Propuesta de Seguridad en Herramienta de Desarrollo Colaborativo

Gustavo Sarubi, Martín Bonamico, Lucas Ambrosi, Ana M. Darcacha y Daniela López de Luise •

Resumen

El presente trabajo propone la extensión de una herramienta de trabajo colaborativo, a fin de inferir la posible intrusión por parte de terceros malintencionados y preservar la confidencialidad de la información involucrada. El análisis se centra en un framework de IBM denominado Jazz. Se presenta la arquitectura general conceptual, y se compara con alternativas, aplicando técnicas del área Web Mining.

Abstract

The present work proposes the extension of a collaborative work tool, in order to infer the possible intrusion of third-part ill-disposed, and to preserve the confidentiality of the involved information. The analiysis is centered on a IBM framework called Jazz. The general conceptual architecture is presented, and is compared with alternatives, applying technologies(skills) of the Web Mining area.

***Keywords*:** Computer security, Site security monitoring, Collaborative work.

• Universidad de Palermo; AIGROUP, Buenos Aires, Argentina. sarubi_gustavo@hotmail.com | martinbonamico@gmail.com | ambrosi@gmail.com | lopezdeluise@yahoo.com.ar

G. Sarubi es analista de sistemas de ORT, actualmente alumno de Licenciatura en Sistemas del Plan Complementación en UP, CABA, Argentina. Desde Octubre del 2004 al presente se desempeña como IT Speciallist en EDS para la cuenta Sprint en un grupo OffShore -entorno corporativo globalizado- desde Argentina dando soporte y desarrollo en aplicaciones de entorno midrange (e-mail: gustavo_sarubi@hotmail.com).

I. Introducción

Este documento se focaliza en una herramienta de desarrollo colaborativo de fuente abierta que actualmente la empresa IBM pone a disponibilidad de los interesados en el área del desarrollo de software.

En el contexto del trabajo de investigación del AIGroup, como parte del proyecto Jazz, se pretende realizar un aporte original para extender funcionalmente la herramienta del proveedor dando soporte o solución a la problemática de la detección de intrusos.

El objetivo perseguido consiste en desarrollar un esquema de seguridad para prevenir, mitigar, advertir la intrusión o violación de ciertos aspectos que pueden atentar en un entorno de desarrollo colaborativo. A continuación se analizan los antecedentes, objetivo, alcance y problemática. Luego, se describe la propuesta original basada en el estudio detallado a lo largo del paper y dejando planteado el trabajo a futuro para concretar dicha propuesta.

Desde los años ’80 el desarrollo de aplicaciones ha cambiado significativamente, sólo a modo de ejemplo se pueden citar características como el trabajo en desarrollos transaccionales con programación no estructurada, típicamente en equipos como mainframes, y el empleo de lenguajes como Cobol. Si bien ya se hablaba de grupos de desarrollo, el estilo de programación no ayudaba a la interacción de sus integrantes. Ya pasados los ’90 se trabajaba en una nueva idea de programación, la estructurada, que venía a resolver aquellos programas extensos y difíciles de comprender, no sólo por su volumen en líneas de código, sino también por la numerosa cantidad de saltos y bifurcaciones. Esto generaba una dependencia en el autor o conocedor del programa realizado como sus características funcionales. Lo cual, como se ha mencionado, cada uno responsable de cierta funcionalidad, grupo de programas o ciertas partes de la aplicación. Con la llegada de la programación estructurada en los ’90 se acotó a módulos aquellos programas de gran extensión, prevaleciendo los bucles repetitivos y diversificación de funcionalidad en funciones y módulos. Si bien se pensó que se estaba frente a la solución de los inconvenientes citados, en realidad fue meta y punto de partida para una nueva concepción en el análisis y pensamiento en objetos.

Así como en el desarrollo de software empezó a tomar preponderancia la programación basada en objetos, también comenzaron a ser comunes las herramientas de control de versiones de código, elementos básicos como el correo electrónico y herramientas de mensajería instantánea en todo grupo de desarrollo de software o ‘software factories’.

Este conjunto de componentes son los que actualmente resultan indispensables y los que entre otros, forman el entorno o herramienta de desarrollo colaborativo.

Estos entornos atienden las necesidades de sus integrantes como grupo de desarrollo, que pueden o no estar ubicado físicamente en el mismo área, departamento, edificio, ciudad o país, favoreciendo a los grupos de trabajo globales, es decir, grupos de trabajo compuesto por integrantes trabajando a distancia o en forma remota [1].

Por lo tanto, el tema de la seguridad en entornos colaborativos se vuelve un tema de importancia.

El autor no ha encontrado estudios publicados en la actualidad sobre el tema de seguridad en entornos o herramientas de desarrollo colaborativo, dado que es un concepto nuevo y poco explorado. En consecuencia, se trabajó sobre conceptos del área Web Mining (específicamente en el rol del Web Usage Mining) dentro del esquema de seguridad. Esto constituye la herramienta fundamental para la realización de la propuesta o aporte original [5].

La propuesta consiste en hacer una investigación de la herramienta Jazz, presentar las características relevantes de un módulo de seguridad basado en la exploración de datos sobre los log del sistema. A tal fin, se explica cómo funciona y los componentes elementales.

Para cumplimentar el objetivo, se ha realizado un estudio preliminar a partir del análisis de fortalezas y debilidades de una herramienta de desarrollo colaborativo de última generación, la cual se presenta al mercado recientemente.

Es importante destacar que no parecen existir desarrollos de este tipo en organismos de publicaciones registradas como el IEEE [5].

Por último cabe destacar que ciertos conceptos se basan en las estrategias conocidas dentro del ámbito del Web Usage Mining, que aporta a esta propuesta de seguridad con la finalidad de obtener conocimientos para ser utilizados en una etapa posterior que detallamos en el Capítulo 0 Los detalles son analizados en el Capítulo VI.

A. Alcance

La herramienta de desarrollo colaborativo que compone el Proyecto IBM Jazz, se llama Rational Team Concert V1.0 [1]. Esta herramienta o producto es de fuente abierta, y forma parte de la iniciativa impulsada por IBM para presentar el Proyecto Tecnológico Jazz, donde ofrece además del producto Rational Team Concert, documentación de ayuda, la plataforma de funcionamiento (o servidor Web), acceso a la comunidad de investigadores a nivel global para interactuar y la posibilidad de plantear inquietudes, propuestas de mejoras y extensiones a través de un sitio oficial. Todo esto forma parte de la iniciativa tecnológica impulsada por IBM y denominada Jazz Project, que se presenta como toda una plataforma de desarrollo colaborativo

extensible [2]. La propuesta planteada en este trabajo es una extensión que contribuye y aporta seguridad al producto.

La contribución o extensión constituye un módulo inteligente de seguridad y detección de ataques de usuarios a través del reconocimiento de patrones de comportamiento maliciosos.

En el marco de trabajo para llegar a dicha propuesta, se han realizado los estudios e investigación y práctica con la herramienta de desarrollo corporativo, desde su instalación, registración en comunidad del proyecto que IBM lleva a cabo como lanzamiento de tecnología para conseguir información, documentación de la misma, participación de foros, etc.

Una vez analizada la herramienta, se ha procedido a diseñar la arquitectura de la propuesta, tomando como base la información y pruebas de las herramientas a utilizar para llevar a cabo dicha enunciación y la justificación de la misma.

Adicionalmente esta iniciativa fue presentada y publicada para el IBM University Research & Collaboration – IBM Jazz Innovation Awards 2008 presentado en Octubre 2008

La información hoy en día es uno de los activos más importantes, no sólo para las empresas y organizaciones, sino para cada individuo. Por este motivo, la misma requiere ser protegida en forma apropiada [3].

La seguridad de la información es el conjunto de metodologías, prácticas y procedimientos que buscan proteger la información como activo valioso, con el fin de minimizar las amenazas y riesgos continuos a los que está expuesta, a efectos de asegurar la continuidad del negocio, minimizar los daños a la organización y maximizar el retorno de inversiones y las oportunidades del negocio [3].

Unos de los aspectos trascendentes de esta propuesta consiste en superar el nivel de seguridad bajo un control de usuario con perfiles de acceso por roles, ya que mayormente no se mantiene un procedimiento de cambio de claves, de monitoreo o auditoria de accesos.

B. Importancia del Problema

Para poder administrar la seguridad de la información en forma correcta, se deben establecer y mantener políticas y controles que busquen cumplir con los tres requerimientos de mayor importancia [6]: La Confidencialidad, la Integridad y la Disponibilidad de los recursos de las organizaciones, siendo los tres requerimientos básicos definidos como:

- **Confidencialidad.**
  La Confidencialidad busca prevenir el acceso no autorizado ya sea en forma intencional o no intencional de la Información. La pérdida de la Confidencialidad

puede ocurrir de muchas maneras, como por ejemplo con la publicación intencional de Información confidencial de la organización.

- **Integridad.**
  El concepto de Integridad busca asegurar:
  - Que no se realicen modificaciones por personas no autorizadas a los datos, Información o procesos.
  - Que no se realicen modificaciones no autorizadas por personal autorizado a los datos, Información o procesos.
  - Que los datos o la Información sea consistente tanto interna como externamente.

- **Disponibilidad.**
  La Disponibilidad busca el acceso confiable y oportuno a los datos, Información o recursos para el personal apropiado.

La implementación de políticas y procesos de Seguridad invariablemente es más económica si se aplica al principio del proyecto y no al final del mismo.

En las herramientas de desarrollo como Jazz, donde se maneja una diversidad de proyectos de desarrollo, existe un riesgo alto que corren las organizaciones sino se hace foco en asegurar la información que se maneja en este tipo de herramientas, no sólo para resguardar el código fuente, o los datos que una aplicación productiva puede manejar, sino también prevenir las intrusiones maliciosas con el mero objetivo de tomar información confidencial sobre el avance –métricas / panel de comandos- de un proyecto en desarrollo que puede ser interesante sabotear u obtener información por interesados de alguna manera en ellos.

Por citar un ejemplo, en algún desarrollo de índole gubernamental podría no ser conveniente su puesta en funcionamiento para cierto sector de la sociedad, grupo empresario, sector o agrupación o inclusive algún individuo en particular.

C. Aporte Original

La herramienta de desarrollo colaborativo mantiene almacenados registros de logs en el servidor Web, los cuales son explorados y analizados principalmente mediante técnicas que se relacionan con la inteligencia artificial, la minería de datos, la teoría de base de datos, etcétera [10].

La minería de uso de Web consiste en tres fases, a saber: el proceso previo de datos, el descubrimiento de modelo y el análisis de modelo.

Se toma de la herramienta los logs, tratados mediante un procesamiento que se detalla en el Capítulo VI. Permite obtener conocimiento de los eventos registrados en los logs y mediante este estudio alertar a la herramienta sobre el potencial ataque de un usuario malicioso o al menos de un proceder no común de acuerdo al tipo de tareas que realiza comúnmente.

Analizadas las diferentes metodologías existentes en el amplio espectro del Web Usage Mining se comienza haciendo foco en lo más general y comúnmente utilizados, Markov y Clustering, para estudiar sus características, y aplicación a este caso de estudio y su articulación en la propuesta descripta en Capítulo VI.

II. Uso de cadenas de Markov en seguridad

Dado que el autor no ha hallado publicaciones con aplicación de cadenas de Markov a sistemas de trabajo colaborativo, se estudiará un problema similar consistente en predecir el comportamiento de un usuario en un sitio Web. El problema ha ganado importancia debido al crecimiento rápido del contenido existente en Internet y la necesidad de personalizar e influir en la navegabilidad del usuario.

El modelo de Markov y sus variantes son apropiados para encarar este problema. Dentro de las alternativas de estudio de comportamiento, el modelo de Markov generalmente se encuentra el que mejor aplica y muestran más alto coeficiente de exactitud predictiva.

Ha sido usado para estudiar y entender modelos estocásticos, esto es, modelos que evolucionan de forma no determinista a lo largo del tiempo en torno a un conjunto de estados.

Una cadena de Markov representa un sistema que varía su estado a lo largo del tiempo. Los cambios no están predeterminados, aunque sí lo está la función de distribución de probabilidades del próximo estado en función de los estados anteriores. Eventualmente, en una transición, el nuevo estado puede ser el mismo que el anterior y es influir en las probabilidades de la transición actuando adecuadamente sobre el sistema.

En general los datos de entrada son las páginas visitadas por un usuario. Al aplicar Markov se pretende inferir la secuencia de páginas, clicks o Web sites que el usuario potencialmente vaya a elegir [13].

Esta técnica se utiliza para la clasificación de tipos de sesiones según la navegación, llegándose a construir diferentes categorías. En algunos casos no ha sido muy apropiada para predecir comportamiento ya que no realiza procesamiento histórico para discriminar correctamente diferentes patrones observados.

También tiene problemas con el alto grado de crecimiento de la complejidad según la cantidad de estados que son comúnmente utilizados, y si fueran reducidos se perdería la precisión en la predicción.

III. Uso de Clustering en Seguridad

Existen diversos y extensos trabajos en los cuales se hace uso del método agrupamiento (clustering) en el entorno Web, entre los cuales se pueden mencionar los que siguen:

A. Minería sobre los logs del servidor:

La minería sobre los archivos de logs en servidores es el estudio de las características en el uso y navegabilidad de los usuarios de Web Sites es significativamente útil para mejorar el modelo de acceso, la estructura del Web Site, y fundamentalmente, obtener información sobre los intereses de sus usuarios y recomendar links de acuerdo a esos intereses[7].

El objetivo es ayudar al usuario a encontrar lo que busca. Entre otras cosas los logs del servidor Web permiten obtener información de la hora de acceso a un link, el tiempo transcurrido en él, la secuencia de links o sitios visitados, etc. Esta información puede ser analizada, agrupada y guardada con el objetivo de ayudar al usuario en próximas búsquedas. Este tipo de sistemas llamados de recomendación personalizada, se basa en Minería de uso de Web (Web Usage Mining), y frecuentemente usan métodos como minería en reglas de asociación, descubrimiento de patrones de secuencia, clustering de usuario integrado, clasificación y técnicas de minería de asociación [12].

Para el caso de estudio la información extraída mediante los logs son el input del módulo de seguridad y los métodos mencionados los utilizados sobre ellos para inferir conocimiento acerca del uso de la herramienta.

La minería de uso Web con sets de datos en bruto es uno de los objetivos del Web Usage Mining es el de obtener conocimiento sobre el empleo, navegabilidad o comportamiento que el usuario hace mientras utiliza el browser [16]. Podría decirse que la minería de datos en el uso de la Web consiste en el descubrimiento de patrones de secuencia en la navegación, extractados en grandes logs del servidor Web. Ya que comúnmente estos logs son utilizados con fines de depuración de errores. Deberían ser preprocesados antes de aplicar las técnicas del data Mining en ellos.

Además de las técnicas del Web Mining el **pre-procesamiento de los logs** del Web Server son identificados en los siguientes: 1) Limpieza de los datos 2) Identificación de usuario 3) Identificación de la sesión 4) Determinación la secuencia de navegación 5) Formateo.

Estas etapas conforman el pre-procesamiento necesario para aplicar a los grandes logs de un Web Server y antes de utilizar alguna de las técnicas de Web Mining.

Estas técnicas son utilizadas en el módulo 'Data Extractor' del módulo de Seguridad propuesto y explicado en detalle en el Capítulo VI. B.

B. Perfiles de usuarios:

Un punto importante en las investigaciones acerca de Web Usage Mining es el de clustering de usuarios Web basados en propiedades comunes. Un método de cluster o agrupamiento es el de medir similitud de intereses entre patrones de acceso de usuarios Web diferentes. Esto brinda información muy útil acerca de diferentes aspectos de intereses, demográficos, información relativa a características de comportamientos de los usuarios, etc. Todo esto es información que el servidor mantiene y puede ser utilizada.

Esta información ha sido abordada desde diferentes perspectivas, por ejemplo, dos usuarios que acceden a las mismas páginas pueden ser registrados en diferentes grupos de intereses similares si ellos acceden en diferente orden. El método de clustering puede presentar inconvenientes de performance cuando se incrementa el trabajo con gran volumen de información [9].

En este método toma como protagonista los atributos relativos a la sesión, tal como, la frecuencia de la página visitada, la lista de links visitados, el tiempo consumido entre dos links visitados, y el orden de las páginas visitadas.

Estas características de sesión se encuentran en los logs del servidor, en caso contrario pueden ser adicionadas para su posterior utilización.

C. Patrones de navegación:

Las aplicaciones de personalización Web, además de ayudar al usuario a encontrar lo que busca, permiten sugerir opciones, ya sea por un fin comercial o para mejorar la navegabilidad[12]. Basado en técnicas de Web Usage Mining y valiéndose en tiempos de navegación, tiempo de estadía en el sitio, la secuencia de páginas visitadas, entre otras es que se llega a analizar el comportamiento del usuario. Luego se adquieren patrones de comportamiento de acceso de grupos de usuarios. Utilizando esta información es que se puede recomendar al usuario opciones que potencialmente atiendan sus intereses. En el trabajo con esta información se utilizan métodos tales como **minería de reglas de asociación, patrones de búsqueda secuencial, clasificación y agrupamiento.**

El Web Mining basado en la minería de reglas de asociación ha ganado más espacio. Aunque los patrones de asociación han tenido un bajo rendimiento en la predicción de comportamiento de navegabilidad futura debido al escaso acoplamiento en el rango de predicción de comportamiento de navegación de usuarios y roles.

En [12] se propone el uso de logs integrando dos técnicas de Web Mining: el agrupamiento de usuario y la minería de asociación.

IV. Diferentes metodologías de clustering:

En algunos estudios revisados [8] se expone dividir el trabajo en tres etapas: 1) Pre-procesamiento de datos 2) Generación del agrupamiento o cluster 3) Visualización del cluster.

Luego se toman atributos de medidas bien particulares y conocidos: 1) basado en índice de visitas o VOB (Visiting Order Based) es el más apropiado y preciso para hacer predicción o recomendación de navegación al usuario 2) basado en rutas de usuario PAB (Path Angle Based) es el atributo de medida basado en una fórmula que considera la distancia entre rutas. Como así también innovadores como el de similitud de medidas basado en rutas comunes SMCP (Similarity measure based on common paths) que mejoran los dos primeros en un 10% de efectividad.

Otro estudio, utiliza un método el llamado FAP-Mining [10] basado en la construcción de un algoritmo que, primero construye y luego recorre, un árbol de acceso frecuente (FP-Tree). Luego hace minería en esta información buscando patrones de acceso.

Tomando como característica principal fecha y horario dan origen al concepto de temporalidad [14] llamado TCMM (Temporal Cluster Migration Matrices) basado fundamentalmente en tomar la información de los logs por periodos de tiempo en secuencia, con información adicional.

Otra técnica de agrupación [15] llamada PLSA (Probabilistic Latent Semantic Analisis) basado en el algoritmo K-Means modificado que clasifica y agrupa de acuerdo a la relación entre la sesión, la página Web y el factor de latencia. Los factores semánticos o tareas son caracterizados por extracciones de contenido de páginas “dominantes” relacionadas con ciertos factores de latencia.

V. El Entorno Jazz

Como ya se ha mencionado en el Capítulo I. A. el Rational Team Concert es la primera aplicación basada en la tecnología IBM Jazz, la plataforma de tecnología de última generación para la entrega de software colaborativo de IBM. La plataforma Jazz provee una arquitectura ampliable diseñada para hacer el suministro de software más colaborador, productivo y transparente. Adaptado a las necesidades de los equipos globales, combina personas, proyectos y procesos con la automatización para acelerar el ciclo de vida del software y mejorar el manejo del proyecto. Además de los procesos, el uso de métricas y responsabilidades para alcanzar las metas del negocio, manejar prioridades y tomar decisiones con datos. La tecnología Jazz está

construida sobre tecnología Eclipse y servirá de base a la plataforma IBM Rational Software Delivery Platform ya que se prevé a futuro cercano integrar a esta plataforma otras tecnologías de desarrollo de software como la de los sistemas OS/390 así como también proveer módulos orientados a la calidad y otro con foco a los requerimientos [1].

A. El concepto del Jazz y Rational Team Concert:

Las empresas de desarrollo de software están enfrentando una creciente presión para brindar en forma veloz, software de alta calidad en línea con los objetivos comerciales en constante evolución. Se espera que los equipos de trabajo ofrezcan más aplicaciones con igual número o con menores recursos. Muchos de estos equipos está geográficamente separados, lo que hace la colaboración aún más difícil, y muchas empresas encuentran dificultad en la aplicación de estándares y procesos constantes en proyectos y grupos muy diferentes [2].

El concepto de Jazz, como una orquesta donde hay buenos músicos y un director, que sigue una partitura, sin improvisación, como en el desarrollo de software como parte de un equipo de trabajo, es como tocar un instrumento en una banda. Ambos propósitos requieren un equilibrio de experiencia y colaboración. El IBM Rational Team Concert ofrece un entorno de colaboración, framework basado en Eclipse, que auxilia a las personas y a los grupos de trabajo para lograr el máximo rendimiento.

B. Herramientas integradas al entorno:

Posibilita un control integrado de la versión del software a través del Source Control Management, una gestión del espacio de trabajo, y el soporte del trabajo en paralelo para personas individuales y equipo [4].

Está diseñado para interconectar equipos de desarrollo dispersos y aumentar de esta manera la productividad individual y del mismo equipo. Para esto se vale de mensajería instantánea integrada al framework y al correo electrónico.

A través de visualización personalizable del portal Web, los miembros del equipo pueden obtener informaciones globales sobre los proyectos como noticias y acontecimientos, estado actual de compilación, sobre qué se ha estado trabajando y los cambios que hayan sido requeridos. Métricas y dashboards en real time para ser consultadas por niveles gerenciales desde el portal Web desde cualquier browser.

También puede visualizar sobre qué están trabajando sus compañeros de equipo, así como también quien está online, disponible para una eventual colaboración.

También existe el concepto de diferentes visualizaciones en el entorno de trabajo o framework Eclipse, que permite de acuerdo al rol del recurso, ya sea Developer,

Team Leader, Stakeholder (usuario clave y asignado a un proyecto específico) obtener una visión clara de lo que como recurso asignado a ese rol le interesa. También estas vistas o perspectivas son adaptables o customizables por el usuario.

C. Roles de trabajo:

Rational Team Concert crea y rastrea automáticamente el progreso de cada elemento de trabajo individual, de acuerdo al proceso y a las reglas de proyecto del equipo. Esta función permite que los defectos, mejoras, y comunicación fluyan eficazmente por todo el equipo. A medida que avanza el trabajo el framework captura informaciones valiosas como quién, qué, cuándo y porqué, asociadas con cada elemento de trabajo para proveer ese contexto tan necesitado cuando los elementos de trabajo se comparten entre miembros de algún equipo [2].

Provee una recolección de datos automatizada y funciones de información que brindan la visualización en tiempo real que se necesita para administrar el proyecto de software. Esto permite estar fácilmente en pleno conocimiento del día a día con el progreso del proyecto, resolver los problemas del mismo y realizar las correcciones necesarias anticipadamente durante el ciclo de vida del proyecto.

D. Los procesos en el ciclo de vida del software:

Permite automatizar el proceso y las funciones guiadas para la implementación de una determinada cantidad de procesos. Dado que cada empresa u organización mantiene diferentes metodologías, el Rational Team Concert permite configurar las reglas del proceso para satisfacer las necesidades de un proyecto específico o de una empresa específica [2].

Los tres tipos de procesos son:

- Cascada (Waterfall): más estable y confiable, pero entrega más lenta (el cliente tarda más en ver los resultados).
- Interactivo (Interactive): estable y confiable vs entrega más rápida.
- Ágil (Agile) más dinámica, de entrega al cliente rápida y contínua.

Permite a los grupos adoptar los principios más importantes y/o la metodología más adecuada, del ciclo de vida del software, Best Practices, inclusive armar la propia, ya sea con ciclos de repetición más cortos, intervención del cliente más usual y creaciones más frecuentes y automatizadas. Tiene una guía de proceso personalizable que automatiza la implementación de cualquier proceso de suministro de software y soporta metodología Agile y no Agile.

E. La plataforma Jazz, terminología y conceptos:

Muchos aspectos del ciclo de vida en el desarrollo de software está integrados, incluyendo planes de iteración, definición de procesos, control de versionado de código, tracking o registro de defectos, build manager y reporting. Todos estos aspectos están integrados en una sola herramienta que es el framework basado en Eclipse llamado Rational Team Concert o simplemente Jazz.

Una vez instalado el Rational Team Concert, y con el framework Eclipse listo para trabajar, el cliente se conecta al servidor de la Plataforma Jazz, específicamente al repositorio de conexiones especificando un Team Area. Una vez conectado desde el cliente, se visualizan los Artifacts que componen el área de trabajo o entorno colaborativo.

Los Artifacts son almacenados en el Repository, el cual puede ser accedido por los usuarios autorizados.

El repositorio contiene Project Areas las cuales contienen los artefactos de un proyecto. Cada area de proyecto está asociada a un Process, el cual gobierna cómo este proyecto va a desarrollarse en términos del ciclo de vida del software. También va a explicitar cómo será el comportamiento del ciclo de vida de ese proyecto dentro de Jazz.

Este proceso es definido por una especificación y descripción de proceso. La especificación de proceso define las iteraciones del proceso y como este se comportará durante esas iteraciones. La descripción de proceso es la descripta en el Capítulo V. D.

Como se ha mencionado, una vez conectado a la Project Area, el acceso a los Artifacts que la componen es visualizado y permitido.

En las Project Areas se encuentra un conjunto de Team Areas, donde cada una de ellas está compuesta por una lista de miembros del equipo y el rol que ellos cumplen dentro del equipo al cual pertenecen. Un usuario puede ser miembro de más de un grupo. Cada Team Area puede definir sus procesos propios procesos ajustados para su equipo y sub equipos.

Para proyectos simples, todas las actividades se lleva a cabo en un simple proceso principal de una única línea de desarrollo. Adicionalmente se pueden crear líneas o agregar para alguna actividad de mantenimiento que sea necesario contemplar. Cada línea de desarrollo tiene su propia Team Area y sus procesos customizados.

El trabajo planeado es descripto por los Work Item. Existen diferentes tipos de work ítems que son usados en Project Area son definidos por los procesos. Por ejemplo: El proceso Eclipse Way (un tipo de proceso ágil definido por default en la herramienta) define tipos de work item para defects (defectos), task (tareas) y enhancements (mejoras). Cada uno de estos tipos de work item tiene sus propios estados de transición y sus campos ajustables. Los Work Item son categorizados

en Categories, las cuales permiten organizar dichos Work Items por áreas funcionales. Cada Project Area define una lista de categorías disponibles. Cada Team Area is asociada con la categoría para el área funcional de la cual el equipo de trabajo es responsable.

El trabajo en una Project Area se realiza en una secuencia de iteraciones con fechas de comienzo y final, las cuales son definidos en el Estado de Proceso. Una de las iteraciones es definida como la actual por el proceso. Entonces cuando se planifica el trabajo, se le asigna un Work Item para una iteración en particular. Usted puede planificar todo el trabajo que debería entrar en una iteración creando un Iteration Plan.

Cada usuario utiliza un espacio de trabajo personal, el Repository Workspace para trabajar sobre los archivos del proyecto que están bajo el control versionado de código fuente. Al cargarse el Repository Workspace se copian los archivos y carpetas del servidor en su espacio de trabajo del Eclipse cliente. El dicho intercambio de archivos el Jazz Team Server también rastrea todos los cambios hechos a los archivos fuente, guardando por cada uno de ellos información relevante como, nombre de archivo y carpeta, un comentario adjunto, y la referencia al Work Item que motivó el cambio.

El usuario acepta en el check-in para cargar los archivos y carpetas modificadas en su espacio de trabajo de Eclipse (Your Repository Workspace de la Fig.1) al espacio de trabajo de depósito (Team Member's Repository Workspace de la Fig.1). Después que el usuario hizo el check in de los archivos y éstos son almacenados en el depósito, todavía no son compartidos con el resto del equipo de desarrollo hasta que el usuario no distribuya los change sets. El proceso de check-in y distribución de cambios proporciona una protección extra para el control de cambios, dándole la flexibilidad para continuamente hacer cambios sin entregar o provocar conflictos con otros integrantes del Team Area.

VI. Propuesta Security Module para Jazz

El entorno de desarrollo colaborativo está compuesto por varios módulos, inclusive algunos desarrollados por otras empresas y/o partners de IBM [4].

Uno de estos componentes es el Jazz Team Server, el cual es instalado en el servidor y permanece activo y disponible para que los clientes se conecten. El Jazz Team Server es el nexo entre los diferentes módulos que componen el paquete del Rational Team Concert, módulos como soporte a procesos, herramientas de reportes, etc.

Entre una de sus funcionalidades que posee, es un portal Web para administrar y configurar el sistema por usuarios específicos para tal fin [4].

Este servidor Web en su estructura de directorios, contiene un directorio: [installation-path]\jazz\server\tomcat\logs

Donde, entre otros, se almacenan los siguientes logs:

- del framework de desarrollo:
  - jazz.log
- del servidor de aplicaciones Apache Tomcat:
  - catalina.[aaaa-mm-dd].log

Estos archivos de logs son creados por el Jazz Team Server diariamente de acuerdo al nivel de detalle configurado en la herramienta. Esta es la información de entrada o input para este módulo de seguridad que se propone.

La propuesta consiste trabajar con esos logs con el objetivo de reconocer información que podría informar o inferir de ciertos acontecimientos producidos por un usuario conectado y trabajando en el entorno colaborativo de desarrollo y registrado por el mismo framework en el log o bitácora del Jazz Team Server.

A continuación se explica por partes el procesamiento o funcionamiento que el autor propone conformado el módulo inteligente de seguridad y detección de ataques de usuarios a través del reconocimiento de patrones de comportamiento maliciosos que en este apartado se detalla:

A. Información de entrada para el módulo, Activity Logs:

Para trabajar los logs, es necesario crear un módulo que sea capaz de extraer los datos que sean de vital importancia para detectar estos patrones. Para lo cual hace falta, no sólo conocer la definición de los logs y tener una gran experiencia en el trabajo con ellos sino también, saber qué buscar en ellos. La información que se extrae de los logs depende del método de Web Usage Mining a utilizar, de acuerdo a los que el autor ha propuesto en este trabajo, específicamente en los Capítulos II. y III.

En el caso de encontrar algún faltante de información en los logs, es decir, algún dato que sea necesario para poblar la base de datos de patrones maliciosos y no sea encontrado en los logs de la herramienta, en este caso, en primer lugar se procede a incrementar el nivel de detalle de los logs, adicionalmente, agregar marcas o puntos de control ante ciertos eventos en la misma herramienta -si es que ello es necesario- para poder extraer un registro completo con la información que no posee el log en su definición. Esto es posible ya que -en esta versión- el Rational Team Concert es de fuente abierta, por lo tanto esta es una posibilidad totalmente viable en el caso de ser necesario.

Básicamente el siguiente módulo toma como entrada los archivos de logs que la herramienta genera, ya sean los del framework de desarrollo como los del servidor de aplicaciones.

B. El módulo Data Extractor:

Tomando la actividad de dichos logs, se extrae la información que se considera relevante de acuerdo al método de Web Mining a utilizar. El autor propone en este módulo el uso de **Redes Neuronales (NN)** y **Machine Learning (ML)**. Mediante estas herramientas se analiza cierta información que resulta importante como, patrones de comportamiento, perfiles de uso, rutas, navegación de componentes, patrones de uso de recursos, tipos de conexión, etc. Se reduce la información analizada e infiere en diferentes patrones de riesgo, llamativos o destacados, los cuales son almacenados en una base de datos para su posterior consulta.

Con esos datos de entrada extraídos por el **Extractor de Datos**, se evalúan situaciones a través de **Redes Neuronales (NN)** y **Machine Learning (ML)** y se alimenta una base de datos de patrones de conducta peligrosos.

C. El Security Module:

Este módulo estará montado sobre el Jazz Team Server como un proceso back-end, es decir, será un módulo más que es administrado por el Jazz Team Server. Para la parte del front-end, es decir, la iteración con el usuario administrador, el autor propone incluirlo dentro de la funcionalidad del servidor Web, que es utilizada para la administración de usuarios y configuración de la herramienta como una funcionalidad paralela de 'Monitoreo de Seguridad'. El front-end puede variar, puede ser definido a futuro de acuerdo al modo de implementación en la herramienta. Dependerá fundamentalmente de la forma en que sea adaptada la herramienta para manejar la política de seguridad ante la existencia de patrones de comportamiento maliciosos.

La base de datos de Patrones de comportamiento peligrosos es el resultado de los métodos aplicados a los datos relevados, por lo tanto, es utilizado por el **Módulo de Seguridad** para evaluar actividades de trabajo e informar al **Módulo de Jazz** sobre el potencial ataque de usuario.

VII. Conclusiones y trabajo a futuro

Esta contribución agregaría valor a la herramienta en un entorno de desarrollo colaborativo. Y por lo tanto, generaría confianza en las grandes y medianas organizaciones que trabajan en forma globalizada y/o remota a adquirirlo y utilizarlo.

Inclusive trasladando esta confianza tanto en los nuevos desarrollos de software, como en el mantenimiento de sus actuales y futuros clientes, ya que la herramienta no sólo genera software sino que lo hace con calidad y seguridad durante todo el ciclo de vida del proyecto.

Referencias:

Internet:

[1] IBM – Software Rational – Rational Host Integration Solution.
http://www-01.ibm.com/software/awdtools/hostintegration/features/?S_CMP=rnav

[2] Jazz.net – Community site - Learn
https://jazz.net/learn/learn.jsp

[3] ISO 27002
http://www.27000.org

[4] Jazz.net - Rational Team Concert Capabilities
https://jazz.net/learn/LearnItem.jsp?href=content/docs/rtc1.0-capabilities/index.html

[5] IEEE Org.
http://www.ieee.org

Libros:

[6] “Information Security Guideline for NSW Government – Part I Information Security Risk Management. Junio 2003. Office of Information Technology

Papers:

[7] “An Approach of Multi-path Segmentation Clustering Based on Web Usage Mining.” M. Neshati, A. Alijamaat, H. Abolhassani, A. Rahimi and M. Hoseini - 2007 - IEEE.

[8] “A New Similarity Measure for Generalized Web Session Clustering.” Y. Dong, H. Zhang, L. Jiao - 2007 - IEEE.

[9] “Clustering of web users using session-based similarity measures.” H. Yang J. Lei and F. Fu - 2001 - IEEE.

[10] “Discovery of user frequent access patterns on Web usage mining.” N. Khasawneh, C. Chan - 2004 - IEEE.

[11] “Intelligent Decision Support System Based on Data Mining: Foreign Trading Case Study.” J. Xiao, Y. Zhang - 2007 - IEEE.

[12] “Research on Application of User Navigation Pattern Mining Recommendation.” F. Zhang, B. Yang, W. Song and L. Li - 2006 - IEEE.

[13] “Selective Markov Models for Predicting Web-Page Accesses”, M. Jaczynski, B.Trousse – 2000 - University of Minnesota, Department of VcomputerScience/ Army HPC Research Center Minneapolis.

[14] “Temporal Cluster Migration Matrices for Web Usage Mining.” X. Wang, Y. Ouyang, X. Hu, Y. Zhang - 2004 - IEEE.

[15] “Using probabilistic latent semantic analysis for Web page grouping.” P. Lingras, M. Hogo, M. Snorek - 2005 - IEEE.

[16] “Web usage mining using rough sets.” A. Abraham, V. Ramos - 2005 - IEEE.