

Ил. С. С в и р и н, П. А. С и л и н,
В. В. С ю з е в, Е. А. З а р е ц к а я

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ВЗАИМНЫХ БЛОКИРОВОК. КОРРЕКТНОЕ ИСПОЛЬЗОВАНИЕ ИСКЛЮЧАЮЩИХ СЕМАФОРОВ

Одной из основных проблем разработки многопоточного программного обеспечения является взаимная блокировка потоков, которую чрезвычайно трудно выявить. Введена математическая модель взаимных блокировок.

E-mail: silinp@yandex.ru

Ключевые слова: многопоточное программное обеспечение, взаимные блокировки, верификация.

При разработке многопоточного ПО возникает ряд проблем, одна из них — это обеспечение доступа различных потоков к разделяемым ресурсам, для решения которой предоставляются современные средства синхронизации, что, в свою очередь, приводит к возникновению взаимных блокировок — ситуаций, когда потоки ожидают событие, которое никогда не произойдет.

Известно несколько подходов к решению этой проблемы: динамический анализ [1, 2], статический анализ [3–6] и верификация моделей по методу Model Checking [7–10].

Подход, представленный в настоящей статье, относится к верификации моделей. Применение такого подхода позволит разработчику ПО избегать появления потенциальных ситуаций взаимной блокировки в процессе проектирования и разработки ПО.

Настоящая статья содержит описание математической модели, в рамках которой формализованы основные понятия, необходимые для определения взаимных блокировок в многопоточном ПО. В рамках представленной модели доказан критерий отсутствия в ПО потенциальных ситуаций взаимных блокировок. Приведены основные понятия, используемые в модели, введен математический аппарат, необходимый для доказательства критерия отсутствия в ПО потенциальных ситуаций взаимных блокировок (этот критерий доказан), а также указаны направления, в которых можно расширить область действия данного критерия.

Основные понятия модели. Для моделирования ситуаций взаимных блокировок в многопоточном ПО введены следующие понятия: разделяемый ресурс, субъект доступа, средство синхронизации и взаимная блокировка.

Субъект доступа моделируется на основе системы переходов, которая описывает возможные пути выполнения субъекта с точки зре-

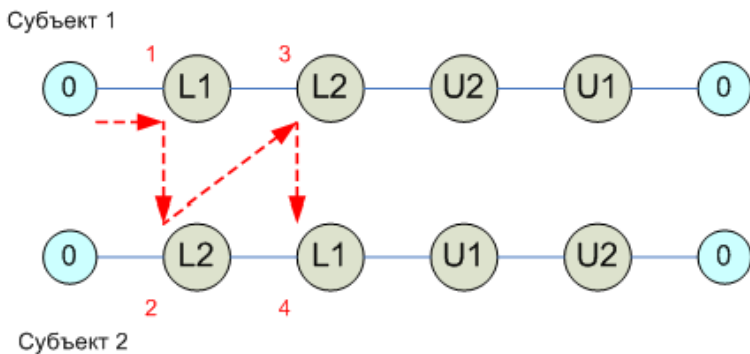


Рис. 1. Пример системы субъектов

ния взаимодействия со средствами синхронизации. На рис. 1 приведен пример графического изображения субъектов доступа.

Для каждого субъекта определены два выделенных состояния — состояние покоя, соответствующее началу произвольного пути выполнения субъекта, и завершающее состояние, соответствующее завершению произвольного пути выполнения субъекта (см. рис. 1, окружность с символом “0” внутри). В рамках модели состояние покоя и завершённое состояние отождествлены, т.е. субъекты доступа являются циклическими.

Рассмотрим всевозможные конечные пути выполнения субъекта, каждый из которых представлен как упорядоченная последовательность различных состояний субъектов (каждое изменение состояния субъекта обуславливается актом взаимодействия со средством синхронизации). Субъекты, изображенные на рис. 1, имеют единственный путь выполнения. Такие субъекты будем называть линейными.

Отметим, что анализируемая система представляет собой произвольную конечную совокупность субъектов доступа.

Средства синхронизации. В данной модели выделяются четыре примитива синхронизации: рекурсивный исключающий семафор, нерекурсивный (обычный) исключающий семафор, сигнальная переменная с памятью и сигнальная переменная без памяти. В данной статье будут рассмотрены взаимные блокировки в системах линейных субъектов, взаимодействующих только с исключающими семафорами.

Время выполнения субъектом оператора взаимодействия полагается равным нулю, поскольку не является существенным параметром с точки зрения модели.

Взаимные блокировки. В системе, изображенной на рис. 1, участвуют два субъекта: “Субъект 1” и “Субъект 2”. Пунктирной стрелкой показана последовательность выполнения операторов субъектами доступа. После выполнения шагов, показанных на рис. 1, оба субъекта находятся в состоянии ожидания — “Субъект 1” не может освободить

семафор, на котором ожидает “Субъект 2”, поскольку для этого ему нужно дождаться ситуации, когда “Субъект 2” освободит семафор, на котором ожидает “Субъект 1”; “Субъект 2” находится в аналогичной ситуации. Таким образом, оба субъекта перешли в состояние ожидания, из которого они не могут быть выведены никакими действиями других субъектов (если бы они были в системе), таким образом, оба субъекта находятся в ситуации взаимной блокировки.

Разделяемые ресурсы. Разделяемые ресурсы присутствуют в предложенной модели неявно. Они определяются совокупностью средств синхронизации, обеспечивающих синхронный доступ субъектов к разделяемым ресурсам.

Исключающие семафоры. Введем понятия локального и глобального времени. Локальное время вводится для каждого пути выполнения каждого из субъектов системы в целях отображения последовательности смены состояний субъекта при прохождении им этого пути. Множество значений, принимаемых локальным временем на пути выполнения, является конечным, поскольку число различных состояний субъекта на пути выполнения также конечно. Порядок моментов локального времени индуцирован порядком следования состояний во время прохождения субъектом пути выполнения. Моменты локального времени будем обозначать $\tau(S_i, O_k)$, где S_i — обозначает номер субъекта (каждый субъект в системе обладает уникальным номером, введенным с целью различать субъекты), O_k — обозначает оператор, который в данный момент выполнен субъектом; $\tau(S_i, U_k)$ означает, что i -й субъект в данный момент времени освободил k -й исключаящий семафор. Между локальными моментами времени существуют естественные операции сравнения, обусловленные порядком смены субъектом состояний во время реализации им данного пути выполнения. Поясним определение примером: рассмотрим систему на рис. 1.

Система состоит из двух субъектов S_1 и S_2 , запишем для них пути выполнения в терминах локального времени:

$$\begin{aligned} \tau(S_1, 0) < \tau(S_1, L_1?) < \tau(S_1, L_1!) < \tau(S_1, L_2?) < \\ < \tau(S_1, L_2!) < \tau(S_1, U_2) < \tau(S_1, U_1) < \tau(S_1, 0), \end{aligned}$$

$$\begin{aligned} \tau(S_2, 0) < \tau(S_2, L_2?) < \tau(S_2, L_2!) < \tau(S_2, L_1?) < \\ < \tau(S_2, L_1!) < \tau(S_2, U_1) < \tau(S_2, U_2) < \tau(S_2, 0). \end{aligned}$$

Символ “0” вместо оператора взаимодействия означает состояние покоя и завершённое состояние. Символ “?” при операторе L_i означает, что субъект в данный момент времени предпринимает попытку захвата исключаящего семафора. Символ “!” при операторе L_i означает, что захват успешно выполнен и субъект может продолжать дальнейшее

выполнение (независимо от того, переходил ли субъект в состояние ожидания или нет).

Введем понятие глобального времени системы. Глобальное время системы нужно для установления отношения порядка между локальными временными моментами различных субъектов, которые пока никак между собой не связаны. Фактически глобальное время системы можно рассматривать как шкалу обычного физического времени (с присоединенной к ней бесконечностью $\mathbf{R} \cup \{\infty\}$), на которую отображены моменты локального времени. Соответственно, та или иная динамика выполнения системы будет эквивалентна группе отображений из локального времени субъекта в глобальное время:

$$\forall i: S_i \in S \quad T_i: \{0, \dots, n_i\} \rightarrow \mathbf{R} \cup \{\infty\}.$$

Здесь под S понимается система субъектов, а под T_i — отображение из локального времени i -го субъекта в глобальное время. Опишем естественные ограничения на данную совокупность отображений.

Во-первых, никакие два различных момента локального времени не должны соответствовать одной точке глобального времени (в рамках модели считается, что события не могут происходить одновременно). Из этого правила существует исключение — если i -й субъект приступает к захвату k -го исключаящего семафора и он в состоянии его захватить, то $t(S_i, L_k?) = t(S_i, L_k!)$, где t обозначает момент глобального времени. Это правило связано с аксиомой модели, постулирующей, что успешный захват (преодоление) средства синхронизации осуществляется субъектом мгновенно. Если же при попытке захвата субъект переходит в состояние ожидания, то эти два момента локального времени соответствуют различным моментам глобального времени.

Во-вторых, каждое из отображений должно быть монотонным, т.е. если $\tau(S_i, O_1) < \tau(S_i, O_2)$, то $t(S_i, O_1) < t(S_i, O_2)$. Это правило отвечает за сохранность естественного порядка событий во времени, т.е. если на пути выполнения одно событие предшествует другому, то и во времени должен сохраниться тот же порядок.

Сформулируем проблему поиска потенциальных взаимных блокировок в терминах локального и глобального времени. Необходимо определить все такие совокупности последовательностей локальных моментов времени (путей выполнения субъектов), которые допускают отображения в глобальное время (динамики выполнения системы), отображающее часть локальных моментов времени в бесконечность, т.е. приводящие к взаимным блокировкам.

Перейдем к определению таких путей для наиболее простого случая — систем линейных субъектов (такие субъекты будем называть простыми), которые оперируют только с исключаящими семафорами.

эти потоки в некоторое общее состояние, характеризующееся тем, что первый исключаящий семафор из двух уже захвачен, а второй — еще нет.

Как следует из определений, условие $\tau(S_k, L_i) \triangleleft_G \tau(S_m, L_j)$ влечет условие $\tau(S_k, L_i) \triangleleft_L \tau(S_m, L_j)$. Обратная импликация не верна, поскольку некоторый набор локальных состояний может не реализовываться одновременно глобально.

Докажем теорему о взаимных блокировках системы простых субъектов, оперирующих с исключаящими семафорами.

Теорема. *В системе простых субъектов $S = \{S_1, \dots, S_k\}$, оперирующих с исключаящими семафорами, существует динамика выполнения системы, приводящая ее в ситуацию взаимной блокировки субъектов, тогда и только тогда, если существуют такие субъекты S_{i1} и S_{i2} и j -й исключаящий семафор, для которых выполнено $\tau(S_{i1}, L_j) \triangleleft_G \tau(S_{i2}, L_j)$. Возможно, $i1 = i2$.*

Доказательство. Пусть в системе простых субъектов $S = \{S_1, \dots, S_k\}$ существует динамика, приводящая систему в ситуацию взаимной блокировки. Это означает, что часть субъектов системы (S_{n1}, \dots, S_{np}) находится в состоянии ожидания при попытке захвата некоторого исключаящего семафора. Обозначим множество исключаящих семафоров, на которых ожидают субъекты, через e_1, \dots, e_f .

Из определения взаимной блокировки следует, что каждый из исключаящих семафоров e_1, \dots, e_f должен быть захвачен одним из субъектов S_{n1}, \dots, S_{np} , участвующих во взаимной блокировке.

Рассмотрим субъект, захвативший e_1 , он принадлежит $\{S_{n1}, \dots, S_{np}\}$, обозначим его S_{r1} ; так как он является участником взаимной блокировки, то он ожидает на исключаящем семафоре из множества $\{e_1, \dots, e_f\}$, обозначим его через q_1 . Если $e_1 = q_1$, то остановим процесс поиска, в противном случае — продолжим. Рассмотрим субъект $S_{r1} \in \{S_{n1}, \dots, S_{np}\}$. Пусть он ожидает на исключаящем семафоре $q_2 \in \{e_1, \dots, e_f\}$. Если $q_2 = q_1$ или $q_2 = e_1$ остановим поиск, в противном случае — продолжим. Множество $\{e_1, \dots, e_f\}$ конечно и его мощность не превосходит числа субъектов, участвующих во взаимной блокировке, поскольку на каждом из средств синхронизации ожидает как минимум один субъект. Следовательно, продолжая процесс поиска, рано или поздно найдем субъект, ожидающий на исключаящем семафоре, который уже был рассмотрен в этом поиске (т.е. уже захвачен каким-то из ранее рассмотренных субъектов). Пусть это будет субъект S_{nu} , ожидающий на исключаящем семафоре e_w , где $e_w = e_y$, который захвачен S_{nx} . Без ограничения общности перенумеруем субъекты и исключаящие семафоры таким образом, что рассмотренная цепочка субъектов от S_{nx} до S_{nu} получит номера по порядку от S_{n1} до S_{nd} , а

$$t(S_{nd}, L_{ed!}) \leq t_* \leq t(S_{nd}, L_{ed+1?}),$$

$$t(S_{i2}, L_{ed+1!}) \leq t_* \leq t(S_{i2}, L_j?).$$

Второе условие означает, что каждый из субъектов подсистемы $\{S_{i1}, S_{n1}, \dots, S_{nd}, S_{i2}\}$ не может захватить исключаящий семафор, находящийся в правой части неравенств для этого субъекта. Докажем это утверждение на примере субъекта S_{i1} . Итак, S_{i1} не может захватить исключаящий семафор e_1 , поскольку тот уже захвачен субъектом S_{n1} . Для того чтобы субъект S_{n1} освободил исключаящий семафор e_1 , ему по первому условию необходимо захватить исключаящий семафор e_2 , который удерживается субъектом S_{n2} . Переходя по цепочке, получаем, что субъекту S_{i2} , чтобы освободить исключаящий семафор e_{d+1} , необходимо первоначально захватить по первому условию j -й семафор, который он захватить не может (с этого начинался проход по цепочке). Получается, что ни один из субъектов не может захватить исключаящий семафор из правой части второй группы неравенств (второе условие). Это свойство не зависит от действий остальных субъектов.

Случай $i1 = i2$ доказывается аналогично. Теорема доказана и в обратную сторону. ►

Выводы. Рассмотрена математическая модель взаимных блокировок, сочетающая в себе формализм и наглядность, необходимые для построения на ее основе системы формальных правил корректного использования средств синхронизации, понятной конечному разработчику ПО. В рамках модели доказан критерий наличия ситуаций взаимной блокировки в системе линейных субъектов, оперирующих с исключаящими семафорами. Поскольку из глобального меньше следует локальное меньше, то из критерия может быть получено важное следствие, существенно упрощающее механизм доказательства отсутствия в модели ПО потенциальных ситуаций взаимной блокировки.

Следствие. Если в системе субъектов S нет таких субъектов S_{i1} и S_{i2} и j -го исключаящего семафора, для которых выполнено $\tau(S_{i1}, L_j) \leftarrow_L \tau(S_{i2}, L_j)$, то в системе нет потенциальных ситуаций взаимной блокировки. Как и ранее, возможно, $i1 = i2$.

Дальнейшая работа будет заключаться в расширении модели для случая сигнальных переменных и субъектов произвольной структуры с последующим получением из данной модели системы правил корректного использования средств синхронизации.

СПИСОК ЛИТЕРАТУРЫ

1. Savage S., Burrows M., Nelson G., Sobalvarro P. and Anderson T. Eraser: A dynamic data race detector for multi-threaded programs // Proceedings of the 16th ACM Symposium on Operating Systems Principles. – 1997. – P. 27–37.

2. Bensalem S. and Havelund K. Dynamic deadlock analysis of multi-threaded programs // Shmuel Ur, Eyal Bin, and Yaron Wolfsthal, editors, Haifa Verification Conference. – 2005. – Vol. 3875. – P. 208.
3. Detlefs D. L., Rustan K., Leino M., Nelson G. and Saxe J. B. Extended static checking. Technical Report 159, Compaq Systems Research Center, Palo Alto, California, USA, 1998.
4. Engler D. and Ashcraft K. RacerX: Effective, static detection of race conditions and deadlocks // Proc. of the 19th ACM Symposium on Operating Systems Principles. – 2003. – P. 237–252.
5. Havelund K. and Pressburger T. Model Checking Java programs using Java PathFinder // International J. on Software Tools for Technology Transfer, 2(4): 366–381, April 2000. Special issue of STTT containing selected submissions to the 4th SPIN workshop, Paris, France, 1998.
6. Artho C. and Biere A. Applying static analysis to large-scale, multi-threaded Java programs // D. Grant, editor, 13th Australian Software Engineering Conference, pages 68–75. IEEE Computer Society, August 2001.
7. Кларк Э., Грамберг О., Пелед Д. Верификация моделей программ: Model Checking. – М.: МНЦМО, 2002.
8. Карпов Ю. MODEL CHECKING. Верификация параллельных и распределенных программных систем. – СПб.: БХВ-Петербург, 2010.
9. Holzmann G. Design and validation of computer protocols // Prentice Hall, 1991.
10. Lamport L. Specifying systems: The TLA+ language and tools for Hardware and Software Engineers // Addison-Wesley, 2002.

Статья поступила в редакцию 15.12.2011