

**ПРОБЛЕМА СОЧЕТАНИЯ ГЕНЕРАТОРОВ ПСЕВДОСЛУЧАЙНЫХ
ВЕЛИЧИН В GPSS-МОДЕЛЯХ****Т. И. Алиев, Г. К. Асафьев (Санкт-Петербург)****Введение**

Точность результатов имитационного моделирования, базирующегося на методе статистических испытаний, в значительной степени зависит от качества применяемых генераторов псевдослучайных величин.

Одной из наиболее апробированных и широко распространенных систем имитационного моделирования является GPSS и, в частности, одна из версий, реализованная в среде Microsoft Windows, GPSS World [1], в которой используется встроенный генератор равномерно распределённых в интервале (0;1) псевдослучайных величин. Последовательность формируемых в процессе моделирования значений псевдослучайных величин зависит от задаваемого в качестве параметра целочисленного начального значения, что позволяет говорить о множестве используемых в GPSS World генераторов случайных величин, которые будем обозначать как 1, 2, 3, Кроме того, в GPSS World имеется более 20 библиотечных процедур для формирования случайных величин с заданными законами распределений, базирующихся на встроенных генераторах равномерно распределённых величин.

Длительная история существования и развития GPSS и её широкое применение в различных областях, казалось бы, даёт все основания полагать, что встроенные генераторы случайных величин обладают отменным качеством и могут использоваться при разработке любых моделей со случайным характером функционирования. При этом считается, что погрешности результатов имитационного моделирования при достаточно большой и обоснованно выбранной длительности моделирования обычно не превышают 3–5%.

В процессе эксплуатации GPSS World было выявлено, что в высокозагруженных моделях массового обслуживания (загрузка 0,8 и более) при некоторых сочетаниях генераторов случайных величин, используемых для формирования значений интервалов между поступающими в систему заявками и длительностей обслуживания заявок, погрешность результатов имитационного моделирования может составлять 30% и более даже для самых простых моделей, причём при использовании других сочетаний встроенных генераторов эта погрешность при той же длительности моделирования, задаваемой числом прошедших через систему заявок, составляет менее 3%.

Для выяснения обстоятельств, связанных с появлением такого эффекта, было проведено исследование генераторов псевдослучайных величин с целью оценки качества используемых генераторов и выявления закономерностей при использовании различных сочетаний, обуславливающих высокую погрешность результатов моделирования. Исследование проводилось в несколько этапов.

На первом этапе были выявлены так называемые проблемные сочетания генераторов, при использовании которых погрешность результатов моделирования оказывается значительной.

Для объяснения возникающей погрешности были протестированы проблемные встроенные генераторы равномерно распределённых случайных величин и библиотечные генераторы на соответствие декларируемым законам распределения [2].

Поскольку большая погрешность моделирования возникает только при использовании некоторых конкретных сочетаний генераторов, необходимо было выполнить корреляционный анализ соответствующих последовательностей случайных значений,

рассчитать коэффициенты корреляции и сравнить эти коэффициенты для проблемных и непроблемных сочетаний генераторов случайных величин.

На последнем этапе была проанализирована зависимость результатов моделирования, а именно среднего времени ожидания заявок в системе, от длительности моделирования, задаваемого количеством заявок, прошедших через модель.

Описание экспериментов

Для исследования погрешностей результатов имитационного моделирования при различных сочетаниях генераторов псевдослучайных последовательностей использовались GPSS-модели одно- и многоканальных систем массового обслуживания (СМО) с накопителем неограниченной ёмкости и однородным потоком заявок.

Для выявления проблемных сочетаний генераторов, при использовании которых погрешность результатов моделирования оказывается значительной, использовались СМО типа M/M/1, M/G/1 и M/M/2, поддающиеся точному аналитическому расчету [2]. При варьировании загрузки систем от 0,1 до 0,9 с шагом 0,1 и пропуске через каждую модель по 1 млн. заявок, было определено, что погрешность результатов моделирования лежит в приемлемых пределах (1–5%) для значений загрузки примерно до 0,7. Начиная со значения загрузки 0,8, при некоторых сочетаниях генераторов случайных величин погрешность начинала расти и составляла 10% и более, а при загрузке системы 0,95 – достигала 40%. Поскольку при исследовании в компьютерных сетях реальных систем, например маршрутизаторов, коммутаторов, серверов, наибольший интерес обычно представляют сведения о качестве функционирования системы при большой загрузке, задача обеспечения приемлемой точности в процессе имитационного моделирования становится весьма актуальной. В связи со сказанным все последующие эксперименты выполнялись для систем, загрузка которых составляла 0,9 и 0,95.

Для указанных значений загрузки системы было выполнено более 10 тысяч экспериментов при различных сочетаниях генераторов с номерами от 1 до 100 и определены сочетания генераторов, при которых достигается наибольшая погрешность результатов моделирования. В процессе экспериментов было установлено, что погрешность результатов моделирования зависит только от загрузки системы и длительности моделирования, задаваемого количеством заявок, прошедших через систему, и не зависит от соотношения составляющих загрузки – интенсивности поступления заявок и средней длительности их обслуживания. Кроме того, было установлено, что использование проблемных сочетаний генераторов случайных величин приводит к значительным погрешностям результатов моделирования для других, более сложных моделей, в частности для многоканальных СМО при любых законах распределений, использующих эти генераторы.

Проблемные сочетания генераторов

На модели M/M/1 с однородным потоком заявок и загрузкой 0,9 было проведено 10 000 экспериментов с целью выявления всех проблемных сочетаний генераторов с номерами от 1 до 100. В каждом эксперименте через модель пропусклся 1 млн. заявок. Относительная погрешность имитационного моделирования рассчитывалась для среднего времени ожидания заявок как $\delta = \frac{w^* - w}{w}$, где w^* – значение, полученное в процессе имитационного эксперимента; w – точное значение, рассчитанное для СМО M/M/1 с коэффициентом загрузки 0,9.

Для большинства сочетаний погрешность не превышала 5%. В то же время более 40 сочетаний выдали погрешность, превышающую 20%. В табл. 1 представлены все проблемные сочетания генераторов случайных величин, при использовании которых

погрешность среднего времени ожидания заявок в имитационной модели превышала 10% и 20%.

Таблица 1

Погрешность, %	Сочетания генераторов							
	10–20	(1;2)	(3;6)	(3;9)	(4;6)	(4;12)	(5;15)	(6;3)
	(7;14)	(7;21)	(8;16)	(8;32)	(8;48)	(9;3)	(10;30)	(11;77)
	(12;4)	(12;6)	(13;26)	(13;65)	(14;7)	(14;42)	(15;3)	(15;5)
	(15;10)	(15;45)	(16;8)	(16;32)	(16;48)	(16;64)	(17;51)	(17;68)
	(18;6)	(19;57)	(20;5)	(20;30)	(20;60)	(20;100)	(21;7)	(21;14)
	(21;63)	(22;44)	(22;66)	(24;48)	(24;72)	(24;96)	(25;50)	(25;75)
	(26;13)	(27;9)	(27;54)	(28;56)	(29;87)	(30;6)	(30;10)	(30;20)
	(30;60)	(32;8)	(32;48)	(33;22)	(33;66)	(33;99)	(36;12)	(38;19)
	(38;76)	(39;78)	(40;8)	(40;10)	(40;80)	(41;82)	(42;6)	(42;14)
	(42;28)	(44;22)	(44;88)	(46;69)	(46;92)	(47;94)	(48;8)	(48;16)
	(48;24)	(48;96)	(50;25)	(51;17)	(60;15)	(60;20)	(60;30)	(63;9)
	(63;21)	(64;16)	(65;13)	(66;22)	(68;17)	(69;46)	(70;35)	(72;24)
	(75;15)	(76;38)	(80;40)	(82;41)	(84;12)	(84;56)	(87;29)	(88;44)
	(92;46)	(93;31)	(94;47)	(96;24)	(96;48)	(99;33)	(100;20)	(100;50)
>20	(2;1)	(4;8)	(5;10)	(8;4)	(9;18)	(10;5)	(10;20)	(12;24)
	(14;28)	(15;30)	(18;9)	(20;10)	(20;40)	(21;42)	(24;12)	(26;52)
	(28;14)	(30;15)	(31;62)	(32;16)	(32;64)	(36;72)	(37;74)	(40;20)
	(42;21)	(43;86)	(45;90)	(49;98)	(52;26)	(54;27)	(57;19)	(62;31)
	(64;32)	(66;33)	(72;36)	(74;37)	(78;39)	(86;43)	(90;45)	(98;49)

Анализ представленных результатов показывает, что наиболее проблемными являются сочетания генераторов, номера которых соотносятся как 1:2, 1:3, 2:1 или 3:1.

Большинство проблемных сочетаний генераторов были проверены также для СМО общего вида G/G/1, в частности U/U/1, и многоканальных СМО. Для сравнения в табл.2 представлены результаты имитационного моделирования при использовании различных сочетаний генераторов в СМО M/M/1, загрузка которой составляла 0,95, а длительность обслуживания была равна 9,5 единиц времени. Точное значение среднего времени ожидания в этом случае равно 180,5 ед.

Таблица 2

Генераторы	Время ожидания, ед.	Погрешность, %
(2; 1)	114,4	-36,6
(2; 2)	173,6	-3,8
(2; 4)	155,8	-13,7
(10; 20)	121,1	-32,9
(20; 10)	106,6	-41,0
(20; 40)	111,8	-38,1
(40; 20)	113,2	-37,3
(110; 40)	179,9	-0,3
(200; 400)	136,5	-24,4
(999; 40)	177,1	-1,9
(2000; 4000)	119,3	-33,9

Как видно из табл. 2, для проблемных сочетаний генераторов случайных величин, используемых для выработки значений интервалов между поступающими заявками и длительности обслуживания соответственно, погрешность результатов моделирования достигала 40%, в то время как у непроблемных сочетаний (2; 2), (110; 40) и

(999; 40) при той же длительности моделирования погрешность не превышала 5%. Более того, использование одного и того же генератора с номером 2 (2; 2) для выработки значений интервалов между поступающими заявками и длительности обслуживания позволяет получить существенно более точный результат, чем использование этого же генератора в сочетании с другим генератором. Перестановка генераторов в сочетании с (10; 20) на (20; 10) незначительно изменяет погрешность моделирования.

Тестирование генераторов

Генератор равномерно распределенных случайных величин в системах ИМ выработывает последовательность значений в интервале (0; 1), которая называется базовой последовательностью. Базовая последовательность используется для формирования случайных величин, распределенных по другим законам и в других интервалах. Поскольку погрешность может появиться как на этапе генерации базовой последовательности, так и на этапе получения экспоненциально распределенной величины, было проведено тестирование базовых последовательностей и последовательностей экспоненциально распределенных случайных величин.

Относительная погрешность математического ожидания для генераторов, входящих в состав проблемных сочетаний, в большинстве случаев составляла более 0,1%, в то время как для других генераторов – менее 0,1%. Этот результат справедлив как для базовых последовательностей, так и для последовательностей экспоненциально распределенных случайных величин.

Для базовой последовательности относительная погрешность среднеквадратического отклонения для всех исследованных генераторов, включая генераторы, входящие в состав проблемных сочетаний, составила менее 0,01%. Для последовательностей экспоненциально распределенных случайных величин эта погрешность оказалась больше 0,01%, а для генераторов, входящих в состав проблемных сочетаний, достигала значений 0,2% и более.

Результаты корреляционного анализа

Для того, чтобы найти ответ на вопрос, почему некоторые сочетания генераторов случайных величин дают большую погрешность результатов моделирования, был выполнен корреляционный анализ базовых последовательностей и последовательностей экспоненциально распределенных случайных величин на выборке по 1 млн значений каждого генератора. Обработка полученных результатов показала, что коэффициенты корреляций проблемных сочетаний генераторов случайных величин в 5–10 раз превышают значения коэффициентов корреляции непроблемных сочетаний, причем они принимают положительные значения, т.е. большему (меньшему) значению первого генератора соответствуют большие (меньшие) значения второго генератора. Это означает, что с большой вероятностью заявке, поступившей в систему с большим интервалом (сформированным первым генератором) от момента поступления предыдущей заявки, будет соответствовать (сформировано вторым генератором) большее время обслуживания, и, наоборот, заявке, поступившей в систему с небольшим интервалом от момента поступления предыдущей заявки, будет соответствовать меньшее время обслуживания. Вероятность же ситуации, когда заявке, поступившей в систему с маленьким интервалом, будет соответствовать большое время обслуживания, оказывается незначительной. Всё это приводит к тому, что результаты имитационного моделирования, в частности среднее время ожидания, оказываются существенно заниженными даже при достаточно длительном моделировании.

Анализ переходных процессов в модели

Для проблемных сочетаний генераторов с целью выяснения характера изменения среднего времени ожидания от длительности имитационного моделирования, задаваемого числом заявок, проходящих через моделируемую систему, были выполнены эксперименты, при которых через систему пропускалось до 100 млн. заявок, а в некоторых моделях до 200 млн. заявок. Изучение переходного процесса в СМО позволило установить, что даже при таком большом количестве заявок, пропущенных через систему, среднее время ожидания не достигало расчетного значения, а погрешность могла составлять до 10%. На рис. 1 представлены результаты имитационного моделирования СМО М/М/1 с коэффициентом загрузки 0,95 при использовании генераторов с номерами 20 и 10 (20;10) для формирования интервалов между поступающими заявками и длительностей их обслуживания соответственно. Отметим, что коэффициент корреляции генераторов 20 и 10 – положительный и равен 0,00247. Представленные графики показывают зависимости времени ожидания заявок и относительной погрешности результатов имитационного моделирования от числа заявок, прошедших через имитационную модель.

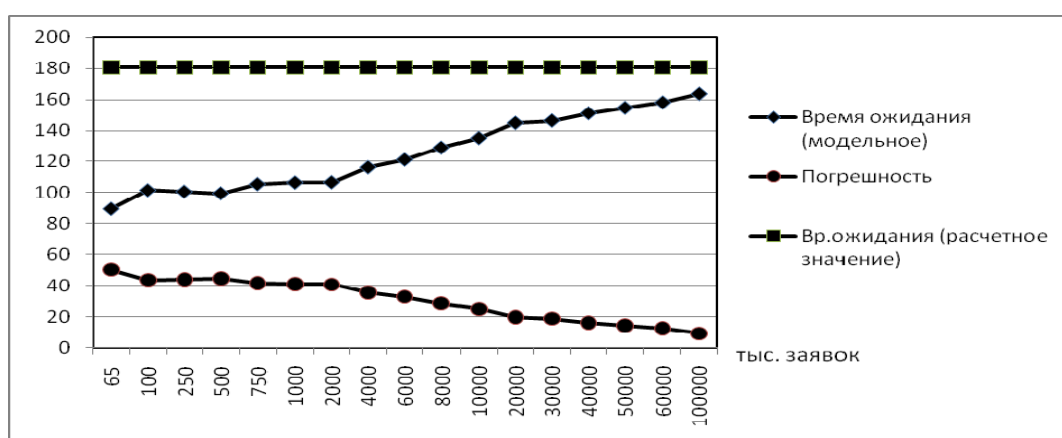


Рис. 1. Сочетание генераторов (20; 10)

Из графика видно, что даже после прохождения через модель 100 млн. заявок (что для такой простой модели непозволительная роскошь) относительная погрешность составляет около 10%.

Заметим также, что использование широко применяемого на практике метода «срезов» для определения вхождения системы в стационарный режим может привести к ошибочным результатам. Например, как видно из графика, среднее время ожидания при пропуске 100 тыс., 250 тыс. и 500 тыс. заявок остается практически неизменным (отличие менее, чем на 1%), откуда можно сделать вывод, что система вошла в стационарный (установившийся) режим и, следовательно, процесс моделирования может быть завершен. При этом погрешность результата моделирования оказывается более 40%.

Для сравнения на рис. 2. показана аналогичная зависимость для той же модели при использовании непроблемного сочетания генераторов (110; 40), откуда видно, что установившийся режим в системе существует уже после пропускания 500 тыс. заявок, и погрешность результатов моделирования в дальнейшем не превышает 4%. Следует обратить внимание на то, что такой характер зависимости модельного времени ожидания по отношению к расчетному может быть объяснен тем, что коэффициент корреляции генераторов 110 и 40 – отрицательный и равен -0,00176.

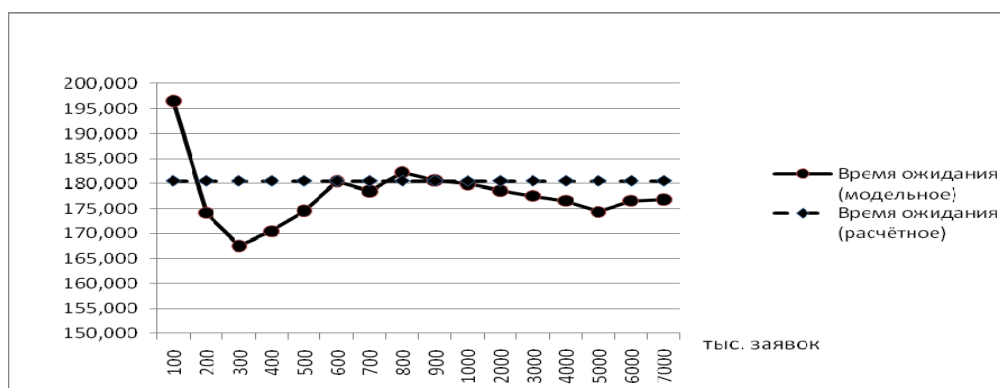


Рис. 2. Сочетание генераторов (110;40)

Выводы

1. Погрешность результатов моделирования СМО с высокой загрузкой (0,8 и выше) в среде GPSS World при использовании определённых сочетаний генераторов случайных величин, используемых для формирования значений интервалов между поступающими в систему заявками и длительностей обслуживания заявок, может оказаться высокой и достигать десятков процентов. Результаты многочисленных (более 10 тысяч) экспериментов, выполненных для различных сочетаний генераторов с номерами от 1 до 100, позволили определить, что наибольшая погрешность достигается для сочетаний генераторов, номера которых соотносятся как 1:2, 1:3, 2:1 и 3:1.

2. Корреляционный анализ показал, что коэффициенты корреляций проблемных сочетаний генераторов случайных величин принимают положительные значения и по абсолютной величине в 5–10 раз превышают значения коэффициентов корреляции непроблемных сочетаний, что в области больших загрузок исследуемой системы оказывает существенное влияние на результаты моделирования.

3. Анализ переходного режима функционирования системы при использовании проблемных сочетаний генераторов случайных величин показывает, что среднее время ожидания заявок медленно увеличивается с увеличением длительности моделирования (числа заявок, пропускаемых через систему), но так и не достигает точного значения, полученного в процессе аналитического расчета, в некоторых случаях даже после прохождения 100 млн. заявок через моделируемую систему.

Литература

1. Бражник А. Н. Имитационное моделирование: возможности GPSS World. СПб.: Реноме, 2006. 439 с.
2. Алиев Т. И. Основы моделирования дискретных систем. СПб.: СПбГУИТМО, 2009. 363 с.