

ОСОБЕННОСТИ РЕАЛИЗАЦИИ ПРОЦЕССОВ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ НА ОСНОВЕ МЕТОДОЛОГИИ DATA FARMING

Е.А. Криковлюк, В.А. Пепеляев, М.А. Сахнюк

Институт кибернетики им. В.М. Глушкова НАН Украины,
03680, ГСП, Киев-187, проспект Академика Глушкова,40,
тел.: (380-44)526 4107, (380-44)526 3308,
e-mail: emk160ik@gmail.com

Рассмотрены основные концепции и особенности применения методологии Data Farming в практике имитационного моделирования сложных стохастических систем. Предложен базирующийся на указанных концепциях подход к повышению эффективности системы распределенного моделирования NEDISOPT_D, разработанной в Институте кибернетики им. В.М. Глушкова.

The basic concepts and features of application of methodology Data Farming in simulation practice of complex stochastic systems are considered. There is proposed based on specified concepts the approach to increase efficiency of distributed simulation system NEDISOPT_D developed in V. M. Glushkov Institute of cybernetics.

Введение

В связи с широким использованием методов и средств имитационного моделирования в практике исследования и проектирования сложных систем актуальными остаются вопросы повышения эффективности указанных методов и разработки новых подходов на их основе.

Проблемы создания систем управления в области космических исследований, транспорта, финансов, маркетинга, логистики, медицины, для различных областей экономики, особенно в условиях глобализации, выдвинули новые требования к расширению функциональных возможностей существующих методов и средств имитационного моделирования. На формирование такого рода требований также существенным образом повлияли задачи принятия, как правило, в критические сроки, соответствующих проектных или управленческих решений при ликвидации различного рода чрезвычайных ситуаций, разработки современных военных стратегий, планов боевых действий, тактик обучения личного состава групп быстрого реагирования в условиях урбанизированных территорий.

В последние годы в мировой практике имитационного моделирования наряду с такими подходами к повышению эффективности, как оптимизационно-имитационная интеграция и использование технологий распределенных вычислений [1, 2], получила развитие методология Data Farming [3].

В основе этой методологии лежит интеграция возможностей методов имитационного моделирования, оптимизации, интеллектуального анализа данных и технологий распределенных вычислений, которые реализуются на высокопродуктивных платформах вычислительной техники (кластерных или сетевых архитектурах). Использование методологии Data Farming обеспечивает высокопродуктивную генерацию и обработку больших пространств параметров решений, что делает возможным оценивание непредвиденных ситуаций (как позитивных, так и негативных) и, как следствие, принятие наиболее правильных решений.

Концепция Data Farming впервые была предложена в 1998 году Гари Хорном (Gary E.Horne) при разработке проекта "Альберт" по заказу военно-морского флота США. Вокруг идеи Data Farming сформировано международное сообщество, которое регулярно проводит конференции International Data Farming Workshop. Основные результаты, полученные в области применения Data Farming, публикуются в работах ежегодной конференции Winter Simulation Conference. В настоящее время сформировано свыше десяти интернациональных групп, которые занимаются вопросами практического применения методологии Data Farming в разных прикладных областях. Наиболее активными участниками являются специалисты из США, Германии, Канады, Сингапура, Швеции, Голландии и Австралии. Кроме практического применения усилия специалистов направлены на разработку методологических и технологических стандартов для Data Farming.

Зарубежными специалистами по Data Farming создан центр SEED (Simulation Experiment and Efficient Design), одним из основных назначений которого является проведение исследований по разработке современных средств планирования и реализации оптимальных экспериментов для повышения эффективности процессов принятия решений [4].

В настоящее время в Институте кибернетики имени В.М. Глушкова НАН Украины выполняется работа по созданию системы распределенного оптимизационно-имитационного моделирования на основе методов и подходов, принятых в методологии Data Farming. При этом в качестве базовой составляющей будет использоваться ранее разработанная в Институте система NEDISOPT_D [5].

©Е.А. Криковлюк, В.А. Пепеляев, М.А. Сахнюк, 2012

Основные концепции методологии Data Farming

Методология Data Farming базируется на следующих двух идеях:

- разработка дистилляционных моделей (distillation), направленных на решение конкретных, поставленных лицом, принимающим решения (ЛПР) задач.

Поскольку методология Data Farming используется при исследовании сложных стохастических систем, для которых, как правило, отсутствуют (или еще не получены) соответствующие формальные описания, возникает необходимость в разработке имитационных моделей. При этом, такие модели не должны усложняться детализацией процессов функционирования исследуемых систем. Процедура “дистилляции” обеспечивает выбор такого уровня детализации, который, в первую очередь, учитывает специфику поставленных ЛПР вопросов и позволяет получить на них корректный ответ. Окончательная версия модели “distillation” погружается в среду Data Farming и многократно прогоняется на огромном пространстве параметров решений с целью выбора оптимального варианта. Эффективный и направленный поиск оптимальной комбинации параметров решений, выступающий в качестве факторов прогоняемой distillation модели, требует планирования экспериментов на основании соответствующих оптимизационных стратегий;

- использование высокопродуктивных платформ вычислительной техники для многократных прогонов дистилляционных моделей на варьируемых начальных условиях для получения возможных распределений результатов, трендов или выбросов.

Реализация любого приложения в среде Data Farming рассматривается как итеративно-командный процесс, в котором должны принимать участие эксперты проблемной области (subject matter experts), разработчики модели (modelers), аналитики (analysts) и ЛПР (decision makers).

На рис. 1. показана схема взаимодействия основных процессов в среде Data Farming [3].

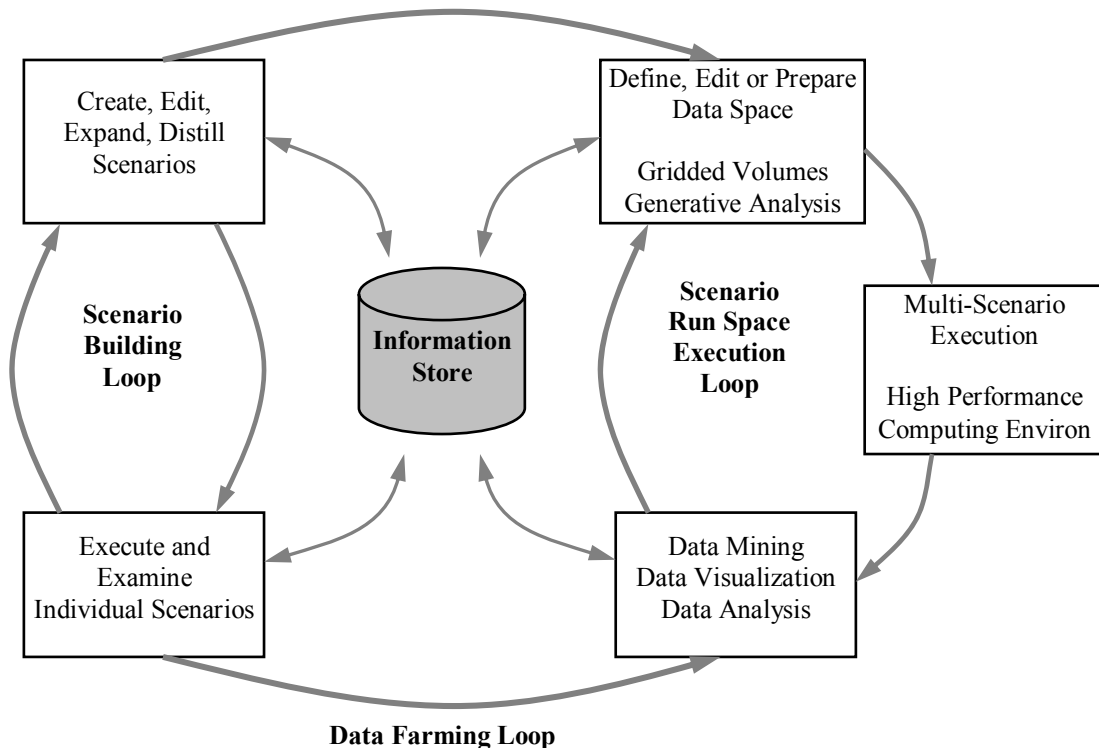


Рис. 1. Взаимодействие процессов в системе Data Farming

Процессы “*Scenario Building Loop*” предназначены для поддержки традиционных процедур имитационного моделирования, связанных с вопросами оценки адекватности построенной дистилляционной модели (верификация, валидация, оценка чувствительности). Это итеративный процесс, который достаточно часто требует коррекции.

Представленный справа на рис. 1 процесс “*Scenario Run Space Execution Loop*” используется командой для разработки сценария исследования, определяющего структуру входных данных для имитационных экспериментов и необходимые для их вариации подходы (методы планирования оптимальных экспериментов).

К наиболее известным оптимизационным стратегиям, используемым в Data Farming, относятся метод полнофакторного эксперимента, Near-Orthogonal Latin HyperCube, а также на основании алгоритмов эволюционных вычислений, базирующихся на определенных пользователем fitness-функциях.

На рис. 2 показаны основные компоненты системы Data Farming и курсирующие между ними потоки данных.

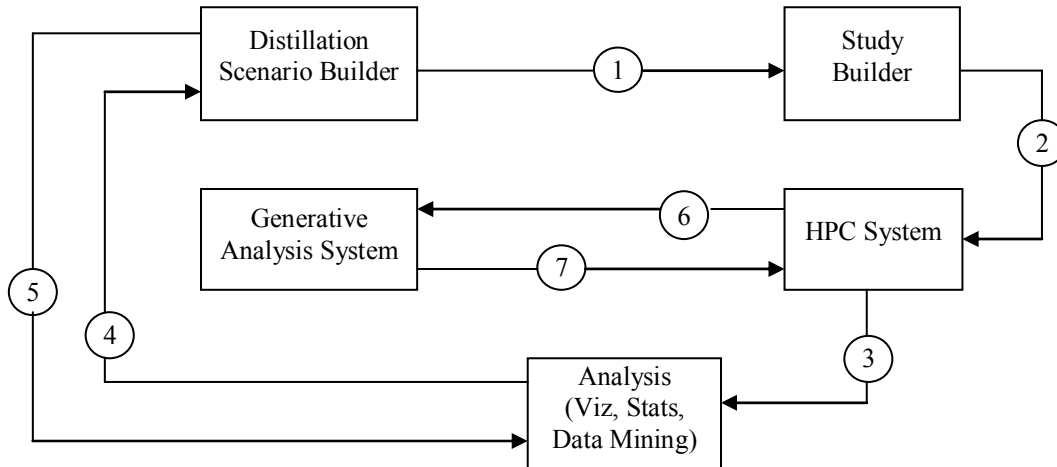


Рис. 2. Основные компоненты и схема потоков данных в системе Data Farming

Разработчики системы Data Farming особое внимание уделили вопросам структурной организации входных и выходных данных для всех компонент системы. Каждая представленная на рис. 2. компонента завершается формированием в специально отведенных файлах выходной информации в заданных форматах, которая будет использоваться компонентой-наследником. Так в результате выполнения модулей *Distillation Scenario Builder* формируются три выходных файла: файл, содержащий описанную на соответствующем алгоритмическом языке моделирования дистилляционную модель (*distillation model*); файл, содержащий описание всех используемых при прогоне модели данных (*roadmap*); пакет данных, используемых при реализации соответствующей модели на HPC (*scenario package*).

Компонента *Study Builder* на основании анализа информации, подготовленной компонентой *Distillation Scenario Builder*, формирует собственный выходной пакет (*study package*), содержащий данные для управления многократными прогонами моделей на HPC.

Компонента *HPC System* посредством соответствующей оптимизационной стратегии поддерживает многократные прогоны дистилляционной модели с последующим накоплением выходных результатов в принятых для Data Farming форматах. Компонента *Analysis* обеспечивает визуализацию и статистический анализ результатов экспериментов с оценкой возможности их завершения. Еще одна составляющая системы *Generative Analysis* (порождающий анализ) в соответствии с заданными критериями оценивает условия останова процесса моделирования или необходимость дополнительных прогонов имитационной модели.

В настоящее время международное сообщество пользователей Data Farming проводит исследования, связанные со стандартизацией форматов данных, используемых основными компонентами системы, с возможностью их автоматической подготовки.

Многие пользователи Data Farming разрабатывают проблемно-ориентированные версии указанной системы и базированные на них приложения [6, 7]. Заслуживает особого внимания разработанная специалистами из Сингапура версия системы Data Farming – Automated Red Teaming (ART), которая на основании концепции "ко-эволюции" (*co-evolution*) позволяет оценивать качество тактических и оперативных решений, принимаемых командами оппонентами: Red Team та Blue Team. Процессы принятия и оценки соответствующих решений реализуются параллельно по схеме *co-evolution* [8].

Структурная организация системы NEDISOPT_DF

Выполняемая в Институте кибернетики разработка системы NEDISOPT_DF базируется на концепциях Data Farming и особенностях структурной организации программной среды системы NEDISOPT_D [9].

Общая схема взаимодействия основных компонент системы NEDISOPT_D показана на рис. 3.

Основные модули компоненты *ОПТИМИЗАТОР* определяют соответствующие оптимизационные стратегии (на основе генетического алгоритма и методов последовательного перебора вариантов) и управляют распределенным поиском оптимальных решений на сетевой архитектуре. На каждом из сетевых компьютеров ($i = P0, \dots, PNA$) размещается компонента *ИМИТАТОР*, включающая систему распределенного моделирования NEDIS_D, которая обеспечивает прогон соответствующих версий имитационных приложений. В свою очередь каждое такое приложение содержит сценарий (*SCNA_i*) управляющий прогоном имитационной модели приложения (*SMA_i*) на множестве факторов, определенных в компоненте *ОПТИМИЗАТОР*.

В системе NEDISOPT_D процесс моделирования рассматривается как процесс поиска оптимальной альтернативы, основные характеристики которой представляются набором соответствующих факторов. При этом, каждый такой набор факторов задает структуру соответствующей хромосомы-решения. Множество таких хромосом определяет пространство поиска оптимальных решений.

Модуль *INTERFACE* обеспечивает взаимодействие между *ОПТИМИЗАТОРОМ* и *ИМИТАТОРОМ* на уровне потоков управления и потоков данных.

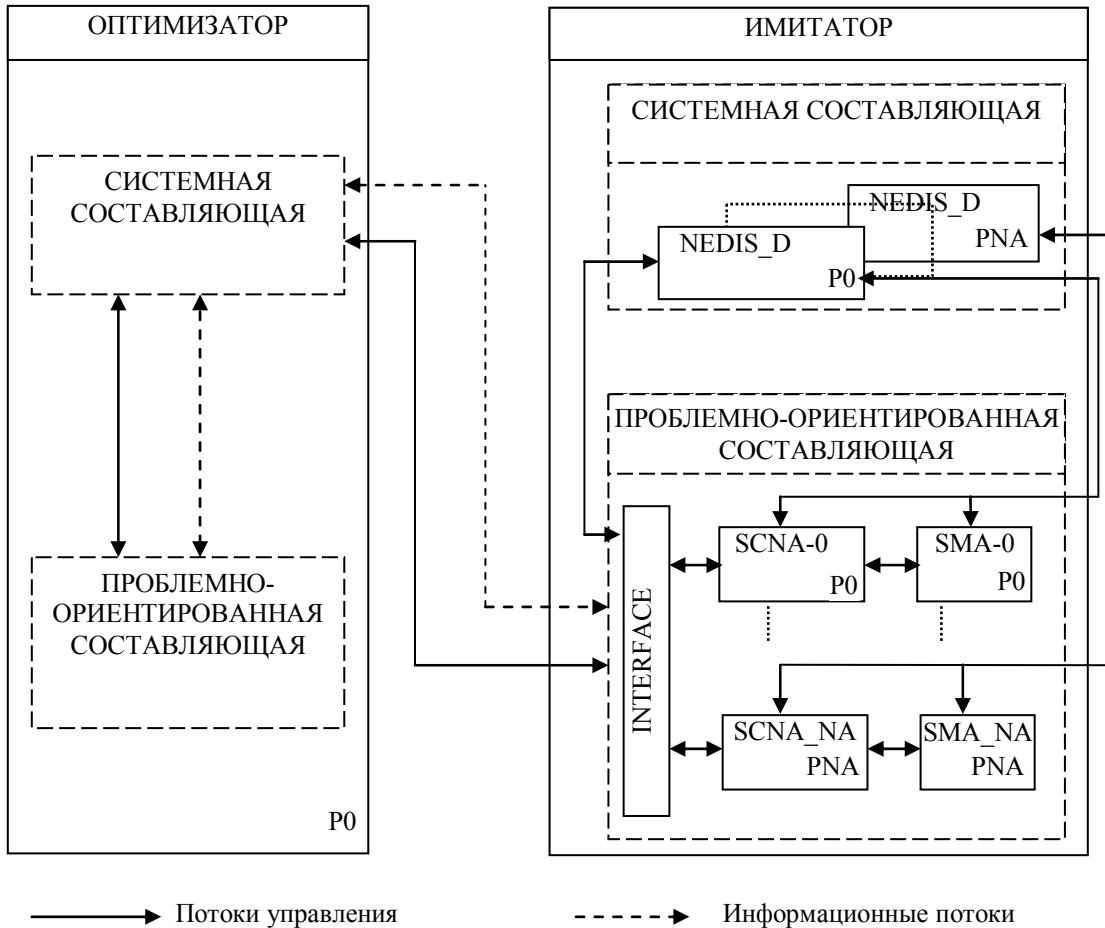


Рис. 3. Виртуальная среда поддержки системы NEDISOPT_D

Система NEDISOPT_D поддерживает различные режимы моделирования: последовательный (на одном процессоре); распределенный с реализацией на сетевой архитектуре. В первом случае система NEDIS_D размещается на одном процессоре с *ОПТИМИЗАТОРОМ*, а прогон имитационных моделей осуществляются циклически.

Схема проекта разрабатываемой системы NEDISOPT_DF показана на рис. 4.

В общем случае каждый узел сети характеризуется своим номером (L0...LN). При этом один из компьютеров каждого узла выбирается в качестве главного и содержит полную систему NEDISOPT_D, а на периферийных компьютерах (по аналогии с системой NEDISOPT_D) размещается компонента *ИМИТАТОР*. Такой подход позволяет распределить пространство параметров решений между компьютерами сети и осуществлять направленный поиск оптимальных решений на каждом узле с возможностью миграции хромосом-решений между узлами сети. При этом главный сценарий системы NEDISOPT_DF обеспечивает управление процессом распределенного моделирования, контролирует шаг тактировки процессов миграции. По завершению каждого шага осуществляется обмен хромосомами-решениями согласно принятому алгоритму миграции.

В процессе реализации экспериментов на каждом из главных компьютеров, принадлежащих узлам сети, осуществляется накопление статистических данных, представленных откликами имитационных моделей и соответствующими fitness-значениями.

По завершенню процесів моделювання кожен із вузлів мережі здійснює експорт отриманих результатів в загальне сховище даних. Аналіз множини таких даних дозволяє вибрати оптимальне рішення.

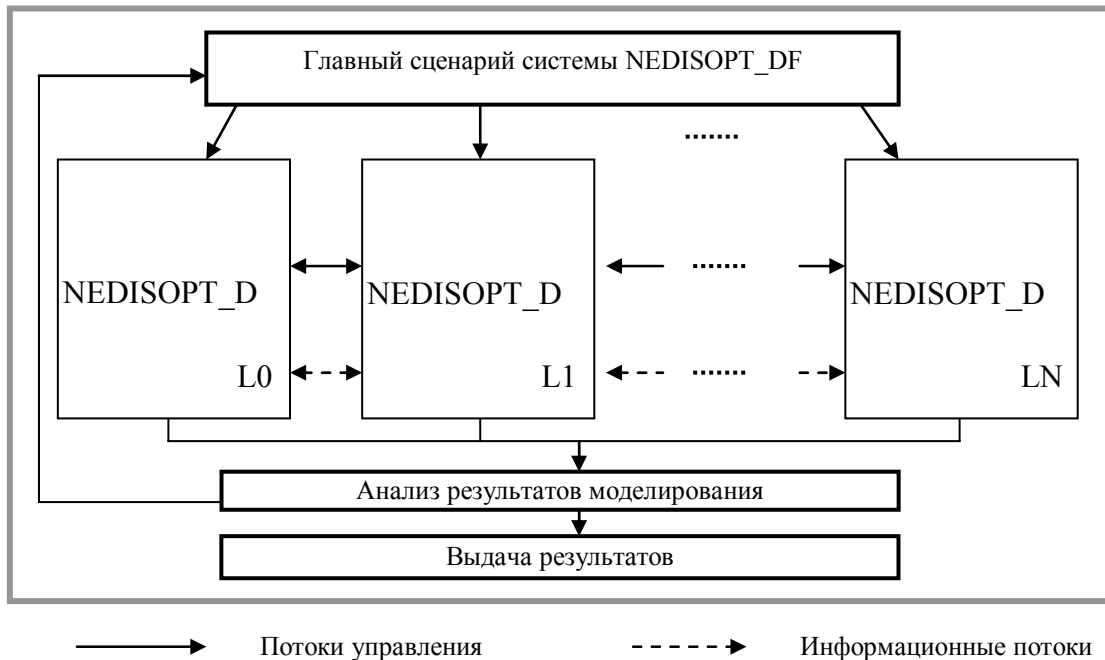


Рис. 4. Структурна схема системи NEDISOPT_DF

Спеціально розроблені інформаційні моделі-шаблони задають управляючу інформацію для експериментів, реалізуємих в форматі сесій моделювання.

Выводы

Поскольку процессы исследования и проектирования сложных систем на базе методологии Data Farming сопровождаются генерированием и накоплением больших объёмов разнородной информации, возникает необходимость в использовании методов и средств интеллектуального анализа данных (в частности Data Mining). Одно из перспективных направлений модификации системы NEDISOPT_DF связано с внедрением и использованием именно таких средств.

1. *Fu M.* Optimization for Simulation: Theory and Practice // *INFORMS J. on Computing.* – 2002. – № 14(3). – P. 192 – 215.
2. *April J., Glover F., Kelly J.P., Laguna M.* Practical introduction to simulation optimization // *Proc. of the 2003 Winter Simul. Conf.* – 2003. – P. 71 – 78.
3. *Horne G. E., Meyer T. E.* Data Farming: Discovering Surprise // *Proc. of the Winter Simulation Conf.*, 2005. – P. 1082 – 1087.
4. *SEED Center for Data Farming* // <http://harvest.nps.edu/>
5. *Галаган Т.Н., Пепеляев В.А., Сахнюк М.А.* Особенности реализации многослойного сценария распределенного поиска оптимальных решений // *Проблеми програмування.* – 2008. – № 2 - 3. – С. 636 – 640.
6. *Barry Ph, Koehler M.* Simulation in context: using Data Farming for decision Support // *Proc. of the Winter Simulation Conf.*, 2004. – P. 814 – 819.
7. *Horne G.E., Schwierz K.-P.* Data Farming around the world overview // *Proc. of the Winter Simulation Conf.*, 2008. – P. 1442 – 1447.
8. *Choo C.S., Ng E.C., Ang D., Chua C.L.* Data Farming: a brief history // *Proc. of the Winter Simulation Conf.*, 2008. – P. 1448 – 1455.
9. *Пепеляев В.А., Сахнюк М.А., Чёрный Ю.М.* Параллельная реализация процессов направленного поиска оптимальных решений // *Проблеми програмування.* – 2010. – № 2 – 3. – С. 572 – 576.