

Threats to privacy

New science called data mining could benefit consumers — OR harass them / **Simson L. Garfinkel**

AN OFFHAND COMMENT last week from a vice president at Sabre Group Holdings sent shivers around the Internet. Sabre runs one of the largest travel reservation systems, used by 50 airlines and 30,000 travel agencies worldwide. And according to an article published by PC Week Online, Sabre was about to start selling the names and destinations of individual travelers — presumably for marketing purposes.

"We're sitting on a wealth of information," Steve Charpet, senior vice president of Sabre Technology Solutions, told the online news service. "Think about how much companies would pay for the names of people who have reservations to go to specific places at specific dates and times."

How much indeed! It's not hard to imagine that Sabre's information would be a treasure trove for marketers. You might make a reservation to fly to Greece, and then be bombarded by telephone calls and mail-ordergrams from tour operators in Athens. Then, when you return from your vacation, you discover that your home has been burglarized by thugs who also had access to the database.

Protests against Sabre's plans followed from consumer groups, privacy activists, and even from several attorneys with the Federal

Trade Commission. Most said the same thing: Making individual flight records available to third parties would pose a tremendous risk to some travelers, and would potentially expose travelers to unwanted marketing efforts.

Then, just two days later, the Sabre Group issued a statement essentially claiming that PC Week had gotten the story wrong. What Sabre actually plans to sell, spokeswoman Jennifer Hudson told me, "is aggregate

anonymous information. We're not talking about personal information or anything that would identify the individual."

What many people don't realize is that carefully processed aggregate information is potentially far more valuable to businesses than individual travel plans, thanks to a variety of sophisticated database analysis techniques that have been developed in recent years. That's because the most you can do with a person's individual travel plans is try to make another sale. But in the right hands, aggregate information can be used to dissect a business's day-to-day operations, pinpoint problems, and improve overall service.

Consider Sabre. For many business travelers, Sabre lists every flight taken, the kind of car rented, and the hotel stayed at. The database also knows how much was paid for the ticket, the car, and the room. It even knows who is paying the bill.

A simplistic way to use this information would be for a hotel to buy the names of all travelers who were planning to stay at competing establishments, and then send them coupons to stay elsewhere for less money. It's this form of marketing that privacy activists are worried about.

Marketing of this type is also problematic from a business standpoint, because it ignores one of the most valuable things in the Sabre system: the historical rec-

ord of how individual preferences and plans have changed over time.

Here's another way that the Sabre database could be used: If a hotel discovered its occupancy rate was down, it could ask Sabre to find out where all of its customers had gone. Using a new generation of on-line analysis tools, Sabre could do a database search for all of the people who stayed at the hotel last year and then see where those people traveled this year.

If they went to other hotels in the same city, then competitive advertising and special deals might make sense. But if a hotel in Boston found that its customers were going to other cities, then it might make more sense to invest in advertising that boosts Boston as a destination.

With a little more processing, a database like Sabre's can even

be used to predict the future. Lots of businesses try to project future sales by comparing orders in the current year with orders in previous years. But using the cross-sectional information that's inside Sabre's data warehouse, it's possible to take many more factors into account, and thus build a statistical model that's far more accurate.

The techniques I'm describing fall under the broad category of data mining. It's a new science made possible by ultra-fast computers that can handle massive amounts of information. Two leaders in the field are Burlington-based Thinking Machines (www.think.com), which has found the market for data-mining software far more lucrative than the massively parallel supercomputers the firm once built; and Virginia-based MicroStrategy (www.microstrategy.com), which sells data-mining products and recently completed its initial offering.

Data mining is an incredibly powerful technology that could benefit businesses and consumers alike. Run the systems on medical records, and one can tell which drugs work better on which cross sections of the population. Run it on credit-card records, and one can learn which restaurants have the most repeat business.

Data mining understandably worries privacy advocates. That's because it requires huge databases of personal information to operate properly. Today there is no guarantee that a data-mining firm will use this information to benefit consumers, rather than to harass us with targeted advertising. Experience has shown that legislation may well be the only way to secure such a guarantee.

Technology writer Simson L. Garfinkel can be reached at plugged-in@simson.net.

