



ISSN 2073-2600

ВОПРОСЫ ЗАЩИТЫ ИНФОРМАЦИИ

1

(100)

В НОМЕРЕ

МУЛЬТИГРАФОВАЯ МОДЕЛЬ АНАЛИЗА
ФУНКЦИОНАЛЬНОЙ УСТОЙЧИВОСТИ
НАВИГАЦИОННО-ИНФОРМАЦИОННЫХ СИСТЕМ

ПОМЕХОУСТОЙЧИВОЕ СТЕГАНОГРАФИЧЕСКОЕ
ВНЕДРЕНИЕ ИНФОРМАЦИИ В ЗВУКОВЫЕ ФАЙЛЫ

АДАПТАЦИЯ И УПРАВЛЕНИЕ РАЗВИТИЕМ СИСТЕМ
ЗАЩИТЫ ИНФОРМАЦИИ

ОБЪЕКТНО-ОРИЕНТИРОВАННЫЙ АНАЛИЗ
ИНТЕГРАЦИИ СРЕДСТВ ЗАЩИТЫ ИНФОРМАЦИИ

Подписывайтесь,

читайте,

пишите в наш журнал

Москва 2013

Все новейшие достижения и современные разработки предприятий оборонного комплекса — в научно-технических журналах ФГУП «ВИМИ»

ФГУП "Всероссийский научно-исследовательский институт межотраслевой информации — федеральный информационно-аналитический центр оборонной промышленности" является учредителем и издателем следующих научных журналов:



Межотраслевой научно-технический журнал

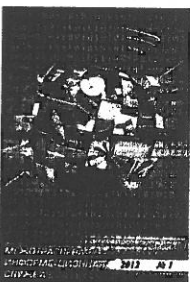
**Оборонный комплекс —
научно-техническому
прогрессу России**
(4 выпуска)

Подписной индекс **79379**
Издается с 1984 года



Научно-технический журнал
**Информационные технологии
в проектировании
и производстве**
(4 выпуска)

Подписной индекс **79378**
Издается с 1976 года



Научно-методический журнал
**Межотраслевая
информационная служба**
(4 выпуска)

Подписной индекс **46144**
Издается с 1977 года



Межотраслевой научно-технический журнал
**Конструкции из композиционных
материалов**
(4 выпуска)

Подписной индекс **80089**
Издается с 1981 года



Научно-практический журнал
Вопросы защиты информации
(4 выпуска)

Подписной индекс **79187**
Издается с 1974 года



Межотраслевой научно-практический журнал по отечественным и зарубежным материалам

**Экология промышленного
производства**
(4 выпуска)

Подписной индекс **80090**
Издается с 1993 года

Все издания ФГУП "ВИМИ":

✓ включены решением ВАК Министерства образования и науки России в перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата и доктора наук;

✓ метаданные выпусков включены в базу данных Российского индекса научного цитирования (РИНЦ).

Более подробную информацию об изданиях, подписке, дополнительных услугах можно получить по тел.: 8 (495) 491-60-87, 8 (495) 491-84-77, 8 (495) 491-60-37, 8 (495) 491-77-20 (подписка); факс 8 (495) 491-68-20.

E-mail: fvimi@yandex.ru, office@vimi.ru
<http://www.vimi.ru>

ВОПРОСЫ ЗАЩИТЫ ИНФОРМАЦИИ

НАУЧНО-ПРАКТИЧЕСКИЙ ЖУРНАЛ

1
(100)

Москва
2013

Основан
в 1974 г.

СОДЕРЖАНИЕ

Техническая защита информации: защита персональных данных, инженерная криптография, управление доступом, доверенная среда, электронная подпись в информационных системах

Тарасов А. А., Королёв А. Н. Мультиграфовая модель анализа функциональной устойчивости навигационно-информационных систем 3

Голобоков В. А. Методика построения рациональной структуры серверов локальных центров управления полетами 9

Аленин А. А., Алексеев А. П. Помехоустойчивое стеганографическое внедрение информации в звуковые файлы 15

Михеев В. А., Кузнецов А. В., Репин М. М. Способ определения степени уязвимости автоматизированной информационной системы в отношении конкретных методов реализации угроз безопасности информации 20

Организационное и правовое обеспечение

Шарамок А. В. О методе разработки модели источника угроз 26

Общие вопросы безопасности информации и объектов

Мазин А. В., Петренко П. Б. Оценка рисков нарушения информационной безопасности и уязвимости в телекоммуникационных системах, построенных на базе интегральных микросхем с ТТЛ-логикой 32

Малюк А. А. Адаптация и управление развитием систем защиты информации 37

Королёв М. В. Определение границ зоны защищенности конфиденциальной информации от утечки за счет ПЭМИ методом критических сечений 40

Васильев Р. А. Исследование особенностей фонетического строя речи и идентификация дикторов по голосу 43

Васин В. А., Ивашов Е. Н., Степанчиков С. В. Особенности распознавания образов в кластерных системах обработки информации 52

Главный редактор *В. Г. Матюхин*,
д-р техн. наук, первый заместитель генерального
директора, научный руководитель ОАО "НИИАС"

Заместитель главного редактора *В. А. Коняевский*,
д-р техн. наук, акад. РАЕН, зам. директора по НИР
ВНИИ ПВТИ

Ответственный секретарь *К. В. Трыкина*,
старший научный сотрудник ФГУП "ВИМИ"

Редакционный совет:

А. Л. Бальбердин, зам. директора Административного
департамента аппарата Правительства РФ; *Е. А. Беллев*,
советник директора Федеральной службы по техническому
и экспортному контролю (ФСТЭК России); *А. С. Голубков*, д-р техн. наук, ведущий сотрудник
ВНИИПВТИ; *Ю. Н. Лаврушин*, канд. техн. наук, зам.
генерального директора службы корпоративной защиты
ОАО "Газпром"; *В. Б. Муравин*, зам. директора Федера-
льной службы по оборонному заказу; *М. П. Сычев*,
д-р техн. наук, профессор, директор РУНЦ "Безопас-
ность" МГТУ им. Н. Э. Баумана

Редакционная коллегия:

М. М. Грунтович, канд. физ.-мат. наук, доцент,
руководитель обособленного подразделения ОКБ
"Сапр"; *С. В. Дворякин*, д-р техн. наук, профессор,
акад. РАЕН, декан факультета кибернетики и инфор-
мационной безопасности МИФИ; *И. Г. Назаров*, канд.
техн. наук, начальник управления Федеральной службы
по техническому и экспортному контролю (ФСТЭК
России); *С. П. Панасенко*, канд. техн. наук, начальник
отделения разработки программного обеспечения
фирмы "Анкард"; *В. Н. Пожарский*, зам. директора РУНЦ
"Безопасность" МГТУ им. Н. Э. Баумана; *Г. В. Росс*,
д-р техн. наук, д-р экон. наук, профессор, замести-
тель директора по учебно-методической работе ФГУП
"ВНИИПВТИ"; *А. А. Стрельцов*, д-р техн. наук, д-р
юрид. наук, начальник департамента аппарата Совета
безопасности РФ; *А. М. Сычев*, канд. техн. наук,
доцент, зам. начальника Главного управления безо-
пасности и защиты информации ЦБ РФ; *А. А. Тара-
сов*, д-р техн. наук, профессор, директор Института
информационных наук и технологий безопасности
РГГУ; *Ю. С. Харин*, д-р физ.-мат. наук, академик
НАН БР, директор НИИ прикладных проблем мате-
матики и информатики БГУ; *И. Б. Шубинский*, д-р техн.
наук, профессор, генеральный директор ЗАО "ИБТранс",
советник генерального директора ОАО "НИИАС".

Свидетельство о регистрации
ПИ № ФС77-35665 от 24.03.2009 г.

© Федеральное государственное унитарное пред-
приятие "Всероссийский научно-исследовательский
институт межотраслевой информации — федеральный
информационно-аналитический центр оборонной
промышленности" (ФГУП "ВИМИ"), 2013

<i>Пителинский К. В.</i> К вопросу об оптимизации топологии системы защиты информации для помещений различной конфигурации	61
<i>Птицын А. В., Птицына Л. К.</i> Объектно-ориентированный анализ интеграции средств защиты информации	69
<i>Самсонов Ф. А.</i> Факторная параметрическая модель и метод определения количественной меры реализации инцидента в АСУ. 77	

Вопросы защиты информации: Науч.-практ. журн./ ФГУП "ВИМИ", 2013. Вып. 1 (100). С. 1—84.

Редактор *Г. А. Никитин*

Корректор *М. А. Николенко*

Компьютерная верстка: *Н. В. Ильина, И. А. Жамальдинова*

Подписано в печать 13.03.2013. Формат 60x84 1/8.
Бумага офсетная. Усл. печ. л. 9,8. Уч.-изд. л. 10,5.
Тираж 1500 экз. Заказ 1782. Цена договорная.
Отпечатано в ООО "МЭЙЛЕР". 127550, Москва,
ул. Прянишникова, д. 19А, стр. 1.
E-mail: vziok@vimi.ru
Индекс 79187.

Особенности распознавания образов в кластерных системах обработки информации*

В. А. Васин, канд. техн. наук; Е. Н. Ивашов, д-р техн. наук;

С. В. Степанчиков, канд. техн. наук

Московский институт электроники и математики НИУ ВШЭ, Россия

Рассмотрена математическая модель функционирования кластерных систем обработки информации, алгоритмы и методы моделирования состояния подобных систем во времени, а также метод определения порога функционирования системы с использованием эволюционной стратегии.

Ключевые слова: кластерные системы обработки информации, распознавание образов, кластеры повышенной производительности, оценка достоверности, эволюционная стратегия, производительность систем обработки информации.

Кластер функционирует как единая система, т. е. для пользователя или прикладной задачи вся совокупность вычислительной техники выглядит как один компьютер. Именно это и является самым важным при построении кластерной системы (КС). В настоящее время кластерные системы обработки информации получают все большее распространение в связи с удешевлением их компонентов, и, как следствие, остро встает вопрос об управлении качеством функционирования подобных систем.

К общим требованиям, предъявляемым к кластерным системам, относятся: высокая готовность, высокое быстродействие, масштабируемость.

Кластерная система обработки информации описывается:

- множеством состояний $\Theta[a_1, \dots, a_n]$, где $a_i \in A_i$ — состояние отдельного элемента системы, A — дискретное множество состояний, которые может принимать i -й элемент системы;
- целевой функцией $F(\Theta_j) = [F_1, \dots, F_m]$, где $\Theta_j \in \Theta$ — некоторое состояние системы; $[F_1, \dots, F_m]$ — вектор целевых показателей, характеризующий систему в целом;
- архитектурой кластерной системы.

В общем случае кластерную систему можно представить в виде графа, узлы которого представляют собой устройства сбора и обработки информации, а ветви — каналы передачи данных. Наиболее часто такой граф имеет древовидную структуру, представленную на рис. 1.

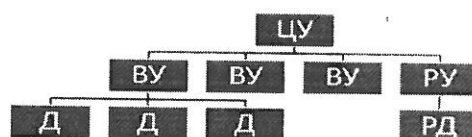


Рис. 1. Древовидная структура кластерной системы: D_1, \dots, D_n — датчики (устройства ввода информации); $РД$ — резервные датчики; $ВУ_1, \dots, ВУ_n$ — вычислительные узлы (устройства обработки информации); $РУ$ — резервные узлы; $ЦУ$ — центральный узел [1]

Архитектура системы находит отражение в целевой функции, так как целевые показатели вышестоящих элементов напрямую зависят от показателей нижестоящих элементов. Исходя из этого, состояние и целевые показатели i -го элемента будут являться функциями от состояний и показателей нижестоящих элементов:

$$a_i = \Theta[b_1, \dots, b_n],$$

где a_i — состояние i -го элемента; $\Theta[b_1, \dots, b_n]$ — функция перехода состояний, учитывающая весовой коэффициент j -го элемента, показывающий его важность для функционирования элемента верхнего уровня и системы в целом:

$$\Theta[b_1, \dots, b_n] = \frac{\sum_{j=1}^n k_j b_j}{n},$$

где $b_j \in B$ — состояние j -го элемента нижнего уровня; k_j — весовой коэффициент элемента.

Целевая функция состояния элемента a_i имеет вид

$$f(a_i) = \sum_{j=1}^n k_j f(b_j),$$

где $f(b_j)$ — целевая функция состояния для элементов нижнего уровня:

$$f(b_j) = [F_1, \dots, F_n].$$

* Статья выполнена членами РВО (Российского научно-технического вакуумного общества).

Рассмотрим целевые показатели обработки информации для кластерных информационных систем.

Показатели кластерной системы:

- Вероятность ложного срабатывания W — это результирующая вероятность программной ошибки в каждом физическом элементе системы.

- Коэффициент доступности (или работоспособности) системы P обусловлен вероятностью полной недоступности системы в связи с аппаратными или программными неполадками. Этот коэффициент выражает количественную меру работоспособности системы.

- Производительность системы обусловлена временем, которое затрачивается системой на решение эталонного задания.

- Время отклика это время, необходимое системе на обработку команды оператора или восприятие новой задачи.

Рассмотрим подробнее каждый целевой показатель.

Чтобы оценить эффективность вероятностных систем обработки информации на основе математического моделирования, можно использовать метод статистических испытаний. Для проведения таких испытаний может служить математическая модель функционирования системы, принципиальная схема которой представлена на рис. 2.

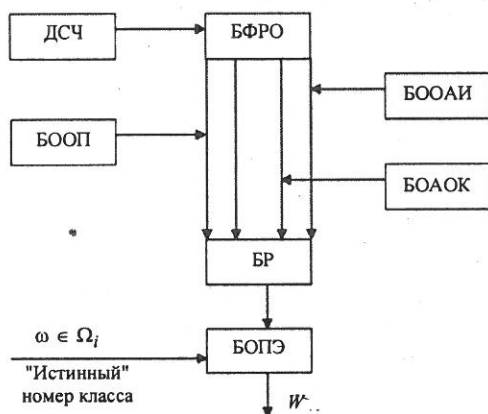


Рис. 2. Принципиальная схема системы обработки информации:

БРО — блок формирования распознаваемых объектов; БООП — блок ошибок определения признаков; БООАОК — блок ошибок априорного описания классов; БООАИ — блок ограничения объема апостериорной информации; БР — блок распознавания; БОПЭ — блок оценки показателя эффективности; ДСЧ — датчик случайных чисел

Принцип действия модели следующий. Для проведения каждого испытания с помощью ДСЧ формируется модель объекта, принадлежность которого к определенному классу заранее известна. Формирование модели объекта производится

заданием совокупности числовых значений признаков x_1, \dots, x_N , которые для объектов из класса Ω_i генерируются как реализации многомерной случайной величины с заданным законом распределения $f_i(x_1, \dots, x_N)$ по одному из известных алгоритмов.

Числовые значения параметров x_1, \dots, x_N , представляющие собой обрабатываемый объект, подвергаются случайному искажению, что имитирует результат воздействия различных помех в процессе определения признаков x_1, \dots, x_N при использовании соответствующих технических средств с определенными точностными характеристиками. Искаженные значения параметров x'_1, \dots, x'_N , представляющие наблюдаемый объект в том виде, в каком его воспринимает система, поступают на вход БР, в котором определяется принадлежность объекта одному из классов $\Omega_1, \dots, \Omega_N$. Блок БОПЭ сопоставляет номер класса, к которому отнесен объект блоком распознавания, с "истинным" номером, т. е. с тем, который задавался на первом этапе формирования объекта, определяет правильность обработки информации и систематизирует соответствующую информацию для подсчета оценок вероятностей верных и ошибочных решений. При обработке объектов из класса Ω_i оценкой p_i вероятности получения правильного решения служит отношение количества правильных ответов $N^i_{пр}$ к общему числу испытаний N^i над объек-

тами класса Ω_i , т. е. $p_i \approx \frac{N^i_{пр}}{N^i}$. Число испытаний

N^i определяется доверительной вероятностью, задаваемой при формулировке задачи исследования.

В зависимости от задачи исследования искажению могут подвергаться также априорные данные о классах объектов, т. е. функции распределения $f_i(x_1, \dots, x_N)$ и $P(\Omega_i)$, информация о признаках x_1, \dots, x_N может урезаться, что соответствует отсутствию некоторых средств определения признаков и т. п.

Если априорные вероятности $P(\Omega_i)$ появления объектов из разных классов известны, то безусловная вероятность правильного решения задачи обработки информации данной системой может быть выбрана в качестве критерия эффективности системы обработки информации:

$$W = \sum_{i=1}^n p_i P(\Omega_i).$$

Рассмотренная статистическая модель позволяет найти зависимость W от вида и количества привлекаемых для обработки признаков и точности $\sigma_1, \dots, \sigma_s$ технических средств, которыми оснащается система обработки информации, т. е.

$$W = W(x_1, \dots, x_N; \sigma_1, \dots, \sigma_s).$$

Сведения, содержащиеся в этом равенстве, — исходные для задач об определении состава технических средств наблюдений системы обработки информации, необходимых точностей их работы, об оптимальном, с точки зрения экономических соображений, распределении точностей по средствам и т. д.

Целевой показатель, коэффициент доступности кластерной системы, определяем с помощью метода соотношений. Суть метода сводится к определению вероятности безотказного функционирования сложной многоуровневой кластерной системы.

Процесс функционирования кластерной системы организован таким образом, что система успешно решает свои задачи при условии, если в исправном состоянии находится хотя бы одно устройство ввода информации, все устройства обработки информации и центральный узел. Данное условие выполнения целевой функции системы можно наглядно представить в форме логической функции:

$$\begin{aligned} F(KC) &= [F_1(D)] \wedge [F_2(BY)] \wedge F_3(CY); \\ F_1(D) &= F_1(D_1 \vee \dots \vee D_n); \\ F_2(BY) &= F_2(BY_1 \wedge \dots \wedge BY_n). \end{aligned} \quad (1)$$

Эти выражения означают, что устройство, указанное в скобках, работает исправно.

Представляет интерес также логическая зависимость, описывающая условия невыполнения системой своих целевых функций:

$$\begin{aligned} \neg F(KC) &= [\neg F_1(D)] \vee [\neg F_2(BY)] \vee \neg F_3(CY); \\ F_1(D) &= F_1(D_1 \wedge \dots \wedge D_n); \\ F_2(BY) &= F_2(BY_1 \wedge \dots \wedge BY_n). \end{aligned} \quad (2)$$

Последнее выражение может оказаться более удобным для решения поставленной задачи определения коэффициента доступности с учетом того, что

$$P[F(*)] = 1 - P[\neg F(*)],$$

где $P[F(*)]$ — вероятность истинности условия $F(*)$;

$P[\neg F(*)]$ — вероятность истинности отрицания истинности данного условия.

Перечисленные элементы КС имеют различное функциональное назначение и соединены так, что надежность каждого из них оказывает непосредственное влияние на работоспособность всей системы в целом. Поэтому в качестве факторов для оценки надежности функционирования КС следует взять вероятности P_i безотказного функционирования устройств в процессе решения системой поставленных задач. В общем случае вероятности P_i могут иметь различные значения. Вероятность $P[F(KC)]$ безотказного функционирования КС, в целом, есть функция от вероятностей безотказного функционирования всех ее элементов, вытекающая из рассмотренных выше логических условий.

Обобщенная схема математической модели, характеризующей безотказность функционирования КС, имеет вид, показанный на рис. 3.

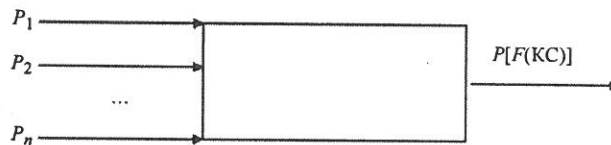


Рис. 3. Обобщенная схема математической модели, характеризующей безотказность функционирования КС

Проблема состоит в том, как из логических условий получить соответствующее выражение для количественного значения вероятности $P[F(KC)]$. Дело в том, что вероятность $P[F(KC)]$ определяется на множестве состояний $\Theta(t)$. Число состояний в данном множестве равно $2^n = N$, где n — число структурных элементов КС. Условие функционирования (1) определяет подмножество состояний системы, обеспечивающих выполнение системой заданных целевых функций, а условие (2) определяет подмножество состояний, в которых система оказывается неработоспособной.

Очевидно, для решения данной задачи таким способом придется осуществить полный перебор всех N -состояний системы или же придумать более эффективный способ определения работоспособных состояний, особенно если учесть, что число состояний системы находится в степенной зависимости от числа ее элементов. Наиболее перспективными методами решения этой проблемы представляются метод имитационного моделирования и формализованный переход от логических функций к соответствующим формулам вероятностей сложных событий.

Рассмотрим производительность систем обработки информации. Общая производительность кластерной системы обработки информации обусловлена производительностью каждого вычислительного элемента системы и определяется экспериментальным путем. Для этого каждому из элементов вычислительной системы дается эталонное задание и определяется время, затраченное на его решение. Исходя из затраченного на решение задачи времени, узлам назначаются весовые коэффициенты, характеризующие производительность вычислительного узла и системы в целом.

В реальных условиях производительность кластерной системы обработки данных зависит не только от производительности вычислительных узлов, но и от надежности и пропускной способности каналов передачи данных. Таким образом, ко времени, затрачиваемому на решение эталонного задания всей системой, добавляется время, не-

обходимое системе на подтверждение принятия задания, и время, затрачиваемое системой на передачу данных между вычислительными узлами к центру.

Производительность систем в реальных условиях можно вычислить следующим образом:

$$\Pi = \sum_{i=1}^M w_i + 2L,$$

где Π — производительность системы;

w_i — весовой коэффициент производительности вычислительного узла;

N — общее количество элементов системы;

M — количество вычислительных узлов;

L — время прохождения сигнала по каналам связи, определяемое по формуле

$$L = \min \left\{ \sum_{j=1}^{N-M} w_j \right\},$$

в которой w_j — весовой коэффициент пропускной способности канала связи.

Одним из важных параметров, описывающих кластерную систему обработки информации, является порог функционирования, т. е. такое значение целевой функции, при переходе через которое система перестает функционировать. Для относительно простых систем обработки информации это значение может быть получено экспериментальным или эмпирическим путем. Однако для систем с большим числом разнородных элементов это представляется затруднительным. Выходом в подобной ситуации может быть моделирование системы с использованием эволюционной стратегии, где критерием отбора будет являться наиболее функциональное состояние системы при максимальном количестве неисправностей. Эволюционные алгоритмы базируются на коллективном обучаемом процессе внутри популяции индивидов, каждый из которых представляет собой поисковую точку в пространстве допустимых решений данной задачи [2]. Наиболее известными из класса эволюционных алгоритмов являются генетические алгоритмы. Генетический алгоритм (рис. 4) может быть легко применен для безусловной оптимизации функций, т. е. для задачи отыскания значений параметров, которые минимизируют или максимизируют заданную целевую функцию, и для безусловной комбинаторной оптимизации, т. е. для задачи отыскания наилучшей комбинации вариантов, которая оптимизирует заданную целевую функцию. Их основные адаптивные процессы концентрируются на идее системы, получающей сенсорную информацию от окружающей среды через бинарные детекторы. В генетических алгоритмах существует строгое различие между фенотипом (решением) и генотипом (представлением решения).

Генетический алгоритм работает только с генотипом, поэтому требуется процесс декодирования генотипа в фенотип и обратно ("обобщенный" рост). Вещественные параметры могут быть представлены числами с фиксированной точкой или целыми числами путем масштабирования и дискретизации. Для вещественных параметров имеет место конфликт между желанием иметь как можно более короткий ген для обеспечения хорошей сходимости и необходимостью получить результат с определенной точностью.



Рис. 4. Генетический алгоритм

Комментарии к генетическому алгоритму

Шаг 1. Генерация начальной популяции. Случайным образом генерируется n уникальных состояний системы (индивидов), для каждого состояния вычисляется значение целевой функции и показатель работоспособности системы.

Шаг 2. Кодирование состояний системы в бинарный код (составление хромосом).

Шаг 3. Оценка пригодности каждого состояния. Для этого состояния ранжируются по значениям показателя работоспособности.

Шаг 4. Репликация состояний, т. е. генерация новой популяции: из $m < n$ состояний попарно генерируются потомки. В нашем случае потомком будет являться результирующее состояние, являющееся следствием событий состояний-родителей.

Шаг 5. Оценка пригодности всех состояний, включая потомков.

Шаг 6. Селекция. Для имитации естественной селекции состояния с более высокой пригодностью должны выбираться с большей вероятностью, поэтому из получившихся состояний выбирается n самых пригодных.

Шаг 7. Проверка конечного условия: если номер поколения $n_{\text{пок}}$ не равен заложенному на этапе инициализации конечному числу поколений $n_{\text{кон}}$, то увеличение $n_{\text{пок}}$ на единицу и переход на Шаг 4.

Если $n_{\text{пок}} = n_{\text{кон}}$, то переход на Шаг 8.

Шаг 8. Декодирование и отображение полученного результата.

Кластеры повышенной производительности

Кластеры повышенной производительности НРС (High performance cluster) позволяют увеличить скорость расчетов, разбивая задание на параллельно выполняющиеся потоки. Используются в научных исследованиях. Одна из типичных конфигураций — набор серверов с установленной на них операционной системой Linux. Такую схему принято называть кластером Beowulf. Для НРС создается специальное ПО, способное эффективно распараллеливать задачу.

Эффективные связи между серверами в кластере позволяют им поддерживать связь и оперативно обмениваться данными, поэтому такие кластеры хорошо приспособлены для выполнения процессов, использующих общие данные.

Физическая схема

- План стойки для каждого их типа (например, управляющие и вычислительные стойки).
- Поэтажный план расположения стоек во время процесса установки системы и при рабочем использовании, если они отличаются.
- Схемы внутренних соединений стоек для сети, цепей питания, пульта оператора и т. д.
- Схема внешних соединений для серверов системы хранения, терминальных серверов и т. д.

Логическая схема

- Схема сети, включая диапазоны IP-адресов, конфигурацию подсетей, соглашения по наименованию компьютеров и т. д.

- CSM-конфигурация по расположению пользовательских сценариев, аппаратные настройки и требования по мониторингу.

- Требования к операционным системам, список специализированных пакетов и параметры конфигурации системы.

- Схема системы хранения данных, включая схему файловой системы, разбиение дисков, параметры репликации и т. д.

Кластер (рис. 5) состоит из компьютеров, работающих на процессорах Intel или AMD, с подключенными подсистемами TotalStorage.

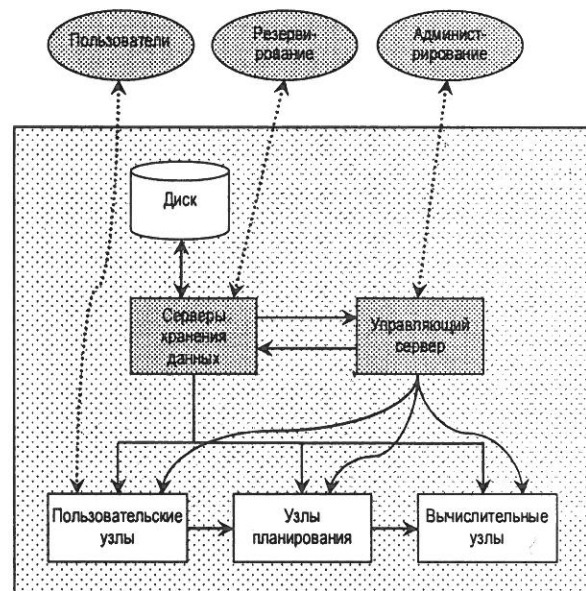


Рис. 5. Пример кластера

Для простоты соединения в кластере выполнены медным кабелем стандарта гигабитный Ethernet. Этот кабель обеспечивает хорошую скорость в большинстве случаев.

Сетевая топология имеет форму звезды — все стойки подключены к основному коммутатору управляющей стойки. В примере используются три сети: первая — для управления/данных (вычислительная сеть), вторая — для кластерной файловой системы (сеть хранения данных) и третья — для администрирования устройств. Первые две сети — это обычные IP-сети. Для большинства задач, включая межпроцессные взаимодействия (например MPI) и управление кластером, используется вычислительная сеть. Сеть хранения данных используется исключительно для доступа и взаимодействия с кластерной файловой системой.

Управляющий сервер

Функция управляющего сервера может выполняться одним сервером или несколькими. В среде

с одним сервером управляющий сервер функционирует в автономном режиме. Можно настроить также управляющие серверы с высокой готовностью. Для этого можно использовать программное обеспечение CSM для поддержки высокой готовности (high-availability — HA), которое будет выдавать тактовые импульсы ("heartbeat") между двумя серверами и поддерживать динамическое восстановление после сбоев при возникновении аварийных ситуаций. Другим возможным методом организации нескольких управляющих серверов является использование репликации, если поддержка HA не важна для вашей среды. В этом случае вы можете резервировать данные управляющего сервера на другую рабочую систему, которую можете при необходимости перевести в оперативный режим вручную. Управляющий сервер — это CSM-сервер, использующийся исключительно для внутреннего управления кластером при помощи CSM-функций: управление установкой системы, мониторинг, обслуживание и другие задачи. В данном кластере присутствует только один управляющий сервер.

Серверы хранения данных и дисковые накопители

Можно подключить несколько серверов хранения данных к организованному на дисковых накопителях хранилищу данных при помощи различных механизмов. Подключить систему хранения данных к серверу можно напрямую: либо через SAN-коммутатор (storage area network — сеть хранения данных) по оптическому волокну или медному кабелю, либо используя оба типа соединений (см. рис. 5). Эти серверы предоставляют совместный доступ к системе хранения данных другим серверам кластера. Если необходимо резервирование базы данных, подключите резервное устройство к серверу хранения данных, используя дополнительное медное или оптическое соединение. В примере кластера хранилище представляет собой единую сущность, обеспечивающую доступ к общей файловой системе в пределах кластера.

Пользовательские узлы

В идеальном случае вычислительные узлы кластера не должны принимать внешние подключения. Они должны быть доступны только для системных администраторов через управляющий сервер. Пользователи системы могут регистрироваться на вычислительных узлах (или узлах регистрации) для выполнения своей работы в кластере. Каждый пользовательский узел состоит из образа с возможностями любого редактирования,

необходимых библиотек разработчика, компиляторов и всего, что необходимо для создания кластерного приложения и получения реальных результатов.

Узлы планирования

Для запуска рабочей нагрузки на кластере пользователи должны передать свою работу узлу планирования. Фоновый процесс-планировщик (scheduler daemon), работающий на одном или нескольких узлах планирования, применяет predetermined политику для запуска рабочих нагрузок в кластере. Аналогично вычислительным узлам, узлы планирования не должны принимать внешних подключений от пользователей. Системные администраторы должны управлять ими при помощи управляющего сервера.

Вычислительные узлы

Эти узлы выполняют рабочую нагрузку кластера, принимая задания от планировщика. Вычислительные узлы — это самые свободные части кластера. Системный администратор может легко переустанавливать или перенастраивать их при помощи управляющего сервера.

Ethernet-коммутаторы

Имеются две физических сети: одна для вычислений, а вторая для хранения данных. Стандартная емкость стойки в 32 узла требует применения двух 48-портовых коммутаторов на каждую стойку, по одному на каждую сеть. В маленьких кластерах в управляющей стойке также необходимо использовать два одинаковых коммутатора. Для больших кластеров 48 портов может оказаться недостаточно, и может потребоваться более мощный центральный коммутатор.

Терминальные серверы играют важную роль в больших кластерных системах, использующих версии CSM, ниже чем CSM 1.4. Кластеры, использующие старые версии, нуждаются в терминальных серверах для сбора MAC-адресов при установке. При совместимости CSM и системных UUID терминальные серверы становятся не так важны для установки более современного кластера. Однако если в большом кластере у вас имеется немного устаревшее оборудование или программное обеспечение, терминальные сервера все еще остаются жизненно важными во время установки системы. Обеспечение корректной настройки терминального сервера само по себе может сэкономить значительное количество времени в дальнейшем в процессе установки системы. Кроме сбора MAC-адресов, терминальные серверы могут

также использоваться для просмотра терминалов из одной точки во время процедуры начального самотестирования (POST) и запуска операционной системы.

Когда компьютер получает DHCP-адрес во время PXE, конфигурационные файлы в /tftpboot/pxelinux.cfg ищутся в определенном порядке, и первый найденный файл используется в качестве загрузочной конфигурации для запрашивающего компьютера. Порядок поиска определяется путем преобразования запрашиваемого DHCP-адреса в шестнадцатиричные цифры и поиска первого подходящего имени файла в конфигурационном каталоге методом расширения подсетей — удаления цифр справа налево на каждом цикле поиска [3].

Коммуникационные библиотеки

Наиболее распространенным интерфейсом параллельного программирования в модели передачи сообщений является MPI. Рекомендуемая бесплатная реализация MPI — пакет MPICH, разработанный в Аргоннской Национальной Лаборатории. Для кластеров на базе коммутатора Myrinet разработана система HPVM, куда также входит реализация Message Passing Interface (MPI, интерфейс передачи данных).

MPI — программный интерфейс (API) для передачи информации, который позволяет обмениваться сообщениями между компьютерами, выполняющими одну задачу.

MPI является наиболее распространенным стандартом интерфейса обмена данными в параллельном программировании, существуют его реализации для большого числа компьютерных платформ. Основным средством коммуникации между процессами в MPI является передача сообщений друг другу. Стандартизацией MPI занимается MPI Forum. В стандарте MPI описан интерфейс передачи сообщений, который должен поддерживаться как на платформе, так и в приложениях пользователя. В настоящее время существует большое количество бесплатных и коммерческих реализаций MPI. Существуют реализации для языков Фортран 77/90, Си и Си++.

Стандарты MPI

Большинство современных реализаций MPI поддерживают версию 1.1. Стандарт MPI версии 2.0 поддерживается большинством современных реализаций, однако некоторые функции могут быть реализованы не до конца.

В MPI 1.1 поддерживаются функции:

- передача и получение сообщений между отдельными процессами;
- коллективные взаимодействия процессов;
- взаимодействия в группах процессов;
- реализация топологий процессов;

В MPI 2.0 дополнительно поддерживаются функции:

- динамическое порождение процессов и управление процессами;
- односторонние коммуникации;
- параллельный ввод и вывод.

Для эффективной организации параллелизма внутри одной SMP-системы возможны два варианта:

1. Для каждого процессора в SMP-машине рождается отдельный MPI-процесс. MPI-процессы внутри этой системы обмениваются сообщениями через разделяемую память (необходимо настроить MPICH соответствующим образом).

2. На каждой машине запускается только один MPI-процесс. Внутри каждого MPI-процесса производится распараллеливание в модели "общей памяти", например, с помощью директив OpenMP.

После установки реализации MPI имеет смысл протестировать реальную производительность сетевых пересылок.

Кроме MPI, есть и другие библиотеки и системы параллельного программирования, которые могут быть использованы на кластерах.

Следует понимать, что использование для программирования OpenMosix — расширение (патч) ядра Linux, позволяющее создать единый кластер, превращает сеть обычных персональных компьютеров в суперкомпьютер для Linux-приложений. Представляет собой полнофункциональную кластерную среду с единой операционной системой (SSI), автоматически распараллеливающую задачи между однородными узлами. Это позволяет миграцию процессов между машинами — узлами сети.

Кластер ведет себя подобно SMP-машине (за исключением любых видов разделяемой памяти). При этом возможно наращивание до тысяч узлов, которые тоже могут быть SMP-машинами. Добавление новых узлов возможно параллельно работе кластера, добавленные ресурсы будут задействованы автоматически. OpenMosix также предоставляет оптимизированную файловую систему (OMFS) для HPC-приложений, которая в отличие от NFS поддерживает кэширование, отметки о времени и ссылки.

OpenMosix — это проект, являющийся продолжением проекта MOSIX, но под свободной лицензией GPL. Последние релизы MOSIX стали про-

приетарными (закрытыми) в конце 2001 г., а проект openMosix стартовал 10 февраля 2002 г. Инициатор проекта — Moshe Bar.

OpenMosix поставляется с набором утилит для администрирования кластера. Для этого имеется также удобное GUI-приложение openMosixview.

При обработке результатов сканирующей зондовой микроскопии часто встает вопрос о принадлежности объекта исследования тому или иному классу объектов. Подобные задачи решаются применением систем распознавания.

Распознавание представляет собой задачу преобразования входной информации, в качестве которой уместно рассматривать некоторые параметры, признаки распознаваемых объектов, в выходную, представляющую собой заключение о том, к какому классу объектов принадлежит распознаваемый объект.

Важнейшей характеристикой принимаемых в процессе распознавания решений и основным показателем качества распознавания является достоверность [4, 5].

В многоуровневых системах распознавания апостериорная информация о признаках определяется на основе косвенных измерений. Для таких измерений используются специализированные локальные системы распознавания. По данным технических средств $T_1, \dots, T_p, \dots, T_n$ определяются признаки $x_1^1, \dots, x_1^k; x_p^1, \dots, x_p^j; x_n^1, \dots, x_n^m$ (первичные признаки), которые используются локальными системами распознавания для определения признаков более высокого уровня, которые, в свою очередь, используются в процессе распознавания неизвестных объектов (рис. 6) [6, 7].

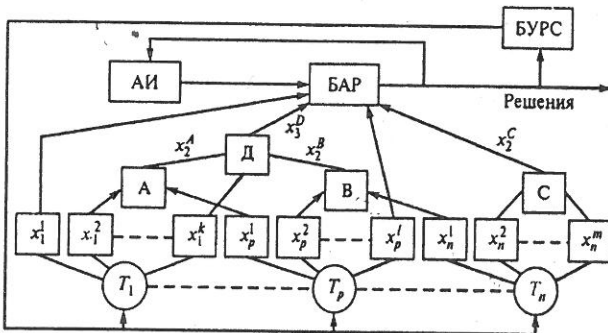


Рис. 6. Сложная многоуровневая система распознавания

Оценка достоверности

Оценка достоверности результатов моделирования может быть произведена различными методами, в частности, непараметрическим.

Непараметрический метод выявления отклонений заключается в ранжировании анализируемых значений $X_1 \leq X_2 \leq \dots \leq X_n$ и вычислении статисти-

стики r для крайних значений [8]. Формулы расчета зависят от числа анализируемых значений n и от того, значение с какого края проверяется на аномальность (самое большее или меньшее) (таблица).

Формула для работы отклонений

Число значений	Проверяется максимальное значение X_n	Проверяется минимальное значение X_1
$3 \leq n \leq 7$	$r10 = (X_n - X(n-1)) / (X_n - X_1)$	$(X_2 - X_1) / (X_n - X_1)$
$8 \leq n \leq 10$	$r11 = (X_n - X(n-1)) / (X_n - X_2)$	$(X_2 - X_1) / (X(n-1) - X_1)$
$11 \leq n \leq 13$	$r20 = (X_n - X(n-2)) / (X_n - X_2)$	$(X_3 - X_1) / (X(n-1) - X_1)$
$14 \leq n \leq 25$	$r21 = (X_n - X(n-2)) / (X_n - X_3)$	$(X_3 - X_1) / (X(n-2) - X_1)$

Полученное значение r сравнивается с критическим r и считается аномальным, если $r > ar$.

После этого процедура проверки повторяется для оставшихся $n - 1$ значений.

Данный критерий оптимален для малых серий наблюдений (данных) и не зависит от числа имеющихся аномальных значений. В то время, как приведенный ранее параметрический критерий (особенно для небольших серий) оптимален, когда имеется лишь одно аномальное значение.

Рассмотренная математическая модель в совокупности с эволюционной стратегией позволяет оценивать критические ситуации для кластерных систем обработки информации и выявлять их последствия, а также моделировать и оптимизировать адаптивными методами показатели качества подобных систем.

Литература

1. Кравчук И. С., Тихоглаз Ю. С., Занг Н. Ч. Математическая модель и алгоритм управления качеством в кластерных системах сбора и обработки информации // Системы управления и информационные технологии. 2008. Вып. 1.2 (31). С. 299—303.
2. Кравчук И. С., Тихоглаз Ю. С., Ву Тхи Туэт Ланг. Эволюционная стратегия управления в задачах распознавания образов // Там же. Вып. 2.3 (32). С. 358—360.
3. Слободин М. Ю., Царев Р. Ю. Компьютерная поддержка многоатрибутивных методов выбора и принятия решения при проектировании корпоративных информационно-управляющих систем. — СПб.: Инфо-да, 2004. — 223 с.
4. Вальков В. М., Вершин В. Е. Автоматизированные системы управления технологическими процессами. — М.: Машиностроение, 1977.
5. Кравченко В. А., Цидилин С. М., Федосеева Т. Л. Алгоритмы решения задач многокритериальной оптимизации: Учеб. пособие. — М.: Изд-во МИЭМ, 1988. — 74 с.
6. Бусленко Н. П. Моделирование сложных систем. — М.: Наука, 1978.
7. Резников Б. А. Методы и алгоритмы оптимизации на дискретных моделях сложных систем. — Л.: ВИКИ им. А. Ф. Можайского, 1983. — 215 с.
8. Рапопорт Э. Я. Анализ и синтез систем автоматического управления с распределенными параметрами: Учеб. пособие. — М.: Высш. шк., 2005. — 292 с.

Specificity of recognition in the cluster systems of information processing

V. A. Vasin, E. N. Ivashov, S. V. Stepanchikov

Moscow Institute of Electronics and Mathematics NRU Higher School of Economics, Russia

In this paper we consider a mathematical model of the cluster systems of information processing, algorithms and methods for modeling of such systems over time, as well as the method for determining the threshold for functioning of the system using evolutionary strategy.

Keywords: cluster systems of information processing, pattern recognition, clusters of high performance, evaluation of the reliability, evolutionary strategy, the performance of data processing systems.

Васин Владимир Анатольевич, старший научный сотрудник.

Ивашов Евгений Николаевич, профессор.

Степанчиков Сергей Валентинович, доцент.

E-mail: ienmiem@mail.ru, vacuumwa@list.ru

Статья поступила в редакцию 16 июля 2012 г.