

УДК 004.7

*Алиев А.И., Гусейнова А.А.*

Институт Информационных Технологий НАНА, Баку, Азербайджан

[amir@lan.ab.az](mailto:amir@lan.ab.az), [ayten@lan.ab.az](mailto:ayten@lan.ab.az)

## **ПУТИ ПОВЫШЕНИЯ ПРОИЗВОДИТЕЛЬНОСТИ КОРПОРАТИВНОЙ СЕТИ И ОБ ОДНОМ КРИТЕРИИ ОБНОВЛЕНИЯ ДОКУМЕНТОВ В КЭШ-СЕРВЕРАХ**

*Статья посвящена вопросам повышения производительности корпоративной сети с ограниченными возможностями. Показаны некоторые пути повышения производительности, основными из которых являются кэширование информации и геометрическое распределение информации по сети серверов (CDN – Content Delivery Network). В сетях CDN одной из острых проблем является своевременное обновление информации в серверах-суррогатах. Для определения критерия обновления текстовой информации предложена мера, основанная на метрике Левенштейна.*

**Ключевые слова:** корпоративная сеть, локальная сеть, кэширование информации, сети CDN, метрика Левенштейна.

### **Введение**

Основу любой корпоративной сети составляют локальные сети [1], в которых активно внедряются новые высокоскоростные технологии передачи данных, но корпоративная сеть не ограничивается только локальной вычислительной сетью (ЛВС), а охватывает множество подразделений, объединенных выделенными каналами или сетью общего пользования Интернет. Очень часто предприятия, имеющие территориально распределенную структуру, сталкиваются с задачей объединения разрозненных, иногда существенно удаленных и отличающихся размерами и оснащенностью локальных сетей.

Для этой цели в корпоративных сетях используют арендованные каналы связи, в которых скорость передачи данных существенно ниже скорости передачи данных в локальной сети. Поэтому рациональное использование арендованных каналов имеет огромное значение для повышения производительности корпоративных сетей.

### **Пути повышения производительности корпоративной сети**

Производительность корпоративной сети можно повысить следующими основными путями [2]:

– **Приобретение дополнительной пропускной способности.**

Расширение каналов за счет приобретения дополнительной пропускной способности решает проблему недостаточной производительности сети лишь на короткое время, поскольку объем трафика быстро увеличивается и снова начинает превышать пропускную способность каналов, по которым он передается.

Объясняется это очень просто. В неоптимизированной сети потребность в полосе пропускания растет линейно по отношению к числу пользователей (при наличии одних и тех же приложений) [3].

– **Управление пропускной способностью.**

Механизмы управления пропускной способностью позволяют лучше использовать имеющуюся полосу пропускания благодаря более эффективному ее распределению между различными приложениями и пользователями.

– **Сжатие данных.**

Сжатие данных работает очень эффективно для некоторых типов файлов (например, для текстовых файлов). Однако оно не приносит большой пользы при передаче «случайных» данных, таких, как предварительно сжатые графические файлы, архивы и т.п. При внедрении средств сжатия данных в реальных сетях с пакетной коммутацией производители и пользователи обнаружили, что их эффективность весьма невысока. В большинстве случаев сжатие трафика Интернет или корпоративных интрасетей увеличивает пропускную способность не более чем на 5-10%.

– **Web-кэширование.**

Традиционная технология кэширования (Web Caching) позволяет разместить статичный web-контент ближе к конечному пользователю. Однако по своей природе этот механизм применим лишь к статичному контенту Интернета. Основная же часть трафика современных корпоративных сетей, напротив, не связана с Интернетом, а контент интрасетей, как правило, чрезвычайно динамичен. Статистические исследования показывают, что кэширование контента Web увеличивает производительность сети приблизительно на 30%.

– **Сети доставки контента – CDN.**

CDN – один из способов ускорения доступа в Интернет, успешно проявивший себя на практике. Они представляют собой наложенные сети, по которым часто запрашиваемые материалы сети перемещаются из ядра ближе к конечному пользователю. Передавая контент на периферию сети, службы CDN смогут быстрее удовлетворять последующие запросы на ту же информацию, поскольку число промежуточных маршрутизаторов глобальной сети сокращается и, соответственно, уменьшается задержка. Контент на периферии помещается в серверах-суррогатах, где хранятся данные, предназначенные для последующего извлечения [4].

Из числа интенсивных путей оптимизации нагрузок на каналы передачи информации серьезное развитие получили последние два подхода: кэширование информации [5, 6] и геометрическое распределение информации по сети серверов CDN [7].

Сети CDN и механизмы кэширования не только обеспечивают повышение быстродействия при доступе в Web, но и снижают нагрузку на сервер – источник данных. Другие пользователи в поисках той же информации будут попадать на ближайший к ним кэш, а не на сервер-источник. При этом потребление пропускной способности глобальной сети также снижается, так как трафик идет от конечного пользователя на периферию сети, а не в ее ядро. К тому же кэширование данных на дисках обходится гораздо дешевле, чем модернизация каналов глобальной сети.

Удаленные офисы предприятия обычно связываются со своими хостинговыми центрами посредством более медленных соединений, поэтому предварительная выборка контента позволяет избежать перегрузок и свести к минимуму потребности в корпоративных ресурсах глобальной сети. Кроме того, сети CDN

доставляют содержимое сервера-источника в кэш одним потоком трафика, после чего тот разделяется на несколько потоков для доставки в настольные системы. В связи с этим требования к пропускной способности арендованного канала корпоративной сети становятся еще более умеренными.

В сетях CDN важное значение имеет обработка динамического контента. То, что доступ к базе данных производится в реальном масштабе времени на периферии сети, означает снижение нагрузки на серверы с исходным контентом. Вы сможете быстро создавать страницы Web с доступом к базе данных в реальном времени, так что их не придется переделывать каждый раз, когда запрашивается обновленный или персонализированный контент. В результате страницы генерируются с меньшей избыточностью, повышается быстродействие узла, обеспечивается защита от всплесков активности и снижается уровень требований к инфраструктуре.

### Определение критерия обновления текстовой информации

В сетях CDN одной из острых проблем является своевременное обновление информации в серверах-суррогатах. Относительно других видов информации (аудио, видео), как правило, чаще всех изменяются текстовые информации. Для определения критерия обновления текстовой информации ниже предлагается мера, основанная на метрике Левенштейна.

Степень изменения документа будем вычислять расстоянием Левенштейна (редактирование). Первоначально расстояние Левенштейна было введено для определения расстояния между словами. Расстояние Левенштейна равно минимальному числу элементарных операций редактирования, необходимых для преобразования одной строки в другую. Набор элементарных операций состоит из операций замены, вставки и удаления одной буквы. Для вычисления расстояния Левенштейна каждое слово должно представляться в виде набора символов. Расстояние Левенштейна между словами  $a = (a_1, \dots, a_N)$  и  $b = (b_1, \dots, b_M)$  вычисляется следующей рекурсивной формулой [8]:

$$L((a_1, \dots, a_N), (b_1, \dots, b_M)) = \begin{cases} N, & \text{если } M = 0 \\ M, & \text{если } N = 0 \\ \min \left\{ \begin{array}{l} L((a_1, \dots, a_{N-1}), (b_1, \dots, b_M)) + 1, \\ L((a_1, \dots, a_N), (b_1, \dots, b_{M-1})) + 1, \\ L((a_1, \dots, a_{N-1}), (b_1, \dots, b_{M-1})) + \bar{\delta}(a_N, b_M) \end{array} \right\} & \text{в других случаях} \end{cases} \quad (1)$$

В формуле (32)  $\bar{\delta}(x, y)$  является отрицанием символа Кронекера:

$$\bar{\delta}(x, y) = \begin{cases} 0, & \text{если } x = y \\ 1, & \text{если } x \neq y \end{cases}$$

Для каждого расстояния  $L((a_1, \dots, a_i), (b_1, \dots, b_j))$ ,  $i = 1, \dots, N$ ,  $j = 1, \dots, M$  используем сокращенную запись  $L_{ij}$ . В частности, имеем  $L(a, b) = L_{NM}$ . Для вычисления  $L_{NM}$  должны вычисляться все расстояния  $L_{ij}$ ,  $i = 0, 1, \dots, N$ ,  $j = 0, 1, \dots, M$ ,  $(i, j) \neq (N, M)$ . Достаточно трудоемко вычислять расстояние Левенштейна по рекурсивной формуле (32). Сложность такого вычисления растет экспоненциально с размером слов. Вместо рекурсивного вычисления (31) расстояние  $L_{ij}$  можно вычислить

итеративно. В итеративном алгоритме сначала инициализируются расстояния  $L_{i0} = i$ ,  $i = 1, \dots, N$  и  $L_{0j} = j$ ,  $j = 1, \dots, M$ . Потом по значениям  $L_{(i-1)j}$ ,  $L_{i(j-1)}$  и  $L_{(i-1)(j-1)}$  последовательно вычисляется расстояние  $L_{ij}$ ,  $i = 1, \dots, N$ ,  $j = 1, \dots, M$ .

Для вычисления близости двух текстовых документов расстояние Левенштейна не является адекватной мерой близости. Дело в том, что перестановка отдельных слов в предложении или предложений в документе не воспринимается человеком как существенное изменение. Поэтому при определении близости двух текстовых документов из набора элементарных операций будем исключать операции замены.

Пусть  $L(D, D^*)$  минимальное число элементарных операций удаления или вставки слов, необходимых для преобразования одного документа в другой. Тогда расстояние похожести между документом  $D$  и его копиями  $D^*$  будем определять так:

$$dist(D, D^*) = \frac{L(D, D^*)}{\lambda + \lambda^*}, \quad (2)$$

где  $\lambda$  и  $\lambda^*$  длина (в словах) документов  $D$  и  $D^*$  соответственно.

Из определения видно, что  $0 \leq dist(D, D^*) \leq 1$ . Между идентичными документами расстояние (2) равно нулю, а между различными документами оно равно единице.

## Заключение

В результате проведенного анализа выявлены основные пути повышения производительности корпоративных сетей с ограниченными эксплуатационными ресурсами.

Для решения проблем своевременного обновления текстовой информации в серверах-суррогатах выведена формула для определения степени изменения документа, основанная определением расстояния Левенштейна.

## Литература

1. «Современные корпоративные сети» <http://www.race.ru/page.asp?id=391>
2. <http://www.sovtel.ru/cgi-bin/products.pl?id=115>
3. Джозеф Якубович. Оптимизация сетевого трафика. Журнал «Сети и системы связи», №10, 2001.
4. Дуг Аллен. Домашние сети доставки контента. LAN, #01/2002.
5. Ari Luotonen, Kevin Altis. World-Wide Web proxies. Proc. 1st Int. WWW Conf., Geneva, May 1994.
6. A. Chankhunthod, P. Danzig, C. Neerdaels, M. F. Schwartz, K. Worrell. A hierarchical Internet object cache. Proc. 1996 USENIX Annu. Tech. Conf., San Diego, Jan 1996, pp.153163.
7. Алгулиев Р.М., Алыгулиев Р.М., Шарифов М.Г. Подход к оптимальному расположению серверов CDN в узлах глобальных сетей //Информационные технологии. 2006, №11, с.20–26.
8. Runkler T.A., Bezdek J.C., Web mining with relational clustering //International Journal of Approximate Reasoning. 2003. Vol. 32. №№2–3. P.217–236.

UOT 004.7+681.3

**Əliyev Ə.İ., Hüseynova A.Ə.**

AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

[amir@lan.ab.az](mailto:amir@lan.ab.az), [ayten@lan.ab.az](mailto:ayten@lan.ab.az)

**Korporativ şəbəkənin məhsuldarlığının yüksəldilməsinə yanaşmalar və keş-serverlərdə sənədlərin təzələnməsinin bir kriteriyası haqqında**

Məqalə məhdud imkanlı korporativ şəbəkələrin məhsuldarlığının yüksəldilməsi məsələsinə həsr edilmişdir. Burada məhsuldarlığın artırılmasının bir neçə yolu araşdırılmış və onların ən vacibi kimi informasiyanın keşləşdirilməsi və informasiyanın serverlər şəbəkəsində coğrafi bölünməsi (CDN şəbəkəsi) qeyd edilmişdir. CDN şəbəkələrində əsas məsələlərdən biri surroqat-serverlərdə informasiyanın vaxtlı-vaxtında təzələnməsidir. Mətn şəkilli informasiyanın təzələnmə kriteriyası üçün Levenşteyn metrikasına əsaslanan ölçü təklif edilmişdir.

*Açar sözlər: korporativ şəbəkə, lokal şəbəkə, informasiyanın keşləşdirilməsi, CDN şəbəkə, Levenşteyn metrikası.*

**Aliyev A.I., Huseynova A.A.**

Institute of Information Technology ANAS, Baku, Azerbaijan

[amir@lan.ab.az](mailto:amir@lan.ab.az), [ayten@lan.ab.az](mailto:ayten@lan.ab.az)

**Approaches increase productivity, and corporate networks on one criteria updated documents in the cache servers.**

This article is devoted to issues of productivity povyscheniya corporate network with limited capabilities. Showing several ways to increase productivity, the main of which are caching data and geometric distribution of the information on the network servers (CDN - Content Delivery Network). In CDN networks one of the most pressing challenges is the timely update of the servers-surrogate. To determine the criterion for updating text proposed a measure based on the Levenshtein metric.

*Key words: corporate network, local network, caching data, CDN networks, Levenshtein metric.*