

PROCEDIMIENTO DE DIMENSIONAMIENTO DE INFRAESTRUCTURA EMPLEANDO GESTIÓN DE REDES BASADA EN POLÍTICAS

Mónica Peña Casanova¹, Anías Calderón Caridad²

¹Universidad de las Ciencias Informáticas, carretera a San Antonio Km 2 1/2, ²Universidad Tecnológica de La Habana, dirección

¹e-mail:monica@uci.cu

RESUMEN

Con la softwerización de centros de datos y de funciones de red, la estimación de la infraestructura subyacente para el despliegue y gestión de servicios, constituye un reto para los operadores de la misma. En el proceso de toma de decisiones, se deben tener en cuenta aspectos como maximizar la disponibilidad y optimizar el empleo de recursos tales como ancho de banda, procesamiento, energía eléctrica y almacenamiento. Un mecanismo eficaz para controlar y coordinar todos los elementos que intervienen en el despliegue de servicios, es el empleo de la Gestión de Redes Basada en Políticas. Sin embargo, dimensionar la infraestructura óptima para hacer eficientes estos sistemas sigue siendo un tema no resuelto por la comunidad científica, por cuanto puede resultar contraproducente automatizar la gestión, si compite con los recursos necesarios para el correcto funcionamiento de los servicios. El presente trabajo propone un procedimiento para la estimación de la infraestructura subyacente en el despliegue y gestión basada en políticas de servicios, el cual contribuye al dimensionamiento eficiente de la infraestructura a desplegar y aporta medidas que facilitan el control de este proceso. Los resultados alcanzados muestran la adecuación de la infraestructura subyacente a las necesidades de los servicios que soportan en cada organización.

PALABRAS CLAVES: Dimensionamiento infraestructura TI, gestión de redes basada en políticas, calidad de servicio.

INFRASTRUCTURE DIMENSIONING PROCEDURE WEARING POLICY BASED NETWORK MANAGEMENT

ABSTRACT

Data centers softwarezation and network functions virtualization has converted the estimation of the underlying infrastructure for the deployment of services and for the management in a challenger issue for IT operators. In the decision-making process, they must take into account concerns related to maximizing availability and optimizing the use of resources such as bandwidth, processing, storage, among others. An effective mechanism to control and coordinate all the elements involved in the deployment of services, is the use of Policy-Based Network Management (PBNM). However, sizing the optimal infrastructure to make these systems efficient is still an issue not solved by the scientific community,

provided the possible counterproductivity to automate the management. This is a case when the automatic size of optimal infrastructures competes with the necessary resources for the proper functioning of services. The present work proposes a procedure to estimate the underlying infrastructure in the deployment of services and their management based on policies which contribute to the efficient sizing of the corresponding infrastructure to be deployed. The results achieved consist of the adequacy of the underlying infrastructure to the organization's services needs support.

KEY WORDS: Estimation of information technology infrastructures, policy-based network management, Quality of service.

1. INTRODUCCIÓN

Los recursos físicos sobre los que se despliegan los servicios TI (Tecnologías de la Información) resultan costosos para las organizaciones y con un ciclo de obsolescencia muy corto. Realizar una estimación eficiente y efectiva de la infraestructura, sobre las que se despliegan dichos servicios, continúa siendo una prioridad para los proveedores de los mismos. La llegada de la virtualización resulta un paso importante en este sentido ya que permite un mejor empleo de las infraestructuras física. Utilizando máquinas virtuales se logra un mayor uso de recursos físicos tales como procesamiento, ancho de banda, almacenamiento y energía eléctrica; colocando en un mismo recurso físico, varios recursos virtuales concurrentes con diferentes sistemas operativos, cada uno de ellos diseñados a la medida para el servicio que soporta.

Muchas soluciones se han desarrollado para lograr el despliegue eficiente de las máquinas virtuales. Entre estas soluciones destaca el hospedaje en función de los recursos físicos que consumen y arranque secuencial en función de la demanda de servicios. La operación de estas soluciones se complejiza por cuanto se debe monitorear el empleo de los recursos físicos, predecir la demanda de servicios y migrar las soluciones virtuales a diferentes recursos físicos sin afectar la disponibilidad de los servicios desplegados sobre ellas y garantizando el cumplimiento de los acuerdos de nivel de servicio.

En el escenario antes descrito, la automatización se convierte en un imperativo. En función de ella, es preciso diseñar un conjunto de políticas TI, a partir de las prioridades de las organizaciones, y automatizar la ejecución de las mismas en función del monitoreo de parámetros de funcionamiento de las infraestructuras y los servicios. Una de las soluciones más extendidas en la consecución de este fin, es la Gestión de Redes Basada en Políticas (PBNM por la sigla en inglés de Policy Based Network Management).

Sin embargo, para automatizar la ejecución de políticas, se requiere contar con recursos de hardware y de conectividad, estos adicionales a los consumidos por los servicios que se prestan a través de las infraestructuras disponibles. Para que no se deterioren los parámetros de rendimiento de esta infraestructura, de calidad de los servicios, ni se afecte el proceso de toma de decisiones, al evaluar condiciones para ejecutar acciones, es necesario, previo a la automatización, evaluar la infraestructura TI de la organización. El proceso de toma de decisiones implica recursos de procesamiento y la ejecución de políticas y actualización de base de datos de gestión de configuración consume recursos de conectividad.

En el presente trabajo se muestra un procedimiento que se elaboró para la estimación de la infraestructura física subyacente para el despliegue de servicios y su gestión. El trabajo está estructurado en tres secciones: en la primera se realiza una caracterización de soluciones con vistas a la estimación de infraestructuras TI y a la gestión de redes basada en políticas; en la segunda se expone el procedimiento desarrollado para la estimación de la infraestructura subyacente necesaria en el despliegue de servicios y de su gestión basada en políticas; finalmente se emplea el procedimiento propuesto, en el despliegue y operación del servicio de voz sobre IP en una organización.

1. DESCRIPCIÓN DE MÉTRICAS PARA EL DIMENSIONAMIENTO DE INFRAESTRUCTURAS SUBYACENTES

La virtualización ha permitido el empleo más eficiente de los recursos físicos pues con ella se logra alojar varias máquinas virtuales concurrentes sobre una misma infraestructura física. Sin embargo, resulta vital lograr un dimensionamiento adecuado de las máquinas virtuales para evitar el sobredimensionamiento de la infraestructura y el despilfarro de recursos de cómputo y energía eléctrica. Algunos autores han abordado en sus investigaciones el modo más eficiente de consolidar varias máquinas virtuales en una misma máquina física basada en diferentes parámetros de rendimiento como: uso de CPU, y características de entrada/salida, así como la conjugación en un mismo servidor físico de máquinas virtuales que alojen servicios con propensión alternar el empleo de memoria o que se pueden complementar en sus patrones de comunicación [1]–[3].

La definición de acuerdos de nivel de servicio es un componente fundamental para el dimensionamiento de la infraestructura que soportará dichos servicios. En la etapa de negociación de SLA (por las siglas en inglés de Service Level Agreement) se seleccionan los parámetros de calidad de servicio (QoS por la sigla en inglés de Quality of Service) con los que se evaluará el cumplimiento de estos [4]. Los parámetros de QoS son las instancias para representar la calidad del servicio a los usuarios finales, así como el nivel de aseguramiento sobre el mismo, ellos varían de un servicio a otro y existen un conjunto de organismos internacionales que se ocupan de normarlos para cada tipo de servicio. Algunos parámetros de calidad de servicio son intrínsecos a los servicios y la tecnología y pueden mapear directamente a una o varias Métricas de Rendimiento de la Red y los Servicios (MRRS), otros son operacionales que no son directamente dependientes del servicio, pueden impactar la tecnología y afectan la experiencia de usuario [5].

El mapeo de los parámetros de QoS a las Métricas de Rendimiento de la Red y sus Servicios, se divide en dos etapas, las de mapeo de la medición y mapeo de la evaluación. En la etapa de mapeo de la medición, se realiza el proceso de correlación de una o varias MRRS a parámetros de QoS. Por ejemplo, la disponibilidad como parámetro de calidad de servicio significa la viabilidad del servicio ante cada solicitud, en cambio, como MRRS, para los servicios de red, se divide en conectividad y funcionalidad. La funcionalidad se mide a través del correcto funcionamiento de cada uno de los elementos de servicios y la conectividad es el enlace físico de los elementos de red. Por otra parte, el mapeo de evaluación permite la verificación de los parámetros de funcionamiento de los servicios a partir de los valores medidos de las MRRS [6]. El valor de las MRRS se mide a través de las herramientas tradicionales de gestión utilizando métodos de monitoreo.

En la figura 1 se representa un esquema genérico de mapeo de parámetros de calidad de servicio, representado con la letra Q, a MRRS, representado con 2N y a las funciones de

evaluación, representadas por E. Estas funciones de evaluación estarán asociadas al tipo de servicio y al contexto en el cual los usuarios harán uso de él. En esta representación, el mapeo de medición se representa como $m(x)$ no es lineal y en él influyen adicionalmente el método de monitorización (activo (A), pasivo (P) o a través de agentes asociados a modelos de gestión (S)), del punto de monitoreo, el período de monitoreo y el tipo de agente (agente de aplicación (aa), agente de red (na), agente de servicio (sa) o de tráfico (ta)) [4]. El mapeo $m(x)$ representa la manera en que las MRRS se emplean para medir cada parámetro de QoS para cada servicio. El mapeo de evaluación se representa como $n(x)$ representa la decisión para verificar los parámetros de QoS a partir de las MRRS medidos.

Los parámetros de calidad de servicio operacionales, están en general en función del tiempo [7], de ahí que pudieran implicar la automatización de determinados procesos para cumplir con los tiempos pactados con los clientes. Por ejemplo, para acelerar los tiempos de restauración o de aprovisionamiento de los servicios, es preciso contar con infraestructura dedicada al respaldo de los servicios o dedicar un gestor a la automatización del proceso de carga. Esta infraestructura variará en función de los requisitos funcionales establecidos y de acuerdo a las necesidades de infraestructura para implementar estos requisitos. El equipamiento destinado, permitirá lograr un nivel de confianza en el funcionamiento de la red, de ahí que la infraestructura dedicada se denomina infraestructura para lograr un nivel de confianza (INC).

Las funciones de mapeo $m(x)$ y $n(x)$ correlacionan requisitos funcionales definidos en los SLA, como parámetros de QoS, con MRRS. Para representar el mapeo de los parámetros de QoS se propone construir una matriz de diseño MD_{ij} para cada servicio. En MD_{ij} se parte de los requisitos funcionales [8], obtenidos a través de los atributos de cada cliente tal que como se muestra en la ecuación (1). Dicha matriz se construye a partir del análisis histórico de los parámetros de rendimiento, de la infraestructura subyacente, lo cual permite la obtención de grupo de parámetros de diseño PD_{ij} , bajo las restricciones que definen los requisitos funcionales para cada cliente de cada servicio (RF_{ij}):

$$MD_{ij} = \frac{\delta RF_{ij}}{\delta PD_{ij}} \quad (1)$$

Donde:

RF_{ij} : Requisitos funcionales para cada cliente (j) de cada servicio (i)

PD_{ij} : Parámetros de diseño para cada servicio (i) que utiliza cada cliente (j)

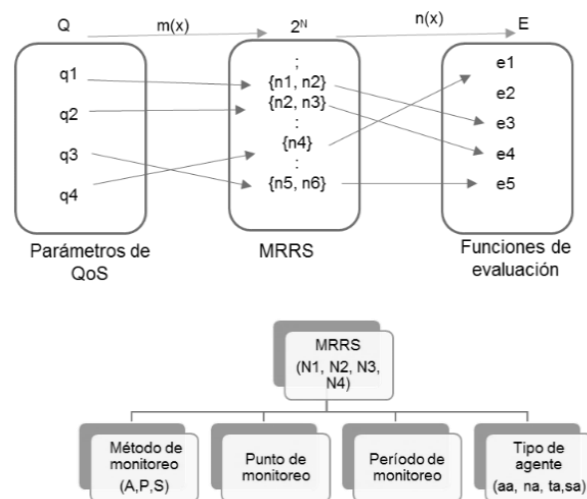


Figura 1 Mapeo de parámetros de calidad de servicio a métricas de funcionamiento de la Red [6].

Cuando los servicios se proveen, exhiben patrones de demanda variantes en el tiempo en los que se alternan períodos de alta demanda con otros de baja demanda [3]. El funcionamiento de la infraestructura subyacente, sobre la que se soportan varios servicios (i) se representa de manera estadística, como una serie temporal y para cada servicio $u(i, t)$. Esto facilita el registro de las MRRS. En esta serie temporal, se hacen corresponder con un conjunto de parámetros de funcionamiento de la red y la infraestructura que la soporta, correlacionados con los acuerdos de nivel de servicio. La función $u(i, t)$ se descompone como se representa en la expresión (2).

Resulta una buena práctica segmentar $u(i, t)$ en períodos de tiempo T , pudiéndose obtener tantas funciones $u(i, t)$ como intervalos estadísticos sean considerados [9], [10]. Las funciones estadísticas resultantes se comparan con predictores basados en datos históricos de funcionamiento lo que permite seleccionar la que más se adecúe a cada intervalo de tiempo. De esta manera, se realiza el muestreo para cada intervalo y cada máquina virtual soportando un servicio o para cada servicio soportado sobre una máquina física, obteniéndose la infraestructura subyacente para el funcionamiento normal (IFN) a través de las expresiones (2) y (3).

$$u(i, t) \rightarrow \tilde{u}(i, t) + \tilde{u}(i, t) + e(i, t) \quad (2)$$

$$I_{FN} = \sum_{t=\gamma}^{\gamma+T} \sum_{i=1}^n u(i, t) \quad (3)$$

Donde:

γ : Tiempo inicial de cada medición

$\gamma+T$: Intervalo durante el que se realiza la medición

i : Máquina virtual soportando un servicio o servicio soportado sobre una máquina física

En dichas expresiones $\tilde{u}(i, t)$ representa la tendencia de funcionamiento de los servicios, $\tilde{u}(i, t)$, representa irregularidades con respecto a la tendencia de funcionamiento de los servicios y $e(i, t)$ representa errores o incertidumbre en el funcionamiento [11].

A partir de considerar un enfoque basado en políticas, se precisa que en su formulación presenta tres elementos básicos: eventos de políticas, acciones de políticas y reglas de políticas. Los eventos de políticas representan los estados del sistema que son relevantes en el contexto de los objetivos de negocio y su realización operacional. Las acciones de políticas son las respuestas deseadas por la organización en caso de que ocurra uno o más eventos de políticas. Las reglas de políticas son los mecanismos que enlazan los eventos de políticas con las acciones de políticas.

Para controlar y coordinar, de manera dinámica, los elementos de red, tomando decisiones de forma automática a través de reglas y peticiones de usuarios o de servicios, se emplea la PBNM. Según el grupo de trabajo del Fuerza de Tarea de Ingeniería de Internet (IETF por la sigla en inglés Internet Engineering Task Force), un modelo de gestión basado en políticas, incluye un contenedor o repositorio de políticas, un punto de decisión de políticas o servidor de políticas (PDP por la sigla en inglés de Policy Decision Point) y uno o varios puntos de ejecución de políticas (PEP por la sigla en inglés de Policy Enforcement Point). En los PEP se aplican o ejecutan las políticas que gobiernan los dispositivos físicos. El agente PDP revisa las políticas almacenadas en el contenedor de políticas y efectúa un proceso de toma de decisiones enviando a los PEP las políticas que se decidan, que son independientes de las características de los dispositivos. Los PEP son los que se encargan de traducir dichas políticas en operaciones o comandos específicos que puedan ser interpretados por la tecnología concreta de los agentes que actúan en los recursos que gestionan. La figura 2 muestra una arquitectura general para un sistema PBNM siguiendo la filosofía cliente-servidor [12], [13].

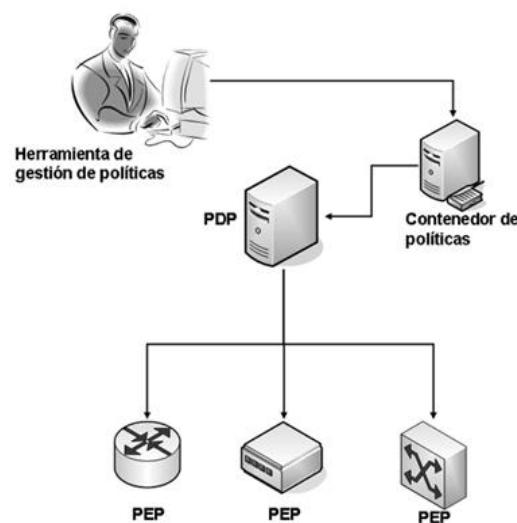


Figura 2 Arquitectura PBNM propuesta por el IETF. Fuente: [13]

2. PROPUESTA DE SOLUCIÓN

La automatización de la gestión en las organizaciones implica el consumo de recursos de adicionales en la infraestructura TI desplegada, que abarcan tanto recursos de red como de cómputo. De no dimensionarse correctamente los recursos necesarios para el despliegue de herramientas de gestión, estas, lejos de ayudar a garantizar el cumplimiento de los acuerdos de nivel de servicio, pueden atentar con el correcto funcionamiento de estos.

Para determinar la infraestructura necesaria para tener un nivel de confianza en la red, debe modelarse la infraestructura subyacente capaz de afrontar insuficiencias temporales o permanentes de los servicios, así como las transiciones alta/baja, falla/reparación y reconfiguración. Esta (INC) se obtiene a partir de un grupo de parámetros de diseño PD_{ij} para cada servicio (i) que utiliza cada cliente (j), tal que $j = \overline{1, m}$; $i = \overline{1, p}$; $m, p \in \mathbb{N}$, bajo las restricciones que definen los requisitos funcionales para cada cliente (RF_{ij}) que se soportan con la infraestructura (INC), la cual se determina como se muestra en la expresión (4).

$$I_{NC} = \sum_{j=1}^m \sum_{i=1}^p MD_{ij} \delta PD_{ij} \quad (4)$$

Como se explicó en el apartado anterior, la infraestructura subyacente para el funcionamiento normal de la red (IFN), se obtiene a partir de modelar el comportamiento de esta en el tiempo y de acuerdo a la demanda. En la modelación se integra la tendencia de funcionamiento de los servicios $\tilde{u}(i,t)$, las irregularidades con respecto a la tendencia de funcionamiento de los servicios $\tilde{u}(i,t)$, y los errores o incertidumbre en el funcionamiento $e(i,t)$. Estos atributos de funcionamiento permiten derivar medidas de rendimiento relacionadas con tiempo de respuesta, rendimiento y utilización a partir del análisis histórico de funcionamiento de la misma y de los errores o incertidumbre en el funcionamiento que podrían ser ajustados a los valores requeridos con la infraestructura dedicada a alcanzar un nivel de confianza (INC).

La infraestructura requerida para alcanzar los niveles de calidad de servicio deseados (I_{exp}) depende de la infraestructura necesaria para alcanzar los atributos de funcionamiento de la red (IFN). En adición, depende de la infraestructura necesaria para conseguir el nivel de confianza que se planifica en la red (INC) como se muestra en la expresión:

$$I_{exp} = I_{NC} \cup I_{FN} \quad (5)$$

Una vez obtenida la infraestructura necesaria para la prestación de servicios con los niveles de calidad diseñados, entonces para automatizar la ejecución de políticas es necesario desplegar una infraestructura adicional (IAP). Esta infraestructura adicional debe ser independiente de la que soporta los servicios, para garantizar la recuperación ante fallos y la respuesta ante eventos que deterioren las prestaciones en los servicios, de ahí que la infraestructura necesaria (IN) se obtiene de la unión de IAP y I_{exp} , tal y como se muestra en la ecuación:

$$I_N = I_{AP} \cup I_{exp} \quad (6)$$

Donde:

IN: es la infraestructura que se necesita

IAP: es la infraestructura necesaria para la automatización de políticas

I_{exp} : es la infraestructura en explotación de los servicios con nivel de calidad deseado [3], [14]

Para determinar la infraestructura que se necesita para la automatización de políticas con una arquitectura PBNM es necesario partir de la evaluación de la gestionabilidad de los PEP como se observa en la figura 3, luego precisar la infraestructura que se requiere en los mismos (IPEP). Esta (IPEP), está determinada por la capacidad de hardware requerido para

la instalación de los agentes en los elementos gestionados que garantizan la ejecución de políticas en estos.

Además, es necesario conocer la cantidad de puntos de decisión de políticas que se necesitan y si se reutilizarán elementos existentes como PDP. La cantidad de elementos PDP dependerá de los dominios de gestión que se establezcan en cada organización, por ejemplo, dominio de gestión de seguridad, dominio de gestión del equipamiento activo de la red, entre otros. Para el cálculo de la infraestructura que se requiere en los PDP (IPDP), debe considerarse la demanda de procesamiento que cada uno de estos tendrán. En el caso de trabajar de manera autonómica, la demanda de procesamiento dependerá del mecanismo que se implemente para la toma de decisiones.

También debe establecerse la infraestructura que se requiere en el repositorio de políticas (IRP), la cual está en función de la capacidad de almacenamiento, la estructura de la base de datos que se utilizará para el almacenamiento de políticas y el modelo de información que se utilice para la representación de estas. Además, el despliegue de la arquitectura de gestión basada en políticas implica consumo de ancho de banda para la ejecución de políticas y para la verificación del funcionamiento de los elementos gestionados (BWPBMM). Considerándose lo antes dicho, se muestra en (7) la expresión para obtener la infraestructura necesaria para la automatización de políticas.

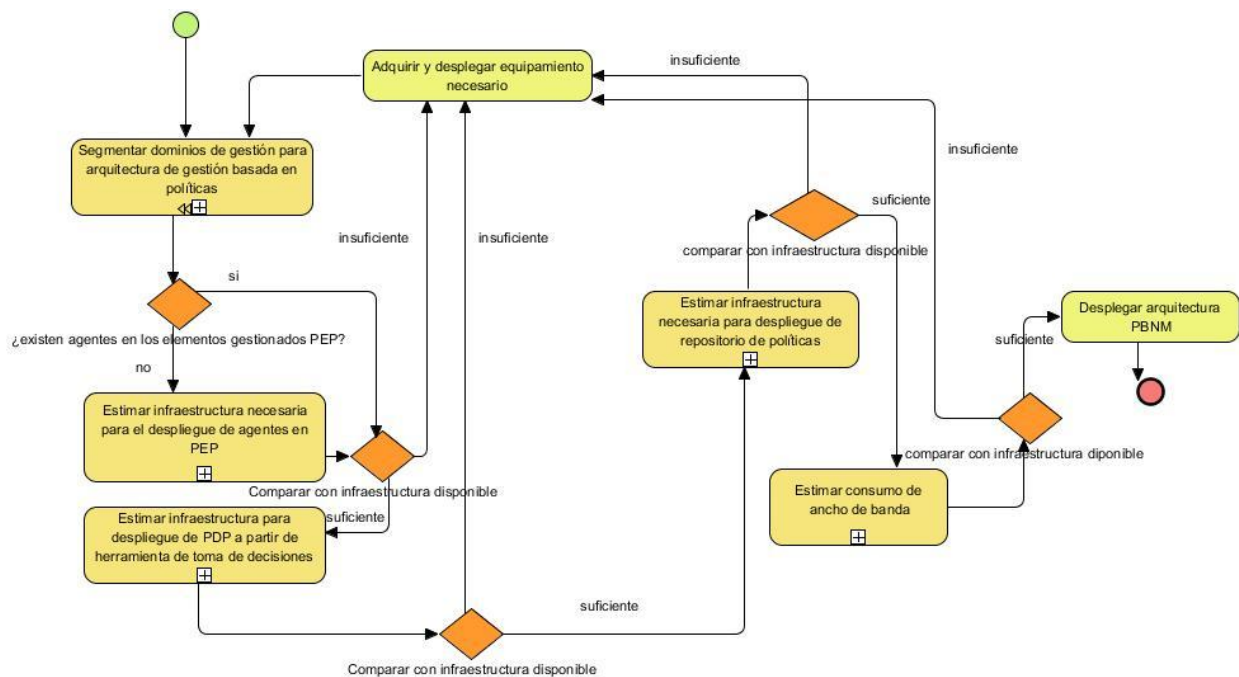


Figura 3 Procedimiento para el dimensionamiento de una arquitectura PBNM.

$$I_{AP} = I_{PDP} \cup I_{PEP} \cup I_{RP} \cup BW_{PBNM} \quad (7)$$

La planificación y organización de políticas, debe realizarse, no solamente para establecer prioridades entre estas, sino para implementar los puntos de decisión de políticas en el lugar óptimo de la infraestructura gestionada. En este sentido surge una relación de compromiso: cuando se determina desplegar una gran cantidad de PDP en la red y se dispersa la toma de decisiones, se hace más certera y efectiva la ejecución de políticas a bajo nivel, pero la evaluación del impacto de las políticas ejecutadas resulta más compleja al igual que la detección de conflictos entre estas. Una práctica bastante extendida es la definición de

puntos de decisión de políticas según sus funciones: de seguridad, de calidad y de tipo de servicios. Se propone crear dominios de aplicación de políticas, estos se pueden segmentar basados en criterios funcionales y no funcionales.

En la figura 4 se observa el procedimiento que se propone para garantizar la ejecución de políticas sobre la infraestructura TI, que parte del dimensionamiento de dicha infraestructura para la organización incluyendo la operación de la misma. Una vez determinados los recursos de cómputo y de red que se requieren para garantizar el funcionamiento de los servicios y el cumplimiento de los acuerdos de nivel de servicio, se logra esto por dos vías: una que incluye la existencia o adquisición de equipamiento para la implementación de una arquitectura de gestión basada en políticas y otra el empleo de mecanismos organizacionales.

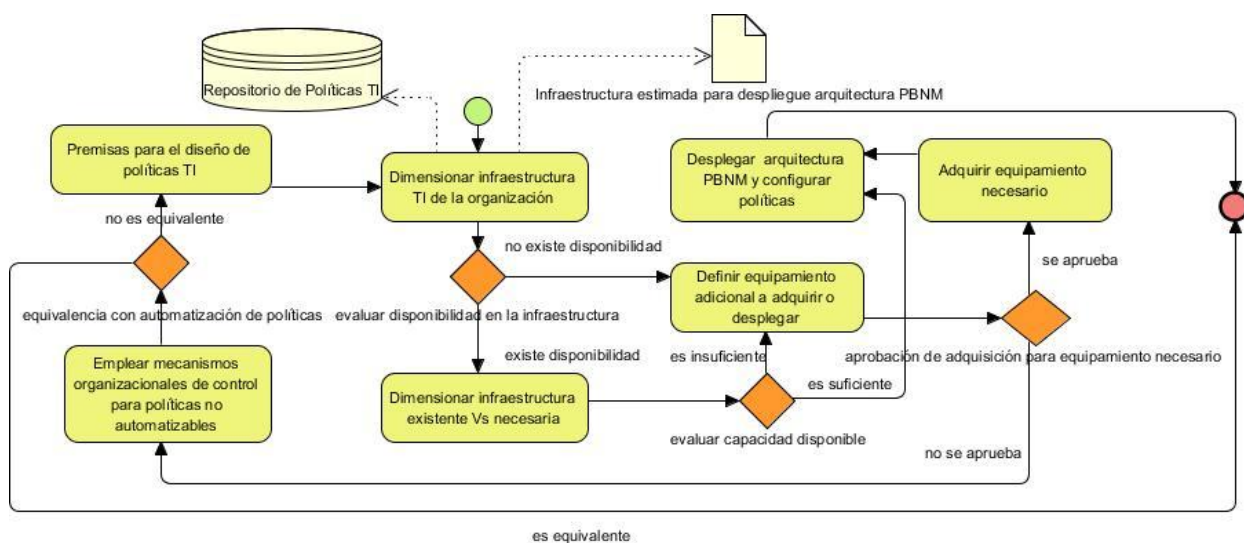


Figura 4 Procedimiento para garantizar la ejecución de políticas sobre infraestructuras TI.

3. ANÁLISIS Y DISCUSIÓN DE RESULTADOS

Para comprobar la factibilidad del procedimiento antes descrito, se aplicó el mismo a la infraestructura que soporta el servicio de voz sobre IP (VoIP) en una organización. Para garantizar el funcionamiento de dicho servicio, se monitoriza el porcentaje de usabilidad de la RAM, la CPU y del disco duro; el tráfico en las tarjetas de red en el servidor así como los principales parámetros de QoS para la VoIP [5]: la pérdida de paquetes, el jitter, la latencia, el estado de las extensiones, así como, la cantidad de llamadas concurrentes y fallidas. Para facilitar el proceso de gestión se efectúa un control sobre cada uno de estos parámetros empleando políticas, establecidas por el administrador de la red. Esto tiene como objetivo lograr que los elementos y servicios que hospedan los servidores de telefonía IP tenga un comportamiento autónomo y se dimensionen en función de la demanda. Además, se generan reportes en los cuales se muestra el comportamiento de los parámetros monitorizados durante un tiempo determinado (es configurable el intervalo de tiempo).

Como se muestra en la figura 5, para el despliegue del servicio de VoIP gestionado, se contará con un servidor de base de datos PostgreSQL 9.4 o superior, con sistema operativo GNU/Linux, Nginx 1.4.6 y framework Django versión 1.7. Siendo consecuentes con el procedimiento para el dimensionamiento de una arquitectura PBNM, descrito en la figura 3, se verificó que todos los terminales de VoIP eran gestionables y que los agentes tenían la

infraestructura necesaria para su funcionamiento (todos eran PC Pentium 4 o superior con 512 MB RAM). Para el punto de decisión de políticas, se determinó que era necesario un servidor con un Microprocesador de 2 núcleos, 4GB de memoria RAM, 250 GB disco duro y una fuente de 800 W como mínimo para lograr el despliegue de la herramienta Zabbix, la cual tiene capacidad de ejecutar monitorización y control de elementos de hardware [15] y funciona como PDP y como repositorio de políticas. Este servidor puede ser desplegado sobre máquinas virtuales o máquinas físicas, según se determine en la organización. La comunicación entre los agentes y el punto de decisión de políticas se implementó a través del protocolo HTTP, verificándose que el ancho de banda disponible resultaba suficiente. Además, los tiempos de sondeo se configuran para no afectar con el correcto funcionamiento del servicio en el momento de establecimiento de llamadas.

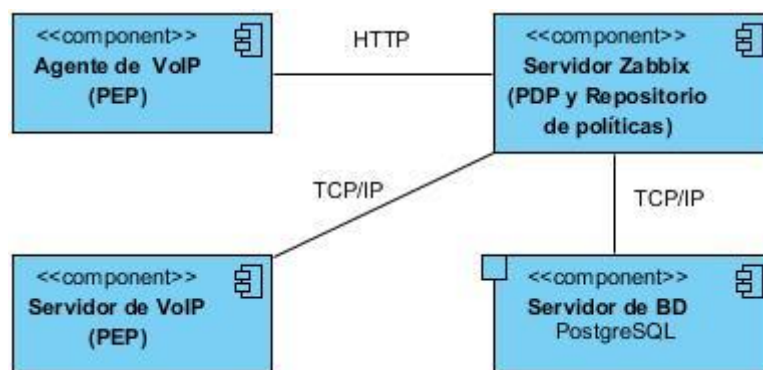


Figura 5. Diagrama de despliegue del módulo Gestión de servidores de telefonía IP

En la figura 6 se muestra la configuración de una política que, a través de parámetros del uso de recursos de cómputo del servidor de VoIP garantiza el cumplimiento de las MRRS y de los parámetros de calidad de este servicio para usuarios concurrentes históricamente registrados en la organización. Los valores de estos parámetros se obtuvieron de promediar los datos históricos de un mes de uso del servicio, empleando como período de monitorización el horario laboral típico, midiendo diariamente las métricas de funcionamiento del servidor calculando los parámetros de QoS.

Nombre	Operador	Porcentaje
RAM	>	60%
CPU	<	65%
RED1	>=	70%
RED 2	<=	75%
HDD 1	<=	80%
HDD 2	=	85%
		90%
		60%

Figura 6. Configuración de una política automatizable a ejecutar en el servidor de VoIP

Al determinar el impacto del despliegue de esta solución de gestión en cuanto a consumo de recursos físicos dentro de la organización, se determinó que era factible automatizar la ejecución de políticas. En la figura 7 se muestra el diagrama entidad relación del repositorio de políticas implementado en Zabbix. Dicho diagrama responde al diagrama de clases de diseño del escenario Gestionar servidor de telefonía IP, que se muestra en la figura 8, en el que está prevista la creación modificación y apagado de servidores de VoIP.

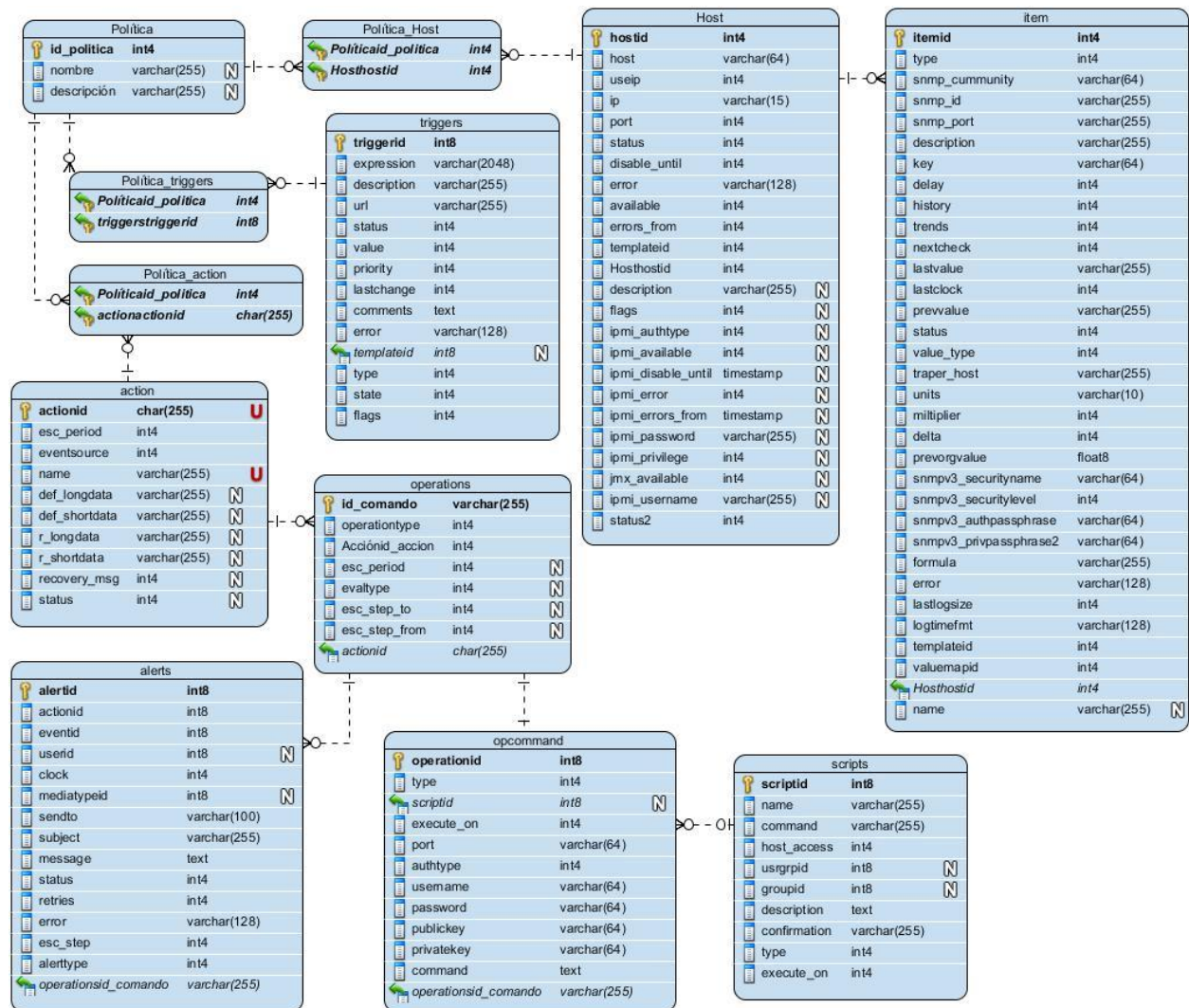


Figura 7. Diagrama entidad relación del repositorio de políticas con Zabbix.

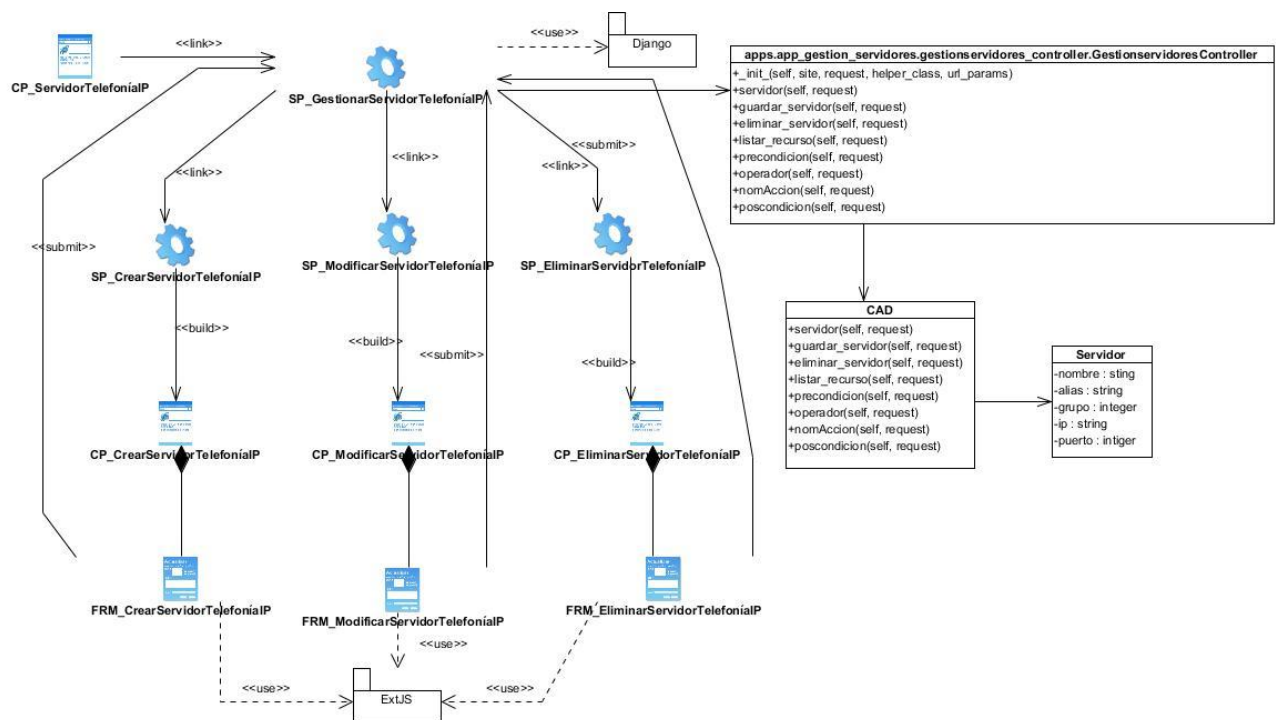


Figura 8. Diagrama de clases de diseño para Gestionar Servidor de Telefonía IP.

La automatización de la gestión tuvo un gran impacto por cuanto permitió garantizar los parámetros de calidad de servicio de la organización. Se crearon varias máquinas virtuales configuradas para funcionar como servidores de VoIP. Al monitorearse en el servidor, que se sobrepasaba los umbrales de rendimiento definidos: 60% de uso de la memoria RAM, 65% de uso del CPU, como promedio más del 70% de uso de las capacidades de red o más de 80% de uso de la capacidad de almacenamiento precisados en la figura 6. Además, se ejecuta como pos condición, un script que crea una máquina virtual para apoyar las peticiones de clientes la cual seguirá funcionando hasta que disminuya la demanda del servicio. En adición, el servidor Zabbix notifica a los administradores del sistema sobre la acción ejecutada, como se observa en la figura 9. Este proceso resulta transparente para el usuario final.

Acción	Parámetros
Enviar Correo	Correo:
Enviar SMS	
Mount	

Figura 9 Configuración de pos condición de la política.

CONCLUSIONES

La automatización de la ejecución de políticas requiere de un dimensionamiento adecuado de la infraestructura que se despliega, ya que para esto es necesario contar con recursos de hardware y de conectividad, adicionales a los consumidos por los servicios que se prestan. En este sentido es necesario, previo a la automatización, evaluar la infraestructura TI de la organización para evitar que se deterioren las MRRS.

El procedimiento que se desarrolló para la estimación de la infraestructura subyacente necesaria en el despliegue de servicios y su gestión basada en políticas, contiene los elementos que, a consideración de las autoras, se deben tener en cuenta. Abarca el dimensionamiento adecuado de la infraestructura subyacente, tanto para el correcto funcionamiento como para la gestión de los servicios.

El empleo del procedimiento en el despliegue y operación de un servicio de VoIP en una organización, permitió su validación. Como resultado de la aplicación del mismo se logró, garantizar los parámetros de calidad de servicio a través de la ejecución de políticas de manera automatizada.

REFERENCIAS

- [1] I. Pietri y R. Sakellariou, «Mapping Virtual Machines Onto Physical Machines in Cloud Computing: A Survey», *ACM Comput Surv*, vol. 49, n.º 3, pp. 49:1–49:30, oct. 2016, DOI 10.1145/2983575.

- [2] A. Mosa y R. Sakellariou, «Dynamic Tuning for Parameter-Based Virtual Machine Placement», en *2018 17th International Symposium on Parallel and Distributed Computing (ISPDC)*, 2018, pp. 38-45, ISBN 978-1-5386-5330-2.
- [3] R. Mijumbi, J. L. Gorricho, J. Serrat, M. Claeys, F. D. Turck, y S. Latré, «Design and evaluation of learning algorithms for dynamic resource management in virtual networks», en *2014 IEEE Network Operations and Management Symposium (NOMS)*, Krakow, Poland, 2014, pp. 1-9, ISBN 978-1-4799-0913-1.
- [4] TM Forum, «GB917, SLA Management Handbook, Release 3.1». 2019.
- [5] TM Forum, «GB988 TM Forum Metrics Definitions R18.5.1». 2019.
- [6] H.-J. Lee, M.-S. Kim, J. W. Hong, y G.-H. Lee, «QoS parameters to network performance metrics mapping for SLA monitoring», *KNOM Rev.*, vol. 5, n.º 2, pp. 42-53, 2002.
- [7] TM Forum, «TMF 701 Performance Reporting Concepts & Definitions». 2001.
- [8] P. Han, Y. Liu, y L. Guo, «QoS Satisfaction Aware and Network Reconfiguration Enabled Resource Allocation for Virtual Network Embedding in Fiber-Wireless Access Network», *Comput. Netw.*, jun. 2018.
- [9] W. Zheng *et al.*, «On dynamic performance estimation of fault-prone Infrastructure-as-a-Service clouds», *Int. J. Distrib. Sens. Netw.*, vol. 13, n.º 7, p. 155, jul. 2017, DOI 10.1016/j.comnet.2018.06.019.
- [10] D. Bruneo, «A Stochastic Model to Investigate Data Center Performance and QoS in IaaS Cloud Computing Systems», *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, n.º 3, pp. 560-569, mar. 2014, DOI 10.1109/TPDS.2013.67.
- [11] X. Meng, C. Isci, J. Kephart, L. Zhang, E. Bouillet, y D. Pendarakis, «Efficient resource provisioning in compute clouds via VM multiplexing», en *Proceedings of the 7th international conference on Autonomic computing*, Washington, DC, USA, 2010, pp. 11-20, ISBN 978-1-4503-0074-2.
- [12] W. Lui *et al.*, «RFC 8338 Policy-Based Management Framework for the Simplified Use of Policy Abstractions (SUPA)». feb-2018.
- [13] A. Westerinen *et al.*, «RFC 3198 Terminology for Policy-Based Management». nov-2001.
- [14] S. S. Mwanje, L. Christoph Schmelz, y A. Mitschele-Thiel, «Cognitive Cellular Networks: A Q-Learning Framework for Self-Organizing Networks», *IEEE Trans. Netw. Serv. Manag.*, vol. 13, n.º 1, pp. 85-98, mar. 2016, DOI 10.1109/TNSM.2016.2522080.
- [15] R. Olups, *Zabbix Network Monitoring*, 2da Edición. Birmingham, UK: Packt Publishing, 2016, ISBN 978-1-78216-128-8.

SOBRE LOS AUTORES

Caridad Anías Calderón, Ingeniera en Telecomunicaciones, Doctora en Ciencias Técnicas, Universidad Tecnológica de La Habana CUJAE, La Habana, Cuba y cache@tesla.cujae.edu.cu

Mónica Peña Casanova, Ingeniera en Telecomunicaciones, Máster en Telemática, Universidad de las Ciencias Informáticas, La Habana, Cuba y monica@uci.cu