Risk in the Real World of InfoSec

Michael Barwise

"Risk" is a much abused word, and most of us use it quite casually without really considering its significance. Statistically (and that's the origin of the concept) risk is the arithmetical product of the likelihood (or statistically speaking, the probability) of an event and the value of the outcome of that event. In the real world, however, the whole secret is to define the value and the likelihood in terms that can multiplied together to yield a result that actually means something useful.

To yield a useful result, two factors have to be taken into account. Firstly, is the likelihood predictable? If it's not, we're not able to determine a risk - only an uncertainty. This is a very, very important distinction. If, for example, we consider the probability of a 30-year old male UK citizen being a specific height, we can refer to masses of data based on actual measurements, from which we can build a probability distribution - a table of the proportion of men aged 30 in the UK who fall into each of a set of ranges of height. This probability distribution can to used to predict the likelihood (or probability) of our subject being within one of the ranges of height. If we're considering a type of accident the severity of which depends on the height of the victim, we can use this probability as one of the factors in assessing the typical outcome provided other relevant factors have been taken into account.

So far so good. But supposing we either don't have enough data or the phenomenon we are considering is completely unpredictable, there's no possibility of creating a probability distribution, so we can't assign a likelihood to the event. What we can often establish, however, is a measure of our uncertainty, which is useful, but not the same thing. You can still use uncertainty as a measure of risk, but it's not the same kind of risk as that based on probability. The former is an indicator of what might happen, but the latter is a measure of our ignorance of the position. Confusing the two is potentially lethal. Let's suppose for example that a critical concentration of flammable gas in a room becomes explosive. A fire-fighter who knows the conditions - the size of the gas pipe, the time it's been leaking, the volume of the room, the air exchange - can establish the risk an electrical spark might present based on probability. But a fire-fighter who knows none of these factors must "take a risk" based on a measure of his lack of knowledge.

The second factor that must be taken into account is the way value is determined. It's not as easy as it might seem, as there are always indirect contributors as well as direct contributors to value. So, for example, corporate customer data may have a direct value in terms of the revenue that derives from its use, but it will probably also have indirect value in terms of the cost of collecting it, regulatory penalties for losing it and so on, and the aggregate indirect value may substantially exceed the apparent direct value.

But now we come to the critical issues. First, which of these two kinds of risk are we dealing with? Second, how are we representing likelihood, uncertainty and value? And third, at what point in the system is the risk being determined?

In infosec, we are primarily dealing with risk based on uncertainty. This is mainly because the common threats are purposive in nature - they're driven by the whims of people and off-chances of exploitable weaknesses in systems and processes being found, and both those factors are pretty much unpredictable. They may go in waves (fashions) but they're mostly the result of other peoples' impulses and other peoples' luck, whether it be of individuals or the criminal fraternity. So we're mostly dealing with uncertainty-based risk, and that's hard, particularly in the common scenario where the risk practitioners don't understand the difference between uncertainty and probability. This often leads to corporate risk assessment being about as reliable as fairground clairvoyance.

Many corporate risk practitioners bin likelihoods and values into three categories - High, Medium, Low - or five at best. This simplifies the assessment process, but if it's done arbitrarily or too early in the determination of likelihood or value it can seriously distort results. Binning should be the very last phase of a rigorous, documented evaluation based on evidence. But even then it must be done with caution. The distribution of value within a corporation is practically never even - it's usually heavily skewed and can have significant outliers at its extremes. And similarly, rare extreme events are by definition outliers. So if you lump these outliers into your High category along with masses of much less extreme events and values, you'll tend to under-emphasise their potential significance. Then there's the not insignificant matter of how you determine the boundaries between your three of five

bins - not only which bin does a particular likelihood or value go into, but are your bins equally disposed across your value and likelihood ranges, or are they exponentially disposed. Or are they just based on wild guesswork, politely labelled "brainstorming"?

Finally, many organisations concentrate on infosec risks at a technical level, whereas they should be considering the impact on the business first and only descending to the technical level if it proves necessary. This approach is more likely to provide robust proactive resilience, whereas concentrating on the technical risk tends to result in reactive security that chases the specifics of emerging attacks. That's a much more expensive and fragile approach to security.

Circulated as a white paper, November 2010