

Ю. ЛИЙВАК

О РАСПРЕДЕЛЕНИИ РЕСУРСОВ В СЕТЯХ ЭВМ

Рассматриваются вопросы разработки структуры технических средств обработки информации на примере сетей ЭВМ. Анализируется влияние стратегий централизованного или децентрализованного расположения ресурсов на такие параметры сетей ЭВМ, как время ответа, надежность и экономическая эффективность.

Характер развития структуры наиболее типичных средств обработки данных — ЭВМ, сетей ЭВМ и технических комплексов АСУ — представлен в таблице. Видна явная тенденция развития от возможности

Развитие структуры технических средств обработки данных

Поко- ле- ния	Режимы обработки данных	ЭВМ	Сети ЭВМ	ТК АСУ
I	Монопольное пользование ресурсами, обработка в автономном режиме	Монопольное использование вычислительных ресурсов, жесткая конфигурация ВнУ	Дистанционное пользование централизованными ресурсами (централизованными сетями ЭВМ, ВЦ КП)	Автономные ВЦ на предприятиях
II	Мультидоступ, мультипрограммирование, мультиплексность; обработка в режиме разделения времени	Пользование вычислительным ресурсом в режиме разделения времени; жесткая конфигурация с фиксированным набором ВнУ	Дистанционное пользование однородными распределенными ресурсами (децентрализованными однородными сетями ЭВМ)	Системное пользование средствами вычислительной техники; системы теледоступа; работа в РМВ; пользование услугами ВЦ КП
III	Мультидоступ к функционально специализированным мультиресурсам; обработка в режиме разделения ресурсов	Мультидоступ к специализированному вычислительным ресурсам через автономный канал, гибкая конфигурация ЭВМ, сменный набор ВнУ, системы ЭВМ	Дистанционное пользование разнообразными распределенными ресурсами через автономную подсеть передачи данных (децентрализованные сети из гетерогенных ЭВМ)	Пользование ресурсами РСВЦ; наличие в локальных АСУ специализированных комплексов ЭВМ для сбора и предварительной обработки данных

IV Эволюция и интеграция различных возможностей систем III поколения

Сокращения: ВнУ — внешние устройства ЭВМ, ВЦ КП — вычислительный центр коллективного пользования, РМВ — реальный масштаб времени, ТК АСУ — технический комплекс автоматизированных систем управления, РСВЦ — республиканская сеть вычислительных центров.

работы с одним пользователем к возможности работы в режиме мультимедиа, от пользования одним ресурсом к использованию мультиресурсами, т. е. переход от работы в автономном режиме к работе в режиме мультимедиа к мультиресурсам — от полной централизации к полной децентрализации. Отдельные аспекты последних проблем затронуты в [1-3]. В данном сообщении эти проблемы анализируются с точки зрения организации структуры технических средств и представляемых к ним требований. Анализ проводится на примере сетей ЭВМ.

Сеть ЭВМ должна обрабатывать, передавать и хранить данные от всех абонентов, а также от локальных абонентных подсетей в режиме разделения ресурсов. Сеть ЭВМ должна отвечать некоторым довольно общим требованиям, к числу которых относятся:

- достаточная производительность и пропускная способность;
- заданное время получения ответа;
- разнообразие ресурсов, гибкость, возможности расширения;
- заданная надежность, живучесть;
- экономическая эффективность.

Посмотрим, какое влияние оказывает стратегия централизации или децентрализации, а также величина (емкость) ресурсов на перечисленные выше требования к сетям ЭВМ.

1. Производительность и общая пропускная способность сети определяются объемом хранимых и обрабатываемых данных. Если, допустим, при централизации существует один ресурс емкостью C операций/сек, то при децентрализации существуют m ресурсов емкостью C/m операций/сек, способных обрабатывать те же V знаков.

2. Время получения ответа $T_{\text{отв}}$ включает время обслуживания $t_{\text{об}}$ и время ожидания $t_{\text{ож}}$ на обслуживание ресурсом R (в одном звене):

$$T_{\text{отв}} = t_{\text{об}} + t_{\text{ож}}. \quad (1)$$

Время обслуживания зависит от емкости ресурса и от длины запроса. Обозначим среднюю длину запроса через $1/\mu$ операций/запрос. Тогда время обслуживания (сек) одного запроса при централизации и децентрализации ресурсов равняется соответственно

$$t_{\text{об.ц}} = 1/\mu C \quad (2)$$

и

$$t_{\text{об.д}} = m/\mu C.$$

Среднее время ожидания (сек) на обслуживание для очереди типа $M/M/1$ (т. е. распределение времени между запросами и распределение времени на обслуживание подчиняются экспоненциальному закону распределения, обслуживающих приборов — один) и для дисциплины обслуживания «первым пришел, первым был обслужен» равняется при централизации ресурса

$$t_{\text{ож.ц}} = \frac{\rho}{\mu C (1 - \rho)}, \quad (3)$$

где $\rho = \frac{\lambda}{\mu C}$ — коэффициент нагрузки (степень занятости ресурса),

λ — интенсивность потока запросов.

Для очереди типа $M/M/m$ с m обслуживающими приборами, т. е. при децентрализации ресурсов, среднее время обслуживания (сек) равняется

$$t_{\text{ож.д}} = \frac{P_m}{\mu C(1 - \rho)}, \quad (4)$$

где P_m — вероятность наличия в системе одновременно m или более запросов.

В [2] доказано, что $P_m < \rho$. Тогда из последних уравнений видно, что $t_{\text{об.д}} < t_{\text{об.д}}$, а $t_{\text{ож.д}} > t_{\text{ож.д}}$. Причем при больших нагрузках (при $\rho \cong 0,7$ эрл) с относительно короткими запросами время ожидания бывает наибольшим. Поскольку при децентрализации время ожидания сокращается, то при больших нагрузках децентрализованные сети предпочтительнее централизованных, а последние предпочтительнее при средних и малых нагрузках с длинными запросами, т. е. в случае обработки больших массивов данных. Если первый режим характерен для диалоговых систем, то второй — для пакетной обработки.

Дополнительного сокращения времени на ожидание ответа можно добиться, увеличивая производительность ресурса, т. е. устанавливая $C' = aC$, где C' — производительность нового ресурса, a — целочисленный коэффициент кратности.

Если вместе с увеличением производительности в a раз увеличивать входной поток запросов λ также в a раз, то, согласно теории очередей [2, 4], получим

$$T_{\text{отв}}(a\lambda, aC) = \frac{1}{a} T_{\text{отв}}(\lambda, C). \quad (5)$$

Из формулы (5) видно, что централизация и укрупнение ресурса могут также уменьшать необходимое для ответа время. Например, одновременное увеличение нагрузки и производительности ресурса в два раза сокращает время на ожидание ответа вдвое. Согласно предыдущим выкладкам, время ожидания, которое особенно затягивается при больших нагрузках, в сетях с распределенными ресурсами уменьшается. Эти противоречивые тенденции устраняются в сетях ЭВМ за счет автономной выделенной подсети передачи данных [5], коммутационные узлы которой работают в условиях больших нагрузок. Поскольку время обработки данных на коммутационных узлах (в режиме коммутации пакетов) относительно короткое, то подсеть передачи данных может быть успешно реализована на базе распределенных мини-ЭВМ. Через подсеть передачи данных осуществляется доступ к сетевым обрабатывающим и накапливающим ресурсам, которые должны быть централизованными и организованными на базе крупных ЭВМ или даже на системах ЭВМ. Далее увидим, что подобная стратегия распределения ресурсов не только уменьшает сроки получения ответа, но и повышает экономическую эффективность сети в целом.

Предыдущие выводы справедливы и для времени передачи данных по каналам передачи данных, которые являются тоже ресурсами. Под емкостью ресурса в этом случае подразумевается пропускная способность канала связи, время обработки равняется времени передачи сообщения по каналу связи, а время ожидания на обслуживание равняется времени ожидания на передачу.

3. Требованиям гибкости и разнообразия ресурсов наилучшим образом отвечают сети с распределенными и специализированными ресурсами.

4. Децентрализация ресурсов, улучшает, как правило, надежность сетей, которая зависит от надежности аппаратных и программных средств, а также от надежности и способности логического обеспече-

ния предотвращать перегрузки. Взаимосвязь надежности и перегрузок вытекает из формулы (3). При увеличении коэффициента нагрузки ($q \rightarrow 1$) время на ожидание обслуживания возрастает до бесконечности и ресурс оказывается в заблокированном состоянии, что равносильно его физическому отказу. По этой причине первостепенной задачей логического обеспечения является устранение перегрузок в сети. Способы устранения в настоящей статье не рассматриваются, однако уже предыдущие выкладки показывают, что сеть с распределенными ресурсами в условиях больших нагрузок работает лучше. Надежность подсети передачи данных, которая представляет собой основную подсеть сети ЭВМ, зависит от вероятности P разъединения отдельных узлов i, j . Под надежностью сети обычно подразумевается живучесть, отражающая степень избыточности сети. Для обеспечения живучести подсети передачи данных любые узлы i, j должны соединяться между собой $r_{i,j}$ путями. Так, в известной сети ARPA $r_{i,j} \geq 2$. Если вероятность отказа отдельных каналов связи $q \ll 1$, то вероятность разъединения отдельных узлов, т. е. живучесть сети, пропорциональна

$$P \approx q^{Q(i,j)}, \quad (6)$$

где $Q(i, j)$ — емкость (количество каналов) минимального сечения между узлами i и j .

В децентрализованных сетях передачи данных, соответствующих требованиям живучести, исключается, таким образом, возможность катастрофических отказов, неполадки отдельных ресурсов приводят лишь к уменьшению эффективности сети. Поэтому для обеспечения надежности, бесспорно, предпочтительнее проектировать сети ЭВМ с распределенными ресурсами.

5. Для обеспечения экономической эффективности сети более приемлема централизация ресурсов. Согласно закону Гроша, производительность ресурса увеличивается пропорционально квадрату стоимости. По отношению к ЭВМ третьего поколения этот закон действителен не для всей ЭВМ, а лишь для памяти и процессора. Однако для наших более или менее общих рассуждений закон Гроша сохраняет силу, так как мы рассматриваем отдельные ресурсы. Исходя из этого закона, можно сказать, что производительность вычислительного ресурса учетверяется, а стоимость лишь удваивается. Значит, использование централизованных высокопроизводительных вычислительных ресурсов экономически оправдывает себя, тем более что при этом улучшается качество обслуживания, уменьшаются затраты на вспомогательное оборудование и на производственную площадь. Экономически более выгодны и централизованные банки данных, которые представляют собой не менее важные ресурсы, чем вычислительные. Использование распределенных банков увеличивает нагрузку в каналах передачи данных из-за необходимости модификации данных. В наиболее неблагоприятном случае, когда все m распределенных ресурсов — банков данных — оказались бы связанными между собой каналами связи каждый с каждым, существовало бы $m(m-1)/2$ каналов связи. Считая, что к стоимости и производительности канала также применим закон Гроша, получим

$$R_6 = K_1(m-1) \sqrt{\frac{m}{2}}, \quad (7)$$

где R_6 — стоимость децентрализованного банка данных, K_1 — коэффициент пропорциональности.

Из формулы (7) видно, что стоимость организации децентрализованного банка данных пропорциональна количеству банков. Следовательно, централизованные банки с минимальным количеством территориально разрозненных ресурсов — экономически наиболее выгодная стратегия для организации банка данных в сетях ЭВМ. Возвращаясь к выражениям (1)—(3), приходим к выводу, что обработка и хранение данных в высокопроизводительных централизованных ресурсах дают наибольшую экономическую эффективность и улучшают свойства сети с точки зрения качества данных и времени получения ответа.

С другой стороны, доступ к централизованным мощным ресурсам дешевле осуществлять с помощью распределенных сетей. Если коммутационные узлы подсети передачи данных, связывающие централизованные ресурсы сети ЭВМ, распределены территориально равномерно, то расстояние от абонента до узла обратно пропорционально количеству коммутационных узлов. Считая, что длина канала пропорциональна стоимости, получим

$$R_c = K_2 \frac{1}{\sqrt{m}}, \quad (8)$$

где R_c — стоимость распределенной сети передачи данных.

Руководствуясь проведенным анализом, структуру технических средств обработки данных (сетей ЭВМ) можно выбирать согласно предъявляемым к ним требованиям в зависимости от конкретных условий. При небольших нагрузках и невысоких требованиях к надежности обработки следует проводить на средствах с централизованной структурой (на централизованных сетях ЭВМ, на ВЦ КП с дистанционным доступом) в пакетном режиме или в режиме запрос — ответ. Для высоконадежных республиканских сетей ВЦ, работающих с большой нагрузкой в таких, например, режимах, как диалог, запрос — ответ, пакетная обработка, т. е. с жесткими требованиями ко времени ответа, оптимальна распределенная структура [5]. Характерной чертой этой структуры является наличие выделенной подсети передачи данных, осуществляющей мультидоступ абонентов и абонентных подсетей к мультиресурсам, т. е. к обрабатывающей и накапливающей подсети со специализированными и укрупненными ресурсами в режиме разделения ресурсов.

Приношу благодарность Х. Салуму за внимание к работе.

ЛИТЕРАТУРА

1. Streeter, D. N., IBM Syst. J., 12, No. 3, 283 (1973).
2. Cleinrock, L., IFIP Cong. proc., 1, 11 (1974).
3. Patrick, R., Datamation, 12, No. 5, 79 (1976).
4. Allen, A., IBM Syst. J., 14, No. 2, 161 (1975).
5. Лийвак Ю., Связь, радио и телевидение, сер. 10, № 8, 1 (1976).

Вычислительный центр
Министерства связи СССР

Поступила в редакцию
18/I 1977

J. LIIVAK

RESSURSSIDE JAOTAMISEST ARVUTIVÕRKUDES

Artiklis on vaadeldud andmetöötlusvahendite struktuuri arengut ressursside jaotamise ja juurdepääsumetodide seisukohast ning analüüsitud mainitud meetodite mõju arvutivõrkudele kui andmetöötlusvahendite eriliigile esitatavaile nõuetele (vastuse saamise aeg, töökindlus, majanduslik efektiivsus).

J. LIIVAK

ABOUT RESOURCE ALLOCATION IN COMPUTER NETWORKS

The author discusses tendencies of the evolution of data processing hardware structure from the viewpoint of resource allocation and access methods. The effect of resource scaling and allocation strategies on the requirements of computer networks (response time, reliability, cost efficiency, etc.) is analyzed.