

# Computer Scientists Find Unexpected Depths In Airfare Search Problem

By Sara Robinson

Visiting a travel agency in the early 1990s, Jeremy Wertheimer, then a graduate student in artificial intelligence at MIT, was astounded at the archaic software used to search for airfares. With some other MIT graduate students, he and Carl de Marcken decided to bring the forces of modern computer science to bear on the problem; the group set out to build software that could quickly and easily find the best airfare for a given set of constraints.

The task they set for themselves, however, proved to be far harder than any of them could have imagined. As de Marcken later discovered, the airline industry's pricing system has created such a complex web of fares and associated rules that the problem of finding the cheapest airfare from point A to point B is unsolvable. Even when the route or the flights are fixed, the problem is computationally hard.

"When we started we assumed it was path planning," says de Marcken, whose PhD is in natural language processing, another complex task. "You think you'll write your algorithm in a day and a half and you're done, but then you see the prices and you realize you have another ten years of work ahead."

The group persisted with their project, however, and today, Wertheimer heads a company called ITA Software, which powers searches for some of the systems used by travel agents, Orbitz (a fare-searching site owned by five of the major airlines), and Delta Airlines. ITA searches a far larger subset of the space of possible fares than other systems, claims de Marcken, now ITA's chief technologist, but it cannot guarantee the cheapest fare: "Nobody can do that," he says.

## The Airline Pricing System

What makes air travel pricing so complicated, airline executives claim, is that there is no one price they can charge for a seat and still turn a profit. Either the price would be too high and the airline would not attract enough passengers to fill the plane, or the plane would be full but the airline would not cover its costs.

As a result, the airlines have developed a system in which business travelers pay more for flexible tickets and convenient times, and bargain seekers willing to accept restrictions fill the remaining seats. "They have separated the cost from the price of flying," de Marcken says. "It's a very strange accounting procedure."

The basic unit of pricing, a "fare," is defined to be the price of one-way travel between two cities, regardless of the number of flights involved. Lists of fares are updated by the airlines ten times daily.

With each fare comes a set of rules for its use. Low fares, for instance, often have requirements like two-week advance purchase, a connection, or travel at an inconvenient time of day.

Rules of another type restrict the way a fare can be combined with other fares. Fares are put together in combinations known in industry lingo as "priceable units." A priceable unit is a collection of fares and associated flights that have one of several possible geometries: a one-way or round trip, for example, or an open-jaw or circle trip. Still other fare rules could require a round trip with a Saturday night stay, or use in combination with an international flight.

Complicating things further, when advertising a new flight or competing with a low-cost regional airline, an airline will often offer steep discounts for a certain route. To compete with the upstart JetBlue, for instance, American and United recently offered \$100 one-way fares between Oakland airport in California and Washington Dulles. For a time, then, a cheap way to fly from Washington to Portland, Oregon, was to make a stop in Oakland.

De Marcken recalls one airline's offer of a cheap last-minute fare to London from the East Coast. Because the fare allowed passengers to leave from Boston and return to New York, travelers searching for bargain last-minute fares from Boston to New York would have done best by including a stop in London. "This sort of thing is common," he says.

As a result of the complex rule system, air travel planning is far more complicated than a shortest-path search. A cheap airfare from A to C cannot necessarily be created from cheap fares from A to B and from B to C—a fare from A to B may not be combinable with a fare from B to C. Even worse, the price of travel from A to C could fall dramatically if the itinerary included a hop to D, a destination far from the path from A to C.

Because the airlines tend to hire information technology specialists rather than mathematicians or computer scientists, de Marcken says, the group from MIT found no one who understood the difficulty of the problem. The ad hoc fare search systems in use often simplified things, by finding only solutions involving a maximum of two flights, say, or by using a preset table of standard connections.

"The people who set the prices are not mathematicians; they don't know how complicated it is," de Marcken says. "They believe that if a computer can't always get the cheapest answer, it must be the fault of the computer programmer."

## A Complex Endeavor

The computational complexity of the airfare search problem is a result of the airlines' complex rule system: Rules associated with a fare used to pay for a single flight can restrict every other fare and flight included on the same ticket. Constraints of this sort make

the air travel problem similar to that of finding a satisfying assignment for a Boolean formula in many variables.

This problem, known as satisfiability, is NP-hard, so that the existence of an efficient algorithm to solve it would imply the existence of efficient algorithms for a whole class of seemingly intractable problems.

When he first looked at the rule system, de Marcken decided that the airlines couldn't possibly know what a tangled mess they had created. For his own amusement, he decided to demonstrate it formally. Using the techniques of complexity theory, de Marcken showed that for a given fare, it can be NP-hard just to know what existing flights satisfy restrictions. But even if the flights are fixed, the problem of choosing fares to cover those flights is NP-hard.

De Marcken then showed that the general version of the search problem, with no restrictions on the route or the number of stops between endpoints, is actually undecidable, meaning that no computer program designed to solve the problem will be guaranteed to halt with the correct answer on all inputs. The undecidability stems from the geometric constraints caused by the grouping of fares into priceable units. This allows the price of one part of a ticket to be affected by others that are seemingly disjoint from it.

While de Marcken's results do make unrealistic assumptions—arbitrarily long flight paths or lists of rules, for instance—the results still indicate that because of its basic structure, the problem defies attempts to find simple solutions. “Even if you put a bound on the number of flights, the constants (in the running time) are very large,” he says.

De Marcken points out, for instance, that airlines can have thousands of fares, including companion fares, with different sets of associated rules for each leg of a trip. Thus, if two people take a round trip together, with three flights going and coming, there can be as many as  $1000^{12} = 10^{36}$  fare combinations.

## A Solution

De Marcken's demonstrations indicated to the researchers that there was no hope of doing an airfare search without significantly restricting the problem. The program they developed, de Marcken says, looks at more options than existing programs do, but it doesn't look at all possibilities.

ITA's algorithms use techniques from natural language processing, the specialty of two of the four founders. Such techniques make frequent use of dynamic programming to break searches into smaller pieces. Dynamic programming is an efficient way to solve optimization problems that can be broken into overlapping subproblems. With the overlapping pieces solved only once and the answers put into a table, the overall computation becomes significantly more efficient.

Air travel planning has no simple breakdown into subproblems; still, clusters of fares with similar rules and other structures make it possible to have some computations done once and then used repeatedly. So many factors must be considered, though, that no simplifications are straightforward, de Marcken says.

He compares the air travel problem to his former specialty, a subfield of artificial intelligence: “In natural language processing, there are lots of nice dynamic programming algorithms that make the assumption that your language is context-free. Of course, real language isn't context-free, and that's where the interesting challenges come up.”

The core engine for ITA's software consists of 200,000 lines of Common Lisp, used by the artificial intelligence group at MIT. Lisp enables the programmer to think at a higher level, but at a cost of efficiency. Thus, to make the algorithms run faster, ITA has optimized the code at a lower level. Thousands of Linux computers run the software at many sites worldwide, answering millions of queries per day.

ITA's software seems to work quite well, judging from a very unscientific test comparing Orbitz with several other popular Internet travel sites: Expedia, Travelocity, Cheap Tickets, and the United and Northwest Web sites. A search for the cheapest possible round-trip ticket from White Plains, New York, to Billings, Montana, departing on May 23 and returning May 30, brought up fares ranging from \$790 on Orbitz (United through Chicago to Seattle, then Alaska Airlines to Billings) to \$1892 from the Northwest Web site. The United Web site was unable to find any ticket at all for this itinerary. Travelocity and Expedia, however, found fares of about \$1400 with only one stop on Northwest, and Cheap Tickets found a \$1200 fare with one stop on United.

Not all search sites have access to the same flight data; Southwest, for instance, doesn't participate in the general search sites, and airlines often offer special deals through a particular site. Still, in this case, separate searches revealed that all the general search sites had access to all the flights and fares found by Orbitz. It seems that their algorithms just didn't put the information together.

## A Move to Simpler Pricing?

The pricing system is so arcane, and the discrepancy in the results of the various search sites so large, that one wonders whether the obscurity is actually a deliberate tactic on the part of the airlines. Asked whether the airlines might be benefiting from the complexity of their system, de Marcken says he thinks not.

At the beginning of their work with the airlines, he says, the ITA researchers worried that the airline executives might not be receptive. What they found was just the opposite. “They would love the problem to be simple so anyone could implement an algorithm, but this is at odds with the need to make the pricing system depend on the market,” de Marcken says.

The airlines' only problem with the new system, he adds, is that it's now harder for them to set their prices. “Now that there's software that lets you put together little puzzle pieces, the overall effect of any small change is harder to understand.” Also, prices are coming down because the airlines are having a harder time competing.

De Marcken doesn't think this will pose problems over time: The airlines can just raise their prices across the board if they are losing money. “Right now, passengers using better software get better prices,” he says. “If you use cutting-edge software your price will be better, but if everyone uses it, individual fare prices may have to rise to keep the aggregate prices level.”

Meanwhile, ITA has taken on new challenges. The company is working with the airlines to help them bring a better understanding of global effects into their pricing system. They've also just finished a beta-version of an international fare finder; one of its features

is a sophisticated interface for showing a very large number of query results. (Orbitz currently uses another company's software for international searches.)

Improvements in search software, exposing the complexity and non-uniformity of the airline pricing system, could (travelers hope) spur the airlines to migrate to more elegant pricing models. In the worst case, though, airlines would be encouraged to exploit the complexity of their ticket system to keep control of the market.

Meanwhile, the pricing situation will continue to challenge computer scientists. "Some might take this as a vindication for algorithms, complexity theory, and computer science in general, but I tend to see it more as a lesson that marketers should never be allowed to design anything," de Marcken says, with a smile.

*Sara Robinson is a freelance writer based in Berkeley, California.*