Cryptography

The non-denial of the non-self

How philosophy can help create secure databases

IN THE 1940s a philosopher called Carl Hempel showed that by manipulating the logical statement "all ravens are black", you could derive the equivalent "all non-black objects are non-ravens". Such topsy-turvy transformations might seem reason enough to keep philosophers locked up safely on university campuses, where they cannot do too much damage. However, a number of computer scientists, led by Fernando Esponda of Yale University, are taking Hempel's notion as the germ of an eminently practical scheme. They are applying such negative representations to the problem of protecting sensitive data. The idea is to create a negative database. Instead of containing the information of interest, such a database would contain everything except that information.

The concept of a negative database took shape a couple of years ago, while Dr Esponda was working at the University of New Mexico with Paul Helman, another computer scientist, and Stephanie Forrest, an expert on modelling the human immune system. The important qualification concerns that word "everything". In practice, that means everything in a particular set of things.

What interested Dr Esponda was how the immune system represents information. Here, "everything" is the set of possible biological molecules, notably proteins. The immune system is interesting, because it protects its owner from pathogens without needing to know what a pathogen will look like. Instead, it relies on a negative database to tell it what to destroy. It learns early on which biological molecules are "self", in the sense that they are routine parts of the body it is protecting. Whenever it meets one that is "not self" and thus likely to be part of a pathogen, it destroys it. In Hempel's terms, this can be expressed as "all non-good agents [pathogens] are non-self".

The analogy with a computer database is not perfect. The set of possible biomolecules is not infinite, but it is certainly huge, and probably indeterminable. The immune system does not care about this, because it has to recognise only what is not in its own database. Make one adjustment, though, and you have something that might work for computers. That adjustment is to define "everything" as a finite set, all of whose members can be known—for instance, all phrases containing a fixed maximum number of characters.

A database of names, addresses and Social Security numbers (a common form of identification in America) might require only 200 characters to contain all possible combinations. That would limit the total number of character combinations. A positive database containing all the data in question would be a small subset of those combinations. The negative counterpart of this database would be much larger and contain all possible names and addresses that were not in the positive database plus a lot of gibberish. But it would not be infinite. By looking at the negative database, it would be possible to deduce what was in the positive database it complemented.

That would not guarantee security against a search for the presence or absence of a particular name and address. Indeed, the whole point is that such searches should be possible. But it would prevent fishing expeditions by making it impossible, for example, to look for the Social Security numbers of all the people living on one street.

Dr Esponda sees great potential for using negative databases when there is a need to look at the intersection of many sets of data owned by different parties. Two or more banks, for example, might wish to work out which transactions they have in common without revealing the whole contents of their databases. Using negative databases to do this would, according to Dr Esponda, provide a robust back-up to traditional cryptography, which relies on codes that can be broken.

An interesting extension of the idea might be to use negative surveys to collect sensitive information privately. Dr Esponda gives the example of a negative survey in which respondents are asked to tick the box of one sexually transmitted disease they do not have. He reckons that this would be sufficient to estimate the population frequency of each disease, without having to ask people whether they actually suffer from such diseases-which is intrusive and also invites lying. As he puts it: "In Hindu philosophy, to find out who you are, you ask what are you not. Then you are left with what you are." •

Supernovas

This is a photograph of Cassiopeia A, the gaseous remains of a supernova that would have been visible from Earth in about 1667. Oddly, no one seems to have noticed it at the time and it was first detected in 1947 as a radio source. Even then, although it is the brightest radio source in the sky outside the solar system, optical telescopes could not locate it for several years. The image was taken by the Hubble space telescope. Comparing it with ones taken a few months ago enables the process of expansion to be studied. Cassiopeia A, despite having been missed by 17th-century astronomers, is a mere 10,000 light years away. Contrast that with a supernova, described in this week's *Nature*, that is 440m light years away. Astronomers have much sharper eyes at their disposal now than they did four centuries ago.

Fonte: The Economist, v.380, n.8493, p.72, 2 Sept. 2006.