**Universidad del Cauca**

**Facultad de Ingeniería Electrónica y Telecomunicaciones**

**Programas de Maestría y Doctorado en Ingeniería Telemática**

**Seminario de Investigación**

**Procesamiento de Eventos Complejos en la Nube**

**Relator: Ing. Oscar Ricardo Valencia, estudiante de Maestría**

**Co-relator: Ph.D. Juan Carlos Corrales Muñoz**

**Protocolante: Mag. Mario F. Solarte S., estudiante de Doctorado**

**Fecha:** 29 de agosto de 2014

**Hora inicio:** 10:10 a.m.

**Hora fin:** 10:50 a.m.

**Lugar:** Salón de posgrado, FIET, Universidad del Cauca, Popayán

**Asistentes:**

Ph.D. Juan Carlos Corrales, coordinador del seminario

Estudiantes de Maestría y Doctorado en Ingeniería Telemática

Estudiantes de pregrado de la FIET

**Orden del día:**

1. Presentación a cargo del relator
2. Intervención del co-relator
3. Discusión
4. Cierre

**Desarrollo**

1. **Presentación a cargo del relator**

Oscar Ricardo Valencia presenta la agenda del día que incluye:

* contexto
* escenario de motivación
* estado del arte
* pregunta de investigación

**Contexto**

El Procesamiento de Eventos Complejos (PEC) es una tecnología en la cual se tiene un conjunto de técnicas y herramientas con el fin de analizar y procesar series de eventos relacionados.

Pero ¿qué es un evento? un evento es algo que ocurre, ha ocurrido o se espera que ocurra dentro de un sistema: ejemplos dentro del ámbito de los dispositivos móviles: llegada de un mensaje en el Whatsapp, mensaje de texto, llamada, descarga de la batería... cualquiera de estos es un evento.

La idea es identificar la serie relacionadas de eventos importantes para nuestro sistema y a raíz de eso hacer el procesamiento, indicando qué se debe hace;, por ejemplo: se descarga la batería... ¿qué se debe hacer?, notificar al usuario por un mensaje, bajar el brillo a la pantalla, desactivar redes wifi, o cualquier tipo de acto seguido a la presentación del evento.

Otro concepto que vale la pena aclarar es la definición de Computación en la Nube (CN): modelo en el cual se permite un acceso conjunto a recursos informáticos, que estos se encuentran compartidos y distribuidos de forma ubicua, se accede a ellos a demanda y de forma conveniente según las necesidades del usuario. La idea es que el usuario haga un esfuerzo mínimo de administración, responsabilidad que pasa al proveedor del servicio.

La CN da un cambio completo en el paradigma en el cual estamos acostumbrados. El usuario es el que de acuerdo a las necesidades, va a adquirir por ejemplo software, infraestructura, o distintas plataformas de acuerdo a lo que va a hacer. Si necesito tener un servidor con determinadas características (capacidad de procesamiento, memoria, etc.), me contacto con un proveedor, doy requerimientos y me responden cuesta tanto, tiene tanta disponibilidad, etc., lo único que hay que hacer es pagar el importe del servicio y se tiene lo que se necesita, lo importante es que es a demanda y depende de las necesidades de cada uno de los usuario.

La CN funciona bajo tres modelos:

* Software como servicio
* Plataforma como servicio
* Infraestructura como servicio

**Software como servicio**, lo que generalmente estamos acostumbrados a hacer cuando usamos herramientas como las de Google, calendar, docs, o Microsoft Office 365, lo único que hacemos es interactuar con el software a través de un navegador

**Plataforma como servicio**, opciones de desarrollo como las que ofrecen Google (apps), Windows Azure, plataformas en las que podemos desarrollar programas en lenguajes como Java, PHP, etc., sin necesidad que la compilación se realice dentro de nuestros equipos. Se programa en un entorno que se puede ver en un navegador y el procesamiento se hace "por detrás".

**Infraestructura como servicio**, ofrece servicios de almacenamiento, alquiler de equipos, de sistemas operativos y todos a través de la nube; se paga para que te den algún tipo de almacenamiento, cuando se contrata un servidor con requerimientos a medida, se paga de acuerdo a las características que se tengan.

Cualquiera de estos modelos puede ser desplegado de una forma pública (para todos los usuarios), privada (para un conjunto de usuarios), o híbrida (una mezcla de las dos anteriores), depende de las necesidades de los usuarios.

En la nube tenemos diferentes servicios disponibles para los clientes; bases de datos, servidores, sistemas operativos y máquinas virtuales, todas pueden acceder a ellas desde sus PC de escritorios, portátiles, smart phones o tablets. ¿Qué pasa con los eventos que ocurren en la nube?, ¿qué pasa si estás trabajando y hay una falla eléctrica, o una falla de comunicación, qué pasa con la seguridad los datos? Hay algunos poblemas como disponibilidad del servicio, etc.

Para un caso concreto, por ejemplo tenemos instalada una estación meteorológica que se encuentra transmitiendo datos que viajan a través de la nube, no vamos a especificar si están puestos como plataforma, como servicio o con cualquier tipo de implementación. Estos datos son almacenados bien sea en la nube o en servidores locales, como los datos se van almacenando se pueden generar diferentes alertas, o se pueden tener datos para predicciones; supongamos que se van a trabajar para el caso de alertas, ¿qué sucede si se están transmitiendo datos de forma errónea en la nube?, ¿qué pasa si ocurren problemas de comunicación? ¿qué pasa si fallan los sensores puestos en la estación?, ¿qué pasa si hay fallas eléctricas o del software? son ejemplos de eventos que pueden ocurrir, almacenar y generar falsos positivos en las alertas que se van a presentar.

La idea es identificar flujos de eventos en un periodo de tiempo constante. Si tenemos fallas en los sensores durante una hora, ¿qué sucede con los datos almacenados durante ese periodo de tiempo?. Si un sensor da temperaturas para Popayán con medidas de 50 grados por más de una hora, lógicamente no es posible pero si hay un fallo son cosas que pueden ocurrir. ¿Cómo tratar ese tipo de eventos para que no afecten la calidad y consistencia de los datos almacenados? pueden finalizar con una falsa alerta o una mala toma de decisiones.

**Estado del arte**

Lo que se ha leído hasta ahora, el PEC tiene que ver con Arquitecturas Dirigidas por Eventos (ADE) y las Arquitecturas Orientadas a Servicio (AOS). Se ha tratado de buscar la conexión en ellas. ADE permiten procesamiento de eventos complejos y AOS estrechamente relacionadas con la CN.

Se tienen cuatro (4) trabajos relevantes sobre ADE y AOS.

* (Ayllón & Reina, 2008): procesamiento y correlación de gran cantidad de eventos con SOA. Se propone un método que integra PEC sobre una arquitectura SOA. Concluyen con la presentación de una pequeña solución basada en Esper (sistema para manejar PEC)
* (Maréchaux, 2006): combinación entre AOS y ADE por medio de Google Servicios Empresariales. Explica cómo se aporta en la integración entre lo que son los eventos y los servicios.
* (Malekzadeh, 2010): ADE y SOA en colaboración. Estudia la integración de ADE que funcionan dentro de AOS. Trabajo a nivel conceptual, que muestra cómo pueden interactuar conceptualmente SOA y ADE encontrando los siguientes tópicos: los objetivos del negocio comunes, las arquitectura basadas en componentes desacoplados, modelos de datos comunes, los dos usan una infraestructura y unas tecnologías comunes.
* (Puig, Bellot, & Bulo, 2011): PEC en entornos SOA, muestra caso de estudio para la detección temprana de epidemias. Ofrece una arquitectura software donde podemos integrar SOA y PEC, utiliza como mediador lo que es el ESB. Hacen una prueba de la arquitectura utilizando un esquema de reglas y un motor de inferencias que han realizado para una detección temprana de brotes epidemiológicos.

Los trabajos que se han encontrado en el momento si bien se centran en la combinación entre ADE y AOS, pero ninguna de ellas hace procesamiento propiamente en la nube. Se encuentran las dos arquitecturas, se tratan de combinar para hacer el procesamiento o para hacer el manejo de eventos, pero no tienen en cuenta la nube.

**Pregunta de investigación**

La pregunta de investigación: ¿cómo procesar los flujos de eventos complejos para mejorar la confiabilidad de los datos en la nube?

1. **Intervención del co-relator**

El Ph.D. Juan Carlos Corrales indica que el trabajo que se está presentando trata de integrar dos conceptos que son bastante conocidos PCE (bastante viejo) y cloud computing (CC) (relativamente nuevo), se empieza en cluod, se baja a AOS y específicamente el trabajo tiene un dominio de aplicación que son los datos climáticos, no es una estación como dijo Oscar, son 120 o 130 estaciones que transmiten información, cada estación tiene nueve variables, de cada variable hay un reporte por minuto, entonces es un gran flujo de datos que se está danto dentro de una arquitectura que es AOS, lo que se intenta es identificar muy bien si hay un dato erróneo, si una estación falló, si es una falla del sensor o un problema de comunicaciones para procesar el evento de la mejor manera.

se quiere con todos esos datos es entregar una alerta al cliente que no sea errónea.

Falta aún todavía por ajustar bastante en el trabajo, identificar exactamente cuál va a ser el aporte, que no va a ser en la parte de arquitectura sino en el componente algorítmico, es decir tener un algoritmo para procesar los eventos sobre una de las plataformas PEC que ya existen.

Los eventos son propios del mundo de las telecomunicaciones, por ejemplo una llamada telefónica ya incluye el disparo de eventos, hay un trabajo asociado a este que está en pregrado, y está trabajando en PEC en JSLEE en servicios convergentes. Estamos tratando de armar un área con un trabajo de maestría y otro de pregrado, y la asesoría de un grupo de la Universidad de la República (Uruguay) que aplica este tema hacia plataformas que hacen bioexperimentos: por ejemplo secuenciación de cadenas de ADN, cuando se detecta un evento hay que procesar una gran cantidad de eventos.

1. **Discusión**

Camilo Corrales: yo estoy más del lado de sistemas, los eventos que se disparan ante una falla, por ejemplo con sensores... esos algoritmos estarían en la estación, en un servidor... no me es claro el disparo de los algoritmos.

OV: hasta ahora no puedo darte una respuesta, lo que hacen las diferentes arquitecturas es tener un motor de inferencia con reglas que están dadas de la forma: si ocurre esto, haga esto. Hay que analizar más a fondo en dónde se debería procesar el evento, si en la estación como tal o al no recibir el dato actúe de esta forma.

CC: si falla la estación o falla el sensor... debería estar en la estación, yo imagino que es un evento que se debe procesar en la estación.

OV: exactamente, ese evento pertenecería directamente a la estación pero cuando ella hace el envío de datos, puede enviar un código que permita identificar que es un evento que se ha disparado por esa razón, y entonces ya sería que concluyera con las reglas que se han hecho o que el motor de inferencia saltara o respondiera de la forma en la cual ha sido programado.

Mario Solarte: pregunta interesante la anterior, pues entendí que el procesamiento de datos se va a hacer en la nube, me dio esa sensación; respondes que no está decidido hacer de esa forma... si los eventos ocurren del lado de la estación, no tendría mucho sentido hacer CC con ellos.

Otra inquietud ¿Tienes caracterizado los eventos complejos?

OV: no, se ha buscado la forma de implementar o trabajos relacionados que traten de combinar de PEC como parte de cloud computing, pero no se tienen trabajos que lo hagan directamente, por eso se ha bajado a AOS; todavía no me he enfocado en los diferentes eventos que puedan existir.

Magaly Cerón: encontrar fallas... de seguridad, mi duda era esa... ¿cómo van a hacer para tomar esas fallas? En telecomunicación se miden varios niveles, es fácil cuando hay una red común y estándar, pero es más complejo medir esos parámetros cuando se aplican en otros tipos de redes. No sé qué han pensado utilizar para tomar esas medidas y medir la probabilidad de cada una de ellas

OV: todavía no se ha estudiado ninguna de las fórmulas. Los eventos que se mostraron son ejemplos de los eventos que se pueden producir. No se ha acotado a algún tipo de eventos porque no se pueden tratar todos los tipos de eventos, habría que centrarse en algunos eventos específicos.

MC: cuando tomas las fallas estás evaluando la calidad, hay diversos tipos de evaluación de la calidad: QoS y sus subcoponentes (si es en el mundo de las telecomunicaciones) y tienes que tener cuidado, cuando hablas de procesamiento de señales debes hacer evaluación de calidad en software, eso es complicado... te lo digo por experiencia propia. Tienes que definir muy bien en dónde te vas a enfocar en la evaluación de tus características. Si lo tomas como explicaste te va a quedar muy muy grande y complejo.

Diego Durán: no conozco mucho del tema, pero las referencias están un poco antiguas, eso se nota en la presentación del estado del arte. Dices que son las principales, pero hay que hacer referencia a publicaciones nuevas, porque ellas te van a dar ideas sobre lo que vas a hacer y qué puedes extraer de esos trabajos también.

Juan Carlos Corrales: Afortunadamente ustedes hicieron varias de las observaciones que iba a hacer. Debo aclara lo siguiente.

1.- Lo que dicen Mario y Camilo es muy cierto... no quiero que el trabajo se centre en aprovechar la capacidad de nube para procesas eventos complejos, Oscar debe entender bien que en la nube hay flujos de información que requieren procesarse.

2.- De acuerdo también con Mario, no se puede decir "voy a hacer PEC si no tengo caracterizados los eventos complejos ni el sistema que está en la nube". Es importante caracterizar los eventos complejos para saber exactamente cuáles de ellos se van a trabajar.

3.- Hay que llegar hasta el nivel de mecanismos, algoritmos para implementar PEC. Ya hay plataformas que implementan PEC, casi todos son árboles de decisión, que tenemos que verificar es si esos algoritmos con los más adecuados o no, y después empezamos a bajar en los niveles de detalle como sugiere Magaly.

Yo te hago la pregunta: ¿cuál es la diferencia entre SOA y computación en la nube?

OV: son complementarias, no se puede hablar de la una sin la otra. SOA son AOS, CC utiliza AOS para presentar nuestros servicios. Cuando tenemos el software como servicio, ellos por debajo lo que tienes es AOS, lo mismo ocurre cuando tenemos la parte de infraestructura.

JCC: ¿Y las ADE?

OV: tienen como base los eventos, y según los eventos reaccionan con flujos de eventos, ahí están los eventos complejos y los eventos simples, de acuerdo a las categorías de eventos resuelven los problemas con árbol de decisión, modelo de reglas, motor de inferencias, etc.

JCC: hay que tener muy claro los conceptos. Sin dar tantas vueltas decir exactamente qué es cada cosa: PCE, ADE, AOS, CC; y en qué parte me estoy ubicando pues parecen al final que son cosas muy similares. Según lo que dice AOS es la materialización de los web services, lo que yo te entiendo es que CC es el concepto y la materialización en la AOS... ¿o no? ¿PCE es la materialización del ADE?, hay que tener muy claros esos conceptos para saber en dónde se va a trabajar, y no caer en los trabajos de pregrado de querer abarcar algo muy grande, en maestría no... hay que ir al punto.

El título hay que variarlo, ese título no es y la pregunta no tiene nada que ver con el título. Observación: apenas sacaste la ADE todo el mundo quedó perdido porque no explicaste lo que era eso.

1. **Cierre**

El coordinador del seminario da fin a la discusión y termina la sesión, invitando a asistir a la sesión de la próxima semana y agradeciendo la asistencia del público.