УДК 004.75

А.В. Скатков, профессор, д-р техн. наук
Севастопольский национальный технический университет
ул. Университетская 33, СевНТУ, Севастополь, Украина, 99053
E-mail: kvt@sevgtu.sebastopol.ua

СТРУКТУРНАЯ ДИНАМИКА И ОБЕСПЕЧЕНИЕ ПРАВ ДОСТУПА В РАСПРЕДЕЛЕННЫХ СИСТЕМАХ ОБРАБОТКИ ДАННЫХ

Рассматривается алгоритмический метод построения информационного процесса структурной динамики системы обработки данных. Предложена модель основной оптимизационной распределительной задачи функционально-полных средств мониторинга. Проанализирована процедура обеспечения прав доступа в условиях вирусных атак на основе методов искусственных иммунных систем.

Ключевые слова: *структурная динамика, права доступа, распределенная обработка.*

Широкое использование распределенных вычислений в качестве опорного средства большинства информационных технологий (в здравоохранении, энергетике, банковских системах, бизнес-процессах и т.д.), а также стремительно растущие требования к обеспечению гарантоспособности вычислительных сервисов, резко обострили проблему диспетчеризации в распределенных вычислительных системах (РВС).

Анализ публикаций [1] и программно-технических решений в этой области позволяет сделать вывод о том, что популярность РВС обусловлена рядом их преимуществ: возможностью распараллеливания вычислений, масштабируемостью, совместной работой пользователей с общими ресурсами и д.р. Закономерно, что перечисленные достоинства распределенной обработки данных порождают собой ряд проблем, связанных с организацией гарантоспособных сервисов по диспетчеризации в условиях различных информационных ситуаций.

Актуальной является задача нахождения методов разрешимости конфликтной ситуации, связанной порой с необоснованным наращиванием вычислительной мощности каналов обработки данных и недостаточной проработанностью вопросов организации диспетчеризации в целом. Предлагается подход к формализации принятия решений для разрешения описанного конфликта на основе рассмотрения структурной динамики и обеспечения прав доступа в распределенных системах обработки данных. Тем самым предполагается построение соответствующей информационной технологии (ИТ), под которой понимается совокупность методов и реализованных на их основе процессов и средств, предназначенных для сбора, обработки, передачи, хранения, поиска и отображения информации в интересах поставленной целевой задачи. Реализация информационных технологий предполагает поиск аппаратных, программных и сетевых решений, являющихся ядром информационно-управляющих систем различного назначения.

Эффективность предлагаемых информационных технологий непосредственным образом влияет на проблему безопасности технологических комплексов критических инфраструктур и во многом определяется динамическими процессами смены состояний. Динамика состояний системы обработки данных может быть задана соответствующей траекторией её состояний. Всё множество возможных состояний можно разбить на ряд непересекающихся классов, и на его основе разработать адекватные методы управления траекториями перехода состояний.

Целью статьи является разработка методов, направленных на повышение эффективности функционирования таких систем на основе учета взаимодействия и взаимной обусловленности информационных и алгоритмически связанных процессов в них. Такие процессы являются нестационарными по своей природе, распределёнными как во времени, так и по узлам обработки данных, т.е. пространственно. Под влиянием информационных субъектов, обладающих определёнными правами доступа к объектам такой системы, в ней развиваются декларативные процессы, а также препятствующие им различные деградационные процессы.

Для успешного функционирования системы обработки данных (СОД) необходимо предусмотреть в ней реализацию процессов реструктуризации и компенсации на основе свойств ресурсной и временной избыточности. Это можно обеспечить за счёт использования нескольких классов моделей, основными функциями которых являются:

– кластеризация информационных потоков;
– синтез решающих правил распределения информационных ресурсов;
– эффективное управление правами доступа субъектов к объектам.

Предлагаемый далее подход ориентирован на применение в условиях сетевых технологий обработки данных, которые обеспечены функционально полными по своему составу мониторинговыми средствами для потоков и основных событий, возникающих в узлах сети.

Постановка задачи. В достаточно общем виде модель оптимизационной распределительной задачи запишется как

$$M_G = \langle A, B, C, M, X, \omega \rangle,$$

где A, B, C – конечное множество сетей; $M = \left\langle M_{AB}, M_{AC}, M_{BC}, m_{AB}, m_{AC}, m_{Bc} \right\rangle$ – конечное множество отношений минимальной и максимальной связанности, построенные на декартовых произведениях соответствующих узлов сетей A, B, C; X – конечное множество функционально полных средств мониторинга; ω – совокупность алгоритмов, позволяющих на полном информационном описании ситуации, связанной с идентификаторами A, B, C, M, X найти конечное множество адресов узлов, в которых необходимо стационарно установить средства мониторинга, в совокупности обладающие функциональной полнотой относительно рассматриваемых событий.

Поиск ω-алгоритмов при несколько иных предположениях рассмотрен в работе [2].

С использованием обозначений, принятых в [2], структура исходных данных приведена в таблице 1.

Таблица 1 – Структура исходных данных для основной оптимизационной распределительной задачи

Типы сетей	Множество видов стационарных установленных средств мониторинга	Множество видов средств мониторинга, подлежащих распределению	Множество адресов узлов со стационарно установленными средствами мониторинга	Множество адресов узлов, в которых необходимо разместить средства мониторинга
Сеть A	DSA_i	DRA_i	SA_i	RA_i
Сеть B	DSB_j	DRB_j	SB_j	RB_j
Сеть C	$DSC_{(ij)}$	$DRC_{(ij)}$	$SC_{(ij)}$	$RC_{(ij)}$

Выполним конструктивное построение локальных алгоритмов разрешения основной распределительной задачи для сетей A_i–типа, $i = \overline{1, p}$. При построении алгоритмов используются следующие основные операнды:

IS_i – множество всех узлов A_i,

I_i – множество адресов A_i,

A_1: A_{11}, … , A_{1s} – узлы,

A_i: A_{i1}, … , A_{is},

A_p: A_{p1}, … , A_{ps},

DS – конечное множество стационарно установленных средств мониторинга,

DR – конечное множество средств мониторинга, подлежащих распределению,

S – конечное множество адресов стационарно установленных средств мониторинга,

R – конечное множество адресов средств мониторинга, подлежащих распределению,

P_{AB} – конечное множество предикатов структурной разрешимости задачи распределения средств мониторинга,

Q_{BC}; Q_{AC} – конечное множество предикатов функциональной разрешимости задачи распределения средств мониторинга,

Δ – отображение DR ставит в соответствие каждому отношению из L_{AB}, M_{AC}, M_{BC} элемент из $P_{AB} \vee Q_{AC} \vee Q_{BC}$.

Далее предполагается, что рассматриваемая сеть является гетерогенной, т.е. для всех $i = \overline{1, p}$ выполняется $S_i = S_a$.

Выделим такую сеть A_i, для которой $\left|DRA_i\right| \neq 0$ (т.е. для сети A_i не решена задача размещения) и пусть $\left|DSA_i\right| \neq 0$, т.е. множество стационарно установленных средств мониторинга непусто. Возьмём некоторый A_{ij} такой, что $I_{ij} \in RA_i$, т.е. для данного узла A_{ij} выполнены одновременно два условия. В нём отсутствует стационарно установленное средство мониторинга. Для него ещё не решена задача размещения. Такой узел назовём свободным.

Локальное окружение свободного узла – это множество узлов $LA_i(A_{ij})$ сети A_i, образованное теми, и только теми узлами A_i, для которых в множестве SA_i определён соответствующий элемент.

При выполнении такта результативного решения задачи размещения для A_i с использованием $LA_i(A_{ij})$ в дискретный момент времени t_n справедливо выполнение следующих рекуррентных соотношений:

$$|DSA_i|_{t_{n+1}} = |DSA_i|_{t_n} + 1, \quad (1)$$

$$|DRA_i|_{t_{n+1}} = |DRA_i|_{t_n} - 1, \quad (2)$$

$$|SA_i|_{t_{n+1}} = |SA_i|_{t_n} + 1, \quad (3)$$

$$|RA_i|_{t_{n+1}} = |RA_i|_{t_n} - 1. \quad (4)$$

Невыполнение равенства (1) является необходимым и достаточным условием невыполнения равенств (2), (3) и (4).

Равенство (1) является определяющим условием результативного разрешения задачи размещения для A_i в $LA_i(A_{ij})$ при $t = t_n$.

Последние четыре элемента кортежа $\langle t_n, A_i, A_{ij}, DSA_i, DRA_i, SA_i, RA_i \rangle$ являются полным информационным описанием для поиска решений в момент времени t_{n+1}.

Упорядоченная последовательность PA_i: $\{\alpha_{i1},...,\alpha_{is}\}|_{t_n}$ образована следующим образом: каждый её элемент α_{ij} –либо пустой элемент, либо элемент множества DSA_i, $\alpha_{ij} \in DSA_i$.

Если $\alpha_{ij} = \alpha$, то каждый элемент множества DRA_i является претендентом для возможного размещения в узле A_{ij} . Для каждого такта времени решения t_n число элементов в последовательности PA_i, для которых выполняется $\alpha_{ij} = \alpha$, в точности равно $|DRA_i|$.

Если в t_n принято решение рассмотреть локальную задачу размещения для сети A_i, то по результатам такого решения последовательность $PA_i|_{t_n}$ трансформируется в такую $PA_i|_{t_{n+1}}$, в которой ровно один символ α заменён на некоторый элемент α_{ij} из $DRA_i|_{t_n}$, причём на следующем такте этот элемент удаляется из множества DRA_i и становится элементом множества $DSA_i|_{t_{n+1}}$. Если для A_i с использованием A_{ij} на шаге t_n не получено результативного решения, то этот элемент отмечается дополнительным индексом $A_{ij}(0,t_n)$ или $A_{ij}(1,t_n)$ в противном случае.

Последовательности вида PA_i можно упорядочить по второму индексу инициирующего элемента из RB_j. Используя $j_{\min} = \min_{j \in R}\{j \in RA_{ij}\}$, $j_{\max} = \max_{j \in R}\{j \in RA_{ij}\}$, а также описанную процедуру формирования PA_i, получим упорядоченную двойную последовательность, которая является конструктивным решением поставленной задачи поиска функционально полного размещения.

Наличие в сетях обработки данных функционально полного состава средств мониторинга позволяет с новых позиций подойти к анализу безопасности управления доступом в компьютерных системах (КС). Весьма перспективным представляется подход к управлению правами доступа на основе методов теории искусственных иммунных систем (ИИС). Плодотворность подхода, связанного с ИИС подтверждена в различных предметных областях использования информационных технологий при управлении вычислительной мощностью, ресурсообеспеченностью и вычислительной устойчивостью прикладных задач.

Модели управления правами доступа оперируют с O_t – множество всех объектов КС в момент времени t; $S_t \subset O_t$ – множество субъектов КС в момент времени t; R_a – множество видов доступа; R_r – множество видов прав доступа; R_f – множество видов информационных потоков.

Аналогично множество видов информационных процессов R_f можно определить как перечень кортежей вида $R_1 : \langle S_1(t)R_1(t)O_1(t)\rangle|\langle S_1(t+1)R_2(t+1)O_1(t+1)\rangle|..|\langle S_k(t)R_c(t)O_n(t)\rangle$. Каждый субъект КС может реализовать особый вид информационного потока, который имеет соответствующую историю одной уникальной инициализации.

Вершины некоторого уровня S не имеют входящих ребер, т.к. не подлежат инициализации, являясь ее источником. Множество ребер, сформированное графом текущих доступов, определяется как элемент декартового произведения $E_t \subset E = S_t \times O_t \times R_a$. Субъект $S_1(t)$ имеет право доступа $R_r(i,t)$ и на основе этого права инициирует доступ $R_a(i,t)$, что порождает информационный поток $R_f(i,t)$, который описывается подграфом текущих доступов $G(i, R_a(i,t))$. Объединение таких подграфов $\bigcup_i G_i \subset G_0 = (O_t, E_t)$, где $E_t \subset E_0 = S_t \times O_t \times R_a$ описывает все множество текущих доступов.

Последовательность графов текущих доступов $G(t=0), G(t=1), ..., G(t=T)$ определяет траекторию функционирования КС с начальным состоянием G_0.

Естественные иммунные системы [3] обладают основным функциональным эволюционным свойством: процессы обучения распознаванию внешних антигенов, а также интеллектуальными функциями, позволяющими выполнить распознавание клеток, содержащихся в организме и их классификацию «свой» / «чужой». Дальнейшее распознавание в естественных иммунных системах (ЕИС) «чужих» с целью стимуляции защитных механизмов соответствующего типа, осуществляется на основе макрофагом – функционально-различимых клеток особого вида со специфической функцией презентации антигенных клеток. В ЕИС основы компетентности и иммунной ответственности реализуются с помощью специальной системы лимфоцитов структуры двух целе- и функционально связанных множеств элементов. Первое множество таких элементов соответствуют клеткам Т-типа с функцией усиления или подавления реакций клеток В-типа. Второе множество таких элементов соответствуют клеткам В-типа с функцией распознавания и обучения эффекторной деятельности.

Функциональная аналогия с искусственными иммунными системами (ИИС) состоит в том, что на основе анализа трафика с помощью специальных программ трассировки и результатов функционально полного мониторинга, полученного на основе рассмотрения основной распределительной задачи можно установить вероятностное описание элементов множеств $R_a, R_к, R_f$ и их декартовых произведений T_a и N_a. Распознавание фактов нарушений прав доступа осуществляется в виде многоэтапных процедур. Принятие решений по распознаванию факта нарушений осуществляется в стохастически определенные упорядоченные промежутки времени.

По аналогии с процессом функционирования системы макрофагов в ЕИС процесс распознавания факта вирусной атаки в ИИС можно описать на основе подсистемы искусственных макрофагов (ПИМ). Динамику процессов в ПИМ будем описывать полумарковским случайным процессом, заданном на конечном фазовом пространстве $E = \{1,2,\ldots N\text{-}1,N\}$. Для этого необходимо задать вероятности переходов вложенной цепи Маркова и времена пребывания в состояниях. Равенство $\xi_n = k$ будем понимать как факт пребывания системы при выполнении задания в момент времени с индексом n на этапе с индексом *k*.

Функционирование ПИМ состоит в следующем:

– в начальный момент времени ПИМ находится на первом этапе распознавания E_1 в течение некоторого случайного времени η_1, после чего переходит на второй этап распознавания E_2, если $\eta_1 < h_k$, или вновь выполняет распознавания на первом этапе E_1, если $\eta_1 \geq h_k$;

– переход ПИМ с этапа *k* на этап (*k*+1) либо на первый этап происходит с вероятностями $P_{k,k+1}$ и $P_{k,1}$, причем $P_{k,1} + P_{k,k+1} \leq 1$ для любого *k*;

– если происходит переход ПИМ с этапа *k* на этап (*k*+1), то распознавание на этапе выполняется в течение случайного времени η_k с функцией распределения $G_{k,k+1}(t)$.

Поведение ПИМ, таким образом, может быть описано двумя последовательностями: $\{\xi_{k,k}\}$ и $\{\eta_{k,k}\}$ или одной двумерной последовательностью $\{\xi_k, \eta_{k,k}\}$. Математическая модель функционирования ПИМ в этом случае может быть задана характеристиками двумерного полумарковского процесса.

Запишем вероятности переходов вложенной цепи Маркова

$$\begin{cases} P\{\xi_{n+1} = k+1 \mid \xi_n = k\} = P\{\eta_k \leq h_k\} = F_k(h_k), \\ P\{\xi_{n+1} = 1 \mid \xi_n = k\} = P\{\eta_k > h_k\} = 1 - F_k(h_k) = \overline{F}_k(h_k), \\ P\{\xi_{n+1} = 1 \mid \xi_n = N\} = 1. \end{cases}$$

Остальные переходные вероятности для последовательной стратегии функционирования ПИМ равны нулю.

Математическое ожидание времени выполнения многоэтапного распознавания вирусной атаки определяется из соотношения

$$t_{1,N} = \sum_{i=1}^{N-1} \frac{h_i - F_i(h_i)[h_i - M\{\eta_i\}]}{\prod_{m=1}^{N-1} F_m(h_m)} \quad ,$$

где N-1 – число этапов задания, h_i – предельно допустимое время выполнения i-го этапа ($i = 1, 2, \ldots, N$-1), η_i – время выполнения задания на i-ом этапе (случайная величина с функцией распределения $F(x)$).

Приведем результаты некоторых экспериментальных исследований, полученных при моделировании и характеризующих динамику ПИМ. Задача анализа состоит в исследовании влияния вида функции распределения времени выполнения этапа задания по распознаванию факта вирусной атаки (рисунок 1). В качестве аргумента будем рассматривать безразмерную величину – относительно операционное время системы мониторинга $\tau = \frac{M(\eta)}{h}$. Основная системная характеристика – относительное время эффективного распознавания факта вирусной атаки $\theta = \frac{\sum h_i}{\sum M_i(\eta)}$.

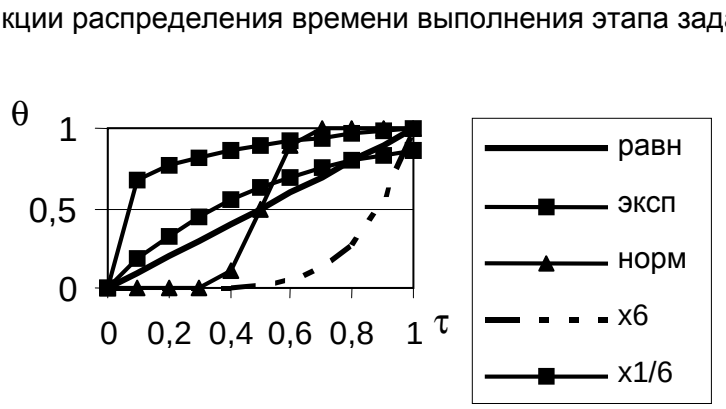

Рисунок 1 – Основные функции распределения, используемые при распознавании факта вирусной атаки

В этой группе экспериментов были установлены следующие параметры: число этапов распознавания $N = 10$; время выполнения i-го этапа η_i распределено по одному их следующих законов: 1) равномерное распределение; 2) показательное распределение с параметром $\lambda = 2$; 3) нормальное распределение с параметрами $m = 0{,}5$ и $\sigma = 0{,}08$; 4) малой интенсивности с показателем $\alpha = 6$; 5) высокой интенсивности с с показателем $\alpha = 1/6$; $M\{\eta_i\} = 0{,}5$.

Анализ результатов модельных экспериментов показывает, что для всех заданных видов функции распределения времени выполнения этапов распознавания фактов вирусной атаки среднее время достигает минимума при максимальном значении предельно допустимого контрольного времени $h = h_{\max} = 1$.

Для значений h, меньших $M\{\eta_i\} = 0{,}5$, наименьшие значения среднее время $t_{1,N}$ имеют место при распределении времени высокой интенсивности. Если же характер распределения таков, что малые значения η_i имеют небольшую вероятность (нормальное распределение и распределение малой интенсивности, то при $h < 0{,}5$ среднее время выполнения распознавания имеет наибольшие значения.

Таким образом, совокупность предложенных моделей структурной динамики, решение основной распределительной задачи средств мониторинга и параметризация оценки эффективности обнаружения вирусных атак на основе методов искусственных иммунных систем являются решением поставленной задачи. Направлением дальнейших исследований является отыскание эффективных методов принятия решений в условиях недостаточной априорной информации о состояниях нагрузки трафика в некоторых характеристических узлах сети.

Библиографический список использованной литературы

1. Топорков В.В. Модели распределенных вычислений / В.В. Топорков. — М.: Физматлит–М., 2004. — 320 с.

2. Скатков А.В. Оптимизационные задачи гарантированного оценивания ситуаций в гетерогенных локальных инфраструктур GRID–систем / А.В. Скатков // Оптимизация производственных процессов: сб. науч. тр. — Севастополь, 2010. — Вып. 12. — С. 9–16.

3. Искусственные иммунные системы и их применение / под. ред. Д. Дасгупты. Пер. с англ. под ред. А.А. Романюхи. — М.: Физматлит, 2006. — 344 с.

Поступила в редакцию 19.08.2011 г.

Скатков О.В. Структурна динаміка і забезпечення прав доступу в розподілених системах обробки даних

Розглядається алгоритмічний метод побудови інформаційного процесу структурної динаміки системи обробки даних. Запропоновано модель основного оптимізаційного розподільного завдання функціонально-повних засобів моніторингу. Проаналізовано процедуру забезпечення прав доступу в умовах вірусних атак на основі методів штучних імунних систем.

Ключові слова: структурна динаміка, права доступу, розподілена обробка.

Skatkov A.V. Structural dynamics and providing access rights to distributed data processing systems

We consider an algorithmic method for constructing the information process for structural dynamics of the data processing system. A model of the main optimization of the distribution problem for functionally-complete monitoring tools is suggested. The procedure for securing access rights under virus attacks is analyzed based on the methods of artificial immune systems.

Keywords: structural dynamic, access, distributed processing.