If only I could see a picture..

T he collection of data, its electronic storage and its conversion into Information is an enormous challenge.

The BT network is one of the largest machines ever built, and collects data in enormous amounts. Approximately 6.8Gbytes of information is collected each day and every three months this is amalgamated into phone bills that need to be completely accurate and easily understood by our customers. Although huge this is a quite simple and easily structured device. But in a system so large, there will be fraud and fraudulent activities – there are enough villains out there and they are smart enough to guarantee it! How do you look for something that you do not know if it exists?

The big problem is to originally analyse exactly what fraud is being perpetrated. Just like in all law and security matters the villains can establish an activity of their choosing, at a time of their choosing, the law enforcers need to be aware of everything and at all times. What is needed is to be able to look for patterns in these enormous amounts of data. Just such a case arose recently.

In the "old days" when creating data bases you had to have a pre-conception as to the types of enquiries you were going to make in the future and then construct the data in a form that could be so questioned. As data bases grow (and this growth is approximately the cube of the number of data entries involved) such pre structuring becomes not only impossible but also inadvisable.

In this case the enquiry of the data was "how many of the phones in the London area share at least 50% of their calls with other London area phones". The resulting data was turned in a coloured picture. The phone line identities being on the circumference of circles and the width of the lines between them indicated the proportion of the calls being made. Very rapidly the pictures developed into recognisable patterns to the human observers. We humans are quite remarkable at recognising patterns. The eyes in the front of our heads can handle 1Gbyte of data each second and we possess a "wet-ware" computer in our heads that can process data at these rates!

The pictures clearly showed a scam of a group setting up a Premium Line number then exercising it themselves. In such arrangements BT bills the customer and a proportion of that money is paid to the owner of the Premium Line Service. We paid the villains on a monthly basis as a business contract; and they would have paid us (honest) on a quarterly basis if they had still been at the address as a domestic contract! The case went to court and the "best evidence" in a UK court, for the first time, was a set of pictures not the written word. The data was so complex that a jury would have been totally unable to comprehend it as a listing

of raw data.. The pictures made the case easily understandable. In a similar way the discovery of people in a war torn part of the world ringing London and auto-call forwarding calls to their colleagues across the war zone has caused this type of behaviour to be constantly monitored throughout the network.

But data bases are getting more complex and dynamic. Just image a system that holds all your preferences, and these are tested against all the preferences of other people within a given distance even when you are moving around town. It could be the ultimate "Dating Agency"! As soon as someone within the given radius fits your desired profile, and you equally fit theirs, you each receive notification via the handset of your 3rd generation mobile phone. Then it is up to you if you take any action, or not.

This is not only an incredibly large database but one which is extremely dynamic as well. One company APAMA (www.apama.com) has inverted that thinking on data bases. The original way of looking at data bases was to have the data as a static unit and dynamically interrogate it by enquiries. But in the larger data bases the data is in practice quite dynamic and the enquiries mostly static. In the "dating agency" above one's preference are unlikely to change that much or that quickly!

By passing the dynamic data past the static enquiries massive interrogation rates can be achieved and handled by relatively simple data base engines.

Just imagine… on any inbound or outbound call you could see all the information of customer, their previous interactions, purchases and preferences as a simple picture.