Proc. 2nd Saudi Sci. Conf.,Fac. Sci., KAU, 15-17 March 2004, ***Part IV,*** pp. 1-17 (2005 A.D./1425 A.H.)

A Doubly-Stochastic Fault-Tree Assessment of the Probabilities of Security Breaches in Computer Systems

ALI M. RUSHDI* and OMAR M. BA-RUKAB**
**Department of Electrical and Computer Engineering,*
King Abdulaziz University,
*and **Department of Computer Technology,*
College of Telecommunications & Electronics, Jeddah, Saudi Arabia

ABSTRACT. The present paper is an initial attempt to adapt the fault-tree methodology of reliability engineering to the quantification of security exposure of computer systems. In this new context, a fault tree can be described as a logic diagram whose input represents breach events at various system levels, and whose vertices represent logic operations or gates. The root or output of the fault tree can be any of the undesired top events. The present paper briefly surveys algorithms for converting the switching (Boolean) expression of the indicator variable for the top event into a probability expression. Once the top-event probability is determined, it can be multiplied by the system's vulnerability to that event to yield a quantified value of the system's exposure to it. The present paper also handles the doubly-stochastic problem of estimating the uncertainty in the top-event probability by using an analytic exact formula relating the variance of the top-event probability to the variances of the basic-event probabilities. An example of a typical computer system is presented wherein numerical estimates are obtained for the top-event probabilities and their variances and also for the importance ranking of the various breach events.

1. Introduction

A new trend in the study of computer system security is to exploit similarities between reliability and security to develop quantitative measures for "operational security" (Brocklehurst *et al.,* 1994). In particular, the fault-tree model, a traditional reliability methodology used in the analysis and design of safety-critical systems, is now being considered also in the analysis and design of security-critical systems (Brooke & Paige, 2003).

The present paper is an initial attempt in the exploration of the possibility of using fault trees in the quantitative assessment of the effect of security breaches on a computer system. Such an assessment can be based on the following

1. Specification of all foreseeable types of **basic breach events**. These events fall under many categories such as catastrophes, hardware and program failures, human

carelessness, malicious damage, malicious programs, and crime. For every single type of breach events, the **probability** of occurrence over a stated period of time should be estimated. However, such an estimation can be made only with a very high uncertainty.

2. Observation of the various types of **security or safeguard measures** introduced during the design and implementation of the system. These may include account numbers and passwords, authentication tables, file access restrictions, encryption techniques, as well as physical, administrative, legal, and societal control (Martin, 1973; O'Gorman, 2003).

3. Definition of the **undesired top events** resulting from a security breach. These may fall into some of the six categories: computer's performance degradation, inability to process, loss of data (files, records, and programs), unauthorized/inadvertent modification of data, unauthorized reading or copying of data, and permanent hardware malfunction. An estimation should be made of the system's **vulnerability** to each of these events which equals the cost incurred by the system if that event took place.

4. Mathematical **modeling** of the logical relations between the aforementioned entities.

To adapt fault trees to the modeling of security exposures of computer systems, a fault tree is viewed as a logic diagram whose inputs represent breach events at various system levels, and whose vertices represent logic operations or gates. The root or output of the fault tree can be any of the undesired top events. For each of these top events, a particular fault tree can be constructed taking into consideration the characteristics of typical present computer systems. As a logic tree, a fault tree produces a switching expression for the indicator variable of its top event in terms of the corresponding variables of its basic events. Usually, the expression is in a sum-of-products (s-o-p) form, and is not readily useful in expressing the top-event probability in terms of basic event probabilities. The present paper briefly surveys algorithms for converting the switching (Boolean) expression of the indicator variable for the top event into a probability expression. These include algorithms that orthogonalize sum-of-products expressions by making their terms disjoint, algorithms that maintain or produce statistically-independent products, and expansion (factoring) algorithms or combinations thereof. Once the top-event probability is determined it can be multiplied by the system's vulnerability to that event to yield a quantitative value of the system's **exposure** to it.

An issue of crucial importance in the study of security breaches is that any predictions of basic-event probabilities will certainly involve relatively high uncertainties. This issue is handled in reliability engineering either by dealing with fuzzy, rather than crisp probabilities (Tanaka *et al.,* 1983; Weber, 1994) or by considering the pertinent probabilities as random variables (Rushdi, 1985). In this latter approach, the problem of analyzing a fault tree is said to be **doubly stochastic**.

The present paper employs the doubly-stochastic approach in estimating the uncertainty in the top-event probability taking into consideration the uncertainties in the basic-event probabilities (Jackson, 1982; Laviron and Heising, 1985; Dezfuli and Modarres, 1985; Rushdi, 1985; Rushdi and Kafrawy, 1988; Kafrawy and Rushdi, 1990). The top-event probability is a **multiaffine** function of the basic-event probabilities (Rushdi, 1983(b)), and hence it has a **finite** multivariable Taylor's expansion. Therefore, an **exact** formula relating the variance of the top-event probability to the variances of the basic-event probabilities can be obtained (Rushdi, 1985). Numerical results are obtained for the variances of the top-event probabilities in typical computer systems, and also for the **importance ranking** of the various breach events.

The rest of the paper is organized as follows. Section 2 presents a detailed list of pertinent notation, while section 3 discusses various issues involved in the study of the security of modern computer systems such as basic breach events, security measures, and undesired top events. Section 4 gives the reader a quick glimpse of fault trees by citing a few small examples. A quick survey of the fundamental families of algorithms used in the analysis of fault trees is presented in section 5, which discusses also how a symbolic expression of the top-event probability can be utilized in the importance ranking of the various basic events. The uncertainty analysis of fault-tree outputs is reviewed in section 6, which presents analytic formulas for the mean and variance of the top-event probability, in the two cases when the basic-event probabilities are statistically dependent or not. Section 7 combines the ideas and concepts of the previous section in a unified numerical example in which a certain security situation is fault-tree modeled, and computations are made for the top-event probability and its variance and also for the importance measures of the basic events. Section 8 concludes the paper.

2. List of Notation

n Number of systems components relevant to the fault tree.

$\overline{X}_i, X_i$ Indicator variables for the occurrence and non-occurrence of basic event i at time t. These are switching (Boolean) random variables; $\overline{X}_i = 1$ and $X_i = 0$ if i occurs, and $\overline{X}_i = 0$ and $X_i = 1$ if i does not occur.

$\overline{S}$ Indicator variable for the existence of the top event at time t.

T Implies the transpose of a vector.

$P(A)$ Probability of event A.

$E(I_A)$ Expectation of random variable I_A (the indicator variable of event A).

q_i Probability of occurrence of basic event i (treated as a random variable); $q_i = P(\overline{X}_i = 1) = E(\overline{X}_i) = 1 - E(X_i) = 1 - p_i$.

Q Top-event probability; also called system unavailability (treated as a random function); $Q = P(\overline{S} = 1) = E(\overline{S}) = 1 - R$.

$\boldsymbol{q}$ N-dimensional vector of basic-event probabilities; $\boldsymbol{q} = [q_1\ q_2 \ldots q_n]^T$.

$\boldsymbol{v}_1$ Mean value of $\boldsymbol{q}$; $\boldsymbol{v}_1 = [v_{11}\ v_{21}\ \ldots\ v_{n1}]^T$.

v_{ij} Central moment j of q_i; $v_{ij} = E\{(q_i - v_{i1})^j\}$, $j \geq 2$.

μ_1 Mean value of Q.

μ_j Central moment j of Q; $\mu_j = E\{(Q - \mu_1)^j\}, j \geq 2$.

$J_{ij \ldots r}$ The joint central moment of the random variables $q_i, q_j, \ldots$ and q_r; $J_{ij \ldots r} = E\{(q_i\ \ v_{i1})(q_j\ \ v_{j1}) \ldots (q_r\ \ v_{r1})\}$.

J_{ij} The joint central moment of the two random variables q_i and q_j; called the covariance of these two variables and denoted by Cov(q_i, q_j).

ρ_{ij} The correlation coefficient between q_i and q_j; and $\rho_{ij} = \mathrm{Cov}(q_i, q_j)/(v_{i2}\, v_{j2})^{1/2}$; $-1 \leq \rho_{ij} \leq 1$; ρ_{ij} is a dimensionless constant that measures the linear interdependence or proportionality between q_i and q_j; $\rho_{ij} = 0$ if q_i and q_j are independent, but the converse is not necessarily true (Trivedi, 2002).

m Median (50*th* percentile) of a log-normally distributed variable.

F Error factor (range factor) of a log-normally distributed variable; $F = 95th$ percentile/$50th$ percentile = $50th$ percentile/$5th$ percentile,

λ, ξ Mean and standard deviation of the natural logarithm of a log-normally distributed variable; $\lambda = \ell n(m)$; $\xi = \ell n(F)/1.645$.

$C(n,i)$ The combinatorial (binomial) coefficient (n choose i) = the number of ways of choosing i objects out of n objects without order or replacement $(0 \leq i \leq n)$.

3. Important Security Issues

Study of the security of modern computer systems deals with many important entities such as basic breach events, security measures and undesirable top events. Table 1 presents our preliminary or first-cut treatment of this sophisticated subject, by citing examples of various types of security exposure in a "typical" computer system. The indices or keys of the columns of Table 1 are indicator variables for certain undesired top events that we thought are of interest (to which the reader may add others of his own). These are :

Y_1 = Performance degradation,

Y_2 = Inability to Process,

Y_3 = Loss of data (files, records, programs),

Y_4 = Unauthorized/Inadvertent modification of data,

Y_5 = Unauthorized reading or copying of data,

Y_6 = Permanent Hardware malfunction.

The indices or keys of the rows i are important basic breach events B_i. An entry a_{ij} at row i and column j of Table 1 means that the probability of occurrence of the event B_{ij} (which is the basic event B_i pertaining to the top event Y_j) is $10^{-a_{ij}}$. Blank positions (in which no entries are given) indicate that the event B_{ij} is virtually impossible, *i.e.* of negligible probability. The basic idea of Table 1 is borrowed from Martin (1973). In fact, the top rows in Table 1 summarize details given in Table 2.1, pp. 12-13 by that author. The bottom rows of Table 1 add breach events due to malicious codes or programs (Stallings, 2003; Alayed *et al.*, 2002; Salah, *et al.*, 2002; Gollmann, 2000). A word of caution is in order. The probabilities entered in Table 1 are very rough estimates that may be easily disputed and that have to be refined through adequate measurement techniques applied to specific systems. To reflect the fact that these probabilities suffer from large uncertainties (that may be severe enough to produce a variability of one order of magnitude), we consider the entries not to represent deterministic variables but rather to represent some central measures of random variables. We further assume these entries to represent the medians of the corresponding variables, which we also assume to have a log-normal distribution (truncated to the [0.0, 1.0] interval), with an error factor F= 10 (Rushdi and Kafrawy, 1988).

Table 1. Examples of types of security exposure (an entry in the table is a rough estimate of the negative of $\log_{10}$ of the pertinent probability of occurrence in one day).

T_j / B_i	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
Catastrophes (Fire, flood, earthquake, war, ...)	4	4	4	-	-	4
Hardware/Software failures	1	2	3	3	-	3
Human Carelessness	3	3	2	0	4	3
Malicious damage (looting, sabotage)	3	3	3	3	3	3
Fraud and Embezzlement	-	-	3	2	2	-
Industrial espionage	-	-	-	-	3	-
Employee betrayal of employer	-	-	-	-	3	-
Malicious Programs:						
Viruses	2	2	2	2	2	3
Worms	3	3	3	3	3	-
Bacteria	3	3	-	-	-	-
Trojan horses	3	2	3	3	3	-
Logic Bombs	-	3	3	3	-	-
Trapdoors	-	-	4	4	4	-

Table 2. Parameters of the basic-event probabilities for top event Y_6.

i	m_i	F_i	ξ_i	v_{i1}	v_{i2}
1	10^{-4}	10	1.3997478	$\times 10^{-4}$ 2.6635156	$\times 10^{-7}$ 4.3234997
2..5	10^{-3}	10	1.3997478	$\times 10^{-3}$2.6635156	$\times 10^{-5}$ 4.3234997

The fault-tree modeling of computer security to be pursued in the following sections cannot be completed without an appropriate understanding of the major security safeguards or measures usually taken in modern systems, since these (in addition to the breach events) are potential candidates as inputs in fault-tree models. Due to space limitations, we refrain from discussing the issue of security safeguards here and refer the reader to some of the texts on this topic (Anderson, 2001; Tanenbaum, 2001; Kaufman *et al.*, 2002).

4. Construction of Fault Trees

Fault-tree analysis is a top-down deductive analysis structured in terms of events (or indicator variables thereof) rather than components. The perspective is on faults or failures rather than successes since a failure is usually easier to define than a non-failure, and there may be far fewer ways in which a failure can occur, as opposed to the numerous ways in which non-failure can occur (Ebeling, 1997). The focus is usually on a significant failure or a catastrophic or undesirable event, which is referred to as the top event since it appears at the top of the fault tree. In construction of a fault tree, logic gates are used to relate the input or basic events and the intermediate events to the top event. Note that a logic gate gives a qualitative description of the causal relationship between its inputs and its output. For example, the output event of an AND gate occurs iff all its input events occur, while the output of an OR gate is caused to occur if at least one of its inputs occurs. Therefore , the indicator variable for the output of an AND (OR) gate is obtained simply by ANDing (ORing) the indicator variables for its inputs. Detailed studies of the construction or synthesis of fault trees is available (Henley & Kumamoto, 1981; Henley & Kumamoto, 1992; Kumamoto, 1993; Aboun-Nour, 1999). A few examples are now presented to demonstrate the construction of fault trees in the computer security arena.

Example 4.1

The situation depicted by Table 1 can be modeled by a multitude of fault trees. Each of the top events Y_j, $1 \leq j \leq 6$ is the output of a fault tree consisting of a single OR gate, the input of which are the pertinent basic events, *i.e.*, those events having contribution to the top event.

Example 4.2

A simple fault-tree model for a message protected by a combination of steganography and cryptography (Stallings, 2003) may consist of a single AND gate with output $\overline{S}$ and inputs $\overline{X}_1, \overline{X}_2$ and $\overline{X}_3$, where:

$\overline{S}$ = The intruder gains access to the plaintext sent,

$\overline{X}_1$ = The intruder manages to intercept the overall message and acquires its ciphertext,

$\overline{X}_2$ = The intruder detects the existence of a secret message despite its concealment within the overall message through steganography,

$\overline{X}_3$ = The intruder succeeds in his cryptanalysis attack and breaks the encryption algorithm.

Example 4.3

Figure 1 shows a fault tree that models system behavior under the attack of a virus that takes place when a certain threshold time is reached. The indicator variable for the top event is $\overline{S}$ = System breach under the virus attack.

While the basic events are :

$\overline{X}_1$ = The system is not protected by an antivirus package,

$\overline{X}_2$ = The dictionary of an added antivirus package lacks a definition of the signature for pertinent virus ,

$\overline{X}_3$ = A virus contamination occurs,

$\overline{X}_4$ = The threshold time is reached,

X_5 = The system is on at the threshold time.

And the corresponding intermediate events are :

$\overline{Z}_1$ = The system is unaware of the virus signature,

$\overline{Z}_2$ = The virus spreads in the system unopposed,

$\overline{Z}_3$ = The threshold time is reached while power is on.

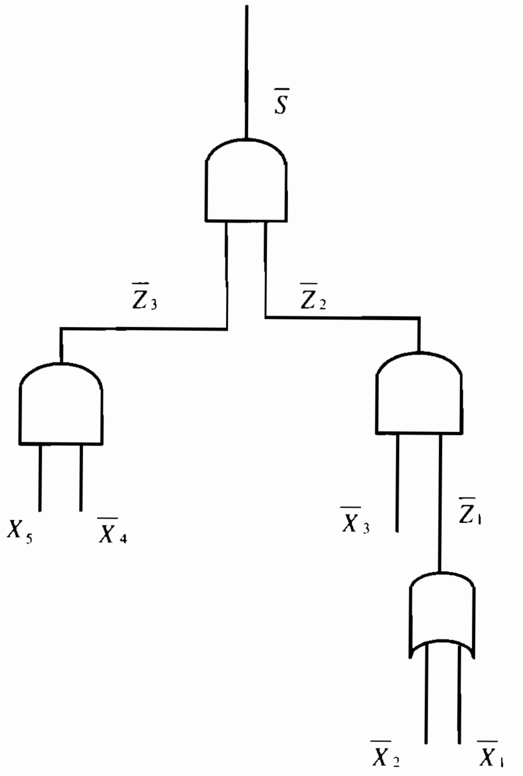

Fig. 1. A fault tree modeling a system breach under virus attack.

5. On the Analysis of Fault Trees

A fault tree is a logical formulation that can be used to express the top event as a logical function of basic events. Noting that the algebra of events (set algebra) is isomorphic to the bivalent or 2-valued Boolean algebra (switching algebra), we may choose to employ this latter type of algebra by considering the inputs and output of a fault tree as indicator variables of the respective events. Hence, the fault tree is to produce a switching or Boolean function for the indicator variable of the top event in terms of the indicator variables of the basic events. Now, it is necessary to move from the Boolean domain to the probability domain so as to obtain the top-event probability as a function of basic-event probabilities. Many algorithms have emerged for converting the switching (Boolean) expression for the indicator variable of the top event into a probability-ready expression (PRE) *i.e.,* into an expression that is directly convertible, on a one-to-one basis, to a probability expression. Note that in a PRE

(a) all ORed terms/(products) are disjoint, and

(b) all ANDed alterms (sums) are statistically independent.

The conversion from a PRE to a probability expression is trivially achieved by replacing Boolean variables by their expectations, AND operations by multiplications, and OR operations by additions (Bennetts, 1975; Rushdi & Abdul-Ghani, 1993). In the following, we give a brief classification of the available algorithms for converting a general switching expression into a PRE.

5.1 Orthogonalization (Disjointness) Algorithms

These algorithms start with a sum-of-products (s-o-p) expression for a switching function and orthogonalize it by making *all* its terms mutually disjoint. The basic internal step for such algorithms is to consider a sum $(T_i \vee T_j)$ of the two terms T_i and T_j that are nondisjoint and are such that neither of them subsumes the other. The term T_j is disjointed with (made orthogonal to) the term T_i by the relation

$$\begin{aligned} T_i \vee T_j &= T_i \vee T_j(\overline{y_1\, y_2 \dots y_e}) \\ &= T_i \vee T_j(\overline{y}_1 \vee y_1 \overline{y}_2 \vee\; y_1 y_2 \overline{y}_3 \;\vee \dots \vee y_1 y_2 y_3 \,\dots\, y_{e-1} \overline{y}_e), \end{aligned} \qquad (1)$$

where $Y = \{y_1, y_2, y_3, \dots, y_e\}$ is the set of literals that appear in T_i and not appear in T_j. Note that T_j is replaced by e terms that are disjoint among themselves beside being disjoint with T_i. In the limiting case of $e = 0$ $(\mathrm{Y} = \varphi)$, T_j subsumes T_i and is absorbed by it, *i.e.*,

$$T_i \vee T_j = T_i \vee T_j(0) = T_i. \qquad (2)$$

The seminal work on orthogonalization (disjointness) is due to Abraham (1979) and to Dotson and Gobian (1979). Visual insight into the process of disjointness can be obtained through the use of logic aids such as the Karnaugh map (Rushdi, 1983(a)). More recent work on orthogonalization involves multiple-variable inversion techniques (Veeraraghavan & Trivedi, 1993) and shellability (Crama & Hammer, 2002).

5.2 Algorithms Based Primarily on Statistical Independence

Orthogonalization algorithms make a natural utilization of the statistical independence of basic events, when such independence can be assumed. There are other algorithms (Rushdi & Goda, 1985; Rushdi & Abdul-Ghani, 1993) that try to make a more direct use of statistical independence, not only through preserving it when it exists, but also by deliberately making it more manifestable through appropriate operations. For example, it is always possible to handle the complement of an expression instead of the expression itself. If one uses the De Morgan identity

$$\overline{(\overset{n}{\underset{i=1}{\vee}} X_i)} = \overset{n}{\underset{i=1}{\wedge}} \overline{X}_i \qquad (3)$$

then statistical independence can be utilized in the ANDed form that appears in the right hand side (RHS) of (3).

5.3 Expansion or Factoring Algorithms

The most powerful class of algorithms producing PREs are algorithms based on repeated use of the Boole-Shannon expansion (Rushdi, 1983(a); Rushdi & Goda, 1985) in which a switching expression is expanded about one of its variables in the form

$$\overline{S}(\overline{\mathrm{X}}) = (\overline{X}_i \wedge \overline{S}(\overline{\mathrm{X}} \mid \overline{X}_i = 1)) \vee (X_i \wedge \overline{S}(\overline{\mathrm{X}} \mid \overline{X}_i = 0)). \qquad (4)$$

Note that (4) represents a good step towards creating a PRE; the two terms in the RHS of (4) are disjoint, and each of them is an ANDing of statistically independent entities (under the assumption that X consists of statistically independent components). Many authors apply (4) directly in the probability domain, in which case it is called the factoring theorem (Page & Perry, 1989) which is simply a version of the total probability theorem, namely

$$Q(\boldsymbol{q}) = q_i \ Q(\boldsymbol{q} \mid 1_i) + (1-q_i)Q(\boldsymbol{q} \mid 0_i) \tag{5}$$

where $Q(\boldsymbol{q} \mid j_i)$, $j=0,\ 1$, is the function $Q(\ \boldsymbol{q}\ )$with q_i set to j while the rest of the elements of $\boldsymbol{q}$ are left intact. Once a symbolic expression of the top-event probability as a function of basic-event probability is obtained, it can be used to derive important measures for the various basic events. One such measure is (Henley & Kumamoto, 1992)

$$\boldsymbol{I}_i = (\partial \boldsymbol{Q} / \partial \boldsymbol{q}_i\ ) = \boldsymbol{Q}(\mathbf{q} \mid \boldsymbol{1}_i\ ) - \boldsymbol{Q}(\mathbf{q} \mid \boldsymbol{0}_i\ ) \tag{6}$$

The variable I_i represents the importance of basic event i. An importance ranking of the basic events is obtained by the following rule: if $I_i > I_j$ then event i is ranked as more important (from the specific top event point of view) than event j.

Example 5.1 (Example 4.3 revisited)

The indicator variable $\overline{S}$ of the top event for the fault tree in Fig. 1 can be expressed in terms of the basic events as

$$\overline{S} = (\ \overline{X}_1 \vee \overline{X}_2)\ \ \overline{X}_3 \overline{X}_4 X_5 \tag{7}$$

Henceforth, the basic events are assumed to be statistically independent. The expression (7) for $\overline{S}$ can be rewritten in the disjoint s-o-p form,

$$\overline{S} = (\ \overline{X}_1 \vee X_1\ \overline{X}_2)\ \overline{X}_3\ \overline{X}_4\ X_5 \tag{8}$$

which corresponds to the following expression for the top-event probability

$$Q = (q_1 + p_1 q_2) q_3 q_4 p_5 \tag{9}$$

Alternatively, the complement of (7) is written as

$$\begin{aligned} S &= X_1 X_2 (X_3 \vee X_4 \vee \overline{X}_5\ ) \\ &= X_1 X_2\ (X_3 \vee \overline{X}_3\ (X_4 \vee \overline{X}_4\ \overline{X}_5\ )) \end{aligned} \tag{10}$$

which results in the following expression for the top-event probability

$$Q = 1 - p_1 p_2\ (p_3 + q_3(p_4 + q_4 q_5)) \tag{11}$$

6. Uncertainty Relations

In reliability analysis of computer systems, models such as reliability block diagrams, fault trees, Markov chains and stochastic Petri nets are built to predict the reliability of the system (Ebeling, 1997; Trivedi, 2002). The parameters in these models are usually obtained from field data, data from systems with similar functionality, and even by expert guessing, and hence are bound to suffer from considerable uncertainty (Yin *et al.*, 2001). The uncertainty problem pertaining to fault-tree outputs has an analytic doubly-stochastic treatment via the method of moments (Rushdi, 1985; Rushdi and Kafrawy, 1988; Kafrawy and Rushdi, 1990). This method of moments utilizes the multiaffine nature (Rushdi, 1983(b)) of the top-event probability as a function of the basic-event probabilities. The basic results of the method of moments, which have been extensively verified through comparison with results obtained via other methods including Monte Carlo simulations, can be summarized as follows.

The mean or expected value of the top-event probability Q is

$$\mu_1 = Q(\boldsymbol{v}_1) + \sum_{1\le i<j\le n}\sum C_{ij}\, J_{ij} + \sum_{1\le i<j<k\le n}\sum\sum C_{ijk}\, J_{ijk} + \ldots + C_{12\ldots n}\, J_{12\ldots n} \tag{12}$$

while its variance (measure of uncertainty) is

$$\begin{aligned}\mu_2 &= \sum_{i=1}^{n} C_i^2\, v_{i2} \\ &+ \sum_{1\le i<j\le n}\sum [\,2 C_i\, C_j\, J_{ij} + C_{ij}^2 (\,J_{ijij} - J_{ij}^2)] + 2\sum_{\substack{1\le i<j\le n \\ \{i,j\}\ \#}}\sum\ \sum_{\substack{1\le k<l\le n \\ \{k,l\}}}\sum C_{ij}\, C_{kl}\, (\,J_{ijkl} - J_{ij}\, J_{kl}) \\ &+ 2\sum_{i=1}^{n}\ \sum_{1\le j<k\le n}\sum C_i\, C_{jk}\, J_{ijk} + 2\sum_{i=1}^{n}\sum\ \sum_{1\le j<k<l\le n}\ \sum C_i\, C_{jkl}\, C_{ijkl} \\ &+ 2\sum_{1\le i<j\le n}\sum\ \sum_{1\le k<l<m\le n}\sum\ \sum C_{ij}\, C_{klm}\, (\,J_{ijklm} - J_{ij}\, J_{klm}) \\ &+ \sum_{1\le i<j<k\le n}\sum\ \sum C_{ijk}^2\, (\,J_{ijkijk} - J_{ijk}^2) + \ldots + C_{12\ldots n}^2\, (\,J_{12..n12..n} - J_{12..n}^2)\end{aligned} \tag{13}$$

which reduce to the following expressions when the components of ***q*** are statistically independent

$$\mu_1 = Q(\boldsymbol{v}_1) \tag{14}$$

$$\begin{aligned}\mu_2 &= \sum_{i=1}^{n} C_i^2\, v_{i2} + \sum_{1\le i<j\le n}\sum C_{ij}^2\, v_{i2} v_{j2} + \sum_{1\le i<j\le n}\sum\ \sum C_{ijk}^2\, v_{i2} v_{j2}\, v_{k2} \\ &+ \ldots + C_{12\ldots n}^2\, v_{i2}\, v_{j2}\, v_{n2}\end{aligned} \tag{15}$$

The coefficients that appear in (12) - (15) are given by

$$C_i = (\partial Q/\partial q_i)_{q=v_1} = Q(v_1|\, 1_i) - Q(v_1|\, 0_i) \tag{16}$$

$$\begin{aligned}C_{ij} &= (\partial^2 Q/\partial q_i \partial q_j)_{q=v_i} \\ &= Q(v_1|1_i,1_j) - Q(v_1|0_i,1_j) - Q(v_1|1_i,0_j) + Q(v_1|0_i,0_j)\end{aligned} \tag{17}$$

$$\begin{aligned}C_{ijk} &= (\partial^3 Q/\partial q_i \partial q_j \partial q_k)_{q=v_i} \\ &= Q(v_1|1_i,1_j,1_k) - Q(v_1|0_i,1_j,1_k) - Q(v_1|1_i,0_j,1_k) + Q(v_1|0_i,0_j,1_k) \\ &- Q(v_1|1_i,1_j,0_k) + Q(v_1|0_i,1_j,0_k) + Q(v_1|1_i,0_j,0_k) - Q(v_1|0_i,0_j,0_k)\ ,etc.\end{aligned} \tag{18}$$

The expressions above for μ_1 and μ_2 are exact closed-form formulas.

Example 6.1

Consider a fault tree consisting of a single AND gate such as the one in Example 4.2. In this case, the top-event probability is a product of the basic-event probabilities

$$Q = \prod_{i=1}^{n} q_i \tag{19}$$

If the basic-event probabilities are assumed to be statistically independent, then the mean and variance of the top-event probability are obtained from (14) and (15) as

$$\mu_1 = \prod_{i=1}^{n} v_{i1} \tag{20}$$

$$\mu_2 = \sum_{i=1}^{n} (\mu_1 / v_{i1})^2 \, v_{i2} + \sum_{1 \le i < j \le n} \sum (\mu_1 / v_{i1} v_{j1})^2 \, v_{i2} \, v_{j2}$$

$$+ \sum \sum_{1 \le i < j < k \le n} \sum (\mu_1 / v_{i1} \, v_{j1} \, v_{k1})^2 \, v_{i2} \, v_{j2} \, v_{k2} + \ldots + \prod_{i=1}^{n} v_{i2} \tag{21}$$

Example 6.2

Consider a fault tree consisting of a single OR gate (such as any of the trees referred to in Example 4.1). By virtue of (3), the top event probability is given by

$$Q = 1 - \prod_{i=1}^{n} p_i = 1 - \prod_{i=1}^{n} (1 - q_i) \tag{22}$$

Again, the basic-event probabilities are assumed to be statistically independent, so that the mean and variance of the top-event probability become

$$\mu_1 = 1 - \prod_{i=1}^{n} (1 - v_{i1}) \tag{23}$$

$$\mu_2 = \sum_{i=1}^{n} ((1 - \mu_1) / (1 - v_{i1}))^2 \, v_{i2} + \sum_{1 \le i < j \le n} \sum ((1 - \mu_1) / (1 - v_{i1})(1 - v_{j1}))^2 \, v_{i2} \, v_{j2}$$

$$+ \sum \sum_{1 \le i < j < k \le n} \sum ((1 - \mu_1) / (1 - v_{i1})(1 - v_{j1})(1 - v_{k1}))^2 \, v_{i2} \, v_{j2} \, v_{k2} + \ldots + \prod_{i=1}^{n} v_{i2} \tag{24}$$

Note that when $v_{i1} << 1, v_{i2} << 1, \forall \, i,$ then (23) and (24) simplify to

$$\mu_1 \approx \sum_{i=1}^{n} v_{i1} \tag{25}$$

$$\mu_2 \approx \sum_{i=1}^{n} v_{i2} \tag{26}$$

Example 6.3

Consider a fault tree in which the top event occurs iff at least k of its n basic events occur. This fault tree simulates a k-out-of-n:F system, *i.e.*, a system that fails iff at least k of its n components fail. If the basic events are assumed to have identical probabilities, the top-event probability is given by

$$\begin{aligned} Q(q) &= \sum_{i=k}^{n} C(n,i)\,q^{i}(1-q)^{n-i} \\ &= \sum_{m=k}^{n} (-1)^{m-k}\,C(m-1,k-1)\,C(n,m)\,q^{m} \end{aligned} \tag{27}$$

Hence, Q is a polynomial function of a single-variable q, and a direct application of the expectation operator to its Taylor expansion leads to the following expression for $E\{Q\}$.

$$\mu_1 = Q(v_1) + \sum_{r=2}^{n} \frac{1}{r!}(\partial^r Q/\partial q^r)_{q=v_1} v_r \tag{28}$$

Similarly, the variance of Q is easily obtained as

$$\begin{aligned} \mu_2 &= (\partial Q/\partial q)^2_{q=v_1} v_2 + [(\partial Q/\partial q)(\partial^2 Q/\partial q^2)]_{q=v_1} v_3 \\ &+ [\frac{1}{4}(\partial^2 Q/\partial q^2)^2 \; + \frac{1}{3}(\partial Q/\partial q)(\partial^3 Q/\partial q^3)]_{q=v_1} v_4 + ... \end{aligned} \tag{29}$$

Expressions (29) and (30) can also be obtained from (12) and (13), if Q is expressed as a function of distinct variables ***q*** (Rushdi; 1993), and later all components of ***q*** are set equal to a single value q.

7. A Sample Numerical Example

Consider the top event Y_6 in Table 1. We recall from Example 4.1 that Y_6 is the output of a fault tree consisting of a single OR-gate of 5 inputs $\overline{X}_i$, $1 \le i \le 5$. The mean and variance of the top-event probability are given by equations (25) and (26) in Example 6.2. The basic-event probabilities are of a log-normal distribution of medians $10^{-4}, 10^{-3}, 10^{-3}, 10^{-3}$, and 10^{-3} respectively and of a common error factor F= 10. For such small medians, there is no need to worry about the tail of the distribution extending out to infinity, and the log-normal distribution is effectively equivalent to a truncated distribution, i.e., to a distribution bounded within the [0.0, 1.0] interval (Rushdi & Kafrawy, 1988). Given the median m and error factor F of a log-normally distributed variable X, then the mean v_1 and variance v_2 of X are

$$v_1 = \exp(\lambda + \xi^2/2) = m \; \exp(\xi^2/2) \tag{30}$$

$$v_2 = v_1^2(\exp(\xi^2) - 1) \tag{31}$$

where $\xi = (\ell n(F)/1.645)$ is the standard deviation for ℓn X (See Appendix A). With the aid of (30), (31) we complete the list of pertinent parameters of the basic-event probabilities as shown in Table 2. The data of Table 2 are now substituted in (25) and (26) to produce the mean and variance of the top-event probability as

$$\mu_1 = 1.0920414 \times 10^{-2},$$

$$\mu_2 = 1.7337234 \times 10^{-4}$$

It is desirable to make some estimate of the potential cost of the damage that might be done by the top event Y_6. If the system vulnerability to Y_6 is estimated to be say SR 10,000, then a quantitative rating for the system exposure to Y_6 is obtained by multiplying the vulnerability by μ_1 to obtain the numerical value of 109.2 SR/day. Can we multiply this number by 365 to obtain the probable average damage per year? The answer is generally no, since the pertinent probabilities are not generally additive, *i.e.*, the per-year probability is not necessarily 365 times the per-day probability.

The importance of the basic event i has a mean value of

$$< I_i > = (1 - \mu_1) / (1 - v_{i1}) \tag{32}$$

and hence we can see that the four basic events 2,3,4 and 5 are of equal importance and are each more important than the basic event 1.

8. Conclusion

A preliminary work on the problem of fault-tree modeling of computer system security is presented. This work is to be more perfected by employing rigorous measurement and statistical techniques in assessing basic-event probabilities for specific computer systems. It is expected that these probabilities will vary significantly according to many factors such as the size, make, model, uses and geographical location of the computer and whether it is internetworked or not. Another important area of future work is to consider more sophisticated fault-tree models that can describe more detailed scenarios in which attention is not restricted to breaches, threats or attacks but also involves a consideration of safeguards or protective measures.

References

Abraham, J.A. (1979), "An improved algorithm for network reliability," *IEEE Transactions on Reliability*, vol. **R-28**, no. 1, pp. 58-62.

Alayed, A., Furnell, S.M., and **Barlaw, I.M**. (2002), *Addressing Internet Security Vulnerabilities*, Chapter 9 (pp. 121-132) in M. A. Ghonaimy *et al.* (Editors), *Security in the Information Society: Visions and Perspectives*, Kluwer Academic Publishers, Boston, USA.

Anderson, R.J. (2001), *Security Engineering*, Wiley, New York, NY, USA.

Bennetts, R.G. (1975), "On the analysis of fault trees," *IEEE Transactions on Reliability*, vol. **R-24**, no. 3, pp. 194-203.

Brocklehurst, S., Littlewood, B., Olovsson, T., and **Jonsson, E.** (1994), "On measurement of operational security," *IEEE Aerospace and Electronic Systems Magazine*, vol. **9**, no. 10, pp. 7-16.

Brooke, P.J. and **Paige, R.F.** (2003), "Fault trees for security system design and analysis, " *Computers and Security*, vol. **22**, no. 3, pp. 256-264.

Crama, Y. and **Hammer, P.H.** (2002), *Boolean Functions*, E-book, Available at http://www.sig.egss.ulg.ac.be/rogp/Crama/ , Accessed on March 20, 2004.

Dezfuli, H. and **Modarres, M.** (1985), "Uncertainty analysis of reactor safety systems with statistically correlated failure data," *Reliability Engineering*, vol. **11**, pp. 47-64.

Dotson, W.P. and **Gobian, J.O.** (1979), "A new analysis technique for probability graphs," *IEEE Transactions on Circuits and Systems*, vol. **CAS-26**, pp. 855-865.

Ebeling, C.E. (1997), *An Introduction to Reliability and Maintainability Engineering*, McGraw-Hill, New York, USA.

Gollman, D. (2000), *Computer Security*, Wiley, Chichester, England, UK.

Henley, E.J. and **Kumamoto, H.** (1981), *Reliability Engineering and Risk Assessment*, Prentice-Hall, Upper Saddle Drive, NJ, USA.

Henley, E.J. and **Kumamoto, H.** (1992), *Probability* Risk Assessment: Reliability Engineering Design and Analysis, *IEEE Press, New York, NY, USA.*

Heyde, C.C. (1963), *"On* a property of the lognormal distribution,*" Journal of the Royal Statistical Society*, vol. **26**, pp. 392-393.

Jackson, P.S. (1982), "A second-order moments method for uncertainty analysis," *IEEE Trans. Reliability*, **R-31**, pp. 382-384.

Kafrawy, K.F. and **Rushdi, A.M.** (1990), "Uncertainty analysis of fault trees with statistically correlated failure data," *Microelectronics and Reliability*, vol. **30**, no. 1, pp. 157-175.

Kaufman, C., Perlman, R. and **Speciner, M.** (2002), *Network Security*, Second Edition, Prentice-Hall, Englewood Cliffs, NJ, USA.

Kumamoto, H. (1993), *Fault Tree Analysis*, Chapter 7 (pp. 249-312) in K. B. Misra (Editor), *New Trends in System Reliability Evaluation*, vol. **16**, *Fundamental Studies in Engineering*, Elsevier Science Publishers, Amsterdam, The Netherland.

Laviron, A. and **Heising, C.D.** (1985), "Error transmission in large complex fault trees using the ESCAF method," *Reliability Engineering*, vol. **12,** pp. 181-192.

Martin, J. (1973), *Security, Accuracy, and Privacy in Computer Systems.* Prentice- Hall, Englewood Cliffs, New Jersey, USA.

O'Gorman, L. (2003), "Comparing passwords, tokens and biometrics for user authentication," *Proceedings of the IEEE*, vol. **91**, no. 12, pp. 2021-2040.

Page, L.B. and **Perry, J.E.** (1989), "Reliability of directed networks using the factoring theorem," *IEEE Transactions on Reliability*, vol. **R-38**, no. 5, pp. 556-562.

Rushdi, A.M. (1983(a)), "Symbolic reliability analysis with the aid of the variable-entered Karnaugh maps," *IEEE Transactions on Reliability*, vol. **R-32**, no. 2, pp. 134-139.

Rushdi, A.M. (1983(b)), "How to hand-check a symbolic reliability expression," *IEEE Transactions on Reliability*, vol. **R-32**, no. 5, pp. 402-408.

Rushdi, A.M. (1985), "Uncertainty analysis of fault-tree outputs," *IEEE Transactions on Reliability*, vol. **R-34**, no. 5, pp. 458-462.

Rushdi, A.M. (1993), *Reliability of k-out-of-n Systems*, Chapter 5 (pp. 185-227) in K. B. Misra (Editor), *New Trends in System Reliability Evaluation*, vol. **16**, *Fundamental Studies in Engineering*, Elsevier Science Publishers, Amsterdam, The Netherlands.

Rushdi, A.M. and **Abdul-Ghani, A.A.** (1993), "A comparison between reliability analyses based primarily on disjointness or statistical independence," *Microelectronics and Reliability*, vol. **33**, no. 7, pp. 965-978.

Rushdi, A.M. and **Goda, A.S.** (1985), "Symbolic reliability analysis via Shannon's expansion and statistical independence," *Microelectronics and Reliability*, vol. **25**, pp. 1041-1053.

Rushdi, A.M. and **Kafrawy, K.F.** (1988), "Uncertainty propagation in fault-tree analysis using an exact method of moments," *Microelectronics and Reliability*, vol. **28**, no. 6, pp. 945-965.

Salah, D., Aslan, H.K. and **El-Hadidi, M.T.** (2002), *A Detection Scheme for the SK Virus*, Chapter 13 (pp. 171-182) in M.A. Ghonaimy, *et al.* (Ed.), *Security in the Information Society: Visions and Perspectives*, Kluwer Academic Publishers, Boston, USA.

Stallings, W. (2003), *Network and Internetwork Security: Principles and Practice*, Third Edition, Prentice-Hall, Englewood Cliffs, NJ, USA.

Tanaka, H., Fan, L.T., Lai, F.S. and **Toguchi, D.** (1983), "Fault-tree analysis by fuzzy probability," *IEEE Transactions on Reliability*, vol. **R-32**, no. 5, pp. 453-457.

Tanenbaum, A.S. (2001), *Modern Operating Systems,* Prentice-Hall, Upper Saddle Drive, NJ, USA.

Trivedi, K.S. (2002), *Probability and Statistics with Reliability Queuing and Computer Science Applications*, Second Edition, Prentice-Hall, Englewood Cliffs, NJ, USA.

Veeraraghavan, M. and **Trivedi, K.S.** (1993), *Multiple Variable Inversion Techniques*, Chapter 2 (pp. 39-74) in K. B. Misra (Ed.), *New Trends in System Reliability Evaluation*, vol. **16**, *Fundamental Studies in Engineering*, Elsevier Science Publishers, Amsterdam, The Netherlands.

Weber, D.P. (1994), "Fuzzy fault tree analysis," *Proceedings of the third IEEE conference on Fuzzy Systems*, Orlando, Florida, USA, vol. **3**, pp. 1899-1904.

Yin, L., Smith, M.A.J. and **Trivedi, K.S.** (2001), "Uncertainty analysis in reliability modeling," Proceedings of the 2001 *Annual Reliability and Maintainability Symposium*, Philadelphia, PA, USA, pp. 229-234.

Appendix A: On the Lognormal Distribution

The lognormal probability density function (PDF) is given by

$$f_X(x) = exp(-((\ell n(x) - \lambda)/\xi)^2/2)/\sqrt{2\pi}\,\xi x, \quad x \geq 0,$$

$$f_X(X) = 0, \quad x < 0 \qquad \text{(A.1)}$$

where λ and ξ are given by

$$\lambda = \mathrm{E}\{\ell n(X)\} \qquad \text{(A.2)}$$

$$\xi^2 = \mathrm{VAR}\{\ell n(\mathrm{X})\} \qquad \text{(A.3)}$$

which means that λ and ξ are the mean and standard deviation of the natural logarithm of the lognornally distributed variable X. They are expressed in terms of the median (50th percentile) m and the error factor F of the lognormal distribution via:

$$\lambda = \ell n(m) \qquad \text{(A.4)}$$

$$\xi = \ell n(\mathrm{F})/1.645 \qquad \text{(A.5)}$$

where F is the ratio of the 95*th* percentile to the median, which also equals the ratio of the median to the 5th percentile.

The moments about the origin and the central moments for the lognormal PDF are given by

$$t_\ell = \mathrm{E}\{\,\mathrm{X}^\ell\,\} = exp\,(\,\lambda\,\ell + \xi^2\,\ell^2\,/\,2), \qquad \ell = 1, 2, \ldots \tag{A.6}$$

$$v_\ell = \mathrm{E}\,\{\mathrm{X}^\ell\,\} = \mathrm{E}\{(X - t_1)^\ell\}, \qquad \ell = 3, 4, \ldots$$

$$= \frac{v_2^{\ell/2}}{(\omega - 1)^{\ell/2}} \sum_{j=0}^{\ell} (\text{-}1)^{\mathrm{j}} \binom{\ell}{j}\, \omega^{(\ell - j)(\ell - j - 1)/2} \tag{A.7}$$

where

$$\omega = \exp\,(\xi^2\,) \tag{A.8}$$

$$v_2 = \omega(\,\omega\ \text{-}\ 1)\exp\,(2\lambda\,) \tag{A.9}$$

Correspondingly, the mean and the lower-order central moments for the lognormal PDF are given by

$$v_1 = \exp\,(\lambda\ +\ \xi^2\,/\ 2)\ =\ m\,\exp\,(\ \xi^2\,/\ 2\ ) \tag{A.10}$$

$$v_2 = v_1^2\ (\exp\,(\xi^2)\ \text{-}\ 1) \tag{A.11}$$

$$v_3 = v_2^{3/2}\ (\exp(\xi^2)\ \text{-}\ 1)^{1/2}\ (\exp\,(\xi^2\,)\ +\ 2) \tag{A.12}$$

$$v_4 = v_2^2\ (\,\exp\,(\,4\ \xi^2)\ +\ 2\exp\ (3\ \xi^2\ )\ + 3\exp\,(2\ \xi^2\ )\ - 3) \tag{A.13}$$

The lognormal PDF is not uniquely determined by the infinite sequence of its moments $\{\,t_\ell\,\}$. If the lognormal PDF (A.1) is multiplied by the factor

$$1 + \varepsilon \sin\,(2\ \pi\ \mathrm{k}\ (\ell\mathrm{n(x)}\ \text{-}\ \lambda)\ /\ \xi^2) \tag{A.14}$$

the resulting PDF possesses the same moments as the lognormal PDF.

التقدير مزدوج العشوائية باستخدام شجرة الأخطاء لاحتمالات الثغرات الأمنية في نظم المحساب

علي محمد رشدي*، و عمر محمد باركب**

** قسم الهندسة الكهربائية وهندسة الحاسبات ، جامعة الملك عبدالعزيز ،*

*و ** قسم تقنية الحاسب ، كلية الاتصالات والإلكترونيات ،*

جدة – المملكة العربية السعودية

المستخلص. تمثل ورقة البحث هذه محاولة أولية لمواءمة أسلوب شجرة الأخطاء المستعمل في هندسة المعولية للاستخدام في التكمية (التقدير الكمي) للانكشافات الأمنية لنظم المحساب. وفي هذا السياق الجديد، يمكن وصف شجرة الأخطاء بأنها مخطط منطقي مدخلاته تمثل أحداث الثغرات عند مستويات النظام المختلفة ، ورؤوسه تمثل عمليات أو بوابات منطقية ، أما خرجه (الذي هو جذر شجرة الأخطاء) فيمكن أن يكون أيًّا من الأحداث الأوجية غير المرغوب فيها. تستعرض ورقة البحث باختصار الخوارزميات المستخدمة لتحويل التعبير التبديلي (البولاني) لمتغير البيان للحدث الأوجي إلى تعبير احتمالي. وبالحصول على قيمة احتمال الحدث الأوجي، يمكن ضرب هذه القيمة في قيمة مجروحية النظام بهذا الحدث للحصول على قيمة كمية لانكشاف النظام أمامه. تعالج ورقة البحث هذه أيضاً المسألة مزدوجة العشوائية الخاصة بتقدير الريبة في احتمال الحدث الأوجي ، وذلك باستخدام صيغة تحليلية دقيقة تربط التباين في احتمال الحدث الأوجي بالتباينات في احتمالات الأحداث الأساسية. يتم عرض مثال لنظام محسابي نمطي يجري فيه الحصول على تقديرات رقمية لاحتمالات الأحداث الأوجية وتبايناتها، كما يجري عمل ترتيب لأحداث الثغرات المختلفة فيها حسب أهميتها.