Risk Mitigation Metric for
Multi-Factor Authentication Systems

BITSystems Commercial Contract No: BITS-13287

Milestone 1 White Paper

ProvenSecure Solutions, LLC

July 19, 2013

Bayes, Maximum Entropy and MrE overview:

There are many ad hoc methodologies that are created or fashioned that 'work' for one problem or another. However, because they are hand crafted for a particular problem, they have little hope of being used for other problems. Worse, is that there is no clear metric for how well they 'work'.

Most current machine learning methodologies that are well understood and have a sound foundation follow one of two trajectories: 1) A Bayesian track such as Bayesian filtering, Bayesian Networks, Hidden Markov Models, Particle filtering and to some degree, Kalman filtering. 2) A Maximum Entropy track such as used largely in image analysis, natural language processing and complex systems. Other methods that are commonly used, such as decision tree learning, neural networks, support vector machines, cluster analysis, etc can all be shown as special cases of one of the above tracks given certain assumptions.

While they both have their merits, recent development of a third methodology, Maximum relative Entropy (MrE) includes not just MaxEnt but also Bayes rule as special case. MrE is capable of reproducing *every aspect* of orthodox Bayesian inference methods and thus opens the door to tackling problems that could not be addressed by either the MaxEnt or orthodox Bayesian methods individually.

Bayesian Methods:

Bayesian methods take advantage of data. When there is very little data, the Bayesian methods uses a priori or "prior" information to their advantage. The proverbial, "Where should one start looking?" is utilized. This is the great advantage of Bayesian methods as compared to the special cases of more traditional statistical methods that must assume uniform or constant priors. This is an advantage if developers could add any data known a priori. The other advantage of Bayesian methods are that they provide clear uncertainties by examining the posteriors. Bayesian methods are well proven to yield excellent results. However, the problem with them that concerns this project is the difficulty of including certain common, so called "coarse grained" information such as an average or a moment. This type of information is important when dealing with high dimensional systems. When this information is known it is extremely time consuming to include it in a Bayesian update. Therefore, this information is not commonly used in Bayesian methods.

Maximum entropy:

Maximum entropy methods (MaxEnt) do just the opposite: They take advantage of constraints such as expectation values (averages) and moments. Because of this, data can be summarized in this form allowing it to be very fast with computation time scaling like log(N), where "N" represents the size of the problem as compared to a pure Bayesian method. Thus, maximum entropy methods are equipped to handle large amounts of information from very complex systems because it takes a very coarse grained, big picture view. The coarser the grain, the faster it can compute a solution. However, this has one obvious drawback: One is always limited to that big picture. We can look at things globally, but not locally. The resolution is inversely proportional to the amount of information needing to be processed within a fixed amount of resources (e.g. time, computing power, etc.) However, this is not

the only limitation. The other limitation is that maximum entropy methods select a particular distribution to use. There is no uncertainty as to model selection, unlike a Bayesian method.

Maximum relative Entropy:

The solution is of course, is to use both, drawing on the strengths of each simultaneously. This can be achieved using the method of Maximum relative Entropy (MrE). This new method, MrE, is capable of reproducing every aspect of orthodox Bayesian inference and proves the complete compatibility of Bayesian and MaxEnt methods. Even more significantly, MrE allows for a dynamic and continuous combination of the two approaches, to enhance inference and learning.

As mentioned, this methodology has already been derived and demonstrated in toy problems. It has been demonstrated in problems where one has very little data (the method converges on a solution faster than Bayes or MaxEnt alone). The impact of this alone should be clear; having all capabilities under one general methodology that can handle any kind of information is the most desirable.

Application to Authentication:

Traditional authentication methods usually calculate the “True Acceptance Rate” and the “False Acceptance Rate”. These “rates” are point estimates that do not have any uncertainty associated with them. Further, multiple factors are typically fused together in an ad hoc manor. To be consistent, as well as to establish uncertainties for our estimates, we use probability theory that will update our estimates and uncertainties as new information presents itself. This is typically called “learning’. However, currently, there is no off the shelf way to calculate the distribution when the information is presented in both constraints and data.

In order to conduct our analysis we break the problem into three pieces: 1) The “bench” which will be the point of access (interface) as well as contain the user information, specifications and potential prior information (such as current state of the system, recent attacks, etc). This is also where the decision process happens as it is user defined. 2) The RMM which is the intelligence portion. This is the portion that will take what the bench gives it in terms of inputs and other related information and fuse it all together. As opposed to simply point estimates, using MrE will allow us to produce distributions that not only incorporate the uncertainties of the authentication devices, but also return an uncertainty regarding its fused answer. This is very important since without this, a point estimate is not very informative. Further, it will compare competing classifications (such as genuine vs. imposter) using Bayes Factors (BF). The importance of this is that we do not accept or reject one model (class), but compare it to others to make a decision. After preliminary review, it was shown using a Receiver Operating Characteristic (ROC) curve (see Figure 1 below) that using BF will always perform *at least* as well as the traditional combining of factors such as used in a voting algorithm and in many cases exceed traditional method results. This will then be passed back to the bench for analysis. In addition, by using BF not only can we fuse the data but we can *suggest* to the bench what factor should be asked next. 3) The MrE engine. This is the labor portion. This portion will be called by the RMM to calculate the probabilities given the various user inputs. There is not currently an easy way to do this

mathematically. However, the MrE engine will return the calculated probabilities to be used with the user scores to determine the BF.

Beyond this particular project, the completion of the MrE engine will have further great practical use in the future as it can be utilized to make other inferences such as determining authentication device dependence and other, higher order, hierarchical decision processes.

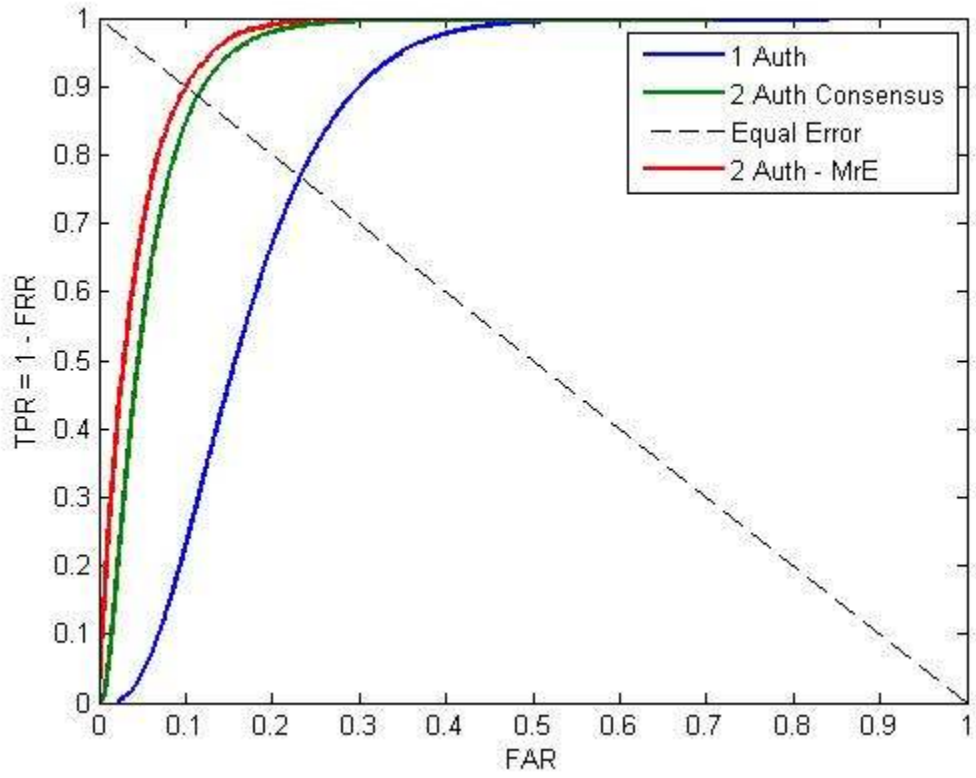

Figure 1: The blue is the curve for one factor. The red and green are using two factors, with the red curve being the MrE solution.

References:

Giffin and A. Caticha, "Updating Probabilities with Data and Moments", AIP Conf. Proc. 954, 74 (2007)

Biometrics Metrics Report v3.0 Prepared for: U.S. Military Academy (USMA) – West Point http://www.usma.edu/ietd/docs/BiometricsMetricsReport.pdf (contains a detail discussion and definitions of the terminology used in this report.)