Quantitative Analysis of Error Injection Covert Channels.

Robert W Harrison[1] and William L. Harrison[2]

[1] Department of Computer Science
Georgia State University
Atlanta GA 30303
[2] Department of Computer Science
University of Missouri, Columbia
Columbia MO 65209

Abstract. Covert channels are mechanisms for communication where a legitimate channel carries a hidden message, and where the hidden message is conveyed using legal operations of the legitimate channel. Quantitative modeling of the response of the covert channel to noise results in an enhanced understanding of the channel and delimits the range of conditions under which a covert channel can operate effectively. The analysis of two covert channels using the confusion matrix and noise models shows that the techniques presented in this paper are widely applicable to covert information flows in noisy channels.

1 Introduction

Quantifying and detecting covert information flow is critical to assuring the security of a computer system. Establishing quantitative bounds on covert channel behavior or establishing quantitative estimates of the side effects of a covert channel is a critical part of analyzing covert information flow. Ideal covert channels are statistically independent from the associated legitimate channel [5, 4], but real channels may not be independent. This paper explores techniques for quantitative analysis of the response of a legitimate/covert channel pair to noise in the channel for both ideal and non-ideal covert channels. The quantitative analysis can be used to either optimize the design of the channel by the attacker to maximize his covert capacity or for optimizing the ability of the legitimate user to minimize covert information flow.

Three model covert channels are examined, an error-injection covert channel [14], a time-delay covert channel [7, 12, 2, 11] and the combination of an error-correction code with a time-delay channel. The error-injection channel has a high capacity relative to the underlying legitimate channel. If the legitimate channel has a capacity of β and a code-word length of c then the injection channel transmits at least one bit per code-word giving a capacity of $\frac{\beta}{c}$. For a 10MB/s channel and a 10-bit code, this would result in a 1MB/s covert channel. The timing delay channel can send one symbol per data-packet. Since the data-packets are

typically larger than error-correcting codewords, this is a lower capacity channel, but still able to transmit significant amounts of information. For example a common packet size on TCP/IP is 1500 bytes so the timing channel would have a capacity of about $\frac{1}{1500}$ times the capacity of the TCP/IP network. A 10MB/s TCP/IP channel could support a 10kB/s time delay covert channel.

The communication channel is modeled as a set of input symbols I, a representation of the possible confusion between symbols introduced by a channel with errors C, and a set of output symbols O that are decoded by the recipient, as in [3, 13]. In an ideal noiseless channel the confusion matrix is the identity matrix, but real channels with errors, will have a diagonally dominant matrix which describes how symbols can be exchanged or confused in the presence of errors. In simple channel models the sets I and O are identical, but in more realistic channel models the set O can include can include symbols representing uninterpretable errors or lost symbols. The simplest model for the confusion is to assume that the errors rates are low and the confusion only happens once. In other words, if the symbol I_1 is converted with a small probability to O_2 in error then the probability of O_2 being converted to some other symbol is negligible[3]. Under these conditions, which are equivalent to a linear model, the confusion is represented by a matrix of conditional probabilities. O = C I or

$$\begin{pmatrix} P(O_1) \\ \vdots \\ P(O_n) \end{pmatrix} = \begin{pmatrix} P(O_1|I_1) & \dots & P(O_1|PI_n) \\ \vdots & \ddots & \vdots \\ P(O_n|I_1) & \dots & P(O_n|PI_n) \end{pmatrix} \begin{pmatrix} P(I_1) \\ \vdots \\ P(I_n) \end{pmatrix}$$

Clearly the structure and properties of C, the matrix of conditional probabilities, channel, or confusion matrix, determine the behavior of the model. Since the probabilities are normalized the sum of each column of C is one. While C does not have to be a square matrix, it can always be represented as a square matrix by introducing dummy or virtual symbols to either O or I to handle destruction/creation of symbols. The probability of a given input symbol I_i being output as the correct output symbol O_i is given by the diagonal element of C or $C_{i,i}$. Therefore $\frac{1}{n}\sum_{i=1}^{n} C_{i,i}$ or the average value of the trace of the matrix gives the average probability of correct transmission of a symbol[4]. (Since the trace is the sum of the eigenvalues of the matrix, and any arbitrary input can be described as a sum of eigenvectors, this result can be put on a more formal basis by averaging over eigenvectors).

[3] If multiple steps are important then $O = (C^n + ... + C)I$, where n would be sufficiently high power to represent multiple transitions. Treating these general cases would complicate the analysis, but not change the results because the sum of all the powers of C would still be a matrix just not the relatively simple one to estimate single transition one that will be used in this work.

[4] For simplicity, this assumes a uniform distribution of input symbols. When the distribution is non-uniform this average would be replaced by the expected value over the distribution of inputs.

In the presence of a covert channel the sets of input and output symbols are expanded to include all possible pairs of symbols. Given the legitimate set I and covert set J then the Cartesian product of I and J ($I \otimes J$) produces the set of symbols for the combined channel. Similarly for the legitimate output set O and covert set Q, the set of possible output is defined by ($O \otimes Q$). Since statistical independence implies that probabilities of events multiply, in other words $P(O_i Q_j | I_k J_l) = P(O_i|I_k)P(Q_j|J_l)$, the estimate for C becomes the Kronecker product of the individual C matrices for legitimate channel and covert channels on their own.

Kronecker matrix products are highly useful models of complicated multi-state systems. Since analytic formula exist for the trace, determinant, and even eigenvector/eigenvalues of Kronecker products in terms of the values for the underlying matrices (for a review see [10, 8]), it is often possible to find analytical approximations for complicated systems that are composed of well defined components - even when the overall system is quite large. In particular,the trace of the Kronecker product is the product of the traces of the two matrices. A Markov chain of N steps, where each step has a corresponding probability or confusion matrix, can be modeled by the Kronecker product of the N individual confusion matrices. The algebraic properties of the Kronecker matrix allow the determination of some properties of this large matrix without explicitly formulating it. If the assumption that the steps are statistically independent is well-founded then the Kronecker model can be quite accurate. This paper shows that with the error-injection covert channel, that the Kronecker model gets the order of dependence of the channel error rate on the underlying bit-rate error correct, while being off by a small constant. When corrected for the lack of statistical independence, with a very small correction (two elements of a matrix become non-zero), the model completely agrees with simulation.

Explicitly deriving expressions for C for the isolated legitimate channel, the isolated covert channel, the covert/legitimate combined channel and comparing them to the Kronecker product of the isolated covert and legitimate channels generates an expression that indicates the degree of dependence of the combined covert/legitimate channel and hence the ability of an adversary to detect it. Similarly, if there are parameters of the covert channel than can be adjusted and effect the error rates, adjusting those parameters to minimize the difference between the explicit expression for C and the Kronecker product will define the conditions or limits on the covert channel where it is possible to be nearly undetectable.

2 Error injection as a covert channel

In an error injection covert channel deliberate errors are made in a communications channel, and the error correcting code is used to detect and correct those errors. The presence/absence of an error is used to send a covert message. Error injection to compromise secure systems by inducing unusual fault behavior is an active area of research [1], however in this paper error injection is simply used

as a mechanism for sending a signal. Developing error-correcting codes that are more resistant to error injection, in order to minimize fault injection in a secure system [15], enhances this channel by reducing the dependence of accuracy for the legitimate channel on the introduced errors.

2.1 3-bit Hamming code - a simple example

A 3-bit Hamming code (majority of three bit) shows how this analysis proceeds. While not particularly sophisticated, this code can correct for any single bit error, and codes like this have been used in error correcting memory. Assuming a bit-rate error of α then the rate of errors with the three bit code is $\frac{1}{2}(3\alpha \times 2\alpha)$ or $3\alpha^2$. Therefore the matrix of conditional probabilities is:

$$\begin{pmatrix} P(O_0) \\ P(O_1) \end{pmatrix} = \begin{pmatrix} 1-3\alpha^2 & 3\alpha^2 \\ 3\alpha^2 & 1-3\alpha^2 \end{pmatrix} \begin{pmatrix} P(I_0) \\ P(I_1) \end{pmatrix} \approx \begin{pmatrix} 1\ 0 \\ 0\ 1 \end{pmatrix} \begin{pmatrix} P(I_0) \\ P(I_1) \end{pmatrix} \tag{1}$$

when the limit of single bit errors is used. The covert channel consists of introducing a single bit error into the message when the value of 1 is to be transmitted and leaving the message intact otherwise. The covert receiver notes whether an error was corrected, before correcting the error and passing on the legitimate message. Since there are 3 single bit errors that could happen with an unmodified message to change the sense of the covert message and one single bit error that could occur to change the sense of a modified message the covert channel follows:

$$\begin{pmatrix} P(O_0) \\ P(O_1) \end{pmatrix} = \begin{pmatrix} 1-3\alpha & \alpha \\ 3\alpha & 1-\alpha \end{pmatrix} \begin{pmatrix} P(I_0) \\ P(I_1) \end{pmatrix}$$

with an average value of the trace of $1-2\alpha$. The Kronecker product of the two isolated channel matrices which corresponds to the independent channel model is:

$$\begin{pmatrix} 1-3\alpha & \alpha & 0 & 0 \\ 3\alpha & 1-\alpha & 0 & 0 \\ 0 & 0 & 1-3\alpha & \alpha \\ 0 & 0 & 3\alpha & 1-\alpha \end{pmatrix}$$

here the average value of the trace is $1-2\alpha$. This value for the average of the trace corresponds to errors in the covert channel alone. The total errors (legitimate and covert) channel will be bounded from below by the errors in the covert channel alone.

Explicit calculation of the confusion matrix gives:

$$\begin{pmatrix} 1-3\alpha & \alpha & 0 & 0 \\ 3\alpha & 1-3\alpha & 0 & 2\alpha \\ 0 & 0 & 1-3\alpha & \alpha \\ 0 & 2\alpha & 3\alpha & 1-3\alpha \end{pmatrix}$$

where the additional 2α corresponds to the two ways that a single bit error in the combined symbols $(legitimate_0, covert_1)$ or$(legitimate_1, covert_1)$ could be converted to two-bit error and change the sense of the legitimate channel by taking

the combinations to $(legitimate_1, covert_1)$ or$(legitimate_0, covert_1)$ respectively. The average value of the trace is now $1 - 3\alpha$ which reflects the increased error probability.

Thus this simple error correcting code covert channel is not fully independent and could be detected by an increase in the rate of errors in the legitimate signal. However it would require additional analysis of the channel output to show that there were errors in the legitimate channel as the symbols themselves would be 'correctly' decoded to incorrect values.

Results from Simulation The Hamming majority of three channel was simulated with 10^8 samples and the results are compared with theoretical estimates in Fig. 1. Fig. 1 contains a log-log plot showing the error rates with covert and legitimate channel. The dotted lines show the theoretical values for the combined channel error 3α and the legitimate channel alone $3\alpha^2$. The square boxes (labeled "decode") correspond to the error correcting code without a covert channel and the other symbols show the error rates in the covert part of the combined channel, the legitimate part of the combined channel, and the total error in the combined channel. Random numbers were generated using the Mitchel-Moore (54,23) irreducible polynomial shift register generator [9]. Results for the legitimate and covert-legitimate channel were kept during the simulation. The effect of the covert channel on the error rate in the legitimate channel is clearly demonstrated by the simulation. The agreement between theoretical analysis and simulation results is pronounced.

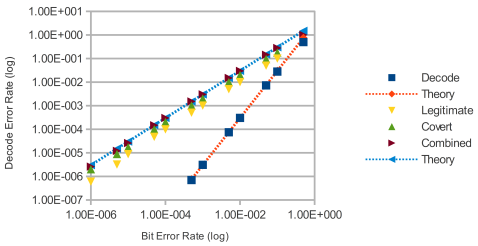

Fig. 1.

2.2 5-bit Linear code - a slightly less simple example

The three-bit Hamming code is a bit simplistic as modern error-correcting codes can detect and correct larger errors. Typical codes can correct one set of errors while detecting a larger set of more severe errors that they cannot correct. A simple example of this is a 5,2 linear code[16] that can correct any one bit error and detect most two bit errors. It is worth examining because it is both small enough to enumerate and possesses the error symbol for two bit error detection. The code:

$$A^t I = \begin{pmatrix} 1\,1\,1\,1\,0 \\ 1\,0\,1\,0\,1 \end{pmatrix}$$

where A is a 3×2 matrix that defines the parity checks and I a 2×2 identity matrix, Each code word differs from the others by at least 3 bits, thus any 1

bit error is uniquely decoded while some 2 bit errors can be detected, but not uniquely decoded. Since a two-bit error is detected the symbol E is included in addition to 0,1,2,3 giving an alphabet of 0,1,2,3,E. When enumerated, the 32 possible five bit strings correspond to 4 exact matches, 20 one-bit errors and 8 two-bit unresolvable errors. Explicitly evaluating the confusion matrix for this code gives (for $x = 1 - 30\alpha^3 - 10\alpha^2$):

$$\begin{pmatrix} P(O_0) \\ P(O_1) \\ P(O_2) \\ P(O_3) \\ P(E) \end{pmatrix} = \begin{pmatrix} x & 10\alpha^3 & 10\alpha^3 & 10\alpha^3 \\ 10\alpha^3 & x & 10\alpha^3 & 10\alpha^3 \\ 10\alpha^3 & 10\alpha^3 & x & 10\alpha^3 \\ 10\alpha^3 & 10\alpha^3 & 10\alpha^3 & x \\ 10\alpha^2 & 10\alpha^2 & 10\alpha^2 & 10\alpha^2 \end{pmatrix} \begin{pmatrix} P(I_0) \\ P(I_1) \\ P(I_2) \\ P(I_3) \end{pmatrix} \approx \begin{pmatrix} 1\,0\,0\,0 \\ 0\,1\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,0\,1 \\ 0\,0\,0\,0 \end{pmatrix} \begin{pmatrix} P(I_0) \\ P(I_1) \\ P(I_2) \\ P(I_3) \end{pmatrix} \quad (2)$$

after dropping all multiple bit errors and which can be expanded to:

$$\begin{pmatrix} 1\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0 \\ 0\,0\,1\,0\,0 \\ 0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,1 \end{pmatrix} \begin{pmatrix} P(I_0) \\ P(I_1) \\ P(I_2) \\ P(I_3) \\ P(I_E) \end{pmatrix}$$

by assuming a probability of an input error symbol.

The confusion matrix for the isolated covert channel is $\begin{pmatrix} 1-5\alpha & \alpha \\ 5\alpha & 1-\alpha \end{pmatrix}$ where the 5 reflects the 5 ways that a single bit error could be introduced to change the sense of the covert message in the 1 in the other column reflects that there is only a single way to change the sense of the message once an error is introduced.

The Kronecker product of the two matrices is:

$$\begin{pmatrix} 1-5\alpha & \alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0\,0 \\ 5\alpha & 1-\alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0\,0 \\ 0 & 0 & 1-5\alpha & \alpha & 0 & 0 & 0 & 0 & 0\,0 \\ 0 & 0 & 5\alpha & 1-\alpha & 0 & 0 & 0 & 0 & 0\,0 \\ 0 & 0 & 0 & 0 & 1-5\alpha & \alpha & 0 & 0 & 0\,0 \\ 0 & 0 & 0 & 0 & 5\alpha & 1-\alpha & 0 & 0 & 0\,0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1-5\alpha & \alpha & 0\,0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 5\alpha & 1-\alpha & 0\,0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\,0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\,0 \end{pmatrix}$$

which has an average trace (not counting the pseudo symbol for E as an input) of $1 - 3\alpha$.

The explicit confusion matrix is:

$$\begin{pmatrix}
1-5\alpha & \alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0\,0 \\
5\alpha & 1-5\alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0\,0 \\
0 & 0 & 1-5\alpha & \alpha & 0 & 0 & 0 & 0 & 0\,0 \\
0 & 0 & 5\alpha & 1-5\alpha & 0 & 0 & 0 & 0 & 0\,0 \\
0 & 0 & 0 & 0 & 1-5\alpha & \alpha & 0 & 0 & 0\,0 \\
0 & 0 & 0 & 0 & 5\alpha & 1-5\alpha & 0 & 0 & 0\,0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1-5\alpha & \alpha & 0\,0 \\
0 & 0 & 0 & 0 & 0 & 0 & 5\alpha & 1-5\alpha & 0\,0 \\
0 & 4\alpha & 0 & 4\alpha & 0 & 4\alpha & 0 & 4\alpha & 0\,0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\,0
\end{pmatrix}$$

which has an average trace (again not counting the pseudo symbol for E) of $1-5\alpha$. The difference between the Kronecker product and the explicit matrix is due to the four possible single bit errors that convert a covert '$(0,1,2,3,1)$' into an undecidable 'E' symbol.

Again the introduction of deliberate errors into the channel as a covert message increases the error rate for the legitimate channel. Most importantly the rate of generation of the error symbol 'E' is dramatically increased.

2.3 Consequences for the design of covert channel detection

It is not surprising that the introduction of deliberate errors increases the overall error rate of the channel as error-correcting codes work by selecting a limited set of points or code words out of a large space and then selecting the closest code word to the observed signal. The introduction of deliberate errors as a message essentially uses up some of that space by increasing the number of signals to be decoded. The analysis shows one important feature with implications in the defense against using an error-injection covert channel. It is critical to have a symbol in the error-correcting code for uncorrectable errors. An increase in the incidence of this symbol to levels above the design specifications is an immediate indicator of covert channel activity.

3 Timing channels

Timing channels are another form of covert channel that is readily analyzed with the confusion matrix. Timing channels have been extensively studied as models of covert channels [7, 12, 2, 11]. These channels can be thought of as error-injection channels, but where the error is injected into the timing between packets rather than the contents of the packet. Timing channels are interesting for this paper since they do not increase the error rate of the legitimate signals, and so it is only necessary to construct the confusion matrix for the covert channel. We posit that the packets arrive at times given by a+nd where a is an invariant message delay and d is an artificial delay. For simplicity d is much larger than the variation in a. This is an example of a message delay channel, other timing channels such as prepending backoff signals in wireless communication [6], or padding packets

with null content blocks would require a different analysis to account for the errors in the null content or padding and the response of both the legitimate and covert channels to those errors.

The confusion matrix, in the presence of random timing errors, for an $n+1$ level channel can be written as:

$$\begin{pmatrix} 1-\sum_{i=1}^{n} p(id) & 0 & 0 & 0 & \dots \\ p(d) & 1-\sum_{i=1}^{n-1} p(id) & 0 & 0 & \dots \\ p(2d) & p(d) & 1-\sum_{i=1}^{n-2} p(id) & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p(nd) & p((n-1)d) & p((n-2)d) & \dots & 1 \end{pmatrix}$$

where d is the delay associated with a signal and p(nd) reflects the probability of a delay longer than nd. C is asymmetric because the only errors are delays which increase the time between messages. If there were variations in the constant part of the timing channel (e.g. the invariant message delay) that were large with respect to d, then the zeros would be replaced with the probabilities that the message delay was short enough to cause an error. The average trace of this matrix $\frac{1}{n}\left(\sum_{i=1}^{n}(1-\sum_{j=1}^{i} p(jd))\right)$ estimates the rate at which symbols are lost due to timing errors.

Converting the trace to a numerical value requires the definition of a distribution of timing errors. Since the choice of a random time delay jitter is an effective counter-measure [7, 12], the properties of a uniform distribution, Gaussian-distributed time delays, and Poisson-distributed time delays are shown.

With the Poisson distribution, $\frac{t}{l}e^{-\frac{t}{l}}$, where l is the characteristic length and t is the time, is chosen, then $\sum_{j=i}^{n} p(jd) = \int_{id-d/2}^{\infty} dt\frac{t}{l}e^{-\frac{t}{l}} = (1+\frac{id-d/2}{l}e^{(\frac{-(id-d/2)}{l})})$ assuming that intervals are rounded to the nearest integer value. The complimentary error function (erfc) is the cumulative distribution for Gaussian-distributed jitter, and a uniform distribution results in a simple linear form.

Figure 2 shows the error rate as a function of size of the variation over the time step used to encode a message for 1,...,10 signals and linear time jitter. Figures 3 and 4 show the same for Poisson and Gaussian distributions respectively. The results are in qualitative agreement with [2] though they present accuracy rates and we present error rates. In all cases the limit at large error rates is 1-1/n where n is the number of signals, which corresponds to accepting only the long time delay as a valid message. Once the time delay is larger than twice the uniform dither, there is no effect of the dither on signal accuracy. The Gaussian distributed time noise decays reasonably promptly to a usable error rate, but the Poisson distributed time noise requires that the time delay be much larger than the its characteristic parameter in order to achieve usable error rates. It is also worth noting that the number of symbols or time delays used has very

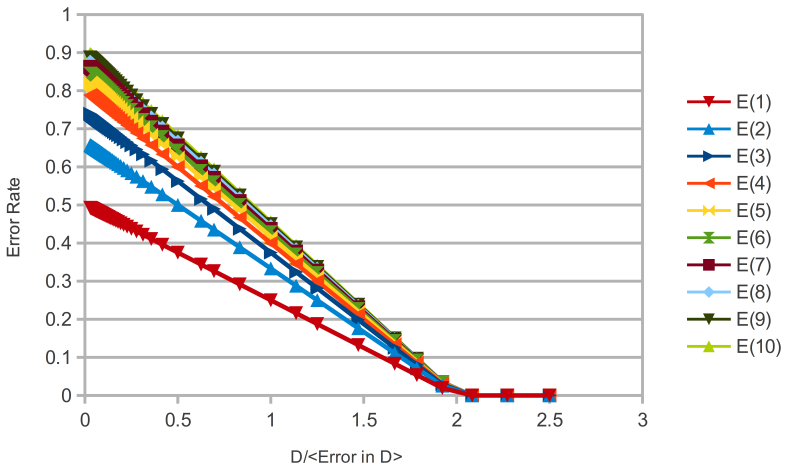

Fig. 2. Error rate of a delay channel with linear or uniformly distributed time jitter. The effects are shown for 2-level through 10-level delay channels. The error rate is shown as a function of the relative size of the time delay used per symbol to the size of the variation in the delay. $\frac{D}{<Derror>} = 1$ implies that the delay is as large as the average time jitter while $\frac{D}{<Derror>} = 0.5$ implies that the average time jitter is twice as big as the delay used as a signal in the channel. It is not surprising with linearly distributed (0,...,1.) jitter that delays larger than a critical threshold result in no errors.

little effect on the rate of improvement in the error rate, especially once the time delay is sufficiently long to result in a usable error rate for the covert channel. Therefore analyzing a single time delay channel is likely to be sufficient.

The results from simulating a multilevel delay channel and normally-distributed noise with 10^8 samples are shown in Fig. 4.

4 Combination of error-correcting codes with delay channels.

The analysis presented in this paper can readily be extended to treat the combination of using a delay channel with an error-correcting code. Equations 1 and 2 give the confusion matrices for the Hamming-3 and linear 5,2 codes in terms of the bit-rate error α. While the two and three-bit errors were ignored for analyzing the injected error channel, they are critical for analyzing the effects of errors in a delay-channel when combined with the error-correction code.

With the Hamming code the average value of the trace of the confusion matrix is $1 - 3\alpha^2$ and with the linear 5,2 code the average value of the trace of the confusion matrix is $1 - 30\alpha^3 - 10\alpha^2$.

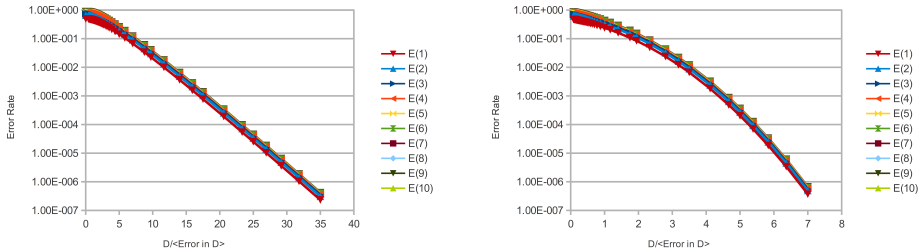

Fig. 3. (Left) Error rate of a delay channel with Poisson distributed time jitter. A log plot is shown to emphasize the slow decay of the error rate with longer time jitter. Note that the number of symbols does not make a large difference in the shape of the curve nor in the error rate. (Right) Error rate of a delay channel with Gaussian or normally distributed time jitter. While the shape of the curve is highly similar to that seen with the Poisson distribution, the spread along the $\frac{D}{<Derror>}$ is much smaller.

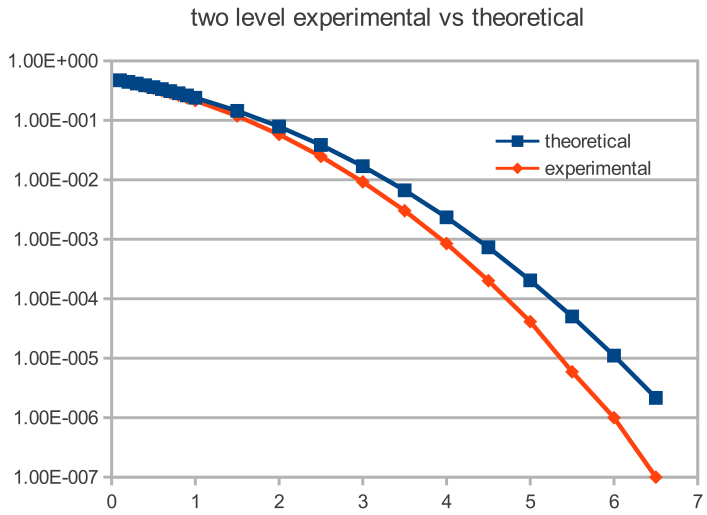

Fig. 4. Comparison of the calculated and simulated error rates of a two-level delay channel with Gaussian or normally distributed time jitter. 10^8 samples were used in single precision so the differences at very low error rates are likely to be numerical artifacts. Normal deviates were generated by the polar method [9]. The polar method tends to underestimate large deviates as one step requires the logarithm of a uniformly distributed pseudo-random number and this step has to be bounded to avoid trying to calculate the logarithm of zero.

The bit-rate error, α, is given by the estimated error for the delay channel, which for a 2-level (n=1) channel and normally distributed noise is simply $\alpha = 1 - \frac{2-erfc(0.5dt)}{2}$ where dt is the time step as a fraction of the standard deviation and 0.5 reflects the level of noise that would cause the signal to be rounded up. Even though this is an extremely simple analysis, the agreement between

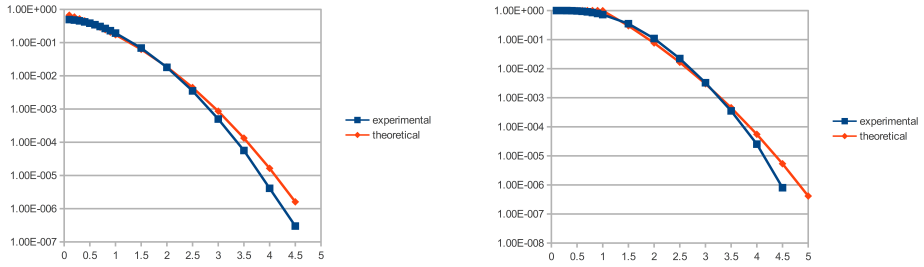

Fig. 5. (Left) Error rate of a delay channel with Gaussian or normally distributed time jitter and using a majority of three Hamming code. Note that the use of an error correcting code significantly reduces the error rates. (Right) Error rate of a delay channel with Gaussian or normally distributed time jitter and the linear 5,2 code.

simulation and theory is striking (Fig. 5). In both cases the error rates are overestimated when there are very large error rates, but show strong agreement elsewhere.

5 Conclusions

This paper explores the use of channel capacity and confusion matrix models for analyzing the behavior of covert channels in the presence of errors. Covert channels constructed both by errors in the symbols and by errors in the channel timing were analyzed. The effects of concatenating error-correcting codes on timing error channels was analyzed as well. The analysis was shown to be consistent with simulation results.

While the analysis is specific to these channel models, although it can be generalized to other channels, it does produce an interesting result for the detection and control of covert communication channels. Put simply, communication errors are your friend. In the absence of errors in the channel, covert and legitimate communications work simultaneously without interference or unexpected results. However, introducing errors produces anomalous effects in the combined covert/legitimate channel that are not present in the legitimate channel alone. The analysis can also be used to select the distribution of errors that has the largest adverse effect on the covert channel while have a small effect on the legitimate channel. Similarly it can be used to quantify the expected behavior

of a system in response to injected errors, thus enabling experimental tests for system infidelities.

References

1. A. Barenghi, L. Breveglieri, I. Koren, and D. Naccache. Fault injection attacks on cryptographic devices: Theory, practice, and countermeasures. *Proceedings of the IEEE*, 100(11):3056–3076, 2012.
2. Serdar Cabuk, Carla E. Brodley, and Clay Shields. Ip covert timing channels: design and detection. In *Proceedings of the 11th ACM conference on Computer and communications security*, CCS '04, pages 178–187, New York, NY, USA, 2004. ACM.
3. P. Cotae, I.S. Moskowitz, and M.H. Kang. Eigenvalue characterization of the capacity of discrete memoryless channels with invertible channel matrices. In *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*, pages 1–6, 2010.
4. V. Crespi, G. Cybenko, and A. Giani. Engineering statistical behaviors for attacking and defending covert channels. *Selected Topics in Signal Processing, IEEE Journal of*, 7(1):124–136, 2013.
5. T. Fine. Constructively using noninterference to analyze systems. In *Research in Security and Privacy, 1990. Proceedings., 1990 IEEE Computer Society Symposium on*, pages 162–169, 1990.
6. R. Holloway and R. Beyah. Covert dcf: A dcf-based covert timing channel in 802.11 networks. In *Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference on*, pages 570–579, 2011.
7. M.H. Kang, I.S. Moskowitz, and D.C. Lee. A network pump. *Software Engineering, IEEE Transactions on*, 22(5):329–338, 1996.
8. Jeremy Kepner. *10. The Kronecker Theory of Power Law Graphs*, chapter 10, pages 205–240.
9. Donald E. Knuth. *The art of computer programming, volume 2 (3rd ed.): seminumerical algorithms*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1997.
10. Jure Leskovec. *9. Kronecker Graphs*, chapter 9, pages 137–204.
11. Xiapu Luo, E.W.W. Chan, and R.K.C. Chang. Tcp covert timing channels: Design and detection. In *Dependable Systems and Networks With FTCS and DCC, 2008. DSN 2008. IEEE International Conference on*, pages 420–429, 2008.
12. I.S. Moskowitz. Variable noise effects upon a simple timing channel. In *Research in Security and Privacy, 1991. Proceedings., 1991 IEEE Computer Society Symposium on*, pages 362–372, 1991.
13. G. Smith. Quantifying information flow using min-entropy. In *Quantitative Evaluation of Systems (QEST), 2011 Eighth International Conference on*, pages 159–167, 2011.
14. R.W. Smith and G. Scott Knight. Predictable design of network-based covert communication systems. In *Security and Privacy, 2008. SP 2008. IEEE Symposium on*, pages 311–321, 2008.
15. Zhen Wang and M. Karpovsky. Algebraic manipulation detection codes and their applications for design of secure cryptographic devices. In *On-Line Testing Symposium (IOLTS), 2011 IEEE 17th International*, pages 234–239, 2011.
16. D.J.A. Welsh. *Codes And Cryptography*. Oxford science publications. Clarendon Press, 1988.