

# Addressing Low Base Rates in Intrusion Detection via Uncertainty-Bounding Multi-Step Analysis

Robert J Cole

*Cyber Security Lab, School of Information  
Sciences and Technology, Pennsylvania  
State University  
rcole@ist.psu.edu*

Peng Liu

*Cyber Security Lab, School of Information  
Sciences and Technology, Pennsylvania  
State University  
pliu@ist.psu.edu*

## Abstract

*Existing approaches to characterizing intrusion detection systems focus on performance under test conditions. While it is well-understood that operational conditions may differ from test conditions, little attention has been paid to the question of assessing the effect on IDS results of parameter estimation errors resulting from these differences. In this paper we consider this question in the context of multi-step attacks. We derive simulated distributions of the posterior probability of exploit given the observation of a series of alerts and bounds on the posterior uncertainty given a particular distribution of the model parameters. Knowledge of such bounds introduces the novel prospect of a confidence versus agility tradeoff in IDS administration. Such a tradeoff could give administrators flexibility in IDS configuration, allowing them to choose detection confidence at the price of detection latency, according to organizational priorities.*

## 1. Introduction

In 1987, Denning [1] introduced the concept of a general-purpose intrusion detection model designed to detect intrusions based on deviations from an existing model of normal system use. This approach inspired much subsequent research in the area of anomaly detection and the development of various anomaly-based intrusion detection systems (IDSs). An example of an early IDS implementation is SRI's prototype rule-based expert system known as Intrusion Detection Expert System [2] (IDES) which was later developed into Next-Generation Intrusion Detection Expert System (NIDES) [3]. Anomaly detection systems like these construct profiles of normal behavior using machine learning techniques where a training phase incorporating attack-free data is used to train the

anomaly detection algorithms. Following completion of the training phase, the system operates in the detection phase in which alarms are raised for events that deviate significantly from the normal profile. Complementary to this approach, another variety of anomaly detector uses some form of specification of normality to construct the normal profile. Various types of specification approaches have been proposed, including specification of application behavior [4] and specification of network protocol behavior [5].

This paper addresses two fundamental IDS issues. The first underlies all of the disparate forms of intrusion detection based on a probabilistic model, the problem of confidence in an inference drawn from an IDS sensor's output. A given probabilistic model used by an IDS, e.g. Bayesian, can be used to draw a quantitative inference over some observation, such as whether an exploit has occurred given a sensor has produced an alert, but what confidence can be assigned to such inferences? The answer depends on the confidence in the model parameters (e.g. false positive rate). We show that high confidence Bayesian inference is infeasible under typical conditions of uncertain model parameters. We argue that model parameters cannot be confidently estimated with arbitrarily high confidence and hence IDS inference uncertainty is an issue to be addressed. Secondly, we consider the low base rate problem in Bayesian IDS inference. This problem, denoted the *base rate fallacy* by IDS researchers, can result in low predictive value for a model inference under very low (i.e. real-world) base rates.

We examine both issues and show that neither is fatal because many real-world attacks involve multiple chained exploits. We refer to probabilistic inference in such cases as *multi-step* inference. The traditional case of inferring the occurrence of an exploit via observation of an alert associated with that one exploit is referred to here as *single-step* inference. Our analysis shows that the two issues of inference confidence and

predictive value, which can be so acute in the single-step case, are alleviated in the multi-step case, given our assumptions. To the best of our knowledge, this is the first work to analyze these issues in the multi-step case.

The biggest limitation of current (profile or specification based) IDS systems is probably that high false alarm rates necessarily result in a lack of trust in IDS output. In fact, this limitation is the primary reason for the reality that very few profile/specification based IDS systems are being seriously used in the field. The significance of our findings lies in the fact that they suggest a new option for security administrators to "evade" this limitation, one allowing for a tradeoff in which trust can be prioritized. In multi-step attacks, trust in IDS output can be gained at the cost of reduced detection speed. The theoretical results demonstrated in this paper enable this tradeoff by characterizing the conditions under which trust can be gained. The practical contribution made in this paper is to take a first step toward characterizing this tradeoff in the context of a linear exploit chain.

The rest of this paper is organized as follows. Section 2 examines the case of single-step Bayesian inference where it is shown that high confidence, highly predictive inference is infeasible. In section 3 we present an overview of our approach. In section 4, the multi-step case is analyzed and it is shown that issues plaguing the single-step can be alleviated under multi-step inference. Section 5 presents a study of a particular multi-step case. In section 6 we discuss our results and argue for the tradeoff described above.

## 2. THE SINGLE-STEP CASE: INFERENCE UNCERTAINTY AND THE BASE RATE FALLACY

In this section we review a fundamental problem with single-step Bayesian inference, a problem termed the *base rate fallacy* [6]. Additionally, we show the effect of parameter uncertainty in this context. Overall, it is clear that the combination of these two effects renders single-step probabilistic inference infeasible.

Intrusion detection systems (IDSs) are characterized by receiver operating characteristic (ROC) [7] curves which plot detection, or true positive (TP) rate versus false alarm, or false positive (FP) rate. In general, detection rate monotonically increases with false alarm rate thus high detection rates often only can be achieved at the expense of high false alarm rates. Since a given IDS system can be tuned to operate at any point along its characteristic ROC curve, a question concerning IDS operation arises: at what point on the ROC curve should the IDS be operated? One

approach is to operate at a point dictated by operational needs [6], such as choosing a tolerable false alarm rate based on the administrative resources available to investigate alarms [3]. Another approach is proposed in [8] where costs of alarm responses are modeled and an operating point is chosen based on minimizing an objective function measuring response cost.

Regardless of how an operational point is chosen, the values of the system parameters represented by the point are likely to have a degree of uncertainty. This is a consequence of the fact that IDS operational parameters are selected based on detector performance in limited test cases, creating a question of ecological validity. If the operational environment differs from that of the test environment, the inevitable result is uncertainty in the results of detection [9]. If one considers IDS parameters in an operational setting to be the true parameters, then the parameters derived under test conditions must be viewed as estimated parameters. These estimates can in principle vary significantly from the true parameters, depending upon the level of accord between the test and operational conditions. Given that bias and variance may exist in parameter estimates, it is reasonable to consider the consequence of such parameter uncertainty on inferences drawn from a probabilistic model. Below we consider this question in the single-step case.

### 2.1 The Problem of Low Base Rate: Revisiting the Base Rate Fallacy

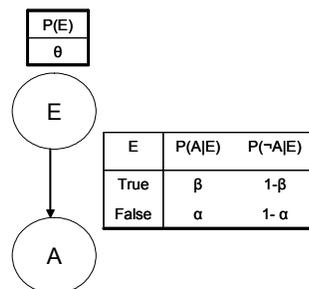


Figure 1. Single-step Bayesian model

A fundamental limiting factor for the performance of an intrusion detection system is the false alarm rate. As noted in [6], the posterior probability of an intrusion given an observed alert can be much lower than the true positive rate when the unconditional probability of intrusion (base rate) is small. This well-known result, termed the base rate fallacy, is often surprising because people typically do not consider the effect of the low base rate on the posterior probability. As a simple example, consider the single-step Bayesian network shown in Figure 1.

This simple belief network consists of a single exploit, E, and a single alert, A. The network is defined by three parameters,  $\theta$ ,  $\alpha$  and  $\beta$ . The unconditional probability of E is  $\theta$ , also known as the base rate (BR). The base rate is an environmental parameter which, for realistic situations, is expected to be very low, on the order of  $10^{-5}$  [10]. The conditional probability of A given that E has occurred (TP rate) is  $\beta$  and the conditional probability of A given that E has not occurred (FP rate) is  $\alpha$ . Two quantities of interest can be derived from this model, the probability that an exploit has occurred given that an alert has been raised and the probability that no exploit has occurred given that no alert has been raised. Following [10], we denote these quantities as positive predictive value (PPV) and negative predictive value (NPV), respectively. The rules of Bayesian inference dictate the following expressions for these quantities:

$$PPV = \frac{p(E, A)}{p(A)} \quad (1)$$

$$= \frac{\beta\theta}{\beta\theta + \alpha(1-\theta)}$$

$$NPV = \frac{p(\neg E, \neg A)}{p(\neg A)} \quad (2)$$

$$= \frac{(1-\alpha)(1-\theta)}{1-\beta\theta - \alpha(1-\theta)}$$

To understand how PPV and NPV relate to the environmental parameter BR and system parameters FP and TP, we show contours of PPV for BR=0.1 (Figure 2a) and BR=0.01 (Figure 2b) and NPV for BR=0.1 (Figure 2c) and BR=0.01 (Figure 2d). In these diagrams, higher values of PPV and NPV are represented with regions of lighter colored shading. Four regions are shown in the diagrams separated by contours with values 0.85, 0.90, and 0.95.

From these contour diagrams, we note the following. First, high values of PPV can be achieved only for very restrictive values of TP and FP. This is indicated by the small area associated with high PPV, area that decreases with decreasing BR. For example, for BR=0.01 and the highest possible detection rate (TP=1.0), the false positive rate must be less than approximately 0.001 for PPV > 0.90. For realistic base rates, the trend evident in these diagrams implies that high a high PPV can only be achieved for impossibly small values of FP. Secondly, the opposite situation exists for NPV: high values of NPV can be achieved for large ranges of TP and FP values and these ranges increase for decreasing base rate. These trends between PPV/NPV and base rate can be understood by

considering the limits of PPV and NPV as base rate goes to zero. From (1) and (2) these limits can be determined to be:

$$\lim_{BR \rightarrow 0} PPV = 0 \quad \lim_{BR \rightarrow 0} NPV = 1$$

Hence, for small base rates, usefully high values of NPV can be obtained but the same is not true for PPV. Given that the base rate of intrusion can easily be on the order of  $10^{-5}$  or smaller, it follows that very small false positive rates are required for useful Bayesian inference in single-step scenarios. For example, given a .999 probability of detection (TP) and a 0.001 base rate, we require a false positive rate on the order of 0.0001 to achieve a 90% PPV. Such unrealistic levels of false alarm rate constitute a well-known fundamental problem with Bayesian inference in intrusion detection [6]. This issue stems from the fact that some observables processed by sensors are produced by both normal and anomalous processes. Consequently such an observable cannot be classified with zero error into either a normal or malicious category and hence a non-zero false alarm rate is an inevitable fact of any IDS based on current approaches [11].

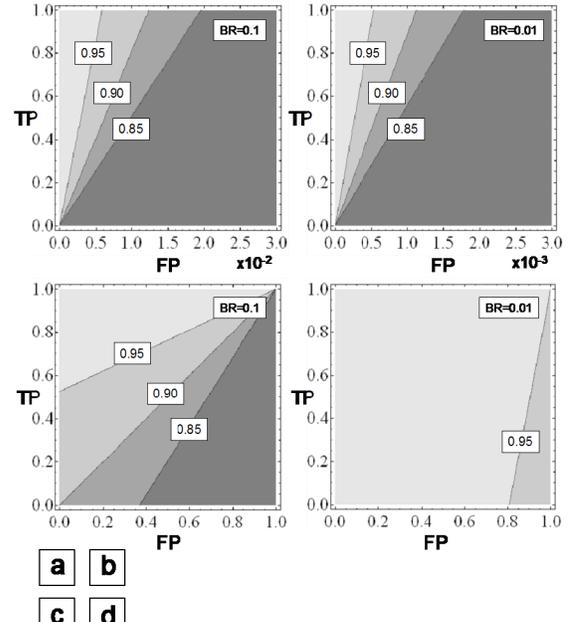


Figure 2 (a,b) PPV contours, (c,d) NPV contours

## 2.2 The Problem of Parameter Uncertainty

Aside from the problem of low PPV in realistic scenarios, there is another problem with Bayesian

inference in an intrusion detection scenario. As observed above, IDS sensor parameters such as detection rate and false alarm rate are typically tuned based upon performance against a test data set. Thus IDS parameters are in fact only estimated values, estimates that could deviate from their corresponding true parameter values. Additionally, environmental parameters, such as unconditional probability of exploit, are also subject to estimation uncertainty.

To investigate the effect of parameter uncertainty, we conduct a sensitivity analysis of the single-step Bayesian model in which the model parameters are treated as random variables and the probability density of PPV is determined. Under this transformation, PPV becomes a random variable whose variance constitutes the uncertainty in the posterior inference.

For this analysis, we assume the base rate  $\theta$  is known. In practice, this environmental parameter is not known but for present purposes only uncertainty in system parameters TP and FP is considered. Variation in TP and FP will produce variation in the inference PPV. This variation in PPV constitutes the uncertainty in the inference drawn from the Bayesian model. Specifically, we define uncertainty as follows:

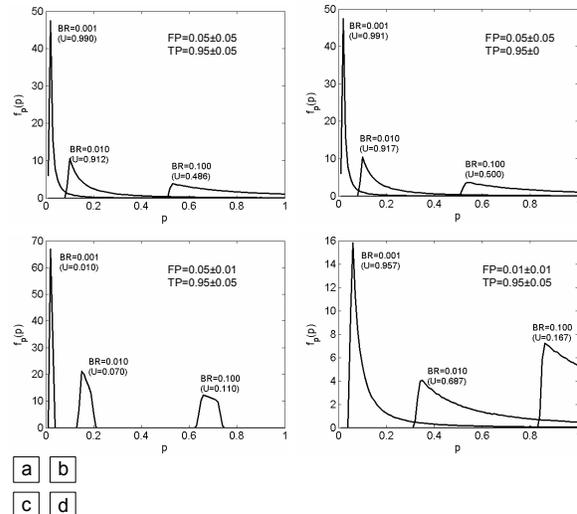
**Definition** Let  $p=PPV$ . The distribution of PPV is given by the density function  $f_p(p)$  with associated range of  $p$ ,  $p_{min} \leq p \leq p_{max}$ . The uncertainty  $U$  in PPV is the range of possible values of  $p$ :

$$U \equiv p_{max} - p_{min} \quad (3)$$

In other words, uncertainty  $U$  is simply the difference between the maximum and minimum possible values of PPV. This uncertainty results from uncertainty in parameter values. In this paper, parameters are assumed to be uniformly distributed over an interval centered on their estimated value. Let the estimated value of parameter  $v$  be  $v_{est}$  and  $v_w$  be the length of the uncertainty interval for  $v$ . Then the distribution of values of parameter  $v$  is given by

$$f_V(v) = \frac{1}{2v_w}, v_{est} - v_w \leq v \leq v_{est} + v_w \quad (4)$$

Figures 3a-d show the PPV density and resulting uncertainty for three base rate values: 0.1, 0.01 and 0.001 and several values of parameter uncertainty. These plots were obtained via simulation as follows. Uniformly random values of TP and FP were selected and PPV was computed according to (1). Normalized histograms of the results were obtained and are plotted in Figure 3.



**Figure 3. PPV distributions**

From these diagrams, we note the following. First, the PPV densities are uni-modal with peaks at decreasing values of  $p$  for decreasing values of BR. This trend is the base rate fallacy effect described above; the expected PPV decreases with BR. Secondly, note that PPV uncertainty can be very large when FP can approach zero. This is evident in Figures 3a,b,d which exhibit long tails in the PPV distribution. In Figure 3c, these long tails are not present because the FP distribution does not approach zero. This result underscores the extreme sensitivity of PPV uncertainty to FP; low values of FP are highly desirable but unless the uncertainty in FP is very small, a large uncertainty in PPV can result.

### 3. OVERVIEW OF APPROACH

In the previous section, we demonstrated two issues in the single-step case. First, we reviewed the well-known result that for realistic base rates, high PPV can only be achieved for very low false positive rates, as illustrated in Figure 2. Secondly, we showed that high PPV uncertainty can result from small parameter uncertainty, as illustrated in Figure 3.

Our observations in the previous section motivate us to understand the issues surrounding low base rate and inference uncertainty in the multi-step case. Our goal in this work is to show that these issues can be mitigated in the multi-step case and to identify the conditions under which useful levels of PPV and PPV uncertainty can be achieved. To accomplish this, we present a general linear multi-step model in section 4 and analyze that model via simulation. As in the single-step case, simulation was performed using uniformly random values of the TP and FP parameters. Uncertainty was obtained following the definition

above by determining the difference between the lowest and highest observed PPV values.

## 4. PROBABILISTIC MODEL FOR UNCERTAINTY-BOUNDING MULTI-STEP ANALYSIS

### 4.1 General Linear Model

In this section we extend the simple model given in Figure 1 to a general linear multi-step exploit chain shown in Figure 4. Our linear exploit chain models a multi-step attack scenario where a linear sequence of  $N$  exploits is executed to accomplish a specific attacker goal. At each step in the chain, the current exploit  $E_j$  can only be attempted if the prior exploit  $E_{j-1}$  was successful. Such a linear dependency between exploits is a particular instance of a more general dependency relation among exploits that can be expressed via an attack graph [12].

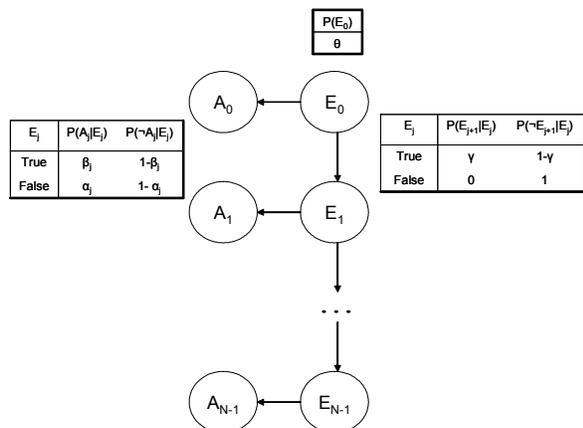


Figure 4. Linear exploit chain model

In the linear exploit chain model, it is assumed that an IDS sensor exists for each exploit  $E_j$ . For each sensor, there is a corresponding false alarm rate  $\alpha_j$  and detection rate  $\beta_j$  associated with the sensor's output  $A_j$ . Because an exploit  $E_{k+1}$  cannot be attempted until the prior exploit  $E_k$  is successful, the conditional probability  $P(E_{k+1}|\neg E_k)$  is zero. We assume that an attacker is free to choose whether to attempt exploit  $E_{k+1}$  after successfully executing exploit  $E_k$ . Thus we consider  $P(E_{k+1}|E_k)$  to be another model parameter  $\gamma$  which is assumed constant for all stages of the attack. Of course we expect that  $\gamma$  will be very high for real-world attacks, perhaps 1 in most cases since an attacker motivated enough to initiate a multi-step attack sequence is presumably sufficiently motivated to pursue the attack until the attack goal is realized. For all analysis in this paper,  $\gamma$  is assumed to be 0.99.

Note that in this section we analyze  $PPV = P(E_0 | A_0, A_1 \dots A_{N-1})$  for an  $N$ -step model. In other words, we assume that all alerts are raised during an attack. Of course, due to the non-zero probability of false negatives, this is only one of  $2^N$  possible result sets. Due to space limitations, we only analyze this one result set; however we note that since detection rates in general are expected to be higher than false negative rates, our analyzed set is the most likely observed sets. Nevertheless, a comprehensive analysis requires examination of all possible sets.

### 4.2 Example

To illustrate our Bayesian model, consider the following example taken from [13]. In this scenario, a web server resides in a DMZ network exposed to the Internet through a firewall. A second firewall segregates the DMZ from the internal LAN which contains a file server and a workstation. The workstation mounts an NFS file share from the file server, a file share containing binary executables. This configuration is shown in Figure 5.

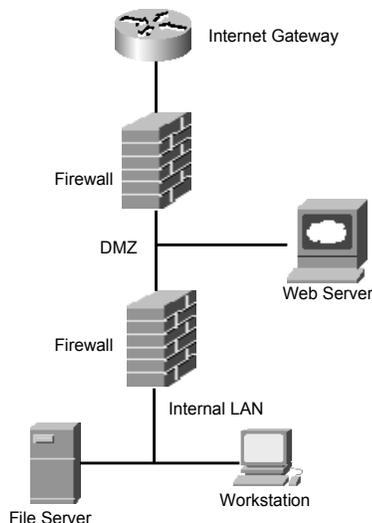


Figure 5. LAN configuration, adapted from [13]

In this example, the attacker's goal is to execute code on the workstation with root privilege. To accomplish this goal, the attacker needs to exploit vulnerabilities in the web server and file server. The sequence of steps executed by the attacker is as follows:

1. The attacker exploits a vulnerability on the web server and gains execute privilege on the web server.
2. Access controls in the firewall separating the DMZ from the internal LAN permit access to the file server from the web server. The

attacker uses execute privilege on the web server to exploit a vulnerability on the file server and gain root access on the file server.

3. The attacker uses file server access to modify arbitrary executable files on the file server which are later executed on the workstation.

We can represent this attack in the form of a 3-step model as shown in Figure 6. This model includes three exploit variables,  $E_0$ ,  $E_1$  and  $E_2$ , corresponding to the three steps described above. For each exploit we assume the existence of an associated IDS sensor with a detection rate of  $\beta_j$  and false alarm rate  $\alpha_j$ . For step 1 in which a web server is compromised remotely, a signature-based IDS sensor programmed to detect buffer overflows is a likely choice. Such a sensor is likely to have a very high detection rate and a very low false positive rate. For steps 2 and 3 we might use anomaly-based sensors having lower detection rates and higher false positive rates than the signature-based sensor 1. Regardless of the sensors used, in principle we can conceptualize them as parameterized by TP and FP rates that can be estimated. Given an estimate of the base rate of attack ( $\theta$ ), and estimates of TP ( $\beta_j$ ) and FP ( $\alpha_j$ ), we can draw inferences from this model given the observations of alerts from the sensors.

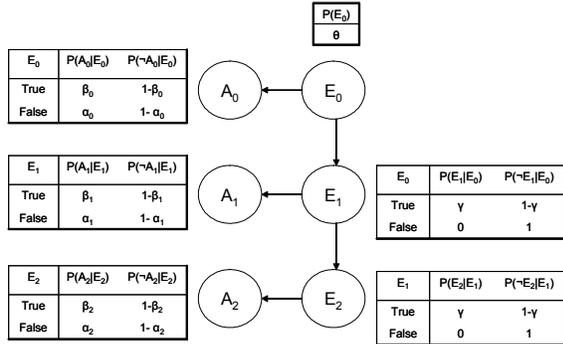


Figure 6. 3-step model

## 5. LINEAR MODEL ANALYSIS

In this section we analyze the linear model in presented in Figure 4. In section 5.1 we examine the effect of varying base rate on PPV with zero parameter uncertainty. In section 5.2 we examine the effect of parameter uncertainty on PPV.

### 5.1 Effect of Base Rate on PPV Under Zero Parameter Uncertainty

Imagine an IDS that outputs alerts along with an associated probability of the associated exploit. From the perspective of a security administrator interpreting

this IDS output, the probability represented by the inference must be high; otherwise it will not motivate one to take action. In the single-step case, it was shown that for low base rates, low PPV will result unless the sensors false positive rate is very low. Our hypothetical IDS will produce low PPV values and thus fail to provide the administrator with useful information.

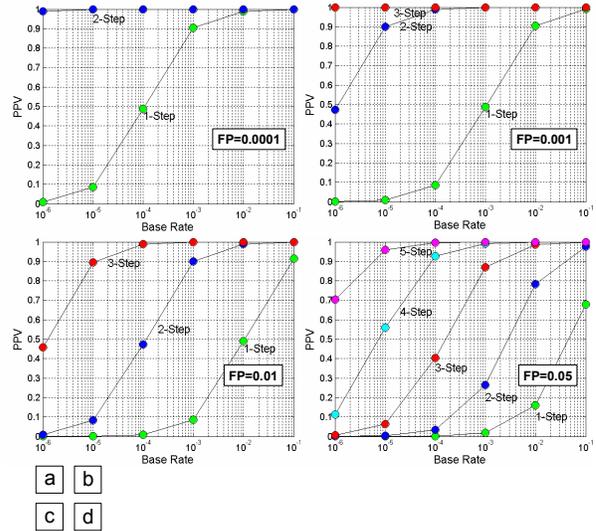


Figure 7. PPV vs. BR

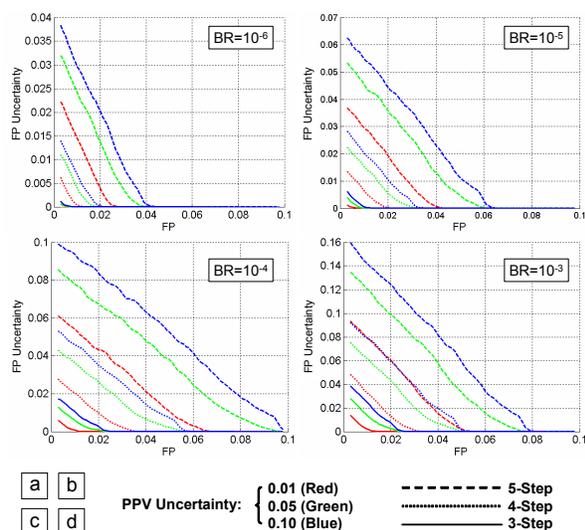
In Figure 7 we present curves of PPV as a function of BR for four different false alarm rates. Note that PPV is highly sensitive to BR, as in the single-step case. For low false positive rates ( $FP < 0.01$ ), high ( $> 0.9$ ) PPV values can be achieved for all model sizes of at least 2 steps if the base rate is no lower than  $10^{-5}$ . For higher FP rates, it can take many steps to achieve a high PPV. Consider for example a sensor with a false positive rate of 0.0001 and a base rate of  $10^{-5}$ . As Figure 7a shows, such a sensor only has a PPV of about 0.10 if operated by itself (single-step). However, a pair of such sensors can provide a PPV of almost 1.0 at the same base rate. However, for a false positive rate of 0.001, the PPV for a pair sensors is only 0.90 but a PPV of nearly 1.0 can be achieved by three sensors (Figure 7b).

### 5.2 Effect of Parameter Uncertainty on PPV Uncertainty

Our analysis has examined the uncertainty in PPV for the linear model given in Figure 4 for model lengths up to five. In the single-step case, shown in Figure 1, low uncertainty can only be obtained under conditions of very low base rate or very low parameter uncertainty. Since low base rate in the single step case implies low PPV, it is only under conditions of very low parameter uncertainty that low PPV uncertainty

can be obtained in the single-step case. Clearly, high-confidence (i.e. low uncertainty) inference is infeasible in the single-step case.

The results of the multi-step case, however, are not equally hopeless. This is due to the relationship between PPV uncertainty and model size shown in Figure 8. This figure plots the maximum FP parameter uncertainty under which a given PPV uncertainty can be obtained as a function of FP. Results for four base rates are shown:  $10^{-6}$ ,  $10^{-5}$ ,  $10^{-4}$ , and  $10^{-3}$  in Figures 8a-d, respectively. In each figure, three curves are shown for each of three model sizes, 3-step, 4-step and 5-step. The three curves shown for each model correspond to three different PPV uncertainty values: 0.01 (shown in red), 0.05 (shown in green) and 0.10 (shown in blue).



**Figure 8. Max parameter uncertainty**

The relationships identified in Figure 8 facilitate IDS design by defining system parameter requirements needed to support a desired PPV confidence level. For example, in the 5-step case for a base rate of  $10^{-5}$ , if a PPV uncertainty of 0.01 (1%) is desired, and the FP rate is estimated to be 0.03, the red 5-step curve indicates that the maximum FP uncertainty under which this PPV uncertainty can be achieved is 0.01. In other words, the true FP rate must lie in the interval  $0.03 \pm 0.01$  to achieve a PPV within a tolerance of 0.01. If our PPV uncertainty tolerance increases to 0.05, the tolerable FP range increases to approximately  $0.03 \pm 0.025$ . For a PPV uncertainty of 0.10, the tolerable FP range becomes approximately  $0.03 \pm 0.035$ . Note that the uncertainty in PPV as defined above is the difference between maximum and minimum PPV values. This is driven by the extreme values of FP. Thus when our uncertainty in FP is low, we can tolerate a higher estimated FP for a given PPV

uncertainty. Conversely, when uncertainty in FP is high, the estimated F must be low.

Note the deleterious effect of small base rate evident in Figure 8. As BR decreases, the maximum FP at which a given PPV uncertainty bound can be achieved decreases. This is the same fundamental base rate issue discussed in the single-step case; it has not disappeared in the multi-step case but instead is mitigated.

### 5.3 3-Step Example Under Heterogeneous Configurations

The above analysis assumed homogeneous configurations where each sensor had parameter characteristics that were identical with all of the other sensors in the model. Of course in real-world situations sensors can have very different characteristics and heterogeneous configurations may be common where some sensors have signature detection characteristics (i.e. low FP, high TP) and some may have anomaly detection characteristics (i.e. higher FP, possibly low TP). In this section we analyze a specific heterogeneous 3-step example.

Table 1.

Parameter	Sensor	Estimate	Uncertainty
FP	1	Variable	$\pm 0.01$
	2	0.001	0
	3	Variable	$\pm 0.01$
TP	1	Variable	$\pm 0.1$
	2	0.99	0
	3	Variable	$\pm 0.1$

This example consists of a 3-step model with as shown in Table 1 for a base rate of  $10^{-5}$ . Sensor 2 is modeled as a high detection rate, low false positive rate sensor, i.e. a sensor with signature detection characteristics. Also, sensor 2 is assumed to have zero-uncertainty parameter estimates. Sensors 1 and 3 are modeled as anomaly style sensors with variable TP and FP rates and non-zero parameter uncertainty.

The results of simulation analysis of this model are presented in Figure 9 which shows PPV (Figure 9a,c) and PPV uncertainty (Figure 9b,d) as a function of FP and TP for two base rates,  $10^{-4}$  and  $10^{-5}$ . Note the lack of sensitivity to detection rate (TP) for both PPV and PPV uncertainty. As Figure 9 shows, higher tolerance for parameter uncertainty can be assumed in this heterogeneous 3-step case than that shown in Figure 8 which reflects the homogeneous 3-step case. At a base rate of  $10^{-5}$ , an FP rate of approximately  $0.01 \pm 0.01$  can be tolerated while still providing a PPV uncertainty of less than 0.05. A wider range can be tolerated at the higher base rate of  $10^{-4}$ .

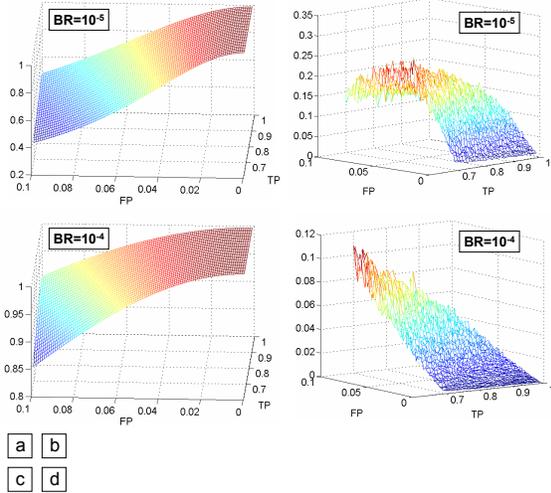


Figure 9. (a,b) PPV, (c,d) PPV uncertainty

## 6. NORMATIVE IMPLICATIONS: A SPEED/ACCURACY TRADEOFF

The results shown in Figure 8 represent the basis for flexibility in IDS administration. Specifically, they define a relationship between speed and accuracy. Speed in this context depends on the number of sensors whose output is used for inference, i.e. the model size. During an attack, we must wait for the attack to progress to observe alerts from the various sensors. This latency represents a loss of detection speed and response agility; however this latency, on the other hand, also enables an increase in PPV inference confidence due to the relationship between posterior uncertainty and model size.

We argue that trading speed for confidence can make sense in situations where the reduced timeliness of detection is still adequate to prevent the attacker's goal. To illustrate this point, consider the 5-step attack represented in Figure 10. Responses to alerts generated by exploit  $E_0$  are, by definition, timely because such alerts represent the earliest evidence of attack. However, such alerts might give little confidence that exploit  $E_0$  has actually occurred due to factors contributing to low PPV and high PPV uncertainty in the single-step case. Given that the attacker's goal cannot be realized until exploit  $E_4$  has been accomplished, it is reasonable to consider decision latencies  $l_1$ ,  $l_2$ , and  $l_3$  as shown in Figure 10. Delaying a decision until one of these latencies has elapsed still leaves a non-zero interval for response prior to the execution of the final exploit and achievement of the attacker's goal. Whether such latencies should in fact be accepted of course depends upon the timing characteristics and risk/exposure of a particular situation. In some cases, the risk associated with increased delays may simply be unacceptable.

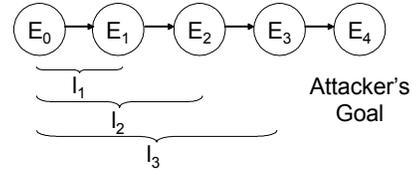


Figure 10. 5-step attack

The optimal choice of delay depends on PPV and associated uncertainty for each choice of delay. For a particular delay one or both of these quantities may have a value that is not useful, necessitating increased delay for an optimal decision. For example, Figure 11 shows PPV and PPV uncertainty versus step count for  $FP=0.01\pm 0.001$  and  $TP=0.09\pm 0.05$ . Note that in this case, it is only at 4 steps that the desirable state of high PPV and low PPV uncertainty is achieved. The right choice of delay for a given situation can in principle be determined by an objective function incorporating costs and expected utility of latency, an analysis that is outside the scope of this paper.

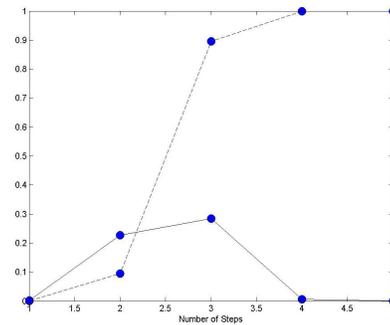


Figure 11. PPV (solid) and PPV uncertainty (dashed) versus model size for  $FP=0.01\pm 0.001$ ,  $TP=0.09\pm 0.05$

## 7. RELATED WORK

Above we discussed implications of low base rates. This problem clearly affects Bayesian inference as discussed above but more generally we expect low base rates to represent a fundamental problem for many intrusion detection systems. The NIDES statistical component [3] is an example of intrusion detection based on detecting departures from a normal profile of user system interaction. The normal profile is based on audit data and is updated daily with more recent audit records being weighted more significantly than older records. A test statistic for each user is computed from a variety of measures and a set of thresholds define a set of alert levels, with each alert level having a corresponding false alarm rate. The design of NIDES is such that alerts are raised when recent behavior is dissimilar to long-term behavior. In

principle, low base rate phenomena can present an issue for any statistical approach, such as NIDES, because such phenomena will not be adequately represented in recent behavior and possibly long-term behavior.

In general, any parametric approach must address the issue of parameter uncertainty. Parameter estimation is briefly discussed in [10]. These authors note that statistical methods can be used to estimate FP and TP of a population based on values measured in a sample. Specifically, the Z-interval for a population proportion is identified proposed for estimation of FP and TP. The Z-interval is given by [14]:

$$E = z \sqrt{\frac{\pi(1-\pi)}{n}}$$

where E is the estimation error, z is a constant based on a desired confidence interval, n is the sample size and  $\pi$  is the sample proportion. Using such an approach, one can simply pick a sufficiently large sample size to achieve a particular estimation error. Such an approach of course requires the test dataset to be very similar to the operational environment. Because real-world attack data is dynamic and difficult to obtain, this approach seems difficult to employ. It is this particular challenge that has motivated the present work: even in the face of large parameter estimation errors, high confidence inferences can be obtained in the multi-step case.

Many probabilistic approaches to intrusion detection have been proposed. A Bayesian approach called eBayes is proposed in [15]. In this approach a Bayesian model is used to obtain belief over a set of TCP session states. Data mining approaches are used in [16] to develop an adaptable and extensible approach to intrusion detection. Association rules and frequent episodes algorithms were adapted to the analysis of audit data. Test results against the 1998 DARPA dataset indicated detection performance as good as that obtained with manual knowledge engineering based approaches. A statistical approach to anomalous system call detection is presented in [17]. This approach evaluates system call arguments using multiple statistical tests. Each test is associated with a model and the set of models is incorporated into a Bayesian network which is used to obtain an overall normal or malicious classification. In [18], simple statistical methods are shown to be effective for detecting intrusions in the 1998 DARPA evaluation dataset. Probes and denial of service attacks were effectively detected using a simple TCP connection volume threshold. A Kolmogorov-Smirnov statistic based on sender/received byte counts was shown to differ significantly between attack and normal telnet data.

All of the approaches discussed above, while different in terms of the observables processed and the methodologies used, are single-step approaches. As such, they all potentially suffer from high inference uncertainty, depending upon the level of uncertainty achievable in their parameter estimates. This work complements these prior works, and many like them, not by proposing a new IDS approach, but rather by taking a first step toward a framework that can be used to tie together multiple individual approaches to accomplish multi-step inference.

## 8. LIMITATIONS

There are several limitations to our work. First, our linear model itself is limited. Clearly, there may be multi-step attacks which follow different model topologies. Also, we make restrictive assumptions regarding model parameters. Specifically, the analysis in section 5.1 focused on the homogeneous case where all sensors had identical characteristics. Also, we assume a constant, zero-uncertainty base rate. We expect that real-world situations include sensors with very different characteristics as well as an uncertain base rate. Additionally, our model assumes conditional independence between different exploits and their corresponding alerts, i.e. exploit<sub>i</sub> can only produce alert<sub>j</sub>, not alert<sub>i</sub>. In general, this may not be the case since similarities between different exploits may lead to the same alert. Also, as noted above, we only analyzed one of the  $2^N$  possible alert sets that can be generated by an N-step model. Although we have examined the most likely set, we note that a comprehensive analysis should cover all possible sets.

Our analysis simplistically assumes alerts can be reliably associated with the appropriate attack stage. Such knowledge must be first obtained via other means, such as alert correlation methods [19]. Such methods are limited when alerts are missed but missed attacks can be addressed through the fusion of complementary alert correlation methods [20]. To further address this limitation, attack graph analysis techniques [12, 21, 22] can be used which can result in a comprehensive set of multi-step attacks through vulnerability analysis. These attack specifications define the sequence of steps belonging to the same multi-step attack. Such specifications can be used together with alert correlation tools to avoid mistakes in correlating alerts and identifying missing alerts.

## 9. CONCLUSIONS AND FUTURE WORK

We have shown that parameter uncertainty in the single-step case renders low uncertainty inference infeasible. We have shown that high PPV and low PPV uncertainty can be achieved in the multi-step case. Thus the inference problems inherent in the single-step case can be mitigated in the multi-step case. This necessarily comes at a price. High PPV confidence can only be achieved at the cost of latency which results from the need to wait for observation of alerts associated with exploits further downstream in the attack chain. Fundamentally this relationship suggests a tradeoff for security administrators. Speed in detection can be traded for accuracy in inference. This tradeoff in our opinion can be warranted in situations where confidence in alerts is sufficiently low that alerts are left ignored. Essentially we assert that a high-confidence decision that is delayed can be preferable to a timely low-confidence decision.

This work represents a first step toward understanding inference uncertainty in multi-step attacks. While this work does give insight into the basic problem, much work remains. We plan to survey multi-step attacks to understand the space of relevant model topologies and to analyze uncertainty in higher dimensions to remove our restrictive parameter assumptions. Also, we plan to extend the analysis to alert sequences not examined here for a more comprehensive treatment. Additionally, further work is required to model the costs of latency to enable an optimized choice of detection latency.

## 10. ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for their valuable comments. This work was supported in part by NSF CNS-0716479, AFOSR MURI: Autonomic Recovery of Enterprise-wide Systems after Attack or Failure with Forward Correction, and AFRL award FA8750-08-C-0137.

## 11. REFERENCES

- [1] D. E. Denning, "An Intrusion Detection Model," *IEEE Transactions on Software Engineering*, vol. SE-13, pp. 222-232, 1987.
- [2] H. S. Javitz and A. Valdes, "The SRI IDES Statistical Anomaly Detector," presented at IEEE Computer Society Symposium on Research in Security and Privacy, 1991.

- [3] H. S. Javitz and A. Valdes, "The NIDES Statistical Component: Description and Justification," SRI International A010, 1994.
- [4] D. Wagner and R. Dean, "Intrusion Detection via Static Analysis," presented at IEEE Symposium on Security and Privacy, 2001.
- [5] R. Sekar, A. Gupta, J. Frullo, T. Shanbhag, A. Tiwari, H. Yang, and S. Zhou, "Specification-Based Anomaly Detection: a New Approach for Detecting Network Intrusions " presented at 9th ACM Conference on Computer and Communications Security (CCS'02), 2002.
- [6] S. Axelsson, "The Base-Rate Fallacy and the Difficulty of Intrusion Detection " *ACM Transactions on Information and System Security (TISSEC)*, vol. 3, pp. 186-205, 2000.
- [7] R. P. Lippmann, D. J. Fried, I. Graf, J. W. Haines, K. R. Kendall, D. McClung, D. Weber, S. E. Webster, D. Wyszogrod, R. K. Cunningham, and M. A. Zissman, "Evaluating Intrusion Detection Systems: the 1998 DARPA Off-line Intrusion Detection Evaluation," presented at DARPA Information Survivability Conference and Exposition (DISCEX '00), 2000.
- [8] J. E. Gaffney, Jr. and J. W. Ulvila, "Evaluation of Intrusion Detectors: a Decision Theory Approach," presented at 2001 IEEE Symposium on Security and Privacy, 2001.
- [9] J. P. Hansen, K. M. C. Tan, and R. A. Maxion, "Anomaly Detector Performance Evaluation Using a Parameterized Environment " in *Recent Advances in Intrusion Detection*, vol. 4219/2006, D. Zamboni and C. Kruegel, Eds.: Springer, 2006, pp. 106-126.
- [10] G. Gu, P. Fogla, D. Dagon, W. Lee, and B. Skorić, "Measuring Intrusion Detection Capability: an Information-Theoretic Approach " presented at 2006 ACM Symposium on Information, Computer and Communications Security, 2006.
- [11] P. Helman and G. Liepins, "Statistical Foundations of Audit Trail Analysis for the Detection of Computer Misuse," *IEEE Transactions on Software Engineering*, vol. 19, pp. 886-901, 1993.