

КОНЦЕПТУАЛЬНЫЙ ПОДХОД К ПОВЫШЕНИЮ ЭФФЕКТИВНОСТИ ИСПОЛЬЗОВАНИЯ ВЫЧИСЛИТЕЛЬНЫХ РЕСУРСОВ КОРПОРАТИВНЫХ СЕТЕЙ ПРИ ПРИМЕНЕНИИ ТЕХНОЛОГИИ ВИРТУАЛЬНЫХ МАШИН

П.А. Рахман

(Москва, Московский Энергетический Институт, Российская Федерация)

На сегодняшний день существует большое множество организаций, имеющих корпоративную сеть, состоящую из множества конечных рабочих мест пользователей и некоторого, так называемого серверного парка. Под серверным парком будем понимать множество физических компьютеров, связанных сетью передачи данных, и множество функционирующих на них логических серверов. Под физическим компьютером будем понимать некоторую ЭВМ, состоящую из реально существующих компонент ее аппаратного обеспечения и предоставляющую некоторые вычислительные ресурсы. Под логическим сервером будем понимать серверную операционную систему со всеми работающими под ее управлением службами, причем в любой момент времени на одном физическом компьютере может функционировать только один логический сервер. Серверный парк предоставляет широкий спектр сервисов [1]: доступ в Интернет, корпоративная почта, антивирусная защита, файловые ресурсы, службы печати и многое другое. Как правило, изначально корпоративная сеть грамотно проектируется специалистами с учетом надежности, безопасности и многофункциональности и руководство организаций крайне отрицательно относится к внесению значительных или даже небольших изменений в сеть, которая уже много лет исправно функционирует и удовлетворяет всем требованиям. Однако, тем не менее руководство всегда интересуют возможности снижения затрат на содержание сети и специалистов, обслуживающих ее, а также получения дополнительной прибыли с используемого технического оборудования. Для конечных рабочих мест характерно то, что, как правило, для них выделяются компьютеры, которые в той или иной степени уступают по техническим параметрам компьютерам, используемым в серверном парке. Кроме того, пользователи используют множество приложений, которые могут на 100% загружать

процессор, “съесть” всю оперативную память, помимо этого, пользователи часто размещают на своих персональных компьютерах данные, не относящиеся к работе, которые могут занимать все дисковое пространство. Наконец, пользователи вправе считать и требовать, что все ресурсы их рабочего компьютера принадлежат задачам и приложениям, используемыми ими. В таких условиях, практически отсутствует возможность и, главное, целесообразность повышения эффективности использования ресурсов на рабочих компьютерах. Использование серверов приложений, когда множество пользователей использует один мощный сетевой вычислительный ресурс для запуска своих приложений, частично решает проблему, но как, показывает практика, рабочие компьютеры все равно остаются достаточно сильно нагруженными. Наконец, по элементарным соображениям информационной безопасности недопустимо размещение каких-либо серверных функций или чужих приложений или данных на пользовательском компьютере. Что же касается, компьютеров серверного парка, то многолетняя практика эксплуатации серверных операционных систем и сетевых служб показала то, что на сегодняшний день большинство компьютеров серверного парка достаточно слабо загружены по ресурсам. Такая ситуация сложилась по следующей причине: рынок аппаратных решений развивается стремительно и производители оборудования очень быстро отказываются поддерживать старое оборудование, которое можно было бы эффективно использовать для размещения некоторых серверных служб. Так, например, один из важнейших элементов хорошо защищенной корпоративной сети – контроллер домена, по сегодняшним меркам крайне малотребователен к ресурсам: 2-3% среднесуточной загрузки процессора класса *Pentium*, сетевой трафик – несколько десятков или сотен килобайт данных, которые передаются не постоянно, а однократно через определенные периоды (от 15 минут до 1 часа или даже реже), памяти требуется не более 64 МБ, дискового пространства требуется порядка 1-1.5 Гб. В тоже время, по соображениям безопасности, на контроллере домена крайне не рекомендуется размещать какие-либо иные серверные службы (файловые ресурсы, WEB-серверы и т.д.), то есть одному лишь контроллеру со столь низкими требованиями необходим целый компьютер. Конечно, можно было бы подобрать адекватную конфигурацию компьютера для такого малотребовательного сервера, однако, тот же жесткий диск емкостью 1.5-2 Гб на сегодняшний день найти достаточно сложно, а новые диски такой емкости не производятся уже давно. Использовать старые диски 10-летней давности крайне неразумно и чревато: они могут выйти из строя в любой момент, а гарантии и техническая поддержка на них отсутствует. Наконец, любой разумный и ответственный человек вряд ли захочет связываться со столь рискованным и устаревшим оборудованием. В таких условиях, руководство фирмы вынуждено приобретать для сервера современные диски емкостью от 40 Гб и выше, модули памяти емкостью от 256 МБ, процессоры класса *Pentium IV* как минимум, прекрасно осознавая, что большая часть каждого из ресурсов сервера будет безнадежно простаивать.

Соответственно, необходимы какие-либо подходы к решению проблемы неэффективного использования ресурсов компьютеров серверного парка. Под эффективностью будем понимать отношение денежных средств, получаемых за счет решения серверным парком некоторого множества полезных задач к совокупной стоимости владения серверным парком. Под множеством полезных задач будем понимать множество задач, функций и сервисов, решение (выполнение, предоставление) которых либо приносят определенный доход, либо снижают определенные виды затрат. Совокупная стоимость владения включает в себя стоимость оборудования, стоимость программного обеспечения, стоимость первоначальной настройки, стоимость аренды помещений, затраты на потребляемую электроэнергию, оплата услуг обслуживающих серверный парк специалистов, затраты на ремонт и замену устаревших компонентов, и она напрямую зависит от объема эксплуатируемого оборудования. Очевидно, что эффективность использования ресурсов можно повысить, либо увеличив объем решаемых “полезных” задач, дающих доход или снижающих какие-либо затраты, либо уменьшив объем оборудования.

Целью работы является анализ существующих подходов к решению проблемы неэффективного использования ресурсов компьютеров и разработка эффективного подхода к решению поставленной проблемы. Задача повышения эффективности использования ресурсов в каждом конкретном случае решается с учетом особенностей и специфики, присущей корпорации.

На сегодняшний день существуют 5 основных подходов к решению поставленной задачи:

- Использование ресурсов для дублирования функций логических серверов или решения дополнительных задач, приносящих прибыль.
- Использование ресурсов для задач сторонних организаций.
- Объединение служб и программного обеспечения разных логических серверов с целью снижения количества логических серверов.
- Применение адекватных аппаратных решений.
- Применение технологии виртуальных машин.

Основным недостатком первого подхода является недостаточная изоляция дополнительных приложений от основных, что порождает проблемы дополнительных уязвимостей с точки зрения информационной безопасности и совместимости приложений. Эти проблемы возникают потому, что в серверную операционную систему некоторого компьютера к существующим приложениям добавляются дополнительные приложения либо службы, дублирующие службы других серверов. Когда в пределах одной ОС оказывается множество работающих приложений, то вопросы безопасности ОС в целом и совместимости приложений неизбежны.

Подход, связанный с использованием вычислительных ресурсов для задач сторонних организаций, имеет те же недостатки, что и первый, и при этом более обостряется проблема информационной безопасности из-за повышения возможности несанкционированного доступа к данным и нарушения функционирования сервисов корпорации, предоставляющих ресурсы в аренду.

Подход, связанный с объединением сервисов, помимо появления проблем совместимости и информационной безопасности, также сказывается на логической структуре сетевой инфраструктуры, и это, как правило, влечет дополнительную работу по перенастройке рабочих мест пользователей. Это происходит потому, что сетевые имена присваиваются не к компьютерам, а операционным системам, функционирующим на них, и перемещение служб с одного компьютера на другой меняют привязку служб от одного сетевого имени сервера к другому.

Подход, связанный с подбором адекватных аппаратных решений, практически нереализуем в условиях современного рынка компьютерного оборудования, поскольку крайне трудно даже приближенно подбирать компоненты компьютеров с учетом требований логических серверов. Кроме того, при использовании старых компонент для логических серверов с невысокими требованиями может существенно снизиться надежность функционирования компьютеров.

Технология виртуальных машин [2] предоставляет новые возможности для построения нового или реорганизации существующего серверного парка. Технология виртуальных машин – это специальная технология, позволяющая с помощью специального программного обеспечения эмулировать виртуальные машины на базе аппаратного обеспечения физического компьютера и обеспечивать их параллельное (псевдопараллельное) функционирование. Виртуальная машина – это некоторая ЭВМ, полноценно эмулируемая некоторым специальным программным обеспечением и состоящая, как из полностью эмулируемых компонент, так и компонент, отображаемых на реально существующие компоненты некоторого физического компьютера. Виртуальная машина функционирует как процесс под управлением некоторой многозадачной операционной системы, называемой базовой ОС, на некотором физическом компьютере. Виртуальная машина предоставляет вычислительные ресурсы, которые в действительности являются частями ресурсов физического компьютера. Также как и на физическом компьютере на виртуальной машине в любой момент времени может функционировать только один логический сервер, но на физическом компьютере возможна параллельная (псевдопараллельная) работы нескольких виртуальных машин. Уровень изоляции виртуальных машин с точки зрения совместимости приложений и информационной безопасности ничуть не хуже, чем у отдельных физических компьютеров. Соответственно, технология виртуальных машин может обеспечить функционирование нескольких изолированных логических серверов на одном компьютере и тем самым обходить проблемы безопасности, совместимости, изменения привязок сервисов к логическим серверам, а также избавляет от необходимости подбора адекватных аппаратных решений, поскольку размещение нескольких логических серверов позволяет существенно повысить эффективность использования ресурсов компьютера. Таким образом, серверный парк со слабой загрузкой ресурсов может быть реорганизован, что в конечном счете должно привести к уменьшению объема используемого оборудования и затрат на его поддержку.

Однако, на сегодняшний день отсутствуют какие-либо рекомендации или подходы к планированию нового или реорганизации существующего серверного парка с целью повышения эффективности использования вычислительных ресурсов при применении технологии виртуальных машин. Реорганизация серверного парка с целью повышения эффективности использования ресурсов компьютеров – это сложная работа, требующая предварительного проектирования с использованием определенных технологий и подходов. В современных условиях жесткой конкуренции реорганизация серверного парка – это далеко не только сбор предварительной информации, постановка задачи реорганизации и ее решение одним из известных методов, но и максимальный учет требований и пожеланий заказчика IT-услуг. Под заказчиком мы будем понимать не только главу какой-либо коммерческой организации, но и, в общем случае, некоторых его IT-специалистов: системных администраторов, аналитиков, экспертов – на которых глава опирается в технических вопросах и которые, в конечном счете, будут оценивать качество выполненных работ и в дальнейшем иметь дело с реорганизованным серверным парком. Как упоминалось ранее, в любой коммерческой организации присутствует своя специфика, которую лучше всего может знать только заказчик, и только лишь тесно взаимодействуя с ним можно получить не только хорошее решение, но и решение, которое будет максимально отвечать интересам заказчика и оправдывать его ожидания.

Таким образом, мы вплотную подходим к тому, что необходимо предварительно оценить целесообразность проведения реорганизации серверного парка, а заказчик, взвесив все “за” и “против” примет окончательное решение о проведении работ по реорганизации. Отсюда следует, что задача реорганизации должна разбиваться на два этапа:

- I) Анализ задачи, сбор первичной информации, разработка первичного проектного решения, оценка качества решения с учетом возможных непредвидимых ситуаций, анализ выгоды, которое дает это решение, и оценка целесообразности проведения реорганизации. Цель первого этапа – максимально обезопасить проектное решение от провала на втором этапе – этапе реализации проектного решения. На первом этапе не допускается внесение каких-либо изменений в серверный парк.
- II) Непосредственная реализация первичного проектного решения, выявление негативных последствий реорганизации. В случае неудовлетворительного качества работы серверного парка – корректировка решения и реализация скорректированного решения, далее повторяется анализ качества. Если на каком-либо шаге корректировка невозможна в силу неприемлемого снижения коммерческой выгоды – то поиск компромиссных решений либо отказ от проекта с возвратом серверного парка в исходное состояние – вероятность избежания такого исхода напрямую зависит от качества анализа на первом этапе.

Рассмотрим подробнее оба этапа.

I) На первом этапе выполняются следующие шаги:

- 1) Сбор первичной информации по серверному парку: получение данных по имеющимся ресурсам физических компьютеров и требованиям программного обеспечения, а также анализ загрузки ресурсов. Учет вопросов по надежности функционирования: выделение логических серверов, дублирующих функции друг друга. Выделение логических серверов, для которых нет дублирующих серверов и в силу их важности желательно создание таких дублей. Составление дополнительных ограничений для задачи распределения, исключающих в дальнейшем размещение виртуальных машин, содержащих логические серверы дублирующих функции друг друга, на один и тот же физический компьютер. Выбор базовой ОС и определение ее требований к ресурсам. Выбор тех типов ресурсов, по которым предпочтительно решение проблемы их неэффективного использования.
- 2) Получение с помощью программного инструмента первичного распределения логических серверов по физическим компьютерам (поиск оптимального распределения по первичным исходным данным).
- 3) Первичный анализ распределения: удаление “бесперспективных” физических компьютеров и логических серверов. “Бесперспективным” считается пара “компьютер – логический сервер” в случае если:
 - Логический сервер в соответствии с распределением размещается на тот же компьютер, на котором он работал изначально, при этом на этот компьютер другие логические серверы не размещаются.
 - Логический сервер в соответствии с распределением никуда не размещен, а компьютер, на котором изначально работал этот логический сервер, оказывается незадействованным.
- 4) Вторичный анализ распределения: прогноз возможных корректировок распределения на случай, если реорганизованный серверный парк не будет отвечать техническим требованиям и пожеланиям заказчика, прогноз количества освобождаемого оборудования с учетом возможных корректировок.
- 5) Предоставление результатов анализа и прогнозов заказчику для оценки выгоды и получение окончательного решения о целесообразности проведения работ по реорганизации.
- 6) Завершение первого этапа.

В случае если на первом этапе было принято положительное решение, то вступает в действие второй этап.

II) На втором этапе выполняются следующие шаги:

- 1) Реализация решения – первичная реорганизация серверного парка в соответствии с полученным на первом этапе распределением: проведение работ по резервному копированию данных для возможности “отката” к исходному состоянию в наихудшем случае (когда, несмотря на самые “радужные” прогнозы на первом этапе, внедрение решения приводит к неудовлетворительной работе серверного парка и внесение всех

возможных корректировок не исправляет ситуации и приводит к тому, что последняя корректировка требует столько же оборудования, сколько было до реорганизации или даже больше или количество освободившегося оборудования снижается настолько, что заказчику уже такой результат просто невыгоден). Подготовка логических серверов к переносу на виртуальную аппаратную платформу. Подготовка задействованных в распределении компьютеров: развертывание базовой ОС, подготовка виртуальных машин. Развертывание логических серверов на виртуальных машинах физических компьютеров в соответствии с первичным распределением.

- 2) Оценка качества работы реорганизованного серверного парка: оценка времени отклика серверов, скорости обработки запросов по ключевым сервисам и т.п. Если все сервисы функционируют и заказчика удовлетворяет время отклика серверных системы, время обработки запросов, то реорганизация считается успешно завершенной и переход к шагу 9. В противном случае переход к шагу 3.
- 3) Выполняется поиск логических серверов, функционирование которых не отвечают тем или иным требованиям. Анализ причин их неудовлетворительной работы и выделение тех типов ресурсов, имеющиеся уровни по которым недостаточны. Выделение логических серверов, для которых размещение на одном компьютере создает проблемы. Повышение требований “проблемных” логических серверов к ресурсам, внесение дополнительных ограничений на одновременное размещение логических серверов на физических компьютерах.
- 4) Получение с помощью программного инструмента оптимального распределения логических серверов по физическим компьютерам по скорректированным исходным данным.
- 5) Первичный анализ скорректированного распределения: удаление “бесперспективных” физических компьютеров и логических серверов. Данная операция проводится совершенно аналогично соответствующей операции первого этапа реорганизации.
- 6) Предоставление результатов анализа заказчику для оценки выгоды и получение окончательного вывода о целесообразности реализации скорректированного решения.
- 7) Если заказчик принимает положительное решение, то проведение коррекционной реорганизации: перераспределение логических серверов с привлечением дополнительных (из числа освободившихся ранее) физических компьютеров в соответствии со скорректированным распределением и переход к шагу 2. В противном случае – к шагу 8.
- 8) Поиск компромиссных вариантов с заказчиком, в худшем случае “откат” к исходному состоянию серверного парка и прекращение работ.
- 9) Завершение второго этапа.

На рисунках 1 и 2 приведены схемы алгоритмов для первого и второго этапа реорганизации серверного парка

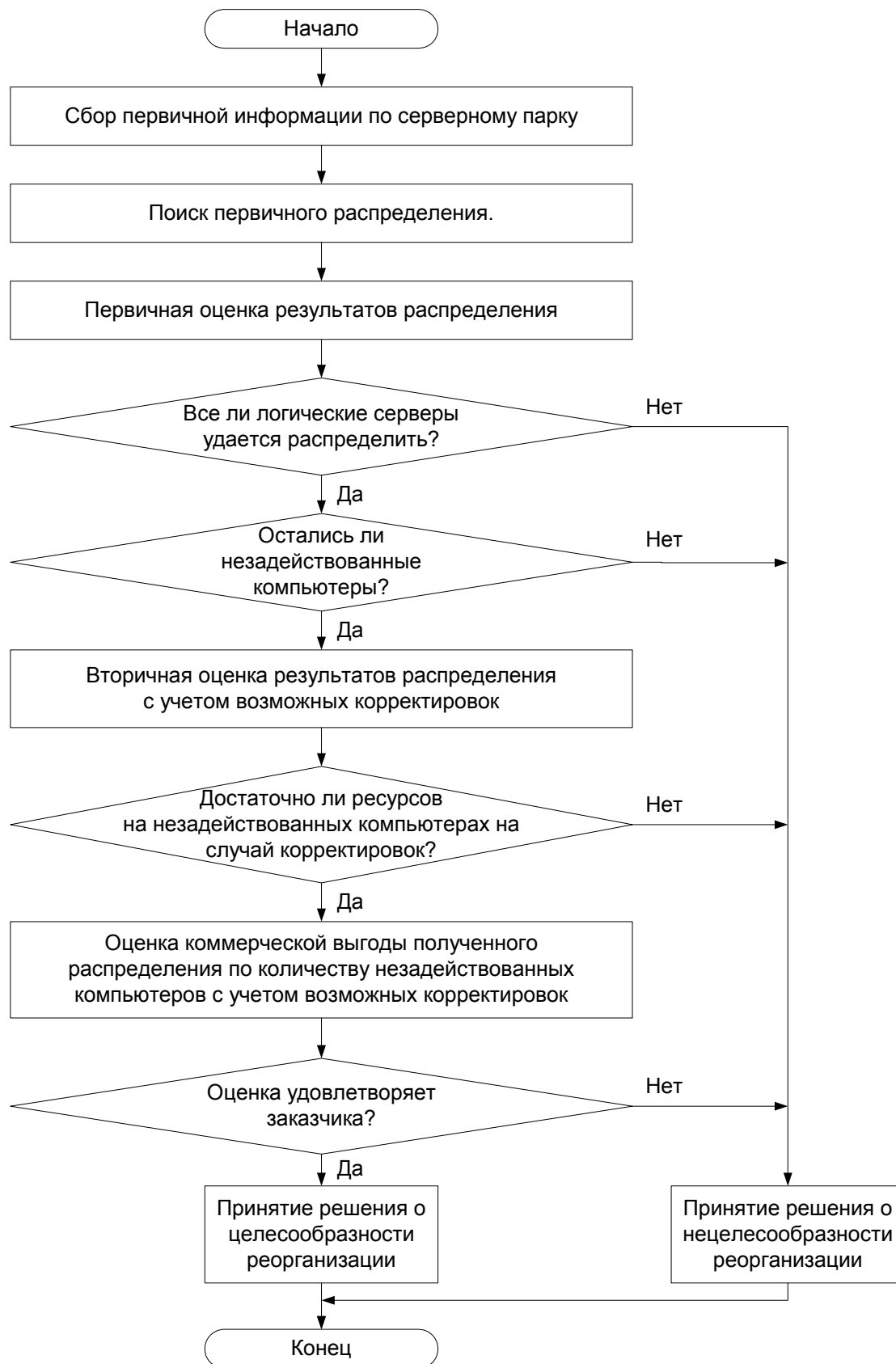


Рис 1. Схема алгоритма для первого этапа реорганизации

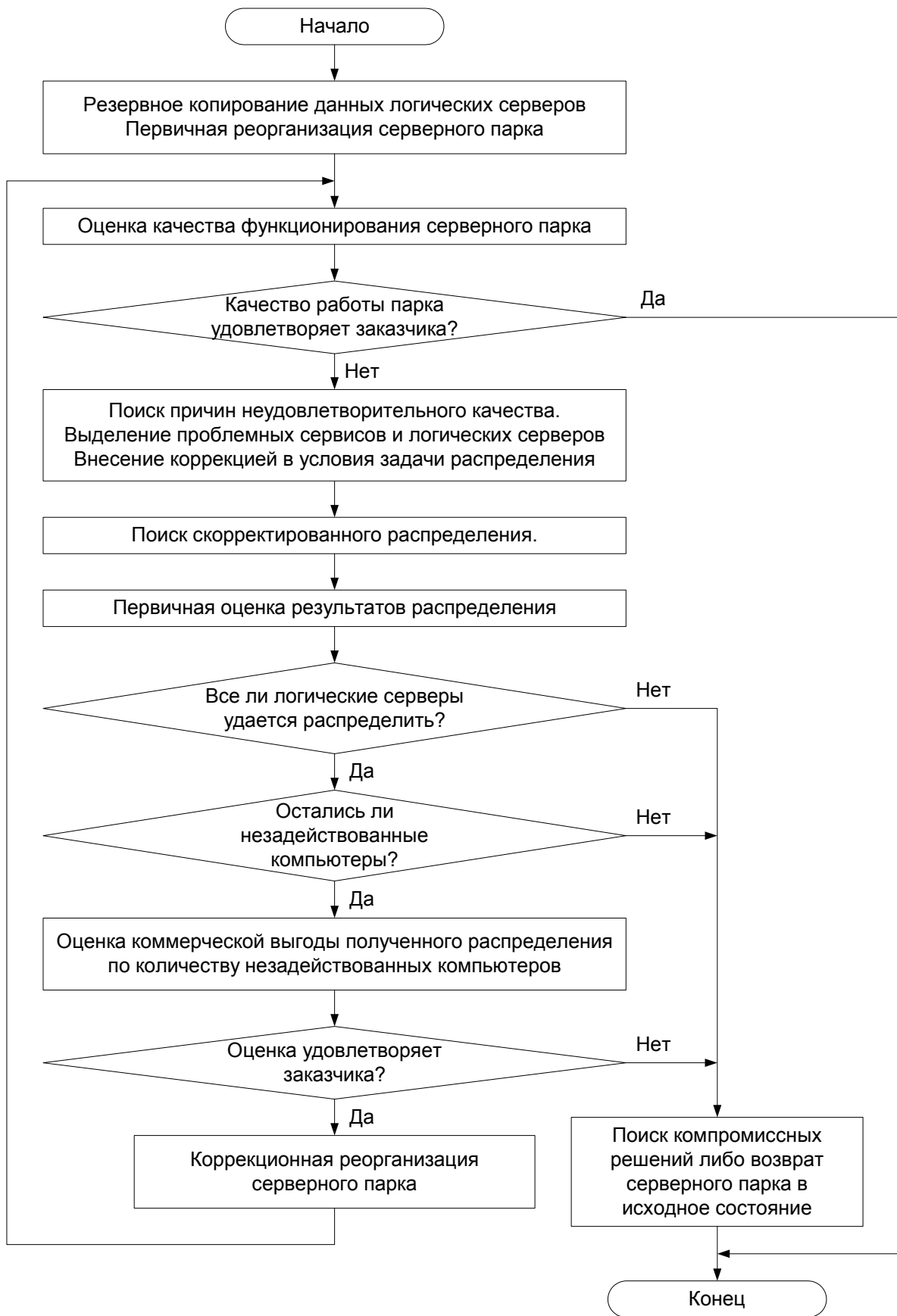


Рис 2. Схема алгоритма для второго этапа реорганизации

Следует отметить, что алгоритмы на рисунках 1 и 2 состоят из “укрупненных” блоков, которые в свою очередь требуют дальнейшей детализации. Однако в рамках данной статьи нет возможности подробно рассмотреть все блоки, поэтому остановимся лишь на блоке поиска первичного (скорректированного) распределения. В основе этого блока лежит математический метод дискретной оптимизации для решения задач распределения ресурсов.

Задача поиска оптимального распределения является достаточно трудоемкой, несмотря на то, что поиск ведется в дискретном булевом пространстве: конкретная виртуальная машина может размещаться либо не размещаться на конкретном компьютере. Однако именно дискретность и усложняет задачу, поскольку существующие точные методы оптимизации, такие как симплекс-метод, рассчитаны на поиск решений в непрерывном пространстве, а целочисленные модификации симплекс-метода имеют определенные сложности из-за погрешностей машинной арифметики, кроме того они предназначены для решения задач небольших размерностей, поскольку в определенных случаях не защищены от полного перебора целочисленных решений. Соответственно, точное решение при решении задач оптимизации в булевом пространстве можно получить только полным перебором. При NS виртуальных машин и NH компьютеров объем полного перебора составляет $(NH+1)^{NS}$, очевидно, что при больших размерностях задачи, поиск решения за приемлемое время невозможен. Соответственно, в данной ситуации используются приближенные методы. В первом приближении, самое простое – это последовательное размещение логических серверов по компьютерам, при рассмотрении каждого компьютера по очереди. При этом для каждого компьютера в отдельности решается задача булевого линейного программирования для нахождения оптимального набора среди еще нераспределенных логических серверов. Таким образом, задача разбивается на два уровня: локальный уровень – повышение эффективности использования ресурсов конкретного компьютера, и глобальный уровень – простое последовательное рассмотрение компьютеров без каких-либо критериев или приоритетов выбора компьютера. Общий перебор при решении задачи распределения в худшем случае составляет $NH*2^{NS}$, в случае, если на локальном уровне подзадачи решаются полным перебором. Однако при таком подходе совсем не проводится оптимизация на глобальном уровне – компьютеры рассматриваются в произвольном порядке и не учитывается то, что выбор каких-либо конкретных компьютеров мог бы быть более лучшим вариантом нежели чем выбор других. Поэтому здесь требуется некоторое компромиссное решение между полным перебором и простым последовательным размещением логических серверов на каждый компьютер до тех пор, пока все серверы не будут распределены. Соответственно, вместо простого последовательного рассмотрения компьютеров, можно на каждом шаге глобального уровня выбирать не произвольный, а наилучший по загрузке ресурсов компьютер. На каждом шаге глобального уровня решаются подзадачи булевого линейного программирования для всех еще незадействованных

компьютеров для размещения на них набора из оставшихся нераспределенными логических серверов, и выбирается наилучший компьютер. Таким образом, если на локальном уровне подзадачи решаются полным перебором, то общий объем перебора при решении задачи распределения составит приблизительно $NH \cdot 2^{NS+1}$. На локальном уровне подзадачи представляют собой задачу булевого линейного программирования (задачу условной псевдобулевой оптимизации), для которой точное решение можно получить только полным перебором. Для решения задач псевдобулевой оптимизации на сегодняшний день существуют как точные методы решения, так и приближенные методы. Точные методы – это известные целочисленные модификации симплекс-метода [3]: метод сечений Гомори и метод ветвей и границ. Такие методы призваны находить точное решение, однако, на реальных практических задачах в силу сложных ограничений, приводят к бесконечному добавлению и удалению дополнительных сечений в методе Гомори и к полному перебору в методе ветвей и границ. Поэтому точные методы на практике пригодны для решения задач небольших размерностей. Приближенные методы [4]: локальный поиск, случайный поиск, гриди-методы, а также применение одновременно нескольких приближенных методов, например, задача решается методом локального поиска, но не однократно, а много раз для множества различных стартовых точек, выбираемых случайным образом. Такие методы позволяют решить задачи больших размерностей с линейными ограничениями любой сложности, причем за приемлемое время, однако, как правило, дают приближенное (субоптимальное) решение. Среди методов для решения подзадач на локальном уровне распределения метод локального поиска является одним из наиболее эффективных при использовании нескольких случайных стартовых точек.

В рамках научных исследований [5] автором был разработан специальный комбинированный метод, состоящий из двух уровней оптимизации: глобальной оптимизации на уровне всего множества компьютеров и локальной оптимизации на уровне размещения определенного подмножества виртуальных машин на конкретном компьютере. На локальном уровне используется метод локального поиска оптимума с использованием функции штрафов, множества случайных стартовых точек и управляемого радиуса зоны локального поиска. На глобальном уровне используется упрощенная схема перебора, выделяющая наилучший компьютер для размещения на нем подмножества виртуальных машин на заданном шаге алгоритма поиска на глобальном уровне. Автором также была разработана программная реализация алгоритма для решения задачи поиска оптимального распределения логических серверов по физическим компьютерам.

Соответственно, имея под рукой программную реализацию для решения задачи поиска оптимального распределения и реализацию вышеизложенного концептуального подхода к реорганизации серверного парка, конечный специалист по сетевым технологиям может эффективно внедрить технологию виртуальных машин и повысить эффективность использования вычислительных ресурсов серверного парка.

За годы эксплуатации серверных систем автором успешно применялась реализация предложенного подхода. В рамках научных исследований [5] автором также была разработана методика по повышению эффективности использования вычислительных ресурсов компьютеров при применении технологии виртуальных машин. Эта методика не только позволяет более эффективно использовать ресурсы компьютера, но и значительно упрощает перенос логических серверов, работающих на базе виртуальных машин, с одного аппаратного обеспечения (физического компьютера) на другое.

ЛИТЕРАТУРА

1. Олифер В. Г., Олифер Н.А. Сетевые операционные системы. – СПб: Издательский дом “Питер”, 2001.
2. Александр Тормасов, Виртуализация операционных систем. – М.: "Открытые системы", 2002, №1
3. Еремин И.И., Астафьев Н.Н. Введение в теорию линейного и выпуклого программирования. – М.: Наука, 1979.
4. Антамошкин А.Н., Масич И.С. Эффективные алгоритмы условной оптимизации монотонных псевдодобулевых функций. Вестник СибГАУ. Выпуск. 4 – Красноярск: СибГАУ, 2003.
5. Рахман П.А. Разработка методики повышения эффективности использования вычислительных ресурсов при применении технологии виртуальных машин. Автореферат дисс. на соиск. уч. ст. канд. техн. наук. – М.: МЭИ, 2005.