

La programación concurrente y el interbloqueo en la asignatura de Sistemas Operativos

Miguel Riesco Albizu, Marián Díaz Fondón

Depto. de Informática
Universidad de Oviedo
e-mail: albizu@lsi.uniovi.es
fondon@correo.uniovi.es

Resumen

La programación concurrente es un tópico dentro de la docencia de los Sistemas Operativos. En este artículo se revisa su papel dentro de la asignatura y su relación, excesivamente estrecha en muchos casos, con otro tema de la misma asignatura como es el del interbloqueo.

1. Introducción

Dentro de los temas típicos que se imparten dentro de la materia de sistemas operativos está el de la programación concurrente. Este tema aparece incluido en los modelos curriculares más utilizados [9, 10], dentro del cuerpo de conocimiento de *Sistemas Operativos* (en concreto en el OS2) en el primer caso y en el módulo 4.3 (*Sistemas operativos y Arquitectura de Computadores I*) en el segundo. Los libros de texto que existen sobre la materia también recogen este tema, con mayor o menor profusión.

Lo mismo ocurre, como no podía ser menos, con la inmensa mayoría de los programas de las asignaturas de Sistemas Operativos que se imparten en las distintas Universidades.

Ante semejante unanimidad, parece lógico considerar que la inclusión de esta materia en el programa de una asignatura de este estilo está fuera de toda duda o discusión. En este artículo comenzaremos discrepando de esta opinión, por muy generalizada que esté, intentando aportar las razones que nos llevan a disentir.

Por otra parte, otro tema que también se incluye en todos los libros y temarios de Sistemas Operativos es el del interbloqueo. Sin discutir en este caso su adecuación con la materia, sí discrepamos de su posición temporal en el desarrollo de

la docencia de la asignatura, generalmente muy ligado al tema de programación concurrente, tanto que muchas veces aparece incluso incluido en él.

2. Programación concurrente y Sistemas Operativos

El tema de programación concurrente, como su propio nombre indica, es un tema eminentemente de programación. Trata de explicar las peculiaridades que presenta la realización de programas concurrentes, los problemas más habituales que aparecen y los medios más utilizados para resolverlos.

Ante este contenido puede surgir la pregunta de por qué se incluye este tema dentro de la asignatura de Sistemas Operativos, dado que, aparentemente, no tiene ninguna relación directa con los contenidos esperados de esta asignatura. A esa pregunta se suelen dar varias respuestas:

1. Con la inclusión de la multiprogramación, el sistema operativo debe manejar procesos concurrentes, con lo que un estudiante de sistemas operativos debe conocer los problemas que pueden aparecer en estos casos, así como la manera de solucionarlos.
2. El sistema operativo suele incluir mecanismos de sincronización y de comunicación entre procesos, por lo que se debe conocer para qué y cómo se van a utilizar estos mecanismos.
3. El propio sistema operativo se implementa muchas veces como un conjunto de funciones concurrentes, con lo que se debe ser capaz de entender, y en su caso desarrollar, programas concurrentes correctos.

Todas estas razones hacen referencia al carácter instrumental de esta materia. Es decir, es neces-

rio conocer programación concurrente para estudiar y comprender la materia específica de Sistemas Operativos, que no es otra que estudiar la estructura interna y el funcionamiento de un sistema operativo. Podemos resumir, por tanto, esas razones para la inclusión de programación concurrente en una sola: “Para la docencia de Sistemas Operativos nos hace falta la programación concurrente; como no se imparte en otro sitio, hay que hacerlo aquí”.

Esta razón podría considerarse académicamente válida si este tema fuera única y exclusivamente de utilidad dentro de la asignatura de Sistemas Operativos. Sin embargo, hay otras asignaturas, donde también son necesarios en mayor o menor grado los conceptos básicos de programación concurrente. Centrándonos en el primer modelo de currículo de los antes citados [9] son necesarios conceptos de programación concurrente en los cuerpos de conocimiento de *Programming Fundamentals* (PF7, en concreto), *Programming Languages* (PL11), *Information Management* (IM5 e IM6) y *Net-Centric Computing* (NC7), además de en el *Operating Systems* (OS2) ya citado.

Por otro lado, además de en las asignaturas que desarrollan estos cuerpos de conocimiento, en muchos Centros se considera la materia como digna de figurar como asignatura independiente en sus planes de estudio, tal y como recomienda el modelo curricular UNESCO-IFIP[10].

Volviendo al razonamiento de “inclusión por necesidad”, si lo aplicamos a otras partes de la asignatura podíamos incluir también un tema sobre programación, dado que es necesaria para entender los algoritmos que se explican, otro sobre estructuras de datos, para poder comprender las estructuras de datos, muchas veces complejas, que usan los sistemas operativos. O incluso podríamos, como se hace en [5] introducir un tema sobre modelado analítico y teoría de colas, que puede ser muy interesante para estudiar y comparar el comportamiento de las políticas que se emplean en distintas partes de un sistema operativo. Como obviamente esto no parece muy adecuado, podemos concluir que el tema de “programación concurrente” no debería estar incluido en la asignatura de Sistemas Operativos.

3. ¿Dónde debería tratarse la programación concurrente?

Centrándonos en las directrices del Consejo de Universidades para las Ingenierías Técnicas en Informática [3, 4] (y podemos extenderlo a las Ingenierías Superiores [2]), podremos ver que el tema de “programación concurrente” no aparece en el descriptor de ninguna asignatura, con lo que podrá impartirse legalmente en cualquiera donde tenga sentido incluirse.

Lo que se pretende con este artículo es iniciar un debate para decidir dónde podría ser más adecuado impartir una introducción a los conceptos básicos de programación concurrente.

Nosotros proponemos que sea en alguna de las asignaturas en la que se estructure la materia de “Metodología y Tecnología de la Programación”, por tres motivos:

1. Es la materia donde más conceptos relacionados con programación concurrente podemos incluir de los expuestos en [9].
2. Tiene casi el triple de carga docente asignada en las directrices del Consejo de Universidades [2, 3, 4] que la de Sistemas Operativos.
3. Independientemente de currículos y de directrices, el tema de programación concurrente es un tema (como su propio nombre indica) de programación.

4. Otro problema: el interbloqueo

Otro tema en principio relacionado con la programación concurrente es el del interbloqueo. En este apartado estudiaremos esta relación, así como el lugar más adecuado para la manera de abordar su docencia en la asignatura de sistemas operativos.

El interbloqueo es una anomalía que puede ocurrir durante la ejecución de procesos concurrentes debido a la competencia por los recursos. Si bien es cierto que prácticamente ningún sistema operativo real incorpora mecanismos de tratamiento de interbloqueo, esto es debido a una cuestión de la pérdida de rendimiento que conlleva su tratamiento para la baja probabilidad que hay de que ocurra. En un sistema operativo ideal, sin embargo, sí deberían incluirse mecanismos para su tratamiento, dado que existen y son bien conocidos.

Podemos considerar, sin temor a equivocarnos, que el tema del interbloqueo, además de ser un clásico en todos los libros, modelos de currículos, etc., sobre sistemas operativos es un tema que por sus características merece estar incluido en el temario de esta asignatura.

Si bien no se discute, por tanto, su adecuación a la asignatura, sí nos vamos a permitir cuestionar su posición en el desarrollo de la materia. La práctica totalidad de los libros de texto [5, 6, 7, 8], los distintos modelos de currículos [9, 10] y los programas de las distintas asignaturas que sobre el tema se imparten colocan este tema inmediatamente después del correspondiente a programación concurrente. Más aún, en algunas ocasiones [9] se llega a incluir el interbloqueo dentro del tema de programación concurrente.

Este hecho nos parece un importante error, por los motivos que expondremos a continuación, que hace por un lado que no se comprenda del todo bien su objetivo, y que, por otra parte, se complique innecesariamente un tema tan árido para los alumnos como es la programación concurrente.

5. La programación concurrente y el interbloqueo

Uno de los problemas con los que se puede encontrar el alumno cuando empieza a realizar sus primeros programas concurrentes es el del interbloqueo. En este contexto el interbloqueo que se produce es debido a una mala programación. Por ejemplo, en los clásicos intentos de solución al problema de la exclusión mutua se muestra este problema debido a un mal planteamiento del problema.

La única solución que hay para solventar los interbloques de este tipo es eliminar el error en la lógica del programa para que los procesos implicados no estén esperando el uno por el otro.

La utilización de mecanismos de sincronización de alto nivel facilita la realización de programas concurrentes libres de errores, de la misma manera que el desarrollo de un programa en un lenguaje de alto nivel es más fácil e introduce menos errores que si lo realizamos en ensamblador. Pero tanto en un caso como en otro sólo una buena programación puede asegurarnos tener éxito en nuestro propósito de construir programas correctos.

Lo que está claro es que, en ninguna circunstancia, podremos utilizar ninguna técnica que nos asegure la realización de un programa concurrente libre de interbloqueo ante un problema dado. Igualmente, tampoco podremos desde el sistema operativo evitar que ese interbloqueo se produzca si los procesos se empeñan en caer en él.

6. Interbloqueo y gestión de recursos

Otro de los campos típicos donde se presenta el problema del interbloqueo es en la gestión de recursos. En este caso varios de los procesos que se ejecutan concurrentemente en el sistema (procesos probablemente independientes entre sí) compiten por el uso de recursos no compartibles, de tal manera que hasta que el proceso que tiene asignado el recurso no lo libere ningún otro podrá utilizarlo. Si otro proceso necesita utilizarlo para continuar su trabajo, lo típico que ocurre es que este proceso suspende su ejecución hasta que el que ocupa el recurso lo libera.

En este contexto el interbloqueo se produce cuando los procesos suspendidos retienen recursos que son necesarios para la continuación de otros procesos, que a su vez retienen los recursos que necesitan los primeros. En definitiva, el problema se debe a que se han asignado los recursos que han ido necesitando los procesos sin ningún criterio, sumado al propio comportamiento de los procesos.

Este tipo de interbloqueo no se da únicamente en los sistemas operativos. Es muy típico también en los sistemas de gestión de bases de datos, por ejemplo, donde los recursos son registros o tablas de una base de datos que los procesos bloquean mientras los manejan si precisan de un acceso exclusivo.

Esta clase de interbloqueo está perfectamente estudiado y puede tratarse sin demasiada dificultad utilizando distintas técnicas, clasificadas normalmente en tres categorías: *prevención, evitación y detección y recuperación*. El problema que tienen estas técnicas es que sacrifican algo a cambio de solucionar el interbloqueo (la asignación de recursos será más lenta, al tener que comprobar si se puede permitir; se exige seguir una serie de normas a los procesos a la hora de solicitar recursos; etc.)

Este es el motivo de que no se suelen apenas utilizar en sistemas reales. Pero en cualquier caso

es un problema muy interesante de gestión de recursos que debe explicarse dentro del tema de Sistemas Operativos¹.

7. Relación entre el interbloqueo, la programación concurrente y los Sistemas Operativos

Como ya hemos indicado anteriormente, en la mayoría de libros, currículos y temarios de asignaturas de Sistemas Operativos el tema del interbloqueo está muy ligado al de programación concurrente. Tanto es así que habitualmente el tema de interbloqueo sigue inmediatamente al de programación concurrente, o incluso en algunos casos [9] aparece incluido en él como un apartado más. A nuestro juicio esto es un grave error, dado que hace parecer que son temas íntimamente relacionados cuando en realidad no es así. El interbloqueo que aparece cuando tratamos con programación concurrente es radicalmente distinto en su origen del que se presenta debido a la gestión de recursos:

- Los procesos implicados en un interbloqueo del primer tipo son procesos cooperantes, que comparten información y se sincronizan para lograr un objetivo común. En el segundo caso los procesos son independientes, y sólo comparten la necesidad de utilizar recursos comunes.
- El primer tipo de interbloqueo es causado por una mala sincronización, debido a una mala programación, al no considerar la situación que lleva al interbloqueo. En el segundo caso el problema aparece por una falta de control en la asignación de recursos que solicitan los procesos por parte del sistema operativo.
- El segundo tipo de interbloqueo sí es un problema que debe resolver el sistema operativo, mientras que el primer tipo es un problema meramente de programación.

De la misma manera, la solución al problema también es completamente distinta:

- El primer tipo de interbloqueo se soluciona mediante la corrección de los programas implicados en los mismos.

- El segundo tipo de interbloqueo se soluciona realizando la asignación de recursos a los procesos según alguna de las técnicas diseñadas al efecto (previniendo o evitando el interbloqueo, fundamentalmente).

En resumen, el segundo tipo de interbloqueo si entraría dentro de la materia de sistemas operativos, mientras que el primero debería resolverse dentro del campo de la programación concurrente.

El modelo curricular de la UNESCO [10] si parece recoger esta sensación, dado que incluye el interbloqueo en dos temas distintos: en el tema de procesos (donde se incluye también el de programación concurrente) en el módulo de *Sistemas Operativos y Arquitectura de Computadores I*, así como en el tema de entrada/salida del módulo de *Sistemas Operativos y Arquitectura de Computadores II*.

8. ¿Y cuál es el problema?

El contenido del tema del interbloqueo que se imparte en la asignatura de Sistemas Operativos se centra única y exclusivamente en el segundo tipo de interbloqueo, como no podía ser de otro modo.

Parece, por tanto, no demasiado adecuada la situación del tema de tratamiento del interbloqueo en el temario de la asignatura, dado que al situarlo a continuación del tema de programación concurrente puede dar la impresión, errónea, de que en ese apartado vamos a ver métodos para resolver los problemas de interbloqueo que nos han ido surgiendo en nuestros programas concurrentes, como en la “Cena de los filósofos” o en el “Problema de la sección crítica”. Esto hace que el alumno no comprenda bien el objetivo del tema y su utilidad dentro de un sistema operativo, creyendo en un principio que los métodos de tratamiento que se exponen van a ser la panacea a los problemas que se le presentan en su incipiente vida de desarrollador de programas concurrentes.

El porqué se sitúa en este lugar el tema del interbloqueo no puede explicarse más que por inercia (porque a alguien se le ocurrió en su momento a hacerlo así y tuvo éxito), dado que, como hemos visto, a pesar de haber “interbloqueos” en programación concurrente éstos nada tienen que ver con los que se tratan en este tema.

¹ No hay que olvidar que una definición clásica de “sistema operativo” lo caracteriza como “gestor de recursos”

9. ¿Dónde debería situarse el tema del interbloqueo?

Dentro de los sistemas operativos los posibles problemas de interbloqueo se suelen dar por la utilización de recursos periféricos. Los distintos modelos curriculares suelen incluir un tema de “Gestión de dispositivos” (o “Gestión de entrada/salida”), donde se explica la forma que tiene el sistema operativo de manejar los distintos dispositivos periféricos.

Nuestra propuesta es, por tanto, situar el tema del interbloqueo tras el tema de “Gestión de dispositivos”, para entender cómo se gestionan los dispositivos periféricos antes de ver cómo se trata el problema del interbloqueo debido a su comparación.

Con esta medida conseguimos, asimismo, alejarlo del tema de programación concurrente, lo cual nos aporta dos ventajas adicionales:

- Aumentamos el tiempo para que los alumnos asimilen los conceptos de programación concurrente, completamente nuevos para ellos, con lo que podrán comprender mejor las diferencias entre los dos tipos de interbloqueo.
- Evitamos confusiones innecesarias al enfatizar la diferencia entre los dos tipos de interbloqueo.

El modelo curricular UNESCO/IFIP [10] apoya, además, esta decisión.

10. Conclusiones

A pesar de que en buena parte de la bibliografía consultada se considera el tema de programación concurrente como una parte importante de la docencia de Sistemas Operativos, consideramos que no debe ser así. Sólo el hecho de que no se quiera incluir en ninguna otra asignatura, donde probablemente sería más adecuado hacerlo, puede justificar que se imparta en ésta, ante la necesidad que tiene de conocer esa materia para comprender los conceptos que le son propios.

En cualquier caso, si se incluye la programación concurrente dentro del temario de la asignatura de Sistemas Operativos consideramos que debe separarse claramente del tema de “tratamiento del interbloqueo”, dado que, por una parte, no tiene prácticamente ninguna relación con él, y que, por otro lado, induce a los alumnos a confun-

dir las dos clases de interbloqueo que se ven en la asignatura, complicando innecesariamente, además, la comprensión del tema de programación concurrente, bastante complejo de por sí.

Referencias

- [1] Carretero Pérez, J. y otros. *Sistemas Operativos. Una visión aplicada*. McGraw Hill, 2001
- [2] Consejo de Universidades. *Directrices generales propias de los planes de estudios conducentes a la obtención del título oficial de Ingeniero en Informática*. Boletín Oficial del Estado, número 278, pag. 34401 a 34403
- [3] Consejo de Universidades. *Directrices generales propias de los planes de estudios conducentes a la obtención del título oficial de Ingeniero Técnico en Informática de Gestión*. Boletín Oficial del Estado, número 278, pag. 34403 y 34404
- [4] Consejo de Universidades. *Directrices generales propias de los planes de estudios conducentes a la obtención del título oficial de Ingeniero Técnico en Informática de Sistemas*. Boletín Oficial del Estado, número 278, pag. 34404 y 34405
- [5] Deitel, H. M. *Sistemas Operativos. Segunda Edición*. Addison-Wesley Iberoamericana, 1993
- [6] Sinlverschatz, A. y Galván, P.B. *Sistemas Operativos*. Quinta Edición. Addison Wesley Longman, 1999.
- [7] Stallings, W. *Sistemas Operativos*. Cuarta Edición. Prentice Hall, 2001.
- [8] Tanenbaum, A. y Woodhull A. *Operating Systems Design and Implementation*, 2ª edición. Prentice-Hall, 1997
- [9] The Joint Task Force on Computing Curricula, *Computing Curricula 2001* IEEE Computer Society, Association for Computing Machinery, 2000
- [10] UNESCO/IFIP, *A Modular Curriculum in Computer Science*. UNESCO, 1994.