



Unidad 1 / Escenario 1

Lectura fundamental

Conceptos generales de los sistemas distribuidos

Contenido

- 1 Definición de los sistemas distribuidos
- 2 Comparación entre sistemas distribuidos y centralizados
- 3 Conceptos de *hardware*
- 4 Conceptos de *software*

Palabras clave: Sistemas, Distribuidos, Servicio, Taxonomía Flynn

Introducción

En esta lectura se abordan aspectos básicos en el estudio de los sistemas distribuidos y se presentan mediante cuatro temas fundamentales que deben leerse y estudiarse de forma ascendente, pues a medida que se avanza en la lectura es necesaria la comprensión de conceptos vistos en temas anteriores.

El primer tema trata de los componentes necesarios para que un sistema distribuido pueda ser utilizado y expone las características de dichos componentes, la manera como están conformados y su efecto dentro del sistema distribuido total.

El segundo tema expone las diferencias entre un sistema distribuido y un sistema centralizado, a partir de algunas características básicas. Estas serán fundamentales en el siguiente Escenario del Módulo.

En el tercero y cuarto tema se exploran conceptos relevantes de *hardware*, mediante un breve repaso de la conocida taxonomía de Flynn, y conceptos de *software*, haciendo énfasis en las diferencias entre el acoplamiento y algunas herramientas fundamentales para el uso, construcción y operación de los sistemas distribuidos.

Antecedentes

Dentro de la sociedad actual, la tecnología juega un papel fundamental en el desarrollo cotidiano de actividades productivas y de ocio. Esto se debe a varios factores; el primero, es que desde hace varios años el precio de la tecnología, de sus componentes y de los servicios asociados ha tenido una baja que ha hecho que estos bienes y servicios sean accesibles para toda la población. Otro factor relevante es que, ante la caída del precio de la tecnología, la demanda de contenido se ha incrementado sustancialmente y, en consecuencia, el acceso a Internet es cada vez más barato, incluso hasta considerar vital el acceso para la población.

El acceso a la información y a desarrollos tecnológicos a menor costo ha hecho que casi cualquier persona con algo de interés y el acceso correcto pueda crear contenidos y ponerlos a disposición de la Internet para que sean consumidos, propiciando el nacimiento de productos y servicios que se ajustan casi de forma individual a las necesidades de las personas. Ejemplos de estos productos son el correo electrónico, las compras electrónicas, el gobierno electrónico, la modalidad de teletrabajo, las redes sociales, etc.

Estos son algunos de los servicios que existen en Internet y es necesario comprender que todo lo que está detrás y permite su funcionamiento está compuesto, mantenido y operado por computadores, es por esta razón por la que el estudio de los sistemas distribuidos es tan importante. Es un conocimiento que trata los pormenores de la puesta en operación de grandes sistemas de cómputo, que buscan prestar un servicio fiable y de valor agregado para los usuarios finales.

1. Definición de los Sistemas Distribuidos

Como definición básica, se podría decir que un sistema distribuido es un conjunto de máquinas de cómputo conectadas a una red de computadores, con el fin de prestar un servicio a un usuario final, tal y como se muestra en la Figura 1. Esta definición sirve para ver el sistema como un todo; sin embargo, y para efectos de este Módulo, es necesario poner especial atención al estudio de los sistemas distribuidos, sus componentes y los paradigmas que los rigen.

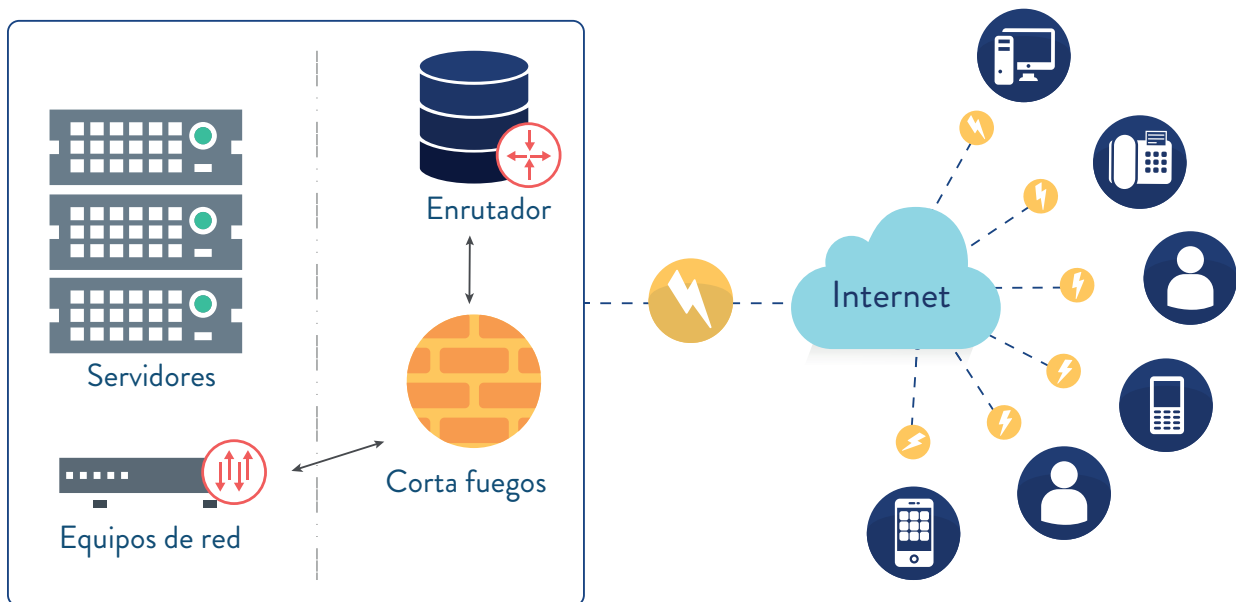


Figura 1: Sistema distribuido básico

Fuente: elaboración propia

Como se aprecia en la Figura 1, en la parte izquierda se encuentran los componentes básicos (equipos de *hardware*) e internamente el *software* requerido para prestar servicios en el Kernel de Internet, como medio para la prestación del servicio. En la derecha aparecen los usuarios que consumen el servicio.

Ahora, a partir de la definición básica de un sistema distribuido y al indagar un poco más sobre su taxonomía, es necesario comprender que un sistema distribuido está compuesto por diferentes partes, y que estas están definidas tanto por componentes físicos, como componentes lógicos que se pueden enumerar inicialmente así:

- Equipos de cómputo y *hardware* adicional
- *Middleware* y *software* transversal
- Aplicación que presta el servicio.

A continuación, se explican en detalle cada uno de los componentes del sistema distribuido.

1.1. Equipos de cómputo y *hardware* adicional

En un sistema de cómputo es fundamental tener en cuenta el componente físico, que hace referencia a los equipos en los cuales van a ejecutarse los programas y aplicaciones necesarias para el funcionamiento del sistema distribuido. Las personas que cumplan las funciones de administradores, arquitectos u operadores deben mantener dentro de su espectro de acción las características de los equipos, con el fin de proceder a trabajar proactivamente para el funcionamiento del sistema. Dentro de estos equipos generalmente se destacan:

1.1.1. Servidores (procesamiento y almacenamiento)

Generalmente, son computadores con prestaciones superiores en procesador, memoria, buses de transmisión y almacenamiento.

1.1.2. Equipos de red (concentradores (*switches*), enrutadores (*routers*) y cortafuegos (*firewalls*))

Los sistemas distribuidos están conectados a una red de computadores que se convierte en un factor fundamental para la prestación del servicio, y estos componentes actúan como un habilitador importante al momento del diseño y operación de cualquier sistema.

Adicionalmente, a los equipos tradicionales pueden sumarse componentes tales como sistemas externos, sensores y otros dispositivos de comunicación.

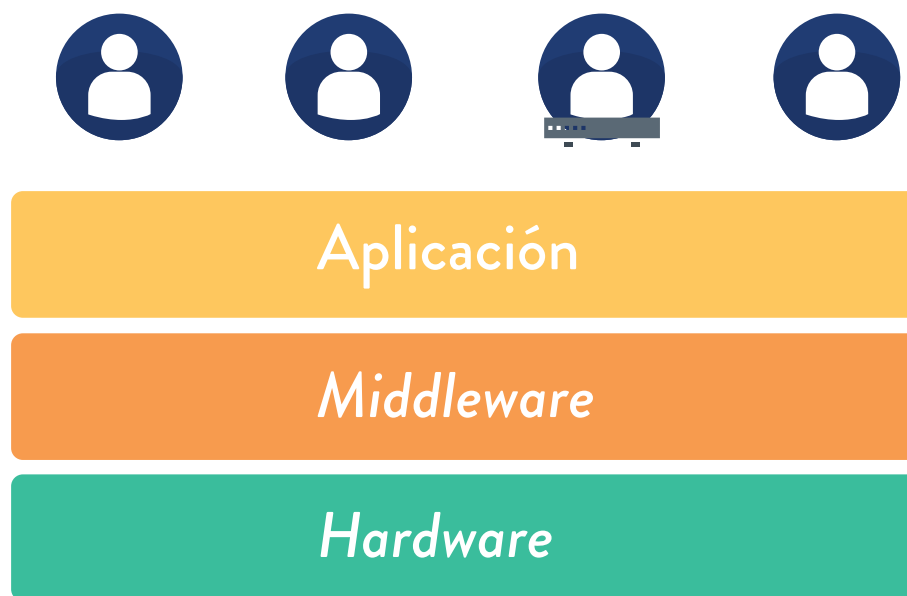


Figura 2: Arquitectura de un sistema distribuido

Fuente: elaboración propia

La arquitectura básica de un sistema distribuido que se muestra en la Figura 2 ubica el *hardware* y los recursos físicos en la parte inferior, en el centro está la capa de *middleware* y servicios transversales, seguidos por la aplicación (el servicio) y, en la parte superior, los usuarios consumidores de la aplicación.

1.2. Aplicación que prestará el servicio

El objetivo del sistema distribuido es la prestación de un servicio mediante su funcionalidad, para generarle un beneficio al usuario final. Este servicio se define como un sistema de *software* compuesto de módulos, creados por un equipo de desarrollo liderado por un arquitecto de *software*, a partir de patrones y buenas prácticas para la construcción.

Estas aplicaciones prestadoras de servicios se asocian a la cadena de valor de una organización y, generalmente, están ubicadas dentro de las siguientes categorías:

- » Correo electrónico
- » CRMs - *Customer Relationship Management*
- » ERP - *Enterprise Resource Planning*
- » Portales corporativos
- » Pasarelas de pagos
- » Tiendas virtuales
- » CMS - Manejadores de contenido
- » Redes sociales

Cabe aclarar que la construcción, diseño y operación de estas aplicaciones están directamente ligadas al proceso de arquitectura, construcción de *software* y ciclo de vida del *software*, los cuales no serán objeto de estudio en este Módulo. Sin embargo, la aplicación sí actúa como insumo en el proceso de definición del diseño y operación de los sistemas distribuidos.

1.3. Middleware y software transversal

Hace referencia al *software* y al servicio interno necesarios para garantizar el normal desempeño de las aplicaciones, las cuales utilizan de forma ordenada y coordinada los equipos de cómputo, con el fin de que los componentes anteriores se puedan visualizar como un todo hacia el usuario final.

De acuerdo con la Figura 2 este componente actúa como una capa intermedia que ayuda con la coordinación de recursos y para ello cuenta con los siguientes elementos:

- » Sistemas operativos de tipo servidor
- » *Middleware*
- » Manejador de transacciones
- » Manejador de sesiones
- » Bus de servicios
- » Manejadores de conexiones a datos
- » Servicios transversales
- » Autenticación
- » Seguridad
- » Relojes

También es el componente en el que se enfoca este Módulo, pues es fundamental para la preparación de diseñadores y arquitectos de tecnologías porque permite entender los retos y las características generales necesarias para una prestación eficaz y eficiente del servicio que se quiere proveer.

1.4. Ejemplos de Sistemas Distribuidos

A continuación, encuentra información general mediante ejemplos de sistemas distribuidos actuales que le permiten identificar el concepto de los sistemas distribuidos y relacionarlo con ejemplos cotidianos o de su entorno.

En cada uno de los ejemplos se enmarcan diferentes tipos de servicios comúnmente utilizados en Internet. Es decir que estos productos dependen de una infraestructura de *hardware* y servicios de *software* que están relacionados de una forma muy estrecha, pero que usan gran cantidad de recursos para prestar dicho servicio.

- » **Buscadores:** Google, Yahoo, Bing, Duck Duck Go, etc.
- » **Redes sociales:** Facebook, Twitter, Instagram, etc.
- » **Comercio:** Amazon, MercadoLibre, Aliexpress, etc.
- » **Finanzas:** bolsas de valores, bancos, bitcoin, etc.
- » **Educación:** Coursera, EdX, MIT, Stanford, Politécnico Gran Colombiano, etc.
- » **Ciencias:** LHC, SETI@Home, ATLAS, LIGO, Hubble, etc.
- » **Entretenimiento:** Netflix, Steam, YouTube, PlayStation Network, etc.

2. Comparación entre Sistemas Distribuidos y Centralizados

Para poder avanzar en el estudio de los Sistemas Distribuidos es necesario realizar una comparación directa entre sus características y las de los sistemas centralizados.

Se puede afirmar que un sistema centralizado (King & Leslie, 1983) es aquel que realiza todas sus operaciones, subsistemas y procedimientos en un solo equipo de cómputo, que se conoce como un *mainframe*.

Un ejemplo clásico de sistema centralizado (*mainframe*) son los sistemas legacy, empleados por empresas de seguros, aerolíneas y bancos, que siguen vigentes debido a la sensibilidad de la información allí guardada y a que la migración a nuevos ambientes distribuidos puede llegar a ser demasiado compleja y costosa.

A continuación, se exploran algunas características esenciales que diferencian los sistemas distribuidos y los sistemas centralizados, enmarcando dichas características, beneficios y perjuicios según la naturaleza de las necesidades de la organización.

2.1. Redes de Comunicaciones

Tanto el sistema distribuido como el sistema centralizado necesitan de una red de computadores para poder prestar el servicio hacia el exterior, pero, en este caso, se explora la dependencia de cada uno a una red interna adicional para la comunicación entre sus subprocesos.

El sistema distribuido tiene una dependencia directa, debido a que la interoperación y coordinación entre sus componentes debe hacerse a través de una red separada de la de servicio externa. Esto se debe a que estar separados en máquinas más pequeñas y manejar datos por la red de servicio externo puede crear cuellos de botella de comunicación.

A diferencia de su contraparte, el sistema centralizado tiene la ventaja de que todos los componentes necesarios para su funcionamiento están dentro del mismo *hardware*.

2.2. Economía

Para comprender esta característica hay que hacer un análisis a partir de dos puntos de vista: el de mantenimiento y el del costo inicial. Con respecto al mantenimiento, es necesario sumar la característica del crecimiento, pues es mucho más económico mantener un solo gran computador que varios, debido a los costos asociados a la energía, personal y equipos de comunicación y respaldo. Sin embargo, esto impone una restricción de crecimiento, que depende del costo inicial.

El decremento del precio de los accesorios (*appliances*) o de los equipos de cómputo hace posible “apilar” poder de cómputo, es decir, construir una sola máquina con características similares a un menor precio.

2.3. Seguridad

En el contexto de los sistemas informáticos y con el uso extendido que le dan las personas a productos o servicios en su vida diaria, la seguridad es una característica indispensable al momento del diseño de un sistema distribuido. Especialmente, porque los paquetes viajan por una red (Connolly, 2003) y se ven expuestos a ataques maliciosos por parte de terceros. Estos sistemas están expuestos por cada punto de conexión de red que tienen, por esto la importancia de tener una red separada para la comunicación interna que evite la mezcla del tráfico de servicio externo con el tráfico de sincronización y de servicio de los procesos internos del sistema.

2.4. Complejidad

Los sistemas distribuidos son más complejos al tener múltiples componentes, esto no solo hace que sea más difícil su operación, sino que también expone el servicio a múltiples tipos de fallos debido a la gran cantidad de puntos de enlace que tienen, no solamente físicos sino también lógicos como, por ejemplo:

- Sesiones de usuario
- Conexiones a datos
- Ingresos erróneos de usuarios
- Indisponibilidades de servicios
- Tiempos de respuesta altos de la red

3. Conceptos de *Hardware*

Con el fin de poder entender los conceptos asociados al manejo masivo de recursos computacionales hay que aclarar algunos conceptos de *hardware*, siguiendo la arquitectura de máquina propuesta por Von Neuman (Godfrey & Hendry, 1993).

3.1. Taxonomía de Flynn

La Taxonomía de Flynn (1972) es una clasificación propuesta por el profesor americano Michael J. Flynn en 1966. En su propuesta separa las arquitecturas de los computadores según los flujos de entrada y salida que existen entre las estructuras de control representadas por las instrucciones de procesamiento y las estructuras de datos representadas en unidades de memoria.

3.1.1. Pocos datos con instrucciones sencillas (*Single-Instruction-Single-Data SISD*)

Este es el tipo de arquitectura más simple de un computador, por cada instrucción que realiza la máquina, se guarda en el almacenamiento un dato que representa el resultado de la operación.



Figura 3. Taxonomía de Flynn - SISD

Fuente: elaboración propia

Para que este computador pueda realizar tareas más complejas es necesario darle un listado de instrucciones sucesivas, que se van ejecutando secuencialmente y, a su vez, se va guardando el resultado de la operación.

3.1.2. Muchos datos con instrucciones sencillas (*Single-Instruction-Multiple-Data SIMD*)

Este tipo de arquitectura incrementa la posibilidad de guardar los resultados de las operaciones en más de un dato, mejorando la posibilidad de mantener más datos en memoria, con el fin de poder realizar operaciones más complejas, permitiéndole a la máquina mejorar su capacidad y la complejidad de sus algoritmos.



Figura 4. Taxonomía de Flynn - SIMD

Fuente: elaboración propia

Sin embargo, el crecimiento en el espacio de memoria está supeditado al tamaño del bus de comunicaciones interno de la máquina. El crecimiento de los flujos de datos o de los espacios de memoria responde a la fórmula exponencial 2^n .

3.1.3. Múltiples Instrucciones con pocos datos – (*Multiple-Instruction-Single-Data MISD*)

Para este caso, la evolución consecuente de las arquitecturas en la taxonomía de Flynn propone que por cada dato hay dos instrucciones.



Figura 5. Taxonomía de Flynn - MISD

Fuente: elaboración propia

Este tipo de arquitectura es ineficiente debido a que una de las múltiples instrucciones asociadas al flujo de datos es la única que está siendo utilizada. Por eso ningún computador posee este tipo de arquitectura.

3.1.4. Múltiples instrucciones con muchos datos (*Multiple-Instruction-Multiple-Data MIMD*)

Para finalizar, la última de las arquitecturas propuestas junta varias veces la arquitectura SIMD, con el fin de que varios procesos complejos puedan ser ejecutados al mismo tiempo en la máquina, introduciendo el concepto de paralelismo real en los computadores.

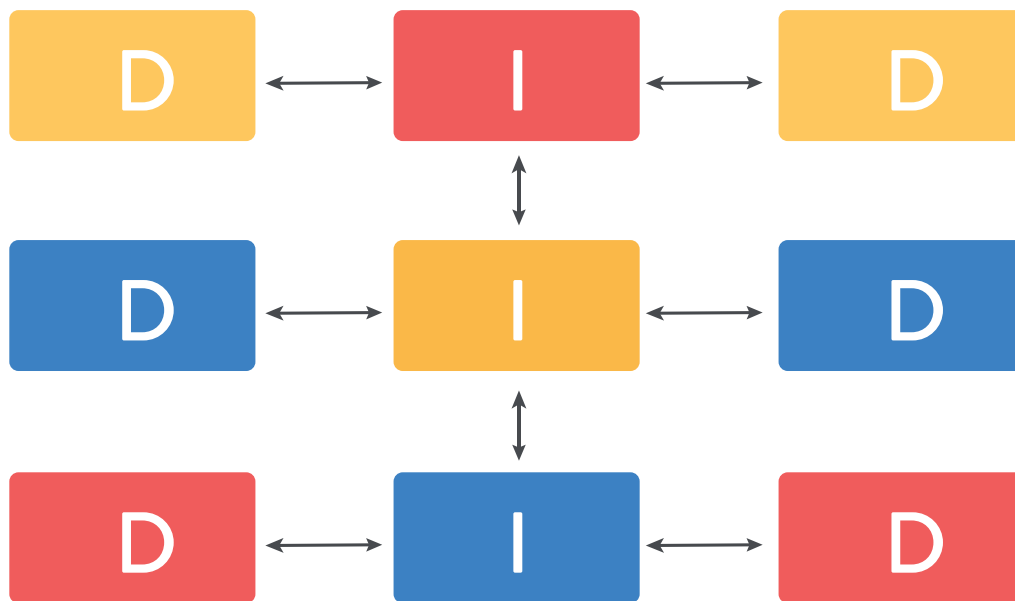


Figura 6: Taxonomía de Flynn - MSMD

Fuente: elaboración propia

Así como en la arquitectura SIMD, el crecimiento del flujo de datos por instrucción está dado por la función exponencial 2^n ; el crecimiento de instrucciones por máquina, dado por la capacidad de almacenamiento de los transistores (procesadores) en la misma unidad de espacio, también corresponde a la función 2^n . Los computadores y dispositivos contemporáneos responden a esta arquitectura.

4. Conceptos de Software

Ahora que los conceptos fundamentales de *hardware* están claros, es necesario revisar algunos conceptos relacionados con los procesos lógicos del computador, enfocados en dos parámetros relacionados con el estudio de los sistemas distribuidos: el acoplamiento y los protocolos de comunicación.

4.1. Acoplamiento

El acoplamiento de los componentes de *software* hace referencia al diseño de los módulos en el sistema y a la dependencia de los componentes o a la autonomía que tiene cada módulo para ser “autosuficiente” o estar ligado directamente con otro componente o módulo del sistema.

4.1.1. Fuertemente Acoplados

Quiere decir que los componentes tienen una fuerte relación en cuanto a su funcionalidad, y en los mensajes que pasan de un módulo a otro de *software*.

Ejemplo: una aplicación que necesita procesar un gran conjunto de datos ya que necesita cargarlos en memoria y hacerla accesible a diferentes procesos.

4.1.2. Débilmente Acoplado

Hace referencia a que las funcionalidades que realizan dos componentes de *software* están separadas y pueden trabajar más independientemente.

Ejemplo: un ODBC de una base de datos cualquiera y el *software* de reportes que ve el usuario final.

4.2. Protocolos

Los protocolos de comunicaciones son un conjunto de reglas preestablecidas, que se ponen en funcionamiento por medio de mensajes enviados entre los participantes de la comunicación.

El protocolo más importante relacionado con el estudio de los sistemas distribuidos es el RPC o *Remote Procedure Call* (Kantor, 1991) (Birrell & Nelson, 1984), encargado de darle instrucciones a un computador remoto, ya sean acciones que una máquina debe hacer para terminar una tarea específica o acciones con fines de coordinación que sirven a un propósito más grande, como una cadena de procesos, paso de mensajes, relegación de tareas, etc.

Referencias

Birrell, A. D. & Nelson, B. J. (1984). Implementing remote procedure calls. *ACM Transactions on Computer Systems*, 2(1), 39–59. Recuperado de <https://doi.org/10.1145/2080.357392>

Connolly, K. J. (2003). *Law of internet security and privacy*. Aspen Publishers.

Flynn, M. J. (1972). Some Computer Organizations and Their Effectiveness. *IEEE Transactions on Computers*, C-21(9), 948–960. Recuperado de <https://doi.org/10.1109/TC.1972.5009071>

Godfrey, M. D. & Hendry, D. F. (1993). The computer as von Neumann planned it. *IEEE Annals of the History of Computing*, 15(1), 11–21. Recuperado de <https://doi.org/10.1109/85.194088>

Kantor, B. (1991). *BSD Rlogin*. Recuperado de <https://tools.ietf.org/pdf/rfc1282.pdf>

King, J. L., & Leslie, J. (1983). Centralized versus decentralized computing: organizational considerations and management options. *ACM Computing Surveys*, 15(4), 319–349. Recuperado de <https://doi.org/10.1145/289.290>

Moore, G. E. (1965). Cramming More Components onto Integrated Circuits. *Electronics*, 114–117. Recuperado de <http://www.cs.utexas.edu/~fussell/courses/cs352h/papers/moore.pdf>

INFORMACIÓN TÉCNICA



FACULTAD DE
**INGENIERÍA, DISEÑO
E INNOVACIÓN**

Módulo: Sistema Distribuidos

Unidad 1: Conceptos Generales de los Sistemas
Distribuidos – Diseño de un SD

Escenario 1: Conceptos generales de los sistemas
distribuidos

Autor: Alexis Rojas

Asesor Pedagógico: Jeimy Lorena Romero Perilla

Diseñador Gráfico: Katherinne Pineda Rodriguez

Asistente: Maria Elizabeth Avilán Forero

Este material pertenece al Politécnico Gran Colombiano.

Prohibida su reproducción total o parcial.