Privacy

Ben Laurie
(ben@algroup.co.uk)

December 17, 2004

1 What is Privacy?

If you view life as a sequence of transactions, then privacy can be defined as the linkability of the transactions. Absolute privacy occurs when none of the transactions are linkable, and zero privacy occurs when they are all publicly available and completely linkable.

2 What is a Transaction?

A transaction is an interaction between two or more parties (I suppose that if you were to define a calculus of transactions you could always break n-way transactions into at most $n - 1$ 2-way transactions, but I do not intend to construct such a calculus at this stage). For example, buying something from someone, voting for someone (the second party being the vote-counter), reproducing (at least two transactions here - impregnation and birth - plus a series of other transactions by the mother - exchange of nutrients [hmmm – a continuous process], ultrasounds, endless advice from health professionals...), getting paid, and so forth.

3 What is Linkability?

Linkability is the ability to assign a probability that two transactions share a known[1] party.

4 The Asymmetry

Notice that there is an asymmetry between absolute privacy and zero privacy – in the absolute case, the transactions merely need to be unlinkable, whether publicly available or not, but for zero privacy they must be both available and linkable.

This implies that there are two routes to privacy: unlinkability of transactions and unavailability of them.

5 Partial Privacy

Of course, between the extremes is what we all expect to have: partial privacy. We traditionally think of this as being achieved through the unavailability of transactions, but the modern world is driving us towards unlinkability as a mechanism, since it is so difficult to ensure transactions are unavailable.

[1] Known in the sense that we know which of the parties to the transactions is the same.

6 "Authentic Privacy"

I still have not understood what this is supposed to be. Privacy is either privacy or it is not. How can it be unauthentic? Nevertheless, an interesting question is posed: at what point does "more" privacy actually act to our disadvantage.

Examine the extreme position, absolute privacy. If all transactions are unlinkable, this means we have no identity. No reputation. We can't have meaningful elections: not only will we be unable to identify who we are voting for, the vote-counters cannot know who the electorate are, nor whether anyone voted twice. We can't know our children or our parents or our significant others – or perhaps I should call them insignificant others.

Zero privacy is equally unpalatable. Perhaps we can still keep our innermost thoughts to ourselves (after all, at least so far, they are not transactions), but everything else would be available to everyone. Every stupid mistake, every indiscretion, every unguarded utterance.

Where is the happy middle ground? I don't think I know, but it must be founded on an ability to control both the availability and linkability of transactions. Of course, one of the interesting aspects of transactions is that we have no real control over the actions of the other party, and hence if we want certainty of privacy we must use unlinkability as our weapon. But there are many cases where, in practice, we can rely on unavailability.

The two obvious mechanisms to achieve unavailability are regulatory – the other parties are simply not permitted to reveal the transactions, and blackmail – it will cost the other parties more than they gain to reveal the transactions.

Unlinkability and unavailability are achieved through cryptographic means too obvious to this audience to be worth describing.

7 Comparing Privacy

Implicit in the idea that there is a certain level of privacy which is "good" and that beyond this level "bad" things start to happen is the notion that two levels of privacy are comparable. However, it seems pretty obvious that privacy is not a total ordering. For example, given four events, A, B, C and D, it is pretty clear that having A linked to B, B linked to C, and C linked to D is less private than having A linked to B and C linked to D. But how does the latter compare with, say, A linked to C and B linked to D?

Similarly, is A linked to B better or worse than A linked to B and C linked to D, each with 50% probability?

8 A Privacy Calculus?

It seems possible to combine these ideas to provide some kind of formal measure, or family of formal measures, of privacy, and a calculus for expressing them, but is this an interesting thing to pursue? What is the point of "measuring" privacy?