmente, un grupo de expertos en seguridad realizó una evaluación de la aplicación, detallando sus hallazgos en un informe. Los resultados aportan evidencia de que el nivel 1 del estándar ASVS puede ser aplicado por analistas de calidad que no sean expertos en seguridad, mediante análisis manual de código y técnicas de pruebas de penetración con apoyo de herramientas. En el software bajo estudio, las vulnerabilidades de Control de acceso y Autenticación fueron las más frecuentes, de mayor severidad y con mayor esfuerzo de reparación. La evaluación realizada por expertos en seguridad ayudó a comprobar que la cobertura del nivel 1 del estándar fue alta a pesar de haber sido realizada por personal inexperto en pruebas de seguridad.

Keywords: Pruebas de seguridad; OWASP; ASVS; vulnerabilidades.

1. INTRODUCCIÓN

Las aplicaciones web están sujetas a características del ambiente en que se ejecutan como el acceso abierto y conexiones sin estado, por lo que su seguridad es un tema de preocupación creciente tanto en la academia como en la industria [8]. Cuando estas aplicaciones administran información sensible de personas y negocios, el tema de la seguridad se vuelve aún más crítico [2].

Una vulnerabilidad es un defecto de seguridad en el software que puede ser usado maliciosamente por un atacante para ganar acceso al sistema o a la red [1]. En el año 2013 el Proyecto Abierto de Seguridad de Aplicaciones Web (OWASP, por sus siglas en inglés), publicó las diez vulnerabilidades de seguridad más comunes en aplicaciones industriales, con base en un conjunto de datos de ocho

firmas que se especializan en seguridad de aplicaciones. Estas vulnerabilidades son: inyección de código maligno, administración de la sesión, vulneración de la autenticación, *cross-site scripting*, referencias directas a objetos de forma insegura, exposición de datos sensibles, control de acceso, *cross-site request forgery*, utilización de componentes con problemas conocidos, y redireccionamiento y reenvíos no validados [3]. En el año 2014 el OWASP publicó la última versión del estándar para verificación de la seguridad de aplicaciones (ASVS, por sus siglas en inglés), el cual tiene tres niveles de verificación: oportunista (nivel 1), estándar (nivel 2) y avanzado (nivel 3) [4]. El nivel 1 contiene 45 verificaciones de seguridad divididas en 9 requerimientos.

Las pruebas de seguridad pueden ser utilizadas para identificar vulnerabilidades en una aplicación web. Las técnicas más utilizadas son las pruebas de penetración y el análisis de código [1]. Las pruebas de penetración consisten en ejecutar una aplicación web de forma remota para encontrar vulnerabilidades, con los mismos permisos que tendría un usuario final y sin saber cómo funciona la aplicación por dentro. El análisis de código consiste en revisar el código en busca de vulnerabilidades, sin ejecutarlo. Esta técnica es muy efectiva cuando es realizada por expertos en seguridad, aunque también existen herramientas automatizadas que apoyan el proceso.

Los objetivos de este estudio fueron: (1) investigar la aplicabilidad del nivel 1 del estándar ASVS a una aplicación web de la industria financiera, por parte de analistas de calidad del software no expertos en seguridad, (2) caracterizar las vulnerabilidades de seguridad encontradas y (3) analizar la efectividad en la cobertura del estándar. Para lograr estos objetivos, nos planteamos las siguientes preguntas de investigación:

RQ 1 ¿En qué grado se puede aplicar el nivel 1 del estándar ASVS? Esta pregunta se subdividió en tres preguntas:

- RQ 1.1 ¿Qué grado de experticia en seguridad necesita un analista de calidad para aplicar el nivel 1 del estándar ASVS?
- RQ 1.2 ¿Cuánto esfuerzo toma la aplicación del nivel 1 del estándar ASVS?
- RQ 1.3 ¿Cuáles herramientas pueden apoyar la aplicación del nivel 1 del estándar ASVS y en qué medida?

RQ 2 ¿Cómo se caracterizan las vulnerabilidades encontradas al aplicar el nivel 1 del estándar ASVS? Esta pregunta se subdividió también en tres preguntas:

- RQ 2.1 ¿De qué tipo y severidad son las vulnerabilidades encontradas?
- RQ 2.2 ¿Cuánto esfuerzo requiere corregir las vulnerabilidades encontradas?

RQ 3 ¿Qué tan efectiva fue la cobertura del nivel 1 del estándar ASVS? Esta pregunta se subdividió en 2 preguntas:

- RQ 3.1 ¿Cómo se caracterizan las vulnerabilidades descubiertas por los expertos, en términos de su tipo, severidad, nivel del estándar y esfuerzo de reparación?
- RQ 3.2 ¿En qué medida validan los hallazgos de los expertos la cobertura del nivel 1 del estándar ASVS lograda por los analistas de calidad?

2. Metodología

2.1. Contexto

El estudio se realizó entre agosto de 2014 y marzo de 2015, en la empresa que para efectos de este reporte llamaremos ServiciosFinancieros[3], la cual desarrolla sistemas de información financieros (el primer autor laboraba en esta empresa al momento de realizar el estudio). En particular, se escogió una aplicación desarrollada en un 85 %, que usaba C# .NET, Html5, Javascript y Google Angular, una arquitectura RESTFul y una base de datos Oracle. En el contexto de esta aplicación, la evaluación por parte de expertos en seguridad era un requisito para liberar a producción. Debido a que el desarrollo de la aplicación no había considerado aspectos de seguridad desde un inicio, se identificó el riesgo de que la evaluación final de seguridad descubriera vulnerabilidades difíciles de solucionar, comprometiendo la fecha de salida a producción, que ya había sido pactada. Para mitigar este riesgo, se optó porque los analistas de calidad de la empresa realizaran pruebas de seguridad apoyadas en el estándar ASVS, con la consigna de lograr una alta cobertura del nivel 1 de dicho estándar.

2.2. Procedimiento y roles

La aplicación del nivel 1 del estándar estuvo a cargo de dos analistas de calidad sin experiencia previa en pruebas de seguridad (uno de los cuales es el primer autor), quienes utilizaron análisis manual de código y técnicas de pruebas de penetración con apoyo de herramientas en la ejecución del estándar. Una vez terminada la ejecución del nivel 1 del estándar, el equipo de desarrollo procedió a corregir las vulnerabilidades encontradas. Una vez probadas las soluciones a todas las vulnerabilidades, un equipo de expertos en seguridad realizó una evaluación de la aplicación, documentando sus hallazgos en un informe. Dicho informe fue revisado y discutido por el equipo de desarrollo en conjunto con uno de los expertos en seguridad. Los hallazgos se solucionaron en las semanas siguientes. Finalmente, el equipo de expertos colaboró con un ciclo de pruebas adicional para confirmar que las vulnerabilidades reportadas ya no eran reproducibles.

2.3. Recolección de datos

Para almacenar los datos relevantes al estudio, se usaron tres bitácoras (hojas de cálculo). La *bitácora de progreso* especificaba, para cada verificación del estándar, su estado (si pasó o falló), la razón del estado y la técnica o herramienta usada. La *bitácora de vulnerabilidades* registraba, para cada vulnerabilidad encontrada por un analista de calidad o por un experto en seguridad, su categoría, estado, severidad, prioridad, descripción, correspondencia con el estándar

[3] La empresa del caso que se presenta no ha autorizado a difundir su nombre, pero los autores han mostrado suficiente evidencia al comité de programa para afirmar que el reporte se formula a partir de datos recogidos en terreno.

(nivel) y esfuerzo requerido para arreglarla. La *bitácora de esfuerzo* registraba el esfuerzo requerido (en horas) para ejecutar cada requerimiento del nivel 1 del estándar, así como el estado de su ejecución. La información acerca del tiempo requerido para ejecutar el estándar y de las vulnerabilidades reportadas, se extrajo del Microsoft Team Foundation Server, herramienta utilizada para el reporte de defectos y gestión del proyecto.

2.4. Análisis y procesamiento de datos

Para responder RQ 1 se tomaron como insumos la bitácora de progreso y la bitácora de esfuerzo. Se calculó el porcentaje de cobertura del nivel 1 del estándar contando las verificaciones que se pudieron realizar y las que no. Para RQ 1.1, se extrajo un detalle del perfil de los analistas encargados de ejecutar el estándar, el entrenamiento requerido y las limitaciones encontradas. Para RQ 1.2, se graficó el esfuerzo invertido en cada requerimiento del estándar, para visualizar cuáles requerimientos tomaron más tiempo. Para RQ 1.3, se contó la cantidad de verificaciones que fueron apoyadas por herramientas y se calculó la contribución de cada herramienta a la aplicación del estándar (cuántas verificaciones apoyó). Las herramientas fueron seleccionadas con base en las sugerencias de la guía de pruebas de OWASP[6]. Adicionalmente se incluyó AppScan [7], debido al amplio uso que tiene en la empresa donde se realizó el estudio.

Para responder RQ 2 se utilizó la bitácora de vulnerabilidades. Para RQ 2.1, se elaboraron gráficos con las vulnerabilidades encontradas por los analistas, agrupadas por requerimiento (tipo) y severidad. Para RQ 2.2, se graficó el esfuerzo requerido para resolver las vulnerabilidades, agrupado por requerimiento.

Para responder RQ 3 también se utilizó la bitácora de vulnerabilidades. Para RQ 3.1, el análisis fue muy similar al de RQ 2, con la adición de un gráfico de distribución de las vulnerabilidades según el nivel del estándar al que corresponden. Esta distribución permitió abordar RQ 3.2, mediante el conteo de las vulnerabilidades encontradas por los expertos en seguridad que debieron haber sido detectadas por los analistas de calidad.

3. Resultados

3.1. Resultados de RQ 1: aplicación del nivel 1 del estándar ASVS

Para el software estudiado, fue posible ejecutar el nivel 1 del estándar en un 98 %, según el reporte de los analistas. La única verificación que no se pudo ejecutar fue la de seguridad de las comunicaciones, por restricciones de infraestructura impuestas por el departamento de TI. El Cuadro 1 muestra el resultado de ejecutar las 45 verificaciones del nivel 1. Una verificación 'aprobada' significa que sí existía el respectivo control en la aplicación; una 'fallida' significa que se descubrió al menos una vulnerabilidad; y 'no aplica' significa que el control no existía en la aplicación por razones justificadas (no era una vulnerabilidad).

Cuadro 1. Resultados de la ejecución de las verificaciones del nivel 1 del estándar.

Estado de la verificación	Cantidad
Aprobada	24
Fallida	17
No aplica	3
No ejecutada	1
Total	45

RQ 1.1 - Grado de experticia en seguridad. Los analistas que ejecutaron el estándar tenían entre 2 y 3 años de experiencia en pruebas de software, pero no tenían experticia en pruebas de seguridad. Por lo tanto, se les dio un entrenamiento (de dos horas) sobre las diez vulnerabilidades más comunes reportadas por OWASP. Por otro lado, los analistas utilizaron características puntuales de las herramientas ZAP[5], AppScan, RequestMaker[9] y herramientas de desarrollo del navegador web para apoyar la aplicación del estándar. Cada analista tuvo dos días para adiestrarse en el uso de estas herramientas. En síntesis, el grado de experticia en seguridad que necesitaron los analistas de calidad para aplicar el nivel 1 del estándar fue muy poco (capacitación de 2 horas sobre las vulnerabilidades más comunes y 2 días para aprender a usar las herramientas).

RQ 1.2 - Esfuerzo de aplicación del nivel 1 del estándar. La aplicación del nivel 1 del estándar requirió un total de 95 horas. La Fig. 1 muestra la distribución del esfuerzo (en horas) por requerimiento del estándar. Los requerimientos "Administración de la sesión" y "Manejo de entradas maliciosas" representaron conjuntamente un 55 % del esfuerzo total.

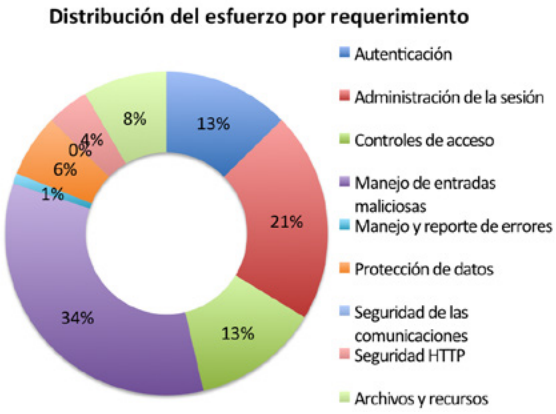

Figura 1. Distribución del esfuerzo requerido para aplicar cada requerimiento del nivel 1 del estándar.

RQ 1.3 - Herramientas de apoyo a la aplicación del nivel 1 del estándar. Las herramientas contempladas en esta investigación apoyaron un 48 % de las verificaciones realizadas. Las herramientas fueron utilizadas con diversos propósitos, entre ellos, dar exhaustividad a las pruebas de inyección de código, atrapar tráfico HTTP, generar peticiones personalizadas al servidor y listar las *urls* del sitio para realizar pruebas de accesos no autorizados. La Fig. 2 muestra la contribución de cada herramienta a la aplicación del estándar, en términos de la cantidad de verificaciones que apoyó. AppScan resultó ser la que apoyó la mayor cantidad de verificaciones. En algunos casos, varias herramientas apoyaron una misma verificación.

3.2. Resultados de RQ 2: vulnerabilidades de nivel 1 del estándar

Se encontraron vulnerabilidades en 6 de los 9 requerimientos evaluados por los analistas de calidad, siendo "Autenticación" y "Control de acceso" los requerimientos con más vulnerabilidades. Es interesante que no se encontraran vulnerabilidades en el requerimiento "Manejo de entradas maliciosas", a pesar de ser el que encabeza la lista de vulnerabilidades más comunes de OWASP. Esto puede atribuirse a que durante el desarrollo de la aplicación, las pruebas funcionales validaron consistentemente que no se permitieran entradas maliciosas tanto a nivel de interfaz de usuario como a nivel de parámetros en las peticiones al servidor.

RQ 2.1 - Tipo y severidad de las vulnerabilidades. La severidad de las vulnerabilidades fue asignada por el Departamento de Seguridad de la empresa. El tipo de la vulnerabilidad está dado por el requerimiento al cual se asocia. La Fig. 3 muestra la cantidad y severidad de las vulnerabilidades encontradas por requerimiento (tipo). Los requerimientos "Autenticación" y "Control de acceso" exhibieron la mayor cantidad de vulnerabilidades y la mayor concentración de severidades críticas. "Control de acceso" concentró las vulnerabilidades de seve-

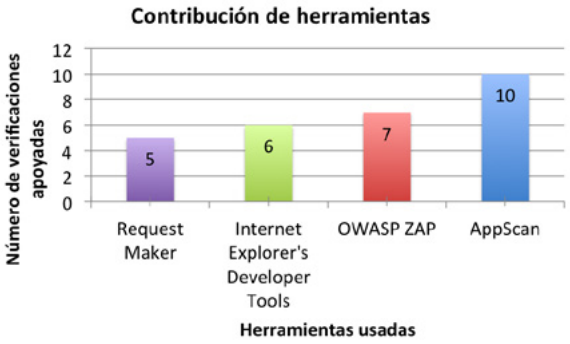

Figura 2. Contribución de cada herramienta a la aplicación del nivel 1 del estándar.

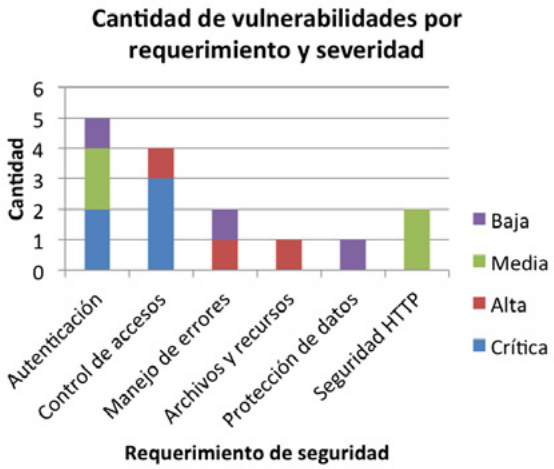

Figura 3. Severidad y cantidad de vulnerabilidades por requerimiento.

ridad más alta: 3 críticas y 1 alta. En total hubo 5 vulnerabilidades de severidad crítica, 3 de severidad alta, 4 de severidad media y 3 de severidad baja.

RQ 2.2 - Esfuerzo de corrección de las vulnerabilidades. La corrección de las 15 vulnerabilidades encontradas por los analistas requirió un esfuerzo equivalente a 55 días de trabajo (con 8 horas diarias). La Fig. 4 muestra la distribución del esfuerzo de corrección de vulnerabilidades por requerimiento. Un resultado notable es que el esfuerzo requerido para solucionar las vulnerabilidades en los requerimientos "Autenticación" y "Control de acceso" suma alrededor de 48 días, es decir, un 90 % del esfuerzo total.

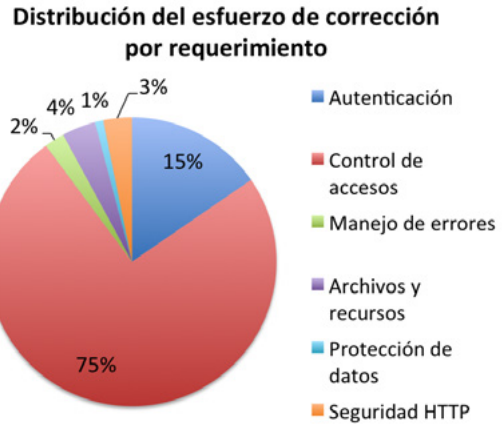

Figura 4. Esfuerzo para corregir las vulnerabilidades reportadas por los analistas.

3.3. Resultados de RQ 3: cobertura del nivel 1 del estándar ASVS

La evaluación realizada por los expertos en seguridad arrojó un total de 11 vulnerabilidades, de las cuales solo 3 correspondían al nivel 1 del estándar (y eran de baja severidad).

RQ 3.1 - Caracterización de las vulnerabilidades descubiertas por los expertos en seguridad. Los expertos en seguridad detectaron vulnerabilidades en los requerimientos de "Autenticación", "Control de accesos", "Lógica del negocio", "Seguridad de las comunicaciones", "Protección de datos", "Seguridad HTTP" y "Administración de la sesión". Una buena parte de las vulnerabilidades descubiertas tuvieron que ver con certificados de seguridad, cifrado de las comunicaciones, protocolos de seguridad en la capa de transporte y configuraciones del servidor de aplicaciones. El Cuadro 2 muestra la distribución de las vulnerabilidades detectadas por nivel del estándar y por severidad. Se observa que la mayoría (55 %) de las vulnerabilidades detectadas fue de nivel 2, y en mucho menor grado se detectaron vulnerabilidades de nivel 1 (27 %) y nivel 3 (18 %). También se observa que las vulnerabilidades de severidad media y baja representaron conjuntamente un 55 % del total, mientras que las de severidad alta representaron un 45 %. Un hallazgo muy interesante fue que el esfuerzo de corrección de estas vulnerabilidades (descubiertas por expertos) fue mucho menor (6,2 días) que el de las vulnerabilidades encontradas por los analistas de calidad (55 días). Esto se debe a que una gran parte de los hallazgos de los expertos se solucionaron a través de modificaciones en la configuración del servidor de aplicaciones y de su entorno de ejecución. Por ejemplo, vulnerabilidades relacionadas con los protocolos SSL, TLS y HTTPS fueron reportadas por los expertos y corregidas con facilidad. Por su parte, algunas vulnerabilidades reportadas por los analistas de calidad requirieron correcciones en todos los controladores de la aplicación, las cuales fueron muy costosas. El requerimiento que más esfuerzo de corrección necesitó fue "Protección de datos", con 2 días. El esfuerzo requerido para corregir las 3 vulnerabilidades de nivel 1 del estándar fue de tan solo 1,75 días, lo cual se considera bastante bueno pues no afectó significativamente el cronograma del proyecto. Estas 3 vulnerabilidades de nivel 1 fueron de severidad baja.

RQ 3.2 - Análisis de cobertura. La revisión por parte de los expertos en seguridad permitió identificar vulnerabilidades en el nivel 1 del estándar que los

Cuadro 2. Cantidad de vulnerabilidades reportadas por los expertos en seguridad, agrupadas por nivel y severidad.

Nivel del estándar	Sev. alta	Sev. media	Sev. baja	Total
Nivel 1 (Oportunista)	-	-	3	3
Nivel 2 (Estándar)	3	2	1	6
Nivel 3 (Avanzado)	2	-	-	2
Total	5	2	4	11

analistas de calidad pasaron por alto. Dicha evaluación también recalcó que las validaciones de los niveles 2 y 3 del estándar son necesarias para aprobar una evaluación de seguridad rigurosa. A continuación se listan los resultados más relevantes de la evaluación realizada por los expertos.

- Los expertos no lograron descubrir vulnerabilidades de severidad alta ni media en el nivel 1 del estándar (solo detectaron tres vulnerabilidades de baja severidad). Esto respalda que las pruebas realizadas por los analistas de calidad fueron altamente efectivas en cubrir el nivel 1 del estándar.
- El esfuerzo requerido para reparar las vulnerabilidades del nivel 1 del estándar detectadas por los expertos fue solo un 3 % del esfuerzo requerido para reparar las vulnerabilidades encontradas por los analistas de calidad. El esfuerzo total de reparación de todas las vulnerabilidades descubiertas por los expertos representó el 11 % del esfuerzo de reparación de las vulnerabilidades encontradas por los analistas. Este resultado fue el más importante desde el punto de vista del proyecto, ya que permitió liberar la aplicación a producción en la fecha prevista y con los arreglos exigidos por los expertos en seguridad.
- Los expertos lograron descubrir cinco vulnerabilidades de alta severidad en los niveles 2 y 3 del estándar, lo que realza la necesidad de someter la aplicación a una evaluación rigurosa por parte de expertos en seguridad, en lugar de conformarse con la cobertura del nivel 1 del estándar alcanzada por los analistas de calidad.

4. Conclusiones

Este trabajo exploró la aplicación del nivel 1 del estándar ASVS de OWASP a una aplicación web financiera. En una primera etapa, dos analistas de calidad sin experiencia previa en seguridad ejecutaron las verificaciones del nivel 1 del estándar, apoyados en herramientas existentes. En una segunda etapa, los desarrolladores corrigieron todas las vulnerabilidades encontradas por los analistas. En una tercera etapa, un grupo independiente de expertos en seguridad evaluaron la aplicación.

Los resultados de este estudio aportan evidencia de que para ejecutar el nivel 1 del estándar ASVS no es necesario tener experticia previa en el área de seguridad, pues los analistas de calidad que nunca habían hecho pruebas de seguridad fueron capaces de completar un 98 % del nivel 1 del estándar en 95 horas de esfuerzo.

En el contexto de la aplicación bajo estudio, la mayoría de las vulnerabilidades de nivel 1 encontradas por los analistas fueron de autenticación y control de acceso (ambas suman 9 de un total de 15 vulnerabilidades). Los requerimientos “Control de acceso” y “Autenticación” del estándar concentraron las vulnerabilidades de mayor severidad y mayor esfuerzo de corrección. Los demás requerimientos del nivel 1 del estándar presentaron vulnerabilidades con costos de corrección más bajos (por debajo de los dos días). Esto da pie para recomendar que las pruebas de control de acceso y autenticación se realicen en fases

tempranas del desarrollo, evitando que sean detectadas tardíamente, cuando el costo de reparación es muy alto.

Asimismo, este caso de estudio revela que el nivel 1 del estándar ASVS deja por fuera una serie de verificaciones importantes (protocolos de seguridad, certificados de seguridad y cifrado), asociadas al requerimiento de "Seguridad de las comunicaciones", que son vitales para la puesta en producción de aplicaciones web con requerimientos altos de seguridad.

La evaluación realizada por el equipo de expertos detectó vulnerabilidades en el nivel 1 del estándar que los analistas de calidad no lograron encontrar. Sin embargo, dichas vulnerabilidades fueron reportadas con una severidad baja y tuvieron un esfuerzo de reparación menor a 2 días.

Referencias

1. Awang, N., and Manaf, A. Detecting vulnerabilities in web applications using automated black box and manual penetration testing. In *Advances in Security of Information and Communication Networks*, Springer Berlin Heidelberg (2013), 230–239.
2. Dukes, L., Yuan, X., and Akowuah, F. A case study on web application security testing with tools and manual testing. In *Southeastcon, 2013 Proceedings of IEEE* (April 2013), 1–6.
3. Foundation, T. O. Owasp top 10 - 2013. los diez riesgos mas criticos en aplicaciones web. `https://www.owasp.org/images/5/5f/OWASP_Top_10_-_2013_Final_-_Espa\%C3\%B1ol.pdf`, 2013. [Online; accessed Dec-2014].
4. Foundation, T. O. Application security verification standard (2014). `https://www.owasp.org/images/5/58/OWASP_ASVS_Version_2.pdf`, 2014. [Online; accessed Dec-2014].
5. Foundation, T. O. Owasp zed attack proxy project. `https://www.owasp.org/index.php/OWASP_Zed_Attack_Proxy_Project`, 2014. [Online; accessed Dec-2014].
6. Foundation, T. O. Testing guide introduction. `https://www.owasp.org/index.php/Testing_Guide_Introduction#Testing_Techniques_Explained`, 2014. [Online; accessed Dec-2014].
7. IBM. Ibm security appscan. `www.ibm.com/software/products/en/appscan`, 2014. [Online; accessed Dec-2014].
8. Mao, C. Experiences in security testing for web-based applications. In *Proceedings of the 2Nd International Conference on Interaction Sciences: Information Technology, Culture and Human*, ICIS '09, ACM (New York, NY, USA, 2009), 326–330.
9. Nurminen, J. Request maker download site. `https://chrome.google.com/webstore/detail/request-maker/kajfghlhfkcocafkcjlajldicbikpgnp`, 2014. [Online; accessed Dec-2014].