LA-UR-98-- 3117

CONF-990120--

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36

TITLE: Evaluating the Risk of Industrial Espionage

AUTHOR(S): T. F. Bott

SUBMITTED TO: 1999 Annual Reliability and Maintainability Symposium
January 18–21, 1999
Washington, DC

Los Alamos

Los Alamos National Laboratory
Los Alamos, New Mexico 87545

Evaluating the Risk of Industrial Espionage

T. F. Bott; Los Alamos National Laboratory, Los Alamos

Key Words: Security, Espionage, Information Protection, Probabilistic Modeling, Industrial Espionage

SUMMARY AND CONCLUSIONS

A methodology for estimating the relative probabilities of different compromise paths for protected information by insider and visitor intelligence collectors has been developed based on an event-tree analysis of the intelligence collection operation. The analyst identifies target information and ultimate users who might attempt to gain that information. The analyst then uses an event tree to develop a set of compromise paths. Probability models are developed for each of the compromise paths that use parameters based on expert judgment or historical data on security violations. The resulting probability estimates indicate the relative likelihood of different compromise paths and provide an input for security resource allocation.

Application of the methodology is demonstrated using a national security example. A set of compromise paths and probability models specifically addressing this example espionage problem are developed. The probability models for hard-copy information compromise paths are quantified as an illustration of the results using parametric values representative of historical data available in secure facilities, supplemented where necessary by expert judgment.

The methodology in this study was developed for national security applications, but the approach is equally applicable to industrial espionage.

Industrial facilities face insider information theft as well as compromise by visitors. The method should be adaptable to most industrial situations with modifications to fit the specific situation and is direct and simple to use. When historical data are not available, expert judgment data can be used as an input. Even in the absence of quantitative data, considerable insight into espionage risk can be gained by developing the compromise paths and their attendant probability models. These qualitative insights may be the greatest benefit gained when applying this methodology.

1. *INTRODUCTION*

The protection of information from espionage is considered vital to our national security interests, but the resources at the disposal of security personnel are not unlimited. One rational criterion for allocating these resources is according to the relative likelihood of different espionage scenarios. This paper presents a methodology for assessing the relative likelihood of espionage scenarios that involve employees of or visitors to a secure facility who are recruited as agents by an outside interest. The method was developed for analyzing espionage attempts on national security information by hostile groups. An analogous situation exists in industrial settings, where information on a product or process may be the target of hostile interests in the form of competitors. This paper is presented in the hope that the methodology developed for a national security setting also will find application in an industrial setting.

An outline of the methodology is shown in Fig. 1. First, a set of espionage scenarios, called compromise paths, are developed using an event tree. These compromise paths are grouped into a classes that can be represented by a common probability model because of shared characteristics. A probabilistic model is developed for each of the classes of compromise paths using parameters that can be estimated using historical experience or expert judgment. Compilation of historical experience and expert elicitation provides a basis for probability-of-occurrence calculations for each compromise path. The compromise probabilities provide a metric for rank ordering the different paths according to relative likelihood.

The compromise probability methodology steps outlined above are explained using a national security application. The compromise paths and probability models in this study are illustrative of the process the analyst would follow when analyzing a specific espionage situation and should be viewed as a tutorial for the method rather than a universally applicable set of logic models and equations.

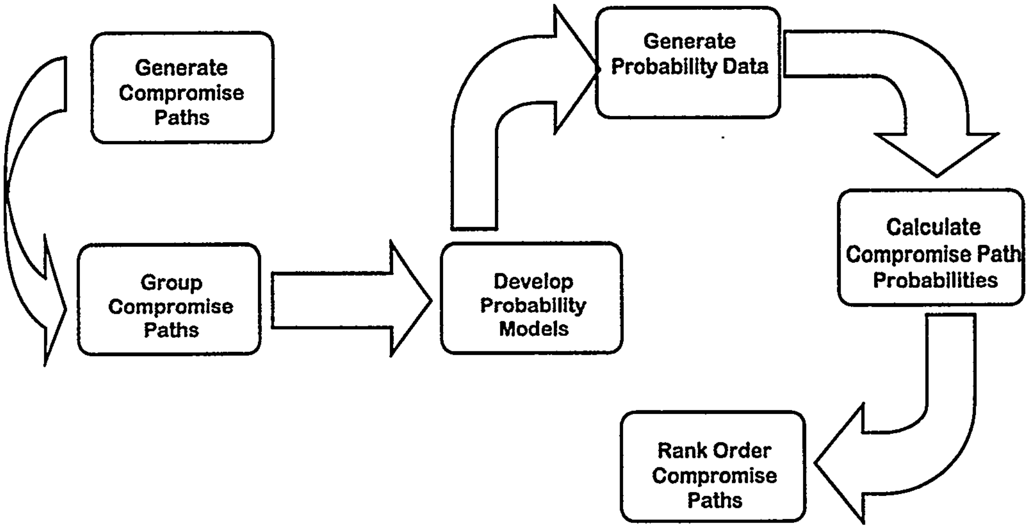
Generate Compromise Paths
Group Compromise Paths
Develop Probability Models
Generate Probability Data
Calculate Compromise Path Probabilities
Rank Order Compromise Paths

Fig. 1. Compromise probability methodology overview

2. GENERATING COMPROMISE PATHS

Possible espionage scenarios for a given situation are modeled as discrete event sequences that we will term "compromise paths." The compromise paths are specific realizations of the process shown in Fig. 2. According to this model, protected information is held in containers within a secure facility. Access to both the facility and the containers is controlled by clearances, physical barriers, and administrative controls. A compromise path begins when an ultimate user identifies target information in a secure facility. The ultimate user and target information largely determine the compromise paths of interest and their characteristics. The ultimate user recruits an agent who gains entrance to the secure facility by an entrance mode that is commensurate with his status. Once inside the facility, the agent's freedom of movement depends on his clearance level. The agent uses an access mode to gain access to the target information. After gaining access, the agent completes his assignment by transmitting the compromised information to the ultimate user using a compromise mode. The points at which security interdiction may occur are noted using dashed lines. Security interdiction is not analyzed in this study but is a standard subject of probabilistic analysis for security problems.

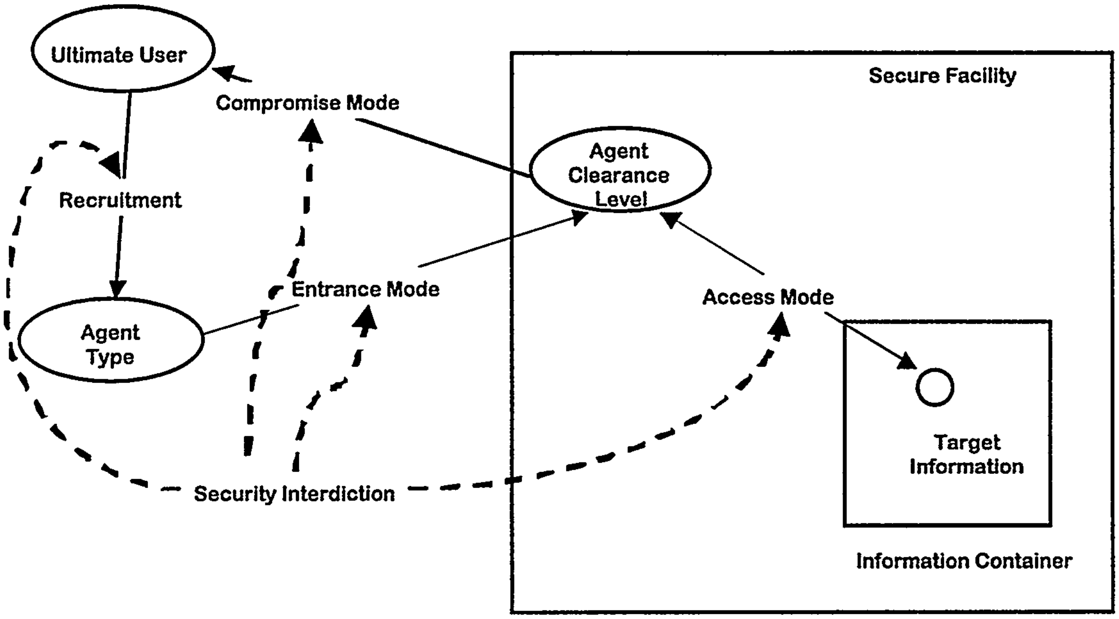

Fig. 2. Generic model of compromise process

The compromise process in Fig. 2 can be visualized as a directed graph with people or objects represented as nodes and actions such as recruitment represented as arcs. A compromise path is a cycle beginning and ending at the ultimate user node and traversing every arc once in each allowed direction. This generic compromise path forms the basis for the generation of specific compromise paths. By replacing the generic elements with possible realizations of each element, a large number of possible compromise paths can be generated. We use the event tree in Fig. 3 as a possibility generator for compromise paths. The generic compromise path along the top of the event tree tells us which elements every compromise path must contain. Starting with the "ultimate user" and working right, the possible realizations for each element in the compromise path are enumerated, with each one forming a branch off an existing branch in the event tree. The taxonomy used to generate the possible specific realizations for each element in the generic compromise path is discussed below.

2.1 Ultimate Users

The ultimate user is treated as the 'initiating event' in the compromise path. The ultimate user is important in determining the agent mode most likely to be used as well as the access modes that are most acceptable in obtaining the desired information. In the national security study, the potential users of national security information included foreign national military establishments and terrorists. Ultimate users in an industrial setting would be competitors.

The types of ultimate users who would be interested in a particular type of information will vary. Determining the type of ultimate user is important in determining the agent mode most likely to be used as well as the access modes that are most acceptable in obtaining the desired information. The probability that an ultimate user is interested in national security information is included implicitly in the historical data collected on the recruitment of agents. In a national security application, these data include all known agents recruited at installations similar to the target facility during specific time periods. We have assumed that the level of interest among potential ultimate users for national security information is the same as that exhibited in the historical data and that agent recruiting capabilities are similar to historical capabilities as represented in the historical data.

The use of historical data to include interest of the ultimate user is supported by an examination of agent occurrence data from other sites. One set of data was collected from a facility that had a particularly high interest to ultimate users of that time. The agent recruit rate for this facility was almost 100 times the average rate for secure facilities. Recruitment rate includes a component of interest as well as the mechanics of recruiting an agent after the target is chosen.

These threats evolve with changing political conditions. Thus, some modification of the uniformity-of-interest assumption may be justified. In an industrial setting, historical rates of interest may be modified as new competitors appear or as political realities bring new potential government-supported espionage threats into the arena.

2.2 *Recruitment Mode*

The recruitment mode does not appear in the generic event sequence for the event tree because no distinction between possible modes was considered important in the national security study.

2.3 *Agent Type*

For the purposes of espionage at a national security facility, the potential agent could be an employee of the target facility, in which case he is called an **insider**. An insider can enter the facility unescorted. A **visitor** to an NS facility is typically an employee of a different site, a domestic contractor, or a foreign delegate who enters the facility under escort.

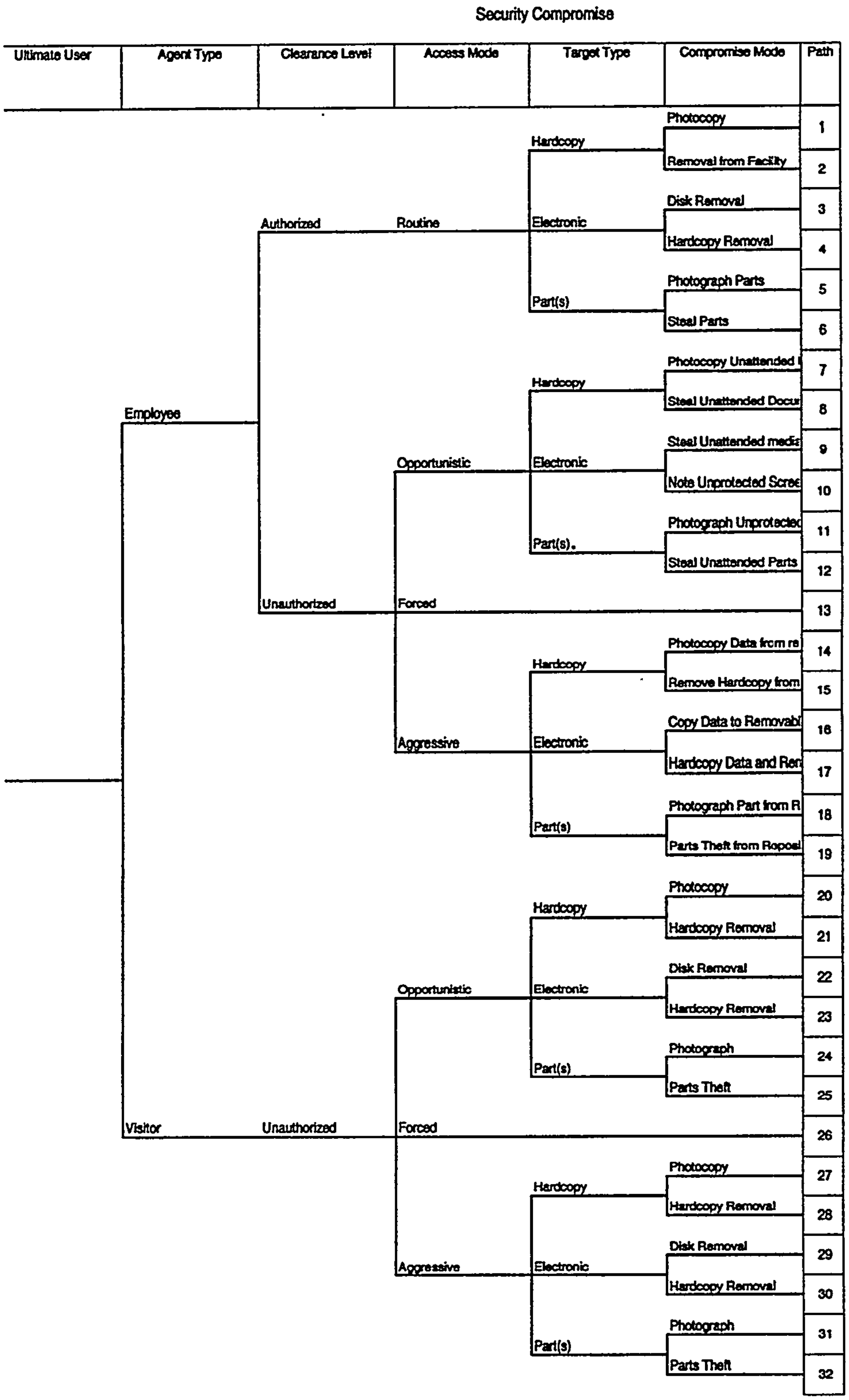

Fig 3. Protected information compromise event tree

2.4 Entrance Mode

The entrance mode describes the manner in which the agent gains entrance to the facility. In the national security study, the entrance mode was considered implicit in the agent type.

2.5 Clearance Level

Access to secure information in a facility is controlled by means of a **clearance level** that determines how the agent moves about the facility and accesses the target material. The taxonomy for clearance level includes **authorized** clearance with open and free access the target material without escort and **unauthorized** clearance level.

2.6 Access Mode

The access mode describes how the agent gains access to the target material once inside the facility. For an authorized insider, the access mode is **routine** by definition. For unauthorized agents, an **opportunistic** agent collects protected material left unsecured and unattended, an **aggressive** agent takes larger risks in checking for unsecured repositories, etc., and a **forced** entry agent could break into repositories or force those with access to open them.

2.7 Target Type

The target type describes the medium of the target material. In the national security example, this includes **hard-copy, electronic media,** and **parts** .

2.8 Compromise Mode

The compromise mode is the means by which the target material is transmitted to the ultimate user. In the example used in this study, the target information was assumed to be too complex or extensive to be merely read and transmitted orally by the agent.

3.0. PROBABILITY MODELS FOR COMPROMISE PATHS

Probability models were developed for the compromise paths in the event tree shown in Fig. 3. The numerical values of frequencies used in the probability formulae in this section were developed from historical experience, principally at government facilities. No data existed for some of the parameters required in the probability formulae, especially those that express the fractions of agent types. In these cases, estimates were generated by expert judgment elicitation (Ref. 1).

One of the major contributions of this methodology is the development of a discrete event model for espionage. This model provides many useful insights into espionage prevention even in the absence of quantitative data. These insights are gained by examining of the effect of different parameters in the probability models. These parameters correspond to characteristics of the espionage scenarios that determine their relative likelihood. By correlating changes in the parameters to changes in security practices, insights about espionage threat reduction can be generated without quantitative data. In our opinion, this type of qualitative analysis is more valuable in most cases than highly uncertain probability estimates. We have included such insights in the discussions of the probability models below.

Several assumptions underlie the models developed in this example. We assumed that historical data on security breaches are applicable to the current situation. We also assumed that security breaches occur at a constant rate per person, document, or repository at risk. Thus, we can use a maximum likelihood estimator to convert security breach data and the population at risk into rates. We then can treat the time to first occurrence of the security breaches as exponentially distributed. We also assumed that the facility starts 'clean' with no sleeper or other agent already in place at time zero.

3.1 *Authorized Insider Compromise Paths*

Compromise paths 1 through 6 on the event tree (Fig. 3) represent compromise paths for an authorized insider agent. The authorized insider has routine access to the target information. In estimating the probability of one of these compromise paths, only the probability of the ultimate user recruiting an authorized insider agent is relevant. Once such an agent is recruited, the agent has unlimited ability to access the information so the clearance level, access mode, and compromise mode are not important.

An estimate of the compromise path probability for authorized insiders can be generated by making a few simplifying assumptions so that an exponential distribution can be used. A preliminary data survey showed that there were potentially useful historical data at a number of secure national security facilities on both the number of cleared people who are at risk of becoming agents and on the actual number who do become agents as a function of time. Under a number of standard assumptions, the data from different facilities may be pooled to estimate a recruitment rate. The number of agents recruited is simply the total number of known agents, and the population is the integrated number of employee-years. If we assume a constant recruitment rate per employee-year in the candidate population, we can make a maximum likelihood estimator (MLE) of a recruitment rate, λ, per person-year (Ref. 2). The probability of recruitment at any given facility is an exponential distribution with the recruitment rate λN dependent on the recruitment rate per employee-year and an employee population N at the facility. According to this model, the probability of at least one recruit in an employee population N during a time T can be found from

$$P(T) = 1 - e^{-\lambda NT} \quad . \qquad (1)$$

This recruitment probability includes the effect of ultimate user interest and the susceptibility of the employees to recruitment through the rate λ. Thus the applicability of the model is highly dependent on the similarity of the facility of interest to the facilities that make up the data base..

Despite the simplicity of Eq. (1), a number of insights into protecting against the authorized insider can be gained from it. One way to reduce the threat from the authorized insider is to restrict the number of people with authorization. This reduces the N in Eq. (1). However, complete reliance on restricted access is not a panacea because the historical record shows that an ultimate user can increase the recruitment effort when interest in a target is high and can offset the reduction in N by increasing λ. A more expensive strategy than restricting access would be to require a "two-man rule" when accessing particularly valuable information. A single insider now must either gain unauthorized access or the ultimate user must recruit two insiders simultaneously. Limiting the time that an employee has access to specific information [reducing T in Eq. (1)] also could reduce the risk from the authorized insider. A final defense against the authorized insider is security interdiction against the compromise mode in the form of exit searches. This analysis clearly shows why the authorized insider is a very difficult threat for security to handle.

Another aspect of industrial sabotage that is different from national security applications is the high occurrence of criminal espoinage. We define criminal espionage as theft of information by someone acting as his own agent who then sells the information to another user. This type of espoinage can be treated using the insider models below where recruitment rate is replaced by occurrence of such criminals.

3.2 Opportunistic Insider Compromise Paths

Compromise paths 7 through 12 on the event tree represent an opportunistic unauthorized insider who is only willing to access information that is left unsecured. All insider agents are assumed to be willing to engage in opportunistic collection, but some fraction is assumed to engage only in opportunistic collection. The compromise paths for the opportunistic insider are more complicated than those for the authorized insider. An unauthorized insider does not have free access to the target information and so must create his own access. An opportunistic insider will access only information that is unattended and unsecured, so he must rely on security lapses to gain access. We have modeled the recruitment of agent followed by the occurrence of a security lapse that allows the agent access to the target material. The compromise probability distribution is a convolution of two exponential distributions (Ref. 3), one for recruitment and one for a security lapse. The compromise probability within time T is given by

$$P(T) = 1 + \frac{\mu M}{\lambda N - \mu M} e^{-\lambda N T} - \frac{\lambda N}{\lambda N - \mu M} e^{-\mu M T} \quad . \qquad (2)$$

In this equation, μ is the security lapse rate per item at risk, M is the number of protected items at risk, λ is the agent recruitment rate per person at risk, and N is the population from which the agent is recruited. We have implicitly assumed in this model that the opportunistic insider will find every opportunity that occurs. This assumption can be easily relaxed by multiplying the security lapse rate by a fraction f_d to reflect only those that the agent discovers.

Several limiting cases for the opportunistic insider are interesting. If the recruitment rate is very high compared with the security lapse rate, then Eq. (2) reduces to an exponential model of the probability of a security lapse. If the security lapse rate is very high relative to the recruitment rate, then Eq. (2) reduces to an exponential model of the recruitment probability. These two limits represent the situation when the agent appears very quickly and waits for a security lapse and when the security lapses occur so often that one occurs almost immediately after any agent is recruited.

The data required for this model can be taken from historical rates for μ and λ. M is the number of target documents at risk, and N is the pool of employees at the facility who do not have authorized access to the protected information but could gain unescorted access to areas where the target information is kept.

An opportunistic insider is totally dependent on security lapses in the protection of information outside repositories or containers to gain access to protected information. Thus, the opportunistic insider may be combated by adherence to good security practices among the employees of a facility. Protecting information by maintaining control and restricting access while the information is in one's possession will reduce the security lapse rate, μ, in Eq. (2).

3.3 *Aggressive Insider Compromise Paths*

Compromise paths 14 through 19 on the event tree are for an aggressive unauthorized insider. We assumed that some fraction f_a, of unauthorized insiders will be willing to take more aggressive actions to obtain the target information. The equation for the probability of an aggressive insider accessing information beyond that which an opportunistic insider would get is similar to Eq. (2). The difference is that now only a fraction, f_a, of the pool are willing to be aggressive insiders, and the items at risk are unsecured repositories, not unattended materials or computer screens. The equation for the probability of an aggressive insider getting information from an unsecured repository or computer is given by

$$P(T) = 1 + \frac{\mu M}{\lambda f_a N - \mu M} e^{-\lambda f_a N T} - \frac{\lambda f_a N}{\lambda f_a N - \mu M} e^{-\mu M T}, \qquad (3)$$

where μ is the rate that repositories are left unsecured and M is the number of repositories at risk.

The aggressive agent is combated effectively by good individual security practices among the employees such as securing safes or other repositories when not attended. Watchfulness and rigorous reporting of any suspicious activities will increase the probability of detection for an aggressive insider who is trying to get into restricted containers. This will reduce the fraction f_a of recruits who are willing to attempt such risky behavior and turn aggressive insiders into opportunistic insiders.

3.4 *Forced-Access Insider Compromise Paths*

A very bold unauthorized insider could attempt forced access to closed containers with high-value information inside. This compromise path is number 13 on the event tree. The model is similar to Eq. (1) with the addition of the fraction f_f of the prospective agent pool who would be willing to attempt forced access. The probability model is given by

$$P(T) = 1 - e^{-\lambda f_f N T} \quad . \tag{4}$$

Forced access may be combated by alarming or otherwise protecting repositories to increase the risk and difficulty in forcing access to them. Good employees practices in reporting suspicious activities also will increase the risk to the agent. These measures will reduce the fraction, f_f, of agents willing to engage in forced-access espionage.

3.5 *Opportunistic Visitor Compromise Paths*

Compromise paths 20 through 25 are for an opportunistic visitor agent. The occurrence of an agent in a single group of visitors to a secure facility was modeled with a binomial distribution. We used a cumulative probability of 1 or more agents occurring among N_v visitors when the probability that any one visitor is an agent is p_a. When the opportunistic visitor agent is admitted, he relies on escort and security lapses to gain access to the target materials. The time during which these events occur is typically short. With this assumption, we simplify the model by multiplying the occurrence probabilities for escort and security lapses occurring during the visit period instead of calculating a convolution in which the security lapse occurs after the escort lapse but within the time of the escort lapse. The probability of one or more attempts by a visitor in a single group to access the target material using our model is

$$P(>1) = \left(1-\left(1-p_a\right)^{N_v}\right) \left(1-e^{-\mu M t}\right) \left(1-e^{-\omega N_v t}\right) , \qquad (5)$$

where N_v is the number of visitors in the group, M is the number of information items at risk, p_a is the probability that any one visitor is an agent, m is the rate of material-unattended security lapses, ω is the rate of escort lapses, and t is the time duration of the visit.

In many cases, a number of separate groups may visit a facility, each potentially including a visitor agent. The probability of one or more compromises of the sort represented by Eq. (5) among n visits is a case of probabilities of overlapping events and was solved in general by Poincaré (Ref. 4):

$$P_N^m(\geq 1) = \sum_{r=1}^{N} (-1)^{r+1} S_r \quad . \qquad (6)$$

In this equation, S_1 is the sum of the probabilities, p_i [in our case probabilities, like Eq. (5) for different visits]. S_2 is the sum of the probabilities p_{ij}, where two compromises occur as a result of two visits and so on. Higher order terms such as S_2 and S_3 are important when the overlap between events is large, but these terms usually are small.

The importance of good security practices in reducing espionage by visitors is graphically illustrated by Eq. (5). Attention to information protection will reduce the material-unattended lapse rate, μ. Careful escort performance will decrease the rate of escort lapses, ω. Both of these rates strongly affect the compromise probability. Minimizing the number of visitors, N_v, looks like it could help, but a determined ultimate user will simply make sure that an agent is among the visitors that are admitted.

Considering both Eqs. (5) and (6), an interesting observation may be made. If the effective escort lapse rate, ωN, can be kept the same, it is better to have a few large groups of visitors rather than an equal number spread over more visits. It can be shown that the compromise probability resulting from a single group of N_v visitors is always smaller than that of N_v single visitors groups. This is because the exposure time during which a security compromise can occur is much larger for the multiple visits. One way to reduce the espionage risk from visitors is to host only one well-escorted group rather than a large number of smaller, but equally well escorted, groups

3.6 Aggressive Visitor Compromise Paths

Compromise paths 27 through 32 are for an aggressive visitor agent. Our model uses Eq. (5) above to estimate the probability of an attack by an aggressive visitor agent by making two modifications. First, the security lapse rate, μ, represents unsecured repositories. Second, the probability of a visitor is multiplied by a fraction, f_a, to account for potentially greater difficulty in recruiting an aggressive agent.

3.7 Forced-Access Visitor Compromise Paths

Compromise path 26 represents a forced-access visitor agent who will gain forced entry to his target. The probability of a visitor who would be willing to obtain forced access to his material was estimated as the probability of occurrence in a visitor population, N_v, of an agent at a probability, $p_a f_a$, where the fraction f_a represents the difficulty in finding a visitor willing to force entry. We assumed that such an operative will not have to wait for an escort lapse but will lose the escort as needed. The probability model for such an attempt is

$$P(>1) = \left(1 - \left(1 - p_a f_a\right)^{N_v}\right) \; . \tag{7}$$

The main tool at the disposal of security in addressing the forced-access visitor is to make forced-access as daunting a task as possible. This has the effect of reducing the fraction f_a in Equation (7).

4. COMPROMISE PROBABILITY ESTIMATES

Table 2 contains a set of compromise probabilities showing the types of results obtained from this method. The compromise paths considered are those shown on the event tree in Fig. 3 for hard-copy target types. Each compromise path is summarized on a separate row. The event-tree path numbers are shown in column one with the compromise paths described in column two. The point estimate of the compromise probability is shown in the last column in each row. The columns between the second and last columns contain parameters used in the probability calculation. This example is based on fabricated data for security reasons and is not meant to represent any specific facility or national security sector. Nevertheless, the general ordering of the compromise paths is representative of actual national security application with insider threats dominating

For clarity and simplicity, only point values are presented in this table. The uncertainties associated with the sample size are available for historical data, and estimates of the uncertainty in the expert judgement can be generated as well. The calculations shown here were performed using spreadsheets, so uncertainties can be propagated readily in the probability estimates using Monte Carlo or Latin hypercube simulation.

REFERENCES

1. M. Meyer, J. Booker, *Eliciting and Analyzing Expert Judgment*, Academic Press, 1991.

2. H. F. Martz, R. A. Waller, "Handbook of Bayesian reliability estimation methods," Los Alamos Scientific Laboratory report LA-6572-MS, 1976 Nov.

3. E. J. Dudewicz, *Bayesian Introduction to Statistics and Probability*, 1970, Holt, Rinehart and Winston.

4. R. von Mises, *Mathematical Theory of Probability and Statistics*, Academic Press, 1964.

Table 2. Probability of Compromise for Any Hard-Copy Item

Event Sequence Numbers	Agent Type	Recruitment Rate or Agent Probability	Access Mode Fraction	Recruit Pool	Document, Part or Repository Pool	Item or Repository Unsecured and Unattended Rate	Time (years)	Probability of Document unattended during visit	Probability of Escort Loss of Control	Compromise Probability
1 and 2	Authorized Insider	4.00E-05	1.00	1250	NA	NA	1.00	NA	NA	0.05
7 and 8	Opportunisitc Insider	2.09E-05	1.00	1250	1.00E+06	1.05E-05	1.00	NA	NA	0.02
14 and 15	Aggressive Insider	2.09E-05	0.10	1250	1.25E+03	5.01E-02	1.00	NA	NA	0.003
13	Forced-Access Insider	2.09E-05	0.01	1250	NA	NA	1.00	NA	NA	0.0003
20 and 21	Opportunisitic Domestic Visitor	1.00E-04	1.00	12000	1.00E+06	1.05E-05	2.74E-03*	2.83E-02	3.56E-02	0.0007
20 and 21	Opportunisitic Foreign Visitor	1.00E-02	1.00	100	1.00E+06	1.05E-05	2.74E-03	2.83E-02	3.56E-02	0.0006
27 and 28	Aggressive Domestic Visitor	1.00E-04	0.10	12000	1.25E+03	5.01E-02	2.74E-03	1.58E-01	3.56E-02	0.0006
27 and 28	Aggressive Foreign Visitor	1.00E-02	0.10	100	1.25E+03	5.01E-02	2.74E-03	1.58E-01	3.56E-02	0.0005
26	Forced-Access Foreign Visitor	1.00E-04	1.00E-03	12000	NA	NA	NA	NA	NA	0.001
26	Forced-Access Domestic Visitor	1.00E-02	1.00E-03	100	NA	NA	NA	NA	NA	0.001

*2.74 x 10^{-3} years is equal to 1 day.