

INTERNET SHOPPING OPTIMIZATION PROBLEM

JACEK BLAZEWICZ *, MIKHAIL Y. KOVALYOV **, JEDRZEJ MUSIAL *, ANDRZEJ P. URBANSKI *,
ADAM WOJCIECHOWSKI *

* Poznan University of Technology
Institute of Computing Science
ul. Piotrowo 2
60-965 Poznan, Poland

e-mail: Jacek.Blazewicz@cs.put.poznan.pl, Jedrzej.Musial@cs.put.poznan.pl,
Andrzej.Urbanski@cs.put.poznan.pl, Adam.Wojciechowski@cs.put.poznan.pl

** United Institute of Informatics Problems
National Academy of Sciences of Belarus
Surganova 6
220012 Minsk, Belarus
e-mail: kovalyov_my@newman.bas-net.by

A high number of Internet shops makes it difficult for a customer to review manually all the available offers and select optimal outlets for shopping. A partial solution of the problem is brought by price comparators which produce price ranking from collected offers. However, their possibilities are limited to a comparison of offers for a single product requested by the customer. The problem we investigate in this paper is a multiple-item multiple-shop optimization problem, in which total expenses of a customer to buy a given set of items should be minimized over all available offers. In this paper, the Internet Shopping Optimization Problem (ISOP) is defined in a formal way and a proof of its strong NP-hardness is provided. We also describe polynomial time algorithms for special cases of the problem.

Keywords: algorithms, computational complexity, combinatorial algorithms, optimization, Internet shopping

1. Introduction

On-line shopping is one of key business activities offered over the Internet. A survey concerning American Internet users behavior published by Pew Internet & American Life Project in February 2008 (Project, 2008) shows that population of on-line customers grows rapidly and systematically from year to year. The number of on-line users either buying or searching for products on-line since 2000 has roughly doubled. While in 2000 the number of 22% of Americans (46% of on-line users) had some experience with buying products in virtual shops, the ratio grew to 39% in 2003 and reached 49% (66% of on-line users) in 2007. The development of on-line shopping is also stimulated by the increasing number of Internet users. E-commerce revenue has grown from \$7.4 billion at the middle of 2000 to \$34.7 billion in the third quarter of 2007. A survey concerning behavior of customers in Poland (Gemius, 2008) confirms the ten-

dency observed in America. Among various business instruments commonly available on-line since 1990s, like auctions (Lee, 1998), (Klein, 2000) and (Vulkan, 2003, p. 149-178), banking and secure payment (Langdon *et al.*, 2000), shopping (Liang and Huang, 1998), electronic libraries (Lesk, 1997), etc., on-line retail remains the service offered by the highest number of providers. A growing number of on-line shops and increased accessibility for customers world-wide by the use of credit card payment in on-line transactions (Langdon *et al.*, 2000) are key attributes that force strong competition on the market, keep prices low (when compared to off-line shopping) and make more and more customers interested in the on-line purchase (Vulkan, 2003, p. 22-27). However, a wide choice of on-line shops makes it difficult to manually compare all the offers and choose optimal providers for the required set of products. Solution of this problem has been supported by software agents (Tolle and Chen, 2000), so-called price comparison sites. The con-

Table 1. Prices of books and delivery costs offered by six internet shops.

Cost	Book a	Book b	Book c	Book d	Book e	Delivery	Total
Shop 1	18	39	29	48	59	10	203
Shop 2	24	45	23	54	44	15	205
Shop 3	22	45	23	53	53	15	211
Shop 4	28	47	17	57	47	10	206
Shop 5	24	42	24	47	59	10	206
Shop 6	27	48	20	55	53	15	218

Table 2. Price comparator solution - the result of the selection process.

	Book a	Book b	Book c	Book d	Book e	Delivery	Total
Price	18	39	17	47	44	45	210
Shop	Shop 1	Shop 1	Shop 4	Shop 5	Shop 2		

cept of price comparator is built on the idea of collecting offers of many on-line shops and building a price ranking on customer's request. This approach is commonly accepted by the customers and, according to Alexa Rank, popular price comparison services belong to the group of 1000 most viewed sites world-wide: shopping.com: 518's place, nextag.com: 533's place, bizrate.com: 600's place, shoplocal.com: 932's place (site popularity results registered in October 2008, www.alexa.com). It is worth to note that price ranking built on-line on customer's request expressed in a text query (product description) is a solution to a specific case of shopping, in which a customer wants to buy a single product. Multiple item shopping is not supported by price comparators available nowadays. Furthermore, price comparison sites, being commercial projects, tend to optimize their incomes from directing customers to particular on-line shops. As a result, price comparison sites play a role of recommender systems (Satzger *et al.*, 2006) which tend to detect customer's preferences and interests in order to suggest products to buy. A side effect of the problems mentioned above is a loss of customer's confidence.

To illustrate the optimization process we would like to consider and assess its benefits, let us consider an example below.

A customer wants to buy five books. Prices of the books and delivery costs in six shops which the customer considers as potential shopping locations are collected in Table 1. Customer's goal is to buy all five books at minimum expense. A support from currently available price comparators allows to build customer's basket based on optimal offers for each individual book. The result of the selection process is presented in Table 2. The shopping performed upon price comparator suggestions would not be optimal (total cost: 210) because simple price comparison does not include delivery cost which grows as the number of shopping locations grows. In order to find the cheapest solution for ISOP illustrated above one can perform a

complete search of all possible realizations of shopping. The simple example we analyze shows that the optimization process may bring some savings. In our example, optimal cost of purchase (in Shop 1 and Shop 4) equals 189 (see Table 3). The problem addressed in this paper is to manage a multiple-item shopping list over several shopping locations. The objective is to have all the shopping done at the minimum total expense. One should notice that dividing the original shopping list into several sublists whose items will be delivered by different providers increases delivery costs. The delivery costs are counted and paid individually for each package (sub-list) assigned to a specific Internet shop in the optimization process.

In the sequel, we consider the above mentioned Internet Shopping Optimization Problem (ISOP) in a more formal way. In Section 2, a formal definition of the problem is given. In Section 3, we prove that ISOP is NP-hard in the strong sense and that it is not approximable in polynomial time. In Section 4, we demonstrate that ISOP is polynomially solvable if the number of products to buy, n , or the number of shops, m , is a given constant. The paper concludes with a summary of the results and suggestions for future research.

2. Problem definition

Notation used throughout this paper is given in Table 4. We study the following problem of Internet shopping. A single buyer is looking for a multiset of products $N = \{1, \dots, n\}$ to buy in m shops. A multiset of available products N_l , a cost c_{jl} of each product $j \in N_l$, and a delivery cost d_l of any subset of the products from the shop to the buyer are associated with each shop l , $l = 1, \dots, m$. It is assumed that $c_{jl} = \infty$ if $j \notin N_l$. The problem is to find a sequence of disjoint selections (or carts) of products $X = (X_1, \dots, X_m)$, which we call a *cart sequence*, such that $X_l \subseteq N_l$, $l = 1, \dots, m$, $\cup_{l=1}^m X_l = N$, and the total product and delivery cost,

Table 3. Optimal purchase cost in selected shops.

	Book a	Book b	Book c	Book d	Book e	Delivery	Total
Cost	18	39	17	48	47	20	189
Shop	Shop 1	Shop 1	Shop 4	Shop 1	Shop 4		

Table 4. Problem definition - table of notation.

Symbol	Explanation
n	number of products
m	number of shops
N_l	multiset of products available in shop l
c_{jl}	cost of each product $j \in N_l$
d_l	delivery cost for shop $l, l = 1, \dots, m$
$X = (X_1, \dots, X_m)$	sequence of selections of products in shops $1, \dots, m$
$F(X)$	sum of product and delivery costs
$\delta(x)$	0-1 indicator function for $x = 0$ and $x > 0$
X^*	optimal sequence of selections of products
F^*	optimal (minimum) total cost

denoted as $F(X) := \sum_{l=1}^m (\delta(|X_l|)d_l + \sum_{j \in X_l} c_{jl})$, is minimized. Here $|X_l|$ denotes the cardinality of the multiset X_l , and $\delta(x) = 0$ if $x = 0$ and $\delta(x) = 1$ if $x > 0$. We denote this problem as ISOP (Internet Shopping Optimization Problem), its optimal solution as X^* , and its optimal solution value as F^* .

3. Strong NP-hardness and inapproximability

In this section we will analyze computational complexity of the ISOP. We will prove its strong NP-hardness by proving strong NP-completeness of its decision counterpart – problem P1. The latter problem has the same input as ISOP plus an additional parameter y , and the question is to determine whether there exists a selection of products with the total cost $F(X) \leq y$.

theorem 1 *Problem ISOP is NP-hard in the strong sense even if all costs of the available products are equal to zero and all the delivery costs are equal to one.*

Proof. We construct a pseudo-polynomial transformation of problem P1 from the strongly NP-complete problem EXACT COVER BY 3-SETS (X3C), see (Garey and Johnson, 1979).

EXACT COVER BY 3-SETS (X3C) can be defined as follows: Given a family $E = \{E_1, \dots, E_L\}$ of 3-element subsets of the set $K = \{1, \dots, 3k\}$, does E contain an exact cover of K , i.e., a subfamily $Y \subseteq E$ such that each $j \in K$ belongs to exactly one 3-element set in Y ? It is clear that if Y is a solution to X3C, then $|Y| = k$.

Given an instance of X3C, we construct the following instance of problem P1. There are $m = L$ shops with available products of the sets $N_l = E_l, l = 1, \dots, L$.

The buyer would like to purchase the set of products $N = K$. The cost of any product available in any shop is equal to zero ($c_{jl} = 0, j \in N_l, l = 1, \dots, m$), the delivery cost from any shop to the buyer is equal to one ($d_l = 1, l = 1, \dots, m$) and the threshold value of the criteria is $y = k$. We show that X3C has a solution if and only if there exists a solution X for the constructed instance of the problem P1 with $F(X) \leq k$. It is easy to see that our transformation is polynomial and pseudo-polynomial at the same time. Therefore, problem P1 belongs to the class NP.

Let Y be a solution to X3C. Construct a solution for problem P1, in which the required products are purchased in k shops determined by their sets of products $N_l = E_l \in Y$, i.e., $X_l = N_l$ if $N_l \in Y$ and $X_l = \phi$ if $N_l \notin Y$. Since Y is an exact cover of K , all the required products are purchased, and the cost of the corresponding solution is $F(X) = k$.

Now assume that there exists a solution X for problem P1 with the cost $F(X) \leq k$. On one hand, for this solution the number of shops with $X_l \neq \phi$ should not exceed k because otherwise $F(X) > k$. On the other hand, the number of these shops should not be less than k because otherwise at least one product $j \in N$ will not be purchased. Therefore, there are exactly k shops with $X_l \neq \phi$. Since the purchased products form the set K , the collections of the shops with $X_l \neq \phi$ represents a solution for X3C. ■

We now discuss the approximability of problem ISOP. Let us consider its special case, in which the cost of any product in any shop is equal to zero, and the delivery cost from any shop to the buyer is equal to one. This special case is equivalent to the following problem MINIMUM SET COVER, see (Crescenzi and Kann, 2008).

MINIMUM SET COVER: Given a collection C of subsets of a finite set S , find a set cover for S , i.e., a subset $C' \subseteq C$ such that every element in S belongs to at least one member of C' , which minimizes the cardinality of the set cover, i.e., $|C'|$.

Due to (Raz and Safra, 1997), problem **MINIMUM SET COVER** is polynomially non-approximable within the ratio $c \cdot \ln |S|$, for some constant $c > 0$. Therefore, the following statement takes place.

statement 1 *There exists no polynomial $(c \cdot \ln n)$ -approximation algorithm for problem ISOP, unless $\mathcal{P} = \mathcal{NP}$.*

4. Polynomial algorithms for special cases

Notice that the intractability of problem **ISOP** is established under the assumption that both the number of products to buy, n , and the number of shops, m , are variable. In this section we present two solution approaches for problem **ISOP**, which are polynomial if either n or m is a constant.

The idea of our first algorithm **SHOP-ENUM** is to enumerate all possible selections of shops containing all the required products, to choose the best cart sequence $X^{(M)}$ for each selection of shops M , $M \subseteq \{1, \dots, m\}$, to calculate the total product and delivery cost for each M , and, finally, to find a cart sequence X^* with the minimum total cost, F^* .

Algorithm SHOP-ENUM

Step 1. Set $F^* = \infty$ and $X_l^* = \phi$, $l = 1, \dots, m$.

Step 2. Consider selections of shops to buy all the required products. Each shop l , $l \in \{1, \dots, m\}$, such that N_l is not empty can be selected or not, therefore, there are at most 2^m selections. For each selection of shops M perform the following computations. If M does not contain all the required products, then skip considering this selection. Otherwise, do the following. Determine a cart sequence $X^{(M)} = (X_1^{(M)}, \dots, X_m^{(M)})$, $X_l^{(M)} \subseteq N_l$, $l = 1, \dots, m$, as follows. For each product $i \in N$, select a shop $l \in M$ in which cost of product i is the cheapest. This can be done in a constant time if costs c_{il} are stored for each product i in a heap. Assign product i to the corresponding multiset $X_l^{(M)}$. Notice that the assignment strategy can eliminate some of the selected shops. This strategy is optimal for a given selection of shops due to the facts that the total product costs are minimized, and that the total delivery cost does not exceed the sum of delivery costs for the selected shops. Calculate the total cost, $F(X^{(M)})$. If $F(X^{(M)}) < F^*$, re-set $F^* := F(X^{(M)})$ and $X^* := X^{(M)}$.

Step 3. Output optimal solution X^* with the minimum cost F^* .

The time complexity of algorithm **SHOP-ENUM** is $O(n2^m)$, which is polynomial (linear) if the number of shops m is a constant.

The idea of our second algorithm **PRODUCT-ENUM** is to enumerate all possible shop choices for each product. Let $S = (S_1, \dots, S_n)$ be a shop sequence such that $S_i \in \{1, \dots, m\}$ is a shop in which product i will be purchased, $i = 1, \dots, n$. The algorithm determines a shop sequence S^* with the minimum total cost, F^* , of the corresponding cart sequence.

Algorithm PRODUCT-ENUM

Step 1. Set $F^* = \infty$ and $S_i^* = \phi$, $i = 1, \dots, n$.

Step 2. Consider shop sequences $S = (S_1, \dots, S_n)$ such that product $i \in N$ will be purchased in shop S_i , $i = 1, \dots, n$. There are at most m^n such sequences. For each sequence S , calculate the cost of the corresponding cart sequence, $F(S)$. If $F(S) < F^*$, re-set $F^* = F(S)$, $S^* := S$ and pass to considering next shop sequence.

Step 3. Output optimal shop sequence S^* with the minimum total cost F^* .

The time complexity of the algorithm **PRODUCT-ENUM** is $O(nm^n)$, which is polynomial if the number of products n is a constant.

Algorithm **PRODUCT-ENUM** can be applied in practice if the number of products to buy, n , is small and the number of shops having these products, m , is large. Alternatively, if the number of shops m is small and the number of the required products n is large, algorithm **SHOP-ENUM** can be efficient.

Conclusions

Internet shopping is often attributed with prices lower than in traditional shops. Another strong advantage of an on-line purchase is a wide choice of alternative retailers which, in general, remain at the same distance from the customer - at least one day for shopping delivery. Postage cost is often non-zero, which makes it reasonable to group purchased products so that each group is ordered in the same shop, and the total purchase and delivery cost is minimized. Changing retail location for the same product can be possible, provided that the customer is guaranteed that product in new location is identical to that in old location. In the case of changing retail location, quantity and quality of the product must be identical. However, total price can change because of different time of delivery, different profit margins, etc. Customers are interested in minimizing total product and delivery cost. Currently available price comparators can be used for these purposes in rare cases, because they do not provide multiple-item basket optimization.

We introduced Internet Shopping Optimization Problem and provided a proof of its strong NP-hardness. Furthermore, two polynomial time algorithms for special cases of ISOP have been described. In the future work, we intend to derive and experimentally test heuristic approaches for ISOP to make the suggested approach applicable for solving complex shopping cart optimization problems in on-line applications. Ideas and algorithmic results given in (Musial and Wojciechowski, 2009) for a simplified version of problem ISOP can be generalized and extended for these purposes.

Acknowledgment

The work was partially supported by the grant from the Ministry of Science and Higher Education No.N519188933. The first author (Jacek Blazewicz) acknowledges a support of the INRIA Rhone-Alpes grant.

References

- Crescenzi, P. and Kann, V. (2008). A compendium of np optimization problems. <http://www.nada.kth.se/~viggo/wwwcompendium/>.
- Garey, M. and Johnson, D. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*, New York, Freeman.
- Gemius, S. (2008). E-commerce in poland. <http://gemius.pl/pl/raporty/2008-06/03>.
- Klein, S. (2000). The emergence of auctions on the world wide web, in S. M. et al. (Ed.), *Handbook on Electronic Commerce*, Springer-Verlag, Berlin-Heidelberg, pp. 627–645.
- Langdon, C., Roghe, F. and Shaw, M. (2000). Consumer mass market on-line payment solutions, in S. M. et al. (Ed.), *Handbook on Electronic Commerce*, Springer-Verlag, Berlin-Heidelberg, pp. 273–288.
- Lee, H. (1998). Do electronic marketplaces lower the prices of goods?, *Communications of the ACM* **41**(1): 73–80.
- Lesk, M. (1997). *Practical Digital Libraries: Books, Bytes and Bucks*, Morgan Kaufmann Publishers, San Francisco, CA.
- Liang, T. and Huang, J. (1998). An empirical study on consumer acceptance of products in electronic markets: A transactional cost model, *Decision Support Systems* **21**.
- Musial, J. and Wojciechowski, A. (2009). A customer assistance system: Optimizing basket cost, *Foundations of Computing and Decision Sciences* **34**(1): 59–69.
- Project, P. I. . A. L. (2008). On-line shopping. http://www.pewinternet.org/~media/Files/Reports/2008/PIP_Onlinepping.pdf.
- Raz, R. and Safra, S. (1997). A sub-constant error-probability low-degree test, and sub-constant error-probability pcp characterization of np, *29th Annual ACM Symposium on Theory of Computing*, pp. 475–484.
- Satzger, B., M., M. E. and Kielssing, W. (2006). A preference-based recommender system, in B. K. et al. (Ed.), *E-Commerce and Web Technologies*, Springer-Verlag, Berlin Heidelberg.
- Tolle, K. and Chen, H. (2000). Intelligent software agents for electronic commerce, in S. M. et al. (Ed.), *Handbook on Electronic Commerce*, Springer-Verlag, Berlin-Heidelberg, pp. 265–382.
- Vulkan, N. (2003). *The Economics of E-Commerce. A Strategic Guide to Understanding and Designing the On-line Marketplace*, Princeton University Press, Princeton, New Jersey.
- Jacek Blazewicz** born in 1951 (M.Sc. in control engineering 1974, Ph.D. and Dr. habil. in computer science in 1977 and 1980, respectively), is a professor of computer science at the Poznan University of Technology. Presently he is a deputy director of the Institute of Computing Science. His research interests include algorithm design and complexity analysis of algorithms, especially in bioinformatics, as well as in scheduling theory. He has published widely in the above fields (over 300 papers) in many outstanding journals, among others in *Operations Research*, *IEEE Transactions on Computers*, *IEEE Transactions on Communications*, *Discrete Applied Mathematics*, *Parallel Computing*, *Acta Informatica*, *Performance Evaluation*, *International Transactions of Operations Research*, *European Journal of Operational Research*, *Information Processing Letters*, *Operations Research Letters*, *Journal of Computational Biology*, *Bioinformatics*, *Computer Applications in Biosciences*. He is the author and co-author of fourteen monographs. Jacek Blazewicz is an editor of the International Series of Handbooks in Information Systems (Springer Verlag) as well as a member of the Editorial Boards of several scientific journals. His science citation index reaches 2300 (according to the ISI data base). In 1991 he was awarded EURO Gold Medal for his scientific achievements in the area of operations research. In 2002 he was elected for the Corresponding Member of the Polish Academy of Sciences. In 2006 he was awarded a Dr.h.c. degree from the University of Siegen.
- Mikhail Y. Kovalyov** born in 1959 (Ph.D. and Dr. habil. degrees in mathematical cybernetics in 1986 and 1999, respectively, Prof. title in computer science in 2004), is a deputy general director of United Institute of Informatics Problems, National Academy of Sciences of Belarus, and a professor of Belarusian State University. His research interests include combinatorial optimization, scheduling, logistics and bioinformatics. He has over 100 publications in refereed scientific journals including *Mathematics of Operations Research*, *Operations Research*, *IEEE Transactions on Computers*, *IEEE Transactions on Computational Biology and Bioinformatics*, *IEEE Transactions on Systems, Man and Cybernetics*, *SIAM Journal on Optimization*, *IIE Transactions*, *Naval Research Logistics*, *Discrete Applied Mathematics*, *Operations Research Letters*, *Information Processing Letters*. His paper coauthored with C.N. Potts was selected for the list of 30 best papers published by *European Journal of Operational Research* in 1975-2005. His citation *h*-index is 14 according to Scopus and it is 16 according to Google Scholar. He is a co-author of one monograph and a chapter in a monograph. Mikhail Y. Kovalyov is an Associate Editor for *Omega* and *Asia-Pacific Journal of Operational Research*, a member of the editorial boards of *European Journal of Operational Research* and *Computers and Operations Research*, a member of EURO working group on production management and scheduling, and vice-president of the Byelorussian Operational Research Society.
- Jedrzej Musial** born in 1982 (M.Sc. in Computer Science from the Poznan University of Technology - 2006) is a Ph.D. student, researcher

and lecturer in the Institute of Computer Science of the Poznan University of Technology. His research interests include electronic commerce, algorithm design and combinatorial optimization. He has published several papers. Since 2008 he is a Board Member of Polish Information Processing Society (Major Poland Branch).

Andrzej P. Urbanski graduated in Computer Science from the Poznan University of Technology, Poland, and obtained his PhD in the same field from the Polish Academy of Science. Actually he is a Researcher and Lecturer in the Institute of Computer Science of the Poznan University of Technology. His research includes the fields of Computer-Aided Design, Artificial Intelligence, and Electronic Commerce. He has been publishing and presenting papers at many European scientific conferences and has also published articles and books popularizing Computer Science, especially for children, with some literary setting. Some of his computer programs have been marketed and were reviewed in popular media, among others by Polish TV networks, Deutsche Rundfunk, BBC, Die Welt, New York Times, and Canadian CRBC.

Adam Wojciechowski graduated from Poznan University of Technology (PUT) in 1995 with M.Sc. in Computer Science. In 1997/98 he used to work at Dublin City University, Ireland. In 2002 he achieved Ph.D. degree in Comp.Sci./Software Engineering at PUT where he currently works as an assistant professor. His research interests include software engineering, electronic business, optimization issues on financial markets and distant education. He is an author or co-author of over 12 papers published in international scientific journals and int'l conference proceedings. He also prepared Polish handbooks for ECDL applicants and published distant education course for webmasters.

Received: