Threat-Based Security Engineering (TBSE)

John Leach

John Leach Information Security Ltd; Innisfree, Stoke Road, Smannell, Andover Hants SP11 6JL England

Abstract: TBSE is a risk modelling technique which models the dynamics of security interactions analytically making it possible to forecast risk in numerical form. It applies non-deterministic techniques to calculating the probability distribution of specific security outcomes as a direct function of the measured threat profile and countermeasure settings. Security engineers can design security solutions which provably meet QoP targets across a specified range of threat levels, and can optimise security settings to minimise cost or operational impact. Technical managers can perform precise cost/benefit analyses and can demonstrate compliance to policy or regulatory mandates objectively. TBSE can be applied to any threat (physical or logical, accidental or wilful, internal or external) and any security measure (technical or non-technical). This paper introduces TBSE and shows some of TBSE's early results. It briefly points to the potential impact of TBSE on the future practice of Information Security.

1 Introduction

Information Security has long been regarded as more an art than a science. We are all familiar with the problems this brings. Business management would like to perform cost/benefit analyses based on meaningful numerical values for protection and risk, and to be able to demonstrate to stakeholders that the organisation is compliant with internal and external mandates. Security practitioners would like to have clear and objective QoP targets set for the security solutions they are charged to provide, and then would like the ability to show whether the security solutions they design satisfy those targets within a specified threat range. None of these things can we do.

People understand from common sense and experience that security measures protect against threats, and they have an intuitive expectation that applying more security, whatever that might mean, should lead to the information assets being better protected, however that might be measured. However, looking beyond intuitive expectation, there are no ready techniques available for calculating the degree of protection provided by security measures and for quantifying the risks which result. Hence, people are left to build security solutions based on experience and on "best" or established practice. Each person is left to decide for themselves which security measures provide the most benefit and to judge what depth of defence is needed to achieve adequate protection.

2 The Form of Modelling Solution Needed

These inabilities and shortcomings arise because of the lack of any general techniques for describing or modelling the dynamics taking place between threats and security measures, the dynamics which lead to security breaches occurring. There is no general method with which to describe how any given security measure engages with the prevailing threats, how it modifies or counters the activity of those threats, and what the probabilities of the various possible outcomes might be as a result.

What form should a potential solution take? It would need to be a general purpose model with which one could describe the end-to-end risk process, i.e. the various processes or interactions by which security attacks are generated, those attacks engage with the target information asset, breaches do or do not occur, and how those breaches could lead to disruptions of the system and operational damage to the asset owner.

A general purpose model of this kind would allow the external inputs (e.g. the threats) to be described in a suitable form, descriptions created of how security measures are deployed, the effects of those measures on the parts of the risk process with which they engage to be quantified, and the arithmetic to be performed as needed to produce numerical results from the threat/security measure interactions.

The numerical results should be in a form from which the probabilities of particular specified outcomes occurring can be calculated. Outcome probabilities might be provided in the form of a single number, i.e. an overall probability value, or more usefully as a function of one or more relevant parameters, i.e. a probability distribution describing how the probability of the outcome varies as a function of relevant attributes of the threat and relevant parameters for the security measures applied.

Such a general purpose model would enable security risks to be measured and managed directly and reliably. One would measure the threats of interest and profile them according to relevant parameters. Using the model, outcome probabilities and appropriate risk indices would be calculated for any given security measure deployment. The security measure parameters within the model would be varied and the effects of those variations on the resulting risk indices calculated.

Information asset owners would set QoP targets for their assets, specifying the limits on outcome likelihood or impact they would be prepared to tolerate. Security engineers would determine the security measures and settings required to satisfy those QoP targets given an expected level of threat. They would calculate how the risk indices would vary across a range of likely threat profiles, and how the countermeasure settings should be adjusted for those QoP targets to remain being satisfied. Once security measures had been deployed, regular measurement of the actual threat profiles to which the information assets were exposed would allow security settings to be adjusted as needed for the information assets to remain continuously protected to the asset owner's satisfaction.

Reaching this goal would be a landmark achievement, enabling security to be deployed with confidence and its benefits to be understood in business terminology. Spending decisions could be made based on reliable data, and security measures implemented to provide a specified level of protection according to business need and budget.

3 TBSE and Some Examples of its Results

The requirement, then, is for a methodology and techniques with which to model the end-to-end dynamics between threats and security measures, the dynamics through which security outcomes and risk are created. There may well be several possible routes by which to achieve this. This paper will describe one approach, TBSE, which is just such a modelling technique and which has been able to achieve the type of results described above using real-life threat data. This paper will describe some of these results, and then describe TBSE in outline and indicate its potential.

TBSE is a general purpose technique for modelling security interactions analytically and calculating outcome probabilities in numerical form. It employs non-deterministic modelling techniques for solving security risk problems. It can be applied to any threat and any security measure, and can model multiple threats and multiple countermeasures working in parallel.

In principle, TBSE can be used to build customised risk models which describe corporate information infrastructures and to calculate a wide variety of risk indices for the many threats and many security measures all working in parallel. In practice, it would be a complex exercise to build such sophisticated models in one step. For this reason, TBSE has, in the first instance, been applied to a number of simpler problems. Three such risk problems and their results will be described in this paper.

3.1 E-mail Viruses and Anti-Virus Software

The first real-life problem to which TBSE has been applied is that of quantifying a user's risk of e-mail virus infection as a function of the way their anti-virus (AV) software is deployed. This has been a collaborative project undertaken with MessageLabs (http://www.messagelabs.com). The results will be freely available in due course on the MessageLabs' web site.

Visitors to the web site will be able to use the MessageLabs "Risk Calculator" to calculate the probability of an infected e-mail making it past their AV software as a function of how that software is deployed. Results are provided in a variety of numerical forms to ensure their accessibility for the untrained user. The visitor will be able to determine their risk and exactly how that risk would vary if they were to change their AV settings. As a result, they will be able to work out accurately how to configure their AV defences to cut their risk by a half, by three quarters, by whatever level they choose.

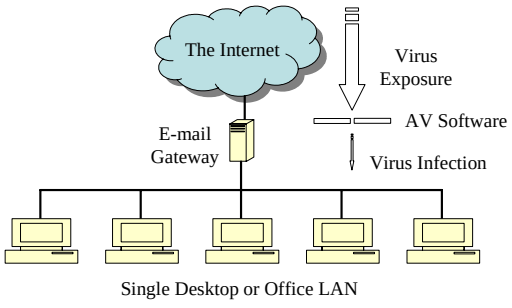

Fig. 1. The Scenario Modelled with MessageLabs

The scenario modelled in this work is shown in Figure 1. The threat from e-mail viruses is carried to the target (e.g. an

office LAN) by e-mails arriving at the e-mail gateway. Even though the gateway hosts some Anti-Virus (AV) software, there is always a small risk that one or other virus will get through to the target. TBSE is used to calculate that risk in objective numerical form as a function of the measured threat and the way the AV software is configured.

MessageLabs starts by measuring the threat and displaying that in the form of a real-time chart, as in Figure 2 below. The threat, the flux of e-mail viruses arriving at the target, is profiled as a function of its relevant attribute, the age of the virus (in hours) at the time the virus is received at the e-mail gateway.

The visitor is invited to enter a few simple details which describe the volume of e-mails they receive on a typical day and how often they check for new virus signatures. TBSE then calculates the probability of an e-mail virus getting past their AV software in those circumstances.

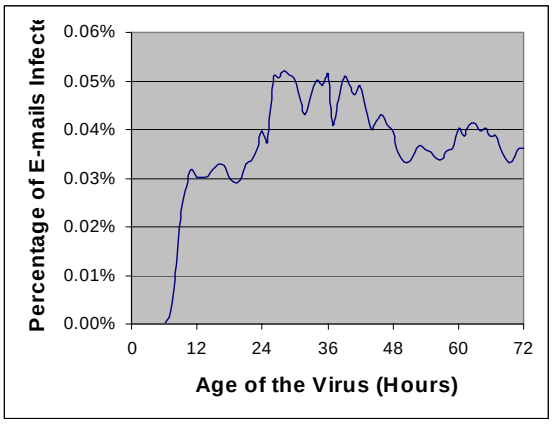

Fig. 2. The E-mail virus threat profile

The probability is given as an actual number, not in the common form of a High / Medium / Low estimate. Rather than showing the raw probability result, which would be a figure such as “one in 357,200 e-mails will carry a virus past your AV software given the description you have just provided”, a result the untrained user might be unclear how to use, the risk result is provided in three alternative numerical forms, with the primary form being the probability (to the nearest whole percentage point) the user would have gone three months without any infected e-mails making it past their AV software based on the measured threat profile. Hence, the user gets a meaningful result they can understand straight away and can work with.

The visitor is then given the opportunity to enter a different value for how often they check for new virus signatures. This lets them see exactly how their risk would fall if they were to strengthen their AV defence in that way. They can continue increasing their signature checking frequency until their risk is pushed down to a level they are comfortable with, at which point they will know exactly how often they need to be checking for new signatures if they are to achieve their desired level of protection. In practice, the visitor will have chosen a QoP target which expresses their tolerance for e-mail virus risk, and will have determined exactly how their AV protection needs to be configured to satisfy that target.

This project with MessageLabs shows how TBSE can be used to help people control their security risk in a direct and simple manner. It gives them an objective numerical figure for a particular risk, and allows them to adjust their risk to their chosen level by adjusting their countermeasure configuration in a specified way.

3.2 Worms and Software Patching

This second example uses TBSE to assess the benefits of software patching. It shows how one's probability of suffering a successful worm intrusion falls as the average time taken to apply a security patch is brought down, based on the measured profile of the worm threat.

The scenario being modelled is broadly similar to that shown for the preceding example except that the threat in this instance is the threat from worms coming in over the target infrastructure's Internet connections. Whereas the first example showed TBSE modelling a security defence which blocks attacks, this example models a security defence which reduces the target's susceptibility to the threat. It would be a simple extension to combine the two models to show the combined effect of the two different types of security measure working in tandem, a problem of obvious interest to every large Internet-connected organisation.

The work behind this example calculated a target's risk of a successful worm attack, for worms which exploit one (or sometimes more) Microsoft software vulnerabilities including, in particular, the LSASS vulnerability. The LSASS vulnerability was very popular with worm writers. The results for the LSASS vulnerability were compared with those for the average of ten other worm-exploitable vulnerabilities each of which was less actively used by worm writers.

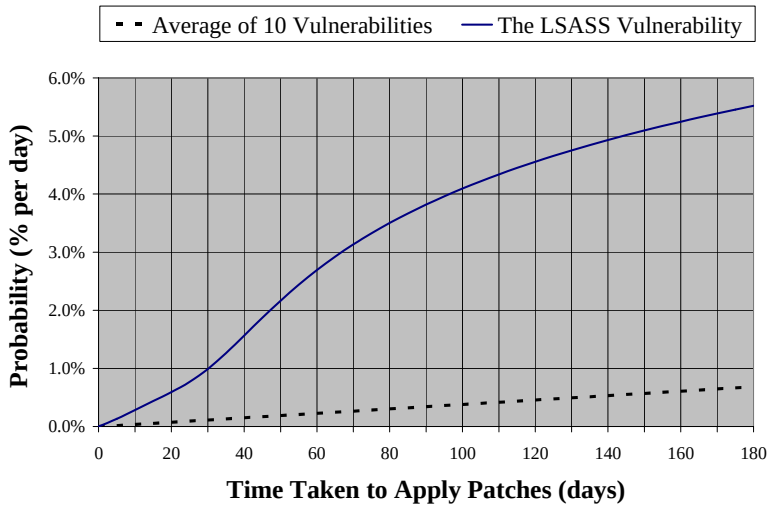

Fig. 3. The probability of a successful worm attack

Figure 3 shows the risk of a system suffering a successful worm attack as a function of the rate at which patches are applied (based on real-life threat data). This shows how the risk rises for an average-risk vulnerability as it waits to be patched, and how much faster the risk rose for the LSASS vulnerability. The modelling presumed a patch management process whereby all vulnerabilities are patched

according to a regular cycle; after a certain time period the system manager rolls up all the patches released since the previous round and applies them all at the same time after an allowance for testing.

For average-risk worm-exploitable vulnerabilities, the IT system manager on a 90-day patching cycle has a probability of suffering a successful worm attack due to an unpatched vulnerability (discounting any other countermeasures they might have in place) of 0.35% per day per exposed IP address. If they were to move to a 30-day patching cycle, their risk would drop to 0.15% . A sample of these results is given in Table 1 below.

Table 1. The probability of a worm successfully exploiting an average-risk software vulnerability for a range of patching cycle lengths

Patching cycle (days)	Probability per day	Probability per month	Probability per quarter	Probability per year
30	0.15%	4.0%	10%	35%
60	0.25%	7.0%	20%	60%
90	0.35%	10%	25%	75%

For a high-risk, i.e. an actively exploited, vulnerability such as the LSASS vulnerability, the risk is, naturally, much higher. The comparable results for the LSASS vulnerability are shown in Table 2.

Table 2. The probability of a worm successfully exploiting the LSASS vulnerability for a range of patching cycle lengths

Patching cycle (days)	Probability per day	Probability per month	Probability per quarter	Probability per year
30	1%	30%	70%	99%
60	3%	60%	93%	100%
90	4%	70%	97%	100%

Hence, if IT system managers do not have a way to identify which vulnerabilities are high risk, then even if they put in the effort to patch their system regularly on a 30-day cycle, they are still placing themselves at considerable risk.

The above scenario was modelled on the basis of the system manager not having a way to identify high-risk vulnerabilities. Of course, many organisations do get advised by their vendors or suppliers in advance whenever a new high-risk vulnerability is about to be publicised. In these situations, IT system managers are able to patch high-risk vulnerabilities on an accelerated patch management path. TBSE has been used to model this variation in the scenario to show by how much adopting a more flexible patch management approach reduces the risks.

Assuming users identify high-risk vulnerabilities and then patch those with only a short delay for testing, the results are as shown in Table 3.

Table 3. The probability of infection by a high-risk vulnerability for a range of patching delays

Time taken to install the high-risk patch (days)	1	2	3	4	5	6	7	10	14	21
Probability of being infected before the patch goes in (%)	1.8	3.6	5.3	7	8.5	10	11.5	15	20	32

These results show clearly the need for high-risk vulnerabilities to be identified and patched as soon as possible, and provide IT system managers with a guide to just how quickly they should aim to have high-risk vulnerabilities patched. These results allow the system manager to explore various options for how they might meet the system owner's QoP target for the system, and to ensure that they have not only adequate resources but also adequate flexibility if they are to adopt a twin-track strategy for patch management.

3.3 Unauthorised Behaviour by Staff

The examples above show TBSE being used to model technical countermeasures and technical threats. TBSE has also been used to model the interactions between non-technical countermeasures and the threat of staff knowingly violating a (written or unwritten) code of authorised behaviour. This threat is a broad one including, at the low-severity end of the spectrum, activity such as the unauthorised use of corporate IT facilities for personal purposes and, at the high-severity end, significant financial fraud. TBSE can be used to assess the effectiveness of different countermeasures at reducing the rate or severity of staff attacks, and can help the security manager select countermeasures to achieve specific effects.

Three countermeasures were analysed to assess their effects on reducing the rate and/or severity of attacks arising from staff misbehaviour. The three were: strengthening the security culture; security vetting; increasing deterrence. The results showed that:

- Strengthening the organisation's security culture reduced the rate of attacks more than it reduced the severity of attacks;
- Security vetting reduced the rate of attacks and the severity of attacks in broadly equal measure;
- Increasing the accuracy of security vetting beyond a moderate level requires more effort by the vetting organisation and is, perhaps, fairer on staff but appears to give almost no benefit in terms of the end results achieved;
- Increased deterrence had a strong effect reducing the rate of attacks but only a small effect reducing the severity of attacks.

Conclusions: Security vetting is very helpful, but only up to a point. Deterrence reduces the risk only to the degree that it reduces the expected rate of attacks.

Strengthening the security culture is the way to achieve the greatest risk reduction if an organisation plans to deploy only a single one of these three countermeasures. The most cost-effective risk reduction comes from deploying a mixture of these countermeasures with the main reliance being placed on security vetting and building a strong security culture.

These are signal results, never before achieved, and show how valuable risk modelling can be even in the absence of precise data.

The results from this third TBSE example were, in the absence of real data, based upon a number of hypotheses regarding the threat profile of staff and how staff modify their behaviour in response to the three different countermeasures. These hypotheses are believed to be sound and to provide a strong basis for the analysis performed. However, it is clear the above results should be treated as indicative, not definitive, until the assumptions on which they are based have been validated.

Those assumptions are eminently testable. By measuring the relevant characteristics of the threat population and by calibrating the way these personnel countermeasures work, TBSE analyses can be conducted using data representative of real situations. The results achieved would then be definitive results of highly significant value.

4 TBSE in Outline

The above examples show that TBSE has the power to address a variety of different scenarios and a variety of different modelling needs. The results can be used to develop numeric QoP targets for specific outcomes or for broad classes of outcome, and can be used to show the level of protection provided by a single security solution or several security measures working together.

TBSE is a fully general model applicable to all types of threat and all types of countermeasure. In the remainder of this paper, we will describe TBSE in outline and indicate how it could support new Information Security services and products.

TBSE is based around a model which allows us to create suitable analytic expressions for the various interacting components and which leads to tractable measurements and calculations. The model describes the creation of security incidents and the resultant impact of those incidents as a series of dynamic processes. This is shown schematically in Figure 4 below.

In the model, a population of Threat Agents generate a population of attacks. That population of attacks creates a population of security breaches, which themselves give rise to a population of system disruptions which lead to a population of damages. Each population is described in terms of a number distribution in relevant attributes. The processes by which one population engenders the next population in the chain can be described by appropriate analytic functions.

The ways in which countermeasures affect the dynamics of the processes in the chain are themselves described by suitable analytic functions. Different countermeasures work at different points of the chain.

– Some work to reduce the population of attacks generated by a population of threat agents. Examples include security vetting, security culture and deterrence, all of which work to reduce the population of attacks generated by misbehaving staff.
– Some countermeasures work to reduce the population of security breaches created by a population of attacks. Examples include firewalls, anti-virus software and software patching.

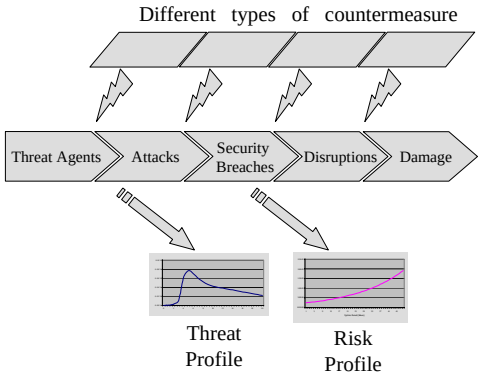

Fig. 4. TBSE Schematic

– Some countermeasures work to reduce the population of disruptions caused by a population of security breaches. Examples include intrusion detection and having a warm standby server. Intrusion detection doesn't stop an intrusion but makes it more likely action can be taken to reduce the severity of any disruptions caused. A warm standby server doesn't stop the main server having an outage but reduces the severity of the disruptions caused each time an outage occurs.
– Some countermeasures work to reduce the damage a population of disruptions leads to. Examples include having contingency arrangements for disrupted business processes, and insurance. Neither reduces the rate or severity of the disruptions which occur but each can reduce the degree of damage (whether measured financially or otherwise) those disruptions might cause.

TBSE can be used to model one link in the chain or several links in the chain, and to model one or more countermeasures operating within the chain. For each of the first two examples above, the threat (the flux of attacks reaching the target system) was measured and TBSE was used to model how the respective countermeasures influenced the likelihood of those attacks creating security breaches. In each example, just one link in the chain was analysed, and just one countermeasure in each link. We did not analyse how other anti-virus measures, e.g., training staff not to open suspicious attachments, or how other anti-worm measures, e.g., a firewall, might further reduce the population of security breaches caused. Each example could easily have been extended to cover additional countermeasures.

We also did not need to start from the top of the chain each time. Lacking information describing the population of virus or worm writers and any description of the rate at which those writers create viruses and worms, we simply measured the populations of attacks experienced. By hypothesising the rate at which those populations create viruses and worms, we could, if we wished, use TBSE to deduce the populations of threat agents needed to generate the profiles of threats observed. Alternatively, if we could measure the populations of threat agents, we could understand the rate at which writers write viruses and worms by deducing the

functions needed to create the measured threat profile from the measured threat agent population.

Thus TBSE allows us to model either just one stage of a particular threat chain or to model the whole chain, and to calculate either upstream or downstream components in the chain depending on our purpose and on which components we can describe or measure.

TBSE does not require the data upon which it works to be of the highest quality; it will work with precise or imprecise data as available. Clearly, the more precise the measurements of the input populations, or the more precisely an analytic function can be formulated, the more accurate the results should be. Some threats can be measured accurately with ease, others less so. Hence, results will be of greater or lesser precision accordingly. However, given the inability of today's non-analytic methods to produce results with any accuracy, even results which are accurate to only ± 50% would represent a significant improvement on the results otherwise available today.

5 E-mail viruses and AV software (Revisited)

Equipped with this initial understanding of TBSE, we shall revisit the first example described above and look more closely at how the results were generated.

Anti-virus software was modelled on the basis that it scans all incoming e-mails looking for any instances of a known virus signature[1]. The user's AV vendor continually releases signatures for new viruses as new viruses are reported, and the list of reference signatures held on the e-mail gateway is updated by the user periodically, sweeping up all the signatures released by the vendor since their previous update. The release of signatures by the vendor and the periodic update of signatures by the user are asynchronous; the user checks for new signatures with a given regularity irrespective of whether their vendor has released none, one or many new signatures in that period.

The probability of the user's AV software detecting a virus in an e-mail is presumed to be 100% provided that virus' signature is held in the software's local signature store. The probability of a virus being detected by the AV software at the moment the target is exposed to that virus then depends on the age of the virus at the time the system is exposed to it and whether, by that time, the vendor has released the relevant signature AND the user has picked up that signature through their signature update process.

If a user is exposed to a new virus before the vendor has released the signature, the probability of the virus getting past the user's AV software is assumed to be 100%. If the user is not exposed to a new virus until well after the vendor has released a signature for it and the user has had time to update their reference store, the probability of the virus getting past the user's AV software is taken to be 0%. In the intervening period, the probability is somewhere between 100% and 0%.

---

[1] The model used in our example contained no heuristic scanning component. This does not mean that TBSE has difficulty modelling heuristic scanning, just that heuristics were not included within this example. AV software which performs both signature-based and heuristic scanning could easily have been modelled with a simple extension of the model.

That probability curve can be calculated accurately and, clearly, it will be a function of the frequency with which the user checks for signature updates. The more frequently the user checks for updates, the more quickly the probability curve will rise from 0% to 100% with increasing age of the virus.

In the work performed with MessageLabs, we collected a large amount of data describing how quickly vendors released signatures after a virus was first detected. The data covered nearly twenty vendors and over thirty virus strains. From this data, we constructed a curve showing the probability of the AV vendor having released a signature as a function of the age of the virus, with the results being weighted according to vendor market penetration. This was combined with the frequency the user checks for new signatures to give the probability of the user having a virus' signature in their local reference store at the moment they are exposed to a virus, for any virus, as a function of two parameters, the age of the virus at the moment of exposure and the frequency with which the user checks for signature updates.

Interestingly, the results showed that, for a user exposed to a new virus before their vendor has released a specific signature for it, the probability of the virus getting past the user's AV software is NOT actually 100%. There is a probability of about 10% that one of the reference signatures already in the local signature file will be sufficient to catch the new virus. Whether or not a new virus is caught by an old signature varies from vendor to vendor, as different vendors build signatures in different ways depending on their differing methods and analysis of how each new virus works.

The probability of the user having a virus of a given age slip past their AV software is then the probability of the user being exposed to a virus of that given age and the probability that, when they are so exposed, their AV software will not have that virus' signature in its local signature list. The latter probability is obtained directly from the probability distribution we have just described. The former, the probability of exposure, is obtained by measurement of the threat.

MessageLabs counts, from the millions of e-mails it manages each hour, the number of e-mails carrying a virus of a given age, for all ages from zero hours upwards. It performs that measurement afresh each hour, each day, to create the threat profile as shown in Figure 2. That profile is combined with the probability distribution described earlier to calculate the probability, per e-mail received by the user, that the e-mail will carry a virus for which the user does not yet have a suitable signature in their local signature file. This is the probability of the user having an infected e-mail slip past their AV software, and it is a direct function of how often the user checks for new virus signatures.

All the user has to do to determine their risk is say how many e-mails they receive in a typical day and how frequently they check for new signatures. TBSE works out their probability of infection based on those two values and the current threat profile.

6 Conclusion

For many years, the Information Security industry has struggled to develop a way to model the interactions which lead to Information Security risk and to forecast

security outcomes in an objective analytical manner. Our inability to achieve this goal has clearly hampered the advance of Information Security as a discipline.

TBSE shows how proven non-deterministic modelling techniques can be applied to the forecasting of security risk. Though it has been suggested in the past that non-deterministic techniques might offer a solution to this modelling problem, we believe that TBSE represents the first time it has been shown in full how such techniques might be applied and the types of results which can be generated. It transpires that this type of modelling is simpler than many people had anticipated and is much simpler and more accurate than relying on huge data mining engines extracting imprecise correlations from terabytes of raw data. This is extremely exciting and opens the prospect for major advances in the security field.

TBSE is of direct benefit to user organisations seeking to improve the effectiveness and efficiency of their security risk management arrangements.

- Users will be able to specify QoP targets for information systems in the form of the maximum levels they will tolerate of threat or outcome populations.
- Security solutions will be designed which can provably satisfy specified QoP targets for a given range of threats, and security evaluations of vendor products will become objective rather than purely relative.
- Dashboards will be created allowing top management to exercise strategic rather than tactical oversight of the organisation's risk management arrangements.
- Threats would be measured continually so that security parameters can be adjusted to ensure QoP targets continue to be met against the changing threat.
- A broad range of metrics will be deployed so that security solutions can be compared objectively against their claims and the effectiveness of locally deployed security practices compared meaningfully across disparate teams and circumstances.

TBSE also creates a wide variety of commercial opportunities for service providers and vendors wishing to support this global user community. Managed services companies can build new services to supply their customers with threat data and the algorithms to turn monthly threat indices into forecast risk indices. Management consultancies can build new services to help their clients tailor risk models and decision-support tools to match their client's particular environments. Security assurance companies can develop services to calibrate security countermeasures and to certify the effectiveness of security deployments. They will address the need for a new range of compliance audits to enable:

- Companies to assure audit committees they have a security programme in place which protects shareholder interests and information assets;
- Companies to demonstrate objectively and measurably to external regulators that they are operating in compliance with regulations and legislation;
- Business partners to be assessed to ensure they do not introduce inappropriate risks to the integrated supply chain.

TBSE will create opportunities for product companies to develop new risk management software tools. TBSE might also be what is needed to kick start an active Digital Risk insurance marketplace by enabling the development of simple Digital Risk insurance products for which premiums can be reliably priced.