

Jakub Swacha

EKONOMICZNE MIARY EFEKTYWNOŚCI KOMPRESJI DANYCH

1. Wstęp

Błyskawiczny postęp informatyzacji, którego jesteśmy świadkami, powoduje szybki przyrost zarówno ilości przechowywanych w przedsiębiorstwach danych, jak i wydatków ponoszonych na ten właśnie cel.

Wielkość samego tylko rynku oprogramowania obsługującego przechowywanie danych zbliżyła się w roku 2004 do 8 miliardów dolarów¹, natomiast rynku nośników danych do 17 miliardów dolarów², co oznacza, że optymalizacja sposobu przechowywania danych może przynieść gospodarce niebagatelne oszczędności.

Znaczne ograniczenie wydatków ponoszonych na przechowywanie i przesył danych osiągnąć można przez zastosowanie metod kompresji danych. Kompresja danych to sztuka zwięzłego zapisu danych, polegająca bądź na wykorzystaniu powtarzalności danych (kompresja bezstratna), bądź na usunięciu z danych nieistotnej ich części (kompresja stratna). Drugi z wymienionych rodzajów kompresji w nieunikniony sposób prowadzi do powstawania w danych przekłamań, na które w wielu przypadkach (np. danych o charakterze ekonomicznym) nie można sobie pozwolić.

Użycie kompresji danych zmniejsza rozmiar danych do przechowania lub przesłania. Oznacza to proporcjonalne do osiągniętego stopnia kompresji zmniejszenie zapotrzebowania na nośniki danych lub łącza transmisyjne. Efekt ekonomiczny tej redukcji nie jest jednak łatwy do oszacowania. Oszczędność może być zarówno zerowa – gdy zaoszczędzone media pozostają niewykorzystane, jak i ponadproporcjonalnie duża – gdy niestosowanie kompresji wymusiło by konieczność pokonania bariery technologicznej. Niemniej, należy uwzględnić różnego rodzaju koszty stosowania kompresji: programowe, sprzętowe i inne.

Nie da się obliczyć efektu ekonomicznego użycia kompresji danych bez odpowiednich miar. Klasyczne miary efektywności kompresji danych pozwalają określić stopień zmniejszenia ilości danych uzyskany wskutek kompresji, co, jak

¹ Źródło: IDC, za: [Jaques 2005].

² Źródło: BCC, Inc., za: [Rajan 2005].

już zauważono, nie wystarcza do ustalenia ekonomicznych konsekwencji zastosowania kompresji danych. Stąd potrzeba ekonomicznych miar efektywności kompresji danych, umożliwiających przeliczenie technicznych parametrów kompresji danych na wartość korzyści ekonomicznych dzięki niej uzyskanych.

Miary takie, dla najbardziej popularnych zastosowań, zostaną przedstawione w niniejszym artykule. Wcześniej jednak omówiona zostanie klasyfikacja metod kompresji danych oraz klasyczne miary efektywności kompresji.

2. Współczesne metody kompresji danych i ich zastosowania

2.1. Klasyfikacja metod kompresji danych

Kryteria klasyfikacji metod kompresji danych

Istnieje wiele kryteriów, według których można sklasyfikować metody kompresji danych. Do najważniejszych spośród nich zaliczamy:

- kryterium wiernego odtworzenia,
- kryterium ogólności stosowania,
- kryterium sposobu działania.

Dwa pierwsze z wymienionych kryteriów mają zasadnicze znaczenie dla użytkownika końcowego: skutkiem niewłaściwego wyboru metody kompresji w kontekście tych kryteriów może być nieodwracalne zniekształcenie danych.

Trzecie kryterium nie wiąże się z takim zagrożeniem, przesądza jednak o parametrach użytkowych programów kompresujących, takich jak czas i efektywność kompresji, jest zatem również istotne dla użytkownika końcowego.

Kryterium wiernego odtworzenia

Podstawowym kryterium podziału metod kompresji jest możliwość wiernego odtworzenia danych oryginalnych z postaci skompresowanej [Swacha 1999].

Jak zostało to już napisane wcześniej, kompresja danych może polegać bądź na wykorzystaniu powtarzalności danych, bądź na usunięciu z danych nieistotnej ich części. W pierwszym wypadku możliwe jest wierne przywrócenie oryginalnej postaci danych.

Metody, które ograniczają się do wykorzystania statystycznych zależności istniejących w danych, nazywane są bezstratnymi (*lossless*), gdyż mimo ich zastosowania nie traci się żadnej części danych.

W wypadku usunięcia z danych pewnej ich części (uznanej za zbędną), oryginalna postać danych może zostać przywrócona tylko w sposób przybliżony. Metody działające w ten sposób nazywamy stratnymi (*lossy*), gdyż ich działaniu towarzyszy bezpowrotne usunięcie części danych.

Kryterium ogólności stosowania

Kryterium ogólności stosowania rozróżnia metody dedykowane – przeznaczone do kompresji danych jednego określonego typu, oraz metody uniwersalne – przeznaczone do kompresji danych dowolnego typu [Swacha 2004].

Metody dedykowane dysponują pewną wiedzą *a priori* o danych przeznaczonych do kompresji. Jeżeli kompresowane dane są zgodne z tą wiedzą, pozwala to na osiągnięcie bardzo wysokich stopni kompresji. Jeżeli jednak kompresowane dane nie należą do właściwego typu, użycie metody temu dedykowanej prowadzi często do ich zniekształcenia lub wydłużenia (zamiast pożądanego skrócenia).

Metody uniwersalne dysponują jedynie ogólnym modelem danych. Pozwala on modelować dane różnego rodzaju, jednak nie tak precyzyjnie jak model dedykowany. Dlatego choć metody uniwersalne pod względem osiąganych stopni kompresji ustępują dedykowanym, ich zaletą jest możliwość użycia ich niezależnie od typu kompresowanych danych.

Kryterium sposobu działania

Kryterium sposobu działania dzieli metody kompresji według sposobu, w jaki wykorzystują one charakter i strukturę danych w celu ich krótszego zapisu.

Kryterium to teoretycznie pozwala wyróżnić tyle kategorii, ile istnieje algorytmów kompresji. W praktyce jednak istnieje tylko kilka technik, które w różnych odmianach wykorzystuje się w większości dostępnych programów kompresujących [Swacha 1999].

Dla kompresji bezstratnej ze względu na sposób działania wyróżniamy metody:

- substytucyjne,
- predykcyjne,
- transformacyjne,
- hybrydowe.

Dla kompresji stratnej ze względu na sposób działania wyróżniamy metody:

- predykcyjne,
- fraktalne,
- kwantyzacji,
- transformacyjne,
- percepcyjne.

Działanie metod substytucyjnych polega na podmianie długich, wielokrotnie występujących ciągów danych ich krótszymi odpowiednikami (np. zastępowanie słów indeksami słownika). Są to metody najszerzej rozpowszechnione, szybkie, lecz – w porównaniu z innymi – najmniej efektywne.

Bezstratne metody predykcyjne próbują przewidzieć kolejny element danych, wykorzystując model statystyczny oparty na wcześniej przetworzonych danych. Skuteczna predykcja wymaga rozbudowanego modelu, co oznacza powolność działania, wynagrodzoną jednak najwyższą efektywnością kompresji [Swacha 1999].

Bezstratne metody transformacyjne przekształcają dane w taki sposób, aby poddawały się one następnie efektywnej kompresji na podstawie prostych modeli statystycznych. Stanowią rozwiązanie pośrednie, są wolniejsze od szybkich metod substytucyjnych i mniej efektywne od najlepszych metod predykcyjnych.

Metody hybrydowe łączą metody substytucyjne z predykcyjnymi lub transformacyjnymi. Osiągają często najlepszy stosunek efektywności do czasu kompresji [Swacha 2004].

Stratne metody predykcyjne, w odróżnieniu od swoich bezstratnych odpowiedników, nie próbują przewidywać dokładnych wartości kolejnych danych, lecz przybliżają ją najbardziej odpowiednią funkcją; reprezentację danych stanowi zatem zbiór współczynników tej funkcji.

Metody fraktalne [Skarbek 1998] opierają się na wykorzystaniu samopodobieństwa występującego w niektórych rodzajach danych, szczególnie pochodzenia naturalnego; mniejsze fragmenty danych opisywane są przez przekształcenia większych.

Kwantyzacja służy do zmniejszenia ilości danych poprzez reprezentację występującego w danych podzbioru wartości jednym z elementów tego podzbioru lub podzbiorem tego podzbioru.

Stratne metody transformacyjne mają tak przekształcić dane, by po poddaniu ich następnie kwantyzacji, w większym stopniu usunięciu uległa nieistotna część danych, aniżeli część istotna.

Kompresja percepcyjna dotyczy tych typów danych, których docelowym odbiorcą jest człowiek. Posługując się wiedzą o niedoskonałości ludzkich zmysłów, usuwa się tę część danych, która i tak zostałaby przezeń niedostrzeżona.

Większość dostępnych algorytmów kompresji stratnej ma charakter hybrydowy, łącząc ze sobą metody kilku typów.

2.2. Przegląd popularnych zastosowań metod kompresji danych

Najbardziej oczywisty obszar zastosowań kompresji to archiwizacja danych. Programy, takie jak Zip czy Rar, pozwalają zmniejszyć przestrzeń dyskową zajmowaną przez dane, jak również oszczędzić czas związany z przesłaniem danych w sieci.

Wiele popularnych formatów wymiany różnego typu danych wykorzystuje metody kompresji w niewidoczny dla użytkownika sposób, np.: archiwa Javy (JAR), pliki pomocy *on line* (CHM), animacje Macromedia Flash (SWF), czy grafika Corel Draw (CDR).

Korzystając z takich programów, jak np. DriveSpace Microsoftu, można skompresować całą partycję dysku, dzięki czemu pomieścić może ona nawet kilkakrotnie więcej danych, niż wynosi jej fizyczna pojemność.

Znaczna część danych, zwłaszcza w systemach informatycznych przedsiębiorstw, przechowywana jest w systemach bazodanowych. Dedykowane metody kompresji pozwalają nie tylko na ograniczenie objętości baz danych, ale i na przyspieszenie szybkości dostępu do danych [Ray, Haritsa, Seshadri 1995].

Ze względu na obszerność baz danych, złożoność wykonywanych zapytań i wymagany szybki czas reakcji, dobrym miejscem dla implementacji metod kompresji są hurtownie danych [Goyal i in. 1999].

Również duża część danych przechowywana jest w postaci tekstowej (np. opartej na XML). Istnieją wydajne metody kompresji tekstu zarówno dla XML [Swacha 2003], jak i oznakowań innego rodzaju [Swacha 2005].

Kompresować można nie tylko dane przechowywane na pamięciach masowych, ale również te znajdujące się w pamięci operacyjnej [Douglass 1993]. Kompresji poddaje się część danych aktualnie nie przetwarzanych, co usprawnia obróbkę wielkich zbiorów danych, które inaczej z trudem mieściłyby się w pamięci operacyjnej, a także zwalnia część pamięci na potrzeby innych aplikacji.

Transmisja danych przez sieci komputerowe bardzo często wiąże się z wykorzystaniem kompresji. Jest ona uwzględniona w specyfikacjach protokołów różnych warstw – począwszy od Point-to-Point Protocol (PPP), przez Internet Protocol (IP), a skończywszy na HyperText Transfer Protocol (HTTP).

Także komutowany dostęp do Internetu przez linię telefoniczną wiąże się z wykorzystaniem kompresji danych. Będące w powszechnym użytku standardy kompresji w komunikacji modemowej to CCITT V42bis oraz Microcom MNP-5 i MNP-7.

Zastosowanie kompresji w systemach *cache* znacznie przyspiesza dostęp do dokumentów przez jednoczesne zmniejszenie średniej wielkości dokumentu i zwiększenie liczby dokumentów mieszczących się w pamięci podręcznej (co zwiększa prawdopodobieństwo trafień *cache*) [Velasco, Velasco Lucianez 1998].

Rozwój biologii molekularnej stworzył duże zapotrzebowanie na algorytmy umożliwiające kompresję łańcuchów RNA i DNA przechowywanych w bazach danych ośrodków badawczych. Wynika ono z dużej objętości tych danych – np. genom człowieka zawiera ok. trzy miliardy znaków.

Obecnie dużą część danych cyfrowych stanowią dane sygnałowe pochodzenia naturalnego, takie jak mowa ludzka, dźwięk, obrazy i film (obrazy ruchome i dźwięk). Zbiory danych tego typu mają często bardzo dużą objętość, co czyni ich kompresowanie techniczną koniecznością [Swacha 2002].

Różnego rodzaju metody kompresji mowy [Skarbek 1998] stosowane są w telefonii komórkowej (GSM), cyfrowej telefonii stacjonarnej (ISDN), telefonii internetowej (Voice over IP), dyktafonach czy w automatach informacyjnych (np. w środkach masowej komunikacji).

Metody kompresji dźwięku szerokopasmowego [Skarbek 1998] znalazły zastosowanie w rozpowszechnianiu muzyki przez Internet (MPEG-1 Layer 3 – MP3), internetowym radiu, radiu cyfrowym, kodowaniu ścieżek dźwiękowych filmów (MP3, AC-3), cyfrowych kasetach (DCC) i płytach (MP3-CD).

Kompresja obrazów to przede wszystkim powszechnie spotykane w Internecie formaty wymiany grafiki – GIF, JPEG i w mniejszym stopniu PNG [Swacha 2001]. Wyszły one poza sferę zastosowań związanych z komputerami, trafiając m.in. do cyfrowych aparatów fotograficznych. Kompresja wykorzystywana jest także przy kodowaniu danych przesyłanych faksem (standardy CCITT, w tym JBIG) i archiwach dokumentów (TIFF, DjVu). Bezstratne metody kompresji znajdują zastosowanie w przypadku zdjęć satelitarnych i różnego typu obrazów medycznych.

Kompresja tekstur w kartach graficznych generujących grafikę trójwymiarową zarówno przyspiesza transmisję danych z pamięcią operacyjną, jak i – dzięki lepszemu wykorzystaniu pamięci na karcie – ogranicza liczbę sytuacji, w których doładowanie tekstur jest potrzebne. Dodatkowo możliwe staje się użycie tekstur o niedopuszczalnym dotąd rozmiarze [Dumont, Pellacini, Ferwerda 2001].

Kompresja obrazów ruchomych obejmuje technologie dla wideotelefonii i wideokonferencji (H261, H263) oraz rozpowszechniania filmów (MPEG) [Sayood 2002], a także cyfrową telewizję satelitarną, telewizję internetową, amatorskie i profesjonalne kamery cyfrowe oraz telefonię cyfrową trzeciej generacji (UMTS).

3. Techniczne miary efektywności kompresji danych

3.1. Klasyczne miary efektywności kompresji danych

Klasyczne miary kompresji danych oceniają efektywność metod kompresji według tego, jak bardzo udaje się dzięki nim zmniejszyć objętość kompresowanych danych. Dają zatem pomiary niezależne od sytuacji i warunków technicznych, jednak niewystarczające do ekonomicznej oceny metod kompresji.

Do najczęściej używanych miar efektywności kompresji zaliczamy stopień kompresji, współczynnik procentowy kompresji i średnią bitową.

Stopień kompresji (*compression ratio* – *CR*) definiuje się jako stosunek długości danych wejściowych (S) do długości danych wyjściowych (L) [Sayood 2002]:

$$CR = \frac{S}{L}. \quad (1)$$

Im wyższa wartość CR , tym lepszą efektywność kompresji osiągnięto.

Z kolei współczynnik procentowy kompresji (*compression percentage* – CP) to stosunek uzyskanego skrócenia danych wejściowych do ich początkowej długości:

$$CP = \frac{S - L}{S} \cdot 100\%. \quad (2)$$

Im wyższa wartość CP , tym lepszą efektywność kompresji osiągnięto. Współczynnik procentowy kompresji przyjmuje wartości ujemne w przypadku, gdy długość danych wyjściowych przekracza długość danych wejściowych, a w miarę wzrostu efektywności kompresji zbliża się do wartości 100%, której wszak nigdy nie osiąga.

Średnią bitową (*bitrate* – BR) nazywamy średnią liczbę bitów przypadającą na pojedynczy symbol (lub inną jednostkę) oryginalnego zbioru danych. W zależności od rodzaju kompresowanych danych za jednostkę BR przyjmuje się bity na znak (bpc), bity na próbkę (bps) czy bity na punkt (bpp). Dla pierwszej z wymienionych jednostek:

$$BR = \frac{8 \cdot L}{S}. \quad (3)$$

Im niższa wartość BR , tym lepszą efektywność kompresji osiągnięto.

3.2. Dynamiczne miary efektywności kompresji danych

Podstawową wadą klasycznych miar efektywności kompresji danych jest pomijanie czynnika czasu. Dlatego do oceny praktycznej przydatności metod kompresji zaproponowano dodatkowe miary [Swacha 2004].

Najprostszą miarą efektywności kompresji uwzględniającą czas jest całkowity czas transmisji:

$$CCT = t_k + t_t + t_d, \quad (4)$$

gdzie: t_k – czas kompresji,

t_t – czas transmisji skompresowanego pliku,

t_d – czas dekompresji.

Jednostką miary CCT jest sekunda (s). Wartość CCT interpretuje się w sposób następujący: jak długo trwa łącznie transmisja danych z wykorzystaniem kompresji.

Wadą tej miary jest konieczność interpretacji pomiarów z uwzględnieniem wielkości pliku. Można tego uniknąć, korzystając z miary, jaką jest całkowita szybkość transmisji:

$$CST = \frac{S}{CCT}, \quad (5)$$

gdzie: S – długość pliku przed kompresją.

Jednostką miary CST jest bajt przez sekundę (B/s). Wartość CST interpretuje się następująco: jak szybko można przysyłać dane z wykorzystaniem kompresji.

Mimo że CST umożliwia porównywanie pomiarów dla plików o różnych długościach, nadal jednak niełatwe jest porównywanie wyników uzyskanych dla różnych prędkości transmisji. Miarą wolną od tej wady jest współczynnik przyspieszenia transmisji:

$$WPT = \frac{CST}{v_t} = \frac{t_{nt}}{CCT}, \quad (6)$$

gdzie: v_t – szybkość transmisji nieskompresowanego pliku,

t_{nt} – czas transmisji nieskompresowanego pliku.

Jednostka miary WPT jest niemianowana. Wartość WPT interpretuje się w następujący sposób: ile razy szybciej dane przesyłane są z wykorzystaniem kompresji, niż byłyby bez niej.

4. Ekonomiczne aspekty efektywności kompresji danych

4.1. Kategorie zastosowań kompresji danych

Najpopularniejsze kategorie zastosowań kompresji danych to:

- 1) elektroniczny przesył danych,
- 2) dystrybucja danych:
 - elektroniczna,
 - na nośnikach fizycznych;
- 3) przechowywanie danych, w tym:
 - przechowywanie danych bieżących,
 - archiwizacja danych historycznych,
 - tworzenie kopii zapasowych danych bieżących.

Ad 1. W przypadku elektronicznego przesyłu danych nadawca kompresuje dane, a odbiorca je dekompresuje. Motywem użycia kompresji danych w tym przypadku może być:

- zmniejszenie czasu transmisji,
- zmniejszenie obciążenia łącza komunikacyjnego.

Konsekwencje ekonomiczne zmniejszenia czasu transmisji to – bezpośrednio – niższy koszt przesyłu danych, w przypadku gdy koszt łącza zależy od czasu korzystania z niego (np. komutowane łącza telefoniczne), a pośrednio – wydłużenie efektywnego czasu pracy (pracownicy mniej czasu spędzają, oczekując na wysłanie lub odebranie danych).

Zmniejszenie obciążenia łącza komunikacyjnego przekłada się zaś na następujące korzyści ekonomiczne:

- niższy koszt przesyłu danych, w przypadku gdy koszt łącza zależy od ilości przesłanych danych (np. GPRS),
- zmniejszenie czasu transmisji dla współużytkowników łącza, gdy jest ono wykorzystywane jednocześnie przez więcej niż jednego użytkownika,
- brak konieczności zwiększania parametrów technicznych łącza, w przypadku gdy transmisja w postaci nieskompresowanej byłaby niepraktyczna dla istniejących parametrów technicznych łącza.

Ad 2. Przy dystrybucji danych dane kompresowane są przez pojedynczego nadawcę, natomiast dekompresowane przez wielu odbiorców, a korzyści wynikające z kompresji obejmują możliwe zwiększenie atrakcyjności danych dla odbiorcy przez:

- przy dystrybucji elektronicznej: krótszy czas otrzymania danych przez odbiorcę,
- przy dystrybucji na nośnikach: dane zapisane na mniejszej liczbie nośników są wygodniejsze w posługiwaniu się nimi,

co w przypadku dystrybucji o charakterze komercyjnym ma efekt ekonomiczny (choć trudny do oszacowania), oraz zmniejszenie kosztów:

- przy dystrybucji elektronicznej: analogicznie do wymienionych dla elektronicznego przesyłu danych,
- przy dystrybucji na nośnikach: niższy koszt mniejszej liczby nośników, niekiedy także ich transportu i opakowania.

Ad 3. W przypadku przechowywania danych stosowanie kompresji pozwala zmniejszyć koszty nośników danych.

Można wyróżnić tu zasadniczo dwie sytuacje:

- typową dla przechowywania danych bieżących, gdy wielkość nośników danych jest ograniczona, a stosowanie kompresji pozwala zmniejszyć tempo powiększania nośników;
- typową dla przechowywania danych archiwalnych, gdzie problemem jest nie tyle wielkość nośników danych, ale ich koszt, a stosowanie kompresji pozwala ten koszt zmniejszyć.

4.2. Jawne i ukryte koszty kompresji danych

W poprzednim punkcie wskazano mnogość korzyści ekonomicznych, jakie przynieść może stosowanie kompresji danych. Oczywiście jest, że stosowaniu kompresji danych muszą towarzyszyć również pewne koszty.

Do kosztów stosowania kompresji danych zaliczyć można:

- koszt oprogramowania służącego do kompresji:
 - pierwotne koszty oprogramowania,
 - koszty zapewnienia zgodności ze skompresowanymi danymi w przypadku zmiany oprogramowania systemowego lub zmiany technologii kompresji;
- koszt obsługi oprogramowania służącego do kompresji:
 - dodatkowe obowiązki dla pracowników odpowiedzialnych za przechowywanie danych,
 - koszty szkoleń, podręczników itp. dla wszystkich pracowników mających bezpośrednio do czynienia ze skompresowanymi danymi;
- koszt wykonywania kompresji:
 - zwiększone wymogi co do mocy obliczeniowej i wielkości pamięci operacyjnej dla sprzętu,
 - czas tracony na oczekiwanie na kompresję i dekompresję danych.

Część z powyższych kosztów może być całkowicie wyeliminowana, np. dostępne jest bezpłatne oprogramowanie kompresujące, kompresja może być zaimplementowana w sposób niewidzialny dla użytkowników końcowych (nie ma więc potrzeby ich szkolenia).

Ponadto część z powyższych kosztów może mieć charakter marginalny, np. proste i szybkie metody kompresji stwarzają niewielkie zapotrzebowanie na zasoby systemowe, a czas ich działania może być dla użytkowników końcowych niezauważalnie krótki.

5. Sformułowanie ekonomicznych miar efektywności kompresji

5.1. Wybór typowych zastosowań

Jak zostało to przedstawione w pkt 4 artykułu, zarówno ekonomiczne korzyści, jak i koszty wynikające z użycia metod kompresji danych mają różną naturę w zależności od rodzaju zastosowania. Co więcej, nie wszystkie z wymienionych wcześniej kategorii zysków i strat mogą być w łatwy sposób mierzone, a nawet szacowane.

Trudno zatem byłoby sformułować pojedynczą ekonomiczną miarę efektywności kompresji danych o uniwersalnym zastosowaniu. Można jednak opracować takie miary dla jednostkowych zastosowań, szczególnie takich, które w praktyce kompresji danych występują często.

Poniżej przedstawione zostaną trzy typowe zastosowania, dla których zostaną następnie opisane ekonomiczne miary efektywności kompresji danych. Są to:

1. Przesył danych łączem komutowanym.
2. Dystrybucja danych na nośnikach fizycznych.
3. Przechowywanie danych bieżących na dyskach twardych.

5.2. Przesył danych łączem komutowanym

Poniższe rozważania dotyczą przypadku komutowanego łącza telefonicznego, gdzie koszt łącza zależy od czasu korzystania z niego. Opłata za korzystanie z łącza pobierana jest w pewnych interwałach czasowych, nazywanych jednostkami taryfikacyjnymi.

Efektywność ekonomiczną kompresji danych w tym przypadku można rozpatrzeć na trzech poziomach szczegółowości:

- uwzględniając jedynie koszt użytkowania łącza,
- uwzględniając także wpływ kompresji na efektywny czas pracy,
- uwzględniając efektywny czas pracy oraz koszty stosowania kompresji.

Na pierwszym z wymienionych, najbardziej ogólnym poziomie zakłada się, że pracownik, oczekując na zakończenie transmisji danych, może w pełni efektywnie zajmować się innymi zadaniami, a koszty stosowania kompresji w stosunku do kosztów użytkowania łącza mają pomijalną wartość.

Na drugim z wymienionych poziomów zakłada się że pracownik oczekujący na zakończenie transmisji danych nie zajmuje się jakimikolwiek innymi zadaniami, a koszty stosowania kompresji w stosunku do kosztów użytkowania łącza mają pomijalną wartość.

Trzeci poziom uwzględnia dodatkowo koszty wykonania kompresji.

W praktyce zapewne częstą sytuacją jest taka, w której okres oczekiwania na zakończenie transmisji, jakkolwiek nie w pełni efektywnie, jest jednak w jakimś stopniu spożytkowany przez pracownika. Precyzyjne określenie tego stopnia byłoby jednak w praktyce zbyt trudne, aby uwzględnić go przy rozpatrywaniu efektywności ekonomicznej kompresji danych.

Podobne utrudnienia wiążą się z ustaleniem faktycznych kosztów wykonania kompresji danych, uwzględniających dodatkowe obciążenie sprzętu i personelu. Najprościej więc ograniczyć ten składnik do kosztów poniesionych na oprogramowanie do kompresji danych, zakładając, że stanowią one zasadniczą jego część.

Oznaczmy przez:

- S – długość danych nieskompresowanych (w bitach),
- L – długość danych skompresowanych (w bitach),
- k_{jt} – koszt jednostki taryfikacyjnej (w jednostce walutowej),
- t_{jt} – czas trwania jednostki taryfikacyjnej (w sekundach),
- v_t – efektywną szybkość transmisji danych (przepustowość łącza) (w bitach na sekundę),
- k_p – godzinową stawkę płacy pracownika oczekującego na zakończenie przesyłu danych (w jednostce walutowej),
- k_{on} – miesięczny koszt oprogramowania komunikacyjnego nie wspierającego kompresji (w jednostce walutowej),
- k_{ok} – miesięczny koszt oprogramowania komunikacyjnego wspierającego kompresję lub sumę kosztów oprogramowania kompresującego i oprogramowania komunikacyjnego nie wspierającego kompresji (w jednostce walutowej),
- T – przeciętną łączną długość danych przesyłanych miesięcznie (w bitach).

Koszt przesłania danych w postaci nieskompresowanej przy użyciu komutowanego łącza telefonicznego wyniesie (w jednostce walutowej):

$$K_{TK1} = \left\lceil \frac{S}{v_t t_{jt}} \right\rceil k_{jt}; \quad (7)$$

- w przypadku dodatkowego uwzględnienia skrócenia efektywnego czasu pracy:

$$K_{TK2} = \left\lceil \frac{S}{v_t t_{jt}} \right\rceil k_{jt} + \frac{k_p}{3600} \frac{S}{v_t}; \quad (8)$$

(gdzie wartość 3600 oznacza liczbę sekund przypadającą na godzinę i wyrażona jest w sekundach);

- w przypadku dodatkowego uwzględnienia skrócenia efektywnego czasu pracy oraz kosztów oprogramowania komunikacyjnego niewspierającego kompresji:

$$K_{TK3} = \left\lceil \frac{S}{v_t t_{jt}} \right\rceil k_{jt} + \frac{k_p}{3600} \frac{S}{v_t} + \frac{k_{on} S}{T}. \quad (9)$$

Koszt przesłania danych w postaci skompresowanej przy użyciu komutowanego łącza telefonicznego wyniesie (w jednostce walutowej):

$$K_{STK1} = \left\lceil \frac{L}{v_t t_{jt}} \right\rceil k_{jt}; \quad (10)$$

- w przypadku dodatkowego uwzględnienia skrócenia efektywnego czasu pracy:

$$K_{STK2} = \left\lceil \frac{L}{v_t t_{jt}} \right\rceil k_{jt} + \frac{k_p}{3600} \frac{L}{v_t}; \quad (11)$$

(gdzie wartość 3600 oznacza liczbę sekund przypadającą na godzinę i wyrażona jest w sekundach),

- w przypadku dodatkowego uwzględnienia skrócenia efektywnego czasu pracy oraz kosztów oprogramowania komunikacyjnego wspierającego kompresję:

$$K_{STK3} = \left\lceil \frac{L}{v_t t_{jt}} \right\rceil k_{jt} + \frac{k_p}{3600} \frac{L}{v_t} + \frac{k_{ok} S}{T}. \quad (12)$$

Oszczędność brutto uzyskana z przesłania danych przy użyciu kompresji wyniesie:

- wyrażona w jednostce walutowej:

$$Z_{TK} = K_{TK} - K_{STK}; \quad (13)$$

- wyrażona procentowo (jako część kosztów przesyłu bez kompresji):

$$ZP_{TK} = 100\% - \frac{K_{STK}}{K_{TK}}. \quad (14)$$

5.3. Dystrybucja danych na nośnikach fizycznych

Przy dystrybucji danych na nośnikach fizycznych celem zastosowania kompresji danych jest zmniejszenie liczby nośników, na których będą dystrybuowane dane (którymi mogą być np. wersje instalacyjne programów).

Precyzyjne określenie wynikających z użycia mniejszej liczby nośników konsekwencji ekonomicznych zwiększenia atrakcyjności danych dla odbiorcy, a także zmniejszenia kosztów transportu i opakowania byłoby zbyt skomplikowane, dlatego rozpatrzona zostanie jedynie oszczędność uzyskana z redukcji kosztów samych nośników.

Oznaczmy przez:

S – długość przeznaczonych do dystrybucji danych (w bitach),

L – długość przeznaczonych do dystrybucji danych po kompresji (w bitach),

N – liczbę dystrybuowanych kopii (jednostka niemianowana),

- k_n – koszt pojedynczego nośnika danych (w jednostce walutowej),
- V – pojemność pojedynczego nośnika danych (w bitach),
- K'_{dn} – całkowity koszt przygotowania dystrybucji bez stosowania kompresji, uwzględniający koszty stosownego oprogramowania i pracy (w jednostce walutowej), bez kosztów nośników,
- K'_{dk} – całkowity koszt przygotowania dystrybucji z zastosowaniem kompresji, uwzględniający koszty stosownego oprogramowania (w tym kompresującego) i pracy (w jednostce walutowej), bez kosztów nośników.

Koszt przygotowania pojedynczej kopii dystrybuowanych danych bez uwzględnienia kosztów pracy i oprogramowania wynosi:

- bez stosowania kompresji danych:

$$k_{dn} = \left\lceil \frac{S}{V} \right\rceil k_n; \quad (15)$$

z zastosowaniem kompresji danych:

$$k_{dk} = \left\lceil \frac{L}{V} \right\rceil k_n. \quad (16)$$

Koszt przygotowania całej dystrybucji z uwzględnieniem kosztów pracy i oprogramowania wynosi:

- bez stosowania kompresji danych:

$$K_{dn} = N \left\lceil \frac{S}{V} \right\rceil k_n + K'_{dn}; \quad (17)$$

- z zastosowaniem kompresji danych:

$$K_{dk} = N \left\lceil \frac{L}{V} \right\rceil k_n + K'_{dk}. \quad (18)$$

Oszczędność brutto uzyskana z kompresji dystrybuowanych danych wyniesie:

- wyrażona w jednostce walutowej:

$$Z_{DK} = K_{dn} - K_{dk}; \quad (19)$$

- wyrażona procentowo (jako część kosztów przygotowania dystrybucji bez użycia kompresji):

$$ZP_{DK} = 100\% - \frac{K_{dk}}{K_{dn}}. \quad (20)$$

5.4. Przechowywanie danych bieżących na dyskach twardych

Ujęcie statyczne

W statycznym ujęciu problemu kosztów przechowywania danych bieżących na dyskach twardych zakłada się, że ilość przechowywanych danych jest stała, a kompresja pozwala na stosowanie dysków o mniejszej pojemności.

Jest ono użyteczne dla sytuacji, w których ilość danych w dłuższym okresie nie ulega zmianie, a także dla wszystkich innych w krótkim horyzoncie czasowym.

Oznaczmy przez:

S' – minimalną pojemność dysku potrzebną do przechowania danych bez zastosowania kompresji (w bitach); wielkość ta jest nie mniejsza od długości nieskompresowanych danych S i zależy od pojemności dostępnych na rynku dysków,

L' – minimalną pojemność dysku potrzebną do przechowania danych skompresowanych (w bitach); wielkość ta jest nie mniejsza od długości skompresowanych danych L i zależy od pojemności dostępnych na rynku dysków,

k_S – koszt dysku o pojemności S' (w jednostce walutowej),

k_L – koszt dysku o pojemności L' (w jednostce walutowej).

Oszczędność brutto uzyskana z kompresji przechowywanych danych wyniesie (pomijając koszty zakupu i eksploatacji oprogramowania kompresującego):

- wyrażona w jednostce walutowej:

$$Z_{PK} = k_S - k_L; \quad (21)$$

- wyrażona procentowo (jako część kosztów przechowywania danych bez użycia kompresji):

$$ZP_{PK} = 100\% - \frac{k_L}{k_S}. \quad (22)$$

Ujęcie dynamiczne

W ujęciu dynamicznym problemu kosztów przechowywania danych bieżących na dyskach twardych przyjmuje się, że ilość przeznaczonych do przechowania danych stale rośnie, w związku z czym wymaga okresowego zwiększania pojemności nośników danych; stosowanie kompresji pozwala zmniejszyć tempo powiększania pojemności nośników. Podejście to jest użyteczne dla wielu sytuacji praktycznych.

Zjawiskiem obserwowanym na rynku dysków twardych, które również należy wziąć pod uwagę, jest trwale postępujący spadek ceny jednostkowej przechowania danych o określonej wielkości. Na przykład, cena jednego gigabajta pojemności

dysku twardego zmalała w latach 1956-2004 z 10 milionów dolarów do 1 dolara i 15 centów [*Historical Notes...* 2004]. Oznacza to dodatkową korzyść z przesunięcia w czasie momentu zakupu nowego sprzętu, uzyskaną dzięki zastosowaniu kompresji.

Oznaczmy przez:

- n_{zn} – ile razy w rozpatrywanym okresie pojemność dyskowa wymagać będzie powiększenia, jeżeli nie zastosuje się kompresji danych (liczba całkowita),
- n_{zk} – ile razy w rozpatrywanym okresie pojemność dyskowa wymagać będzie powiększenia, jeżeli zastosuje się kompresję danych (liczba całkowita); dla każdej prawidłowo dobranej metody kompresji $n_{zn} \geq n_{zk}$,
- k_i – koszt zwiększenia pojemności dyskowej w momencie i -tym, jeżeli nie stosuje się kompresji danych (w jednostce walutowej), gdzie $i \in [1, n_{zn}]$,
- k_j – koszt zwiększenia pojemności dyskowej w momencie j -tym, jeżeli stosuje się kompresję danych (w jednostce walutowej), gdzie $j \in [1, n_{zk}]$,
- K'_{pn} – całkowity koszt obsługi przechowywania danych bez stosowania kompresji, uwzględniający koszty stosownego oprogramowania i pracy (w jednostce walutowej), bez kosztów nośników danych,
- K'_{pk} – całkowity koszt obsługi przechowywania danych z zastosowaniem kompresji, uwzględniający koszty stosownego oprogramowania (w tym kompresującego) i pracy (w jednostce walutowej), bez kosztów nośników danych.

Koszt przechowywania danych w rozpatrywanym okresie, z uwzględnieniem kosztów pracy i oprogramowania wynosi:
bez stosowania kompresji danych:

$$K_{pn} = \sum_{i=1}^{n_{zn}} k_i + K'_{pn}; \quad (23)$$

- z zastosowaniem kompresji danych:

$$K_{pk} = \sum_{j=1}^{n_{zk}} k_j + K'_{pk}. \quad (24)$$

Oszczędność brutto uzyskana z kompresji przechowywanych danych wyniesie:
wyrażona w jednostce walutowej:

$$Z_{PK} = K_{pn} - K_{pk}; \quad (25)$$

- wyrażona procentowo (jako część kosztów przechowywania danych bez użycia kompresji):

$$ZP_{PK} = 100\% - \frac{K_{pk}}{K_{pn}}. \quad (26)$$

6. Podsumowanie

Kompresja danych pozwala na istotne ograniczenie kosztów przechowywania i przesyłu danych. Zaproponowane w artykule ekonomiczne miary efektywności kompresji umożliwiają obliczenie wielkości uzyskanych oszczędności dla najbardziej popularnych sytuacji, w których stosuje się kompresję.

Opisane miary mogą znaleźć zastosowanie do oceny opłacalności implementacji metod kompresji danych w systemie informatycznym przedsiębiorstwa, a także do porównania efektywności ekonomicznej konkurencyjnych technologicznie rozwiązań.

Niewątpliwą zaletą przedstawionych miar jest ich przydatność, także w przypadku nieznanych wartości niektórych parametrów: im więcej dostępnych jest szczegółowych danych, tym bardziej precyzyjnie zostaną oszacowane koszty i zyski wynikające z użycia kompresji. Szacunki będą jednak użyteczne nawet przy znajomości tylko podstawowych parametrów (takich jak np. koszt nabycia nośnika danych), które da się ustalić w prawie każdej sytuacji.

Literatura

- Douglis F., *The Compression Cache: Using On-line Compression to Extend Physical Memory*, „Proceedings of USENIX Conference” 1993.
- Dumont R., Pellacini F., Ferwerda J.A., *Perceptually-based Texture Caching for Hardware Rendering*, „Proceedings of the Eurographics Workshop on Rendering” 2001.
- Goyal K.B., Ramamritham K., Datta A., Thomas H.M., *Indexing and Compression in Data Warehouses*, „Proceedings of International Workshop on Design and Management of Data Warehouses” 1999.
- Historical Notes about the Cost of Hard Drive Storage Space*,
<http://www.littletechshoppe.com/ns1625/winchest.html>, wydanie z 17 kwietnia 2004.
- Jaques R., *Storage software sales shoot skywards*, „IT Week”, 16 marca 2005,
<http://www.itweek.co.uk/2126991>.
- Rajan M., *Global data storage market to reach nearly \$39 billion by 2010*, Data Storage Media: Materials, Technologies, and Markets, RGB-317, maj 2005,
<http://www.bccresearch.com/editors/RGB-317.html>.
- Ray G., Haritsa J.R., Seshadri S., *Database Compression: A Performance Enhancement Tool*, [w:] *Advances in Data Management*, Tata McGraw-Hill, Pune 1995.
- Sayood K., *Kompresja danych – wprowadzenie*, Wydawnictwo RM, Warszawa 2002.
- Skarbek W. (red.), *Multimedia. Algorytmy i standardy kompresji*, Akademicka Oficyna Wydawnicza PLJ, Warszawa 1998.
- Swacha J., *Popularne standardy kompresji danych*, „Pro Dialog” 1999 nr 9.
- Swacha J., *Formaty graficzne*, „Software 2.0”, 2001, nr 79 (lipiec).
- Swacha J., Bandosz M., Radliński Ł., *Zaawansowane multimedia na stronach WWW*, [w:] *Multimedialne i sieciowe systemy informacyjne*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2002.
- Swacha J., *Kompresja dokumentów XML metodą substytucyjno-predykcyjną*, Studia Informatica nr 17, Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, Szczecin, 2003.
- Swacha J., *Predykcyjno-substytucyjna metoda bezstratnej kompresji danych*, Rozprawa doktorska, Politechnika Wrocławska, Wrocław 2004.

- Swacha J., *Compressing Annotated Natural Language Text*, [w:] *Human Language Technologies as a Challenge for Computer Science and Linguistics*, Wydawnictwo Poznańskie, Poznań 2005.
- Velasco J.R., Velasco Lucianez L.A., *Benefits of Compression in HTTP Applied to Caching Architectures*, „Proceedings of the International WWW Caching Workshop” 1998.

ECONOMIC MEASURES OF DATA COMPRESSION EFFICIENCY

Summary

Data compression allows for considerable reduction of data storage and transmission costs. The measures proposed in the article make it possible to calculate savings for the most popular data compression applications.

These measures may be used to estimate the profitability of implementing data compression methods in company's information system, as well as to compare economic effectiveness of technologically competitive solutions.

An important advantage of the presented measures is that they may be used even if not all the cost-specific data is available. In that case, the calculated costs, though not so precise, will still be useful.

Jakub Swacha – dr, adiunkt w Instytucie Informatyki w Zarządzaniu Uniwersytetu Szczecińskiego.