

# Detection for Security

Thomas E. Daniels  
Iowa State University

November 5, 2007

## Abstract

Abstract goes here.

## 1 Detection and Security

## 2 Motivation for Model

## 3 The Model

### 3.1 Truth and Detection

Let  $E$  be the set of all unique events that can be input to a detection mechanism.

We define on those events a function  $\forall e \in E, T(e) \rightarrow \{0, 1\}$  that defines our ground truth. Without loss of generality, we will treat 1 values of  $T()$  as our positive events, i.e. the ones we wish to detect.

To detect events, we define a detection function,  $D(e, P) \rightarrow \{0, 1\}$  where  $e \in E$  and  $P$  is a set of parameters suitable to  $D$ . This models some specific detection process that operates on events and the process's configuration. In our use, we will often drop  $P$  from the notation when it is clear that we are using the same parameter settings throughout the context.

The notation now becomes expository useful as it is easy to use the two functions and basic boolean algebra to define the classical error types and begin to talk about rates and the base rate fallacy.

This rather handily transfers into convenience functions:

$$FP(e) = D(e) \wedge \overline{T(e)}$$

Table 1: Definition of Errors using  $T()$  and  $D()$

	$\overline{D(e)}$	$D(e)$
$\overline{T(e)}$	True Negative	False Positive
$T(e)$	False Negative	True Positive

$$FN(e) = \overline{D(e)} \wedge T(e)$$

It also is quite convenient to use the notation where the same function can be applied to a set to represent a subset. For instance,

$$FP[E] = \{e : e \in E \wedge FP(e)\}$$

Problem: Write the function for True Positives and True Negatives.

### 3.2 Error Probabilities

If we now take  $F \subseteq E$ , as some subset of the events we wish to consider, it makes sense to define the probability of error in that set.

$$P(D(e) \neq T(e)) = \frac{|FP[F] \cup FN[F]|}{|F|}$$

Unfortunately, this metric doesn't tell us much about what kind of errors are occurring. Sometimes, FP's and FN's are much more likely than the other and may have radically different cost implications. Hence, in  $F$  it makes sense to define error probabilities for each.

The obvious, but less useful versions are the *overall rates*,  $FPR_1$  and  $FNR_1$  where  $FPR_1 + FNR_1 = P(D(e) \neq T(e))$ .

$$FPR_1 = \frac{|FP[F]|}{|F|} \quad FNR_1 = \frac{|FN[F]|}{|F|}$$

Both of the above overall rates correspond to the probability that such an error will occur events drawn randomly from  $F$  and their sum is the overall error raet because an event can not be both a FP and a FN. The problem with these rates is that they are dependent on the number of positives and negative cases in  $F$ . If  $F$  does not represent the environment accurately, then these rates will not apply

*The specific error rates* shown below are much more useful because they do not depend so much on the model of the environment. In measuring these rates, it is very important to have a large denominator and to make the events in  $F$  as representative of the typical environment as possible.

$$FPR_2 = \frac{|FP[F]|}{|T[F]|} \quad FNR_2 = \frac{|FN[F]|}{|T[F]|}$$

These are fine formula, but consider that the frequency of occurrence of an event we want to detect, e.g. a virus in a program or an intrusion, may not be equal. In other words, there is some Base Rate of events that occur in the environment. If  $F$  is a truly representative sample from the environment and is large enough, then we can define a the probability that an event is a true event

Figure 1:

for that environment. Further, we can define a time period  $t$  where  $F$  is the set of events observed during the period. We then have BR defined as:

$$BR(F) = \frac{|T[F]|}{|F|} \quad BR_t(F) = BR(F) \frac{|F|}{t} = \frac{|T[F]|}{t} \text{ events/sec}$$

Note that the Base Rate is a measure of the environment and not the detection mechanism. We can use this to understand the effects of the environment on the user's workload.

First, what is the total number of Alarms that a detector  $D$  produces from a set of events,  $F$ , representing the environments behavior:

$$D(F) = FPR_2 * (1 - BR[F]) + (1 - FNR_2) * BR[F]$$

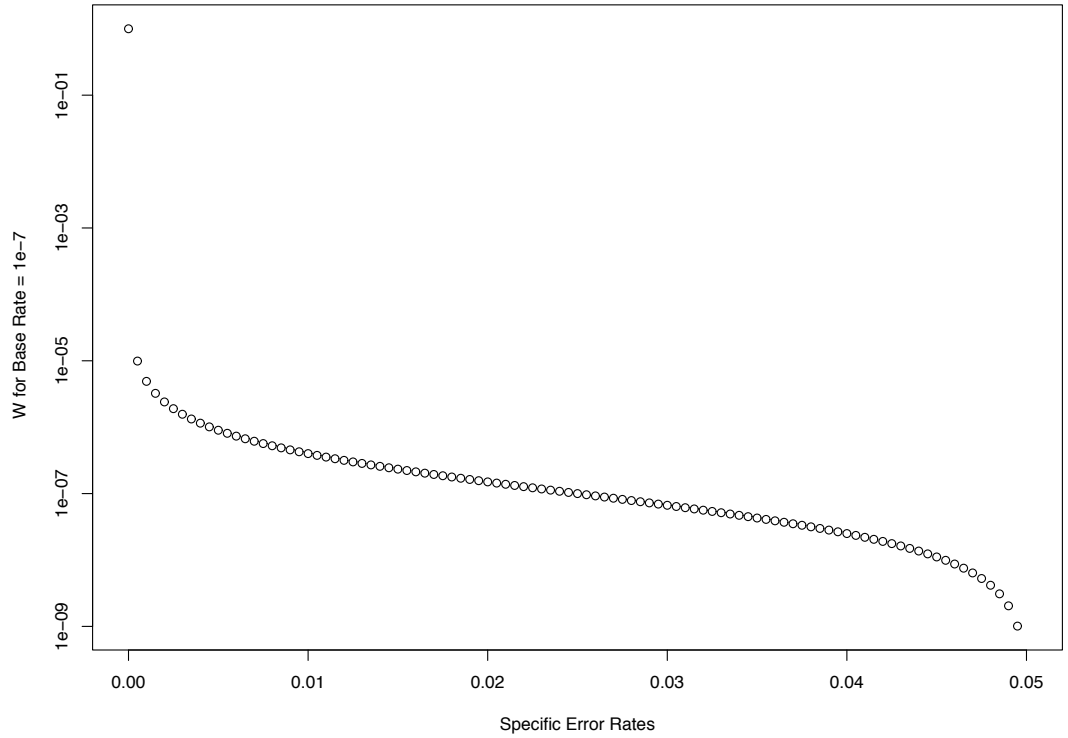
Now, what proportion of alarms are real:

$$W(D, F) = \frac{(1 - FNR_2) * BR(F)}{FPR_2 * (1 - BR(F)) + (1 - FNR_2) * BR[F]}$$

It is tempting to further simplify this, but not wise. *The error rates  $FPR_2$  and  $FNR_2$  are properties of  $D$  while the Base Rate is a property of the environment! Also, note that  $BR_t(F)$  could be used in place of  $BR(F)$  in  $W(D, F)$  as the  $t$ 's will cancel out.*  $W$  is important because it defines how effective the work load of someone using the detection system can be.

Consider the case where the specific error rates are equal,  $FNR_2 = FPR_2 = x$ , and the base rate is  $k$  what happens to  $W(D, F)$ ? It is shown in the below figure.

$$W(D, F) = \frac{(1 - x)k}{x(1 - k) + (1 - x)k}$$



Finally, let's consider a desirable value for the workload effectiveness, say  $W() = 1/2$  and some reasonable base rate,  $k = 10^{-7}$ . In other words, one in 10 million events need to be detected and half of your work is still devoted to chasing false positives. What must be true of the error rates?

$$W(D, F) = \frac{(1 - FNR_2) * BR(F)}{FPR_2 * (1 - BR(F)) + (1 - FNR_2) * BR[F]}$$

$$\frac{k(1 - FNR_2) - Wk(1 - FNR_2)}{W(1 - k)} = FPR_2$$

Simplifying to:

$$\frac{(1 - FNR_2)k(1 - W)}{W(1 - k)} = FPR_2$$

So, if we choose  $W=1/2$ , the workload cancels out and  $FPR_2 = \frac{(1-FNR_2)k}{1-k}$ , and as  $k$  is near zero,  $FPR_2 \approx (1-FNR_2)k$ . So the false positive rate should be approximately the True Detection Rate times the Base Rate! Hence, to achieve  $W = .5$ , and detect half of the attacks, the FPR must be 1/2 of the Base Rate!!!

### 3.3 The Base Rate Fallacy

Why does the base rate play such an important role in security? In much of the historical detection work unrelated to security, it is common to find problems where the Base Rate is about 0.5. For instance, in communications, it is common for certain types of errors to be equally likely. In these cases, the error types are not that important to distinguish. Historically, it was thought that a FPR of 99% was a good result, but clearly this isn't true if the base rate is very low.

#### 3.3.1 Problems:

1. What is a typical Base Rate for detecting security issues and Why
2. Assume you have time to handle a limited number of events per day. How does  $W$  determine your effectiveness?

### 3.4 ROC Curves

The base rate is important in that it tells you something about the environment and how it affects your work, but there is a fundamental tradeoff in error rates that