# Related-Key Linear Cryptanalysis

Poorvi L. Vora
Department of Computer Science
George Washington University
Washington, DC 20052, USA
Email: poorvi@gwu.edu

Darakhshan J. Mir
Department of Computer Science
George Washington University
Washington, DC 20052, USA
Email: mir_d@gwu.edu

***Abstract*— A coding theory framework for related-key linear cryptanalytic attacks on block ciphers is presented. It treats linear cryptanalysis as communication over a low capacity channel, and a related key attack (RKA) as a concatenated code. It is used to show that an RKA, using n related keys generated from k independent ones, can improve the amortized cost – in number of plaintext-ciphertext pairs per key bit determined – over that of k single key attacks, of any linear cryptanalysis, if k and n are large enough. The practical implications of this result are demonstrated through the design of an RKA, with k=5 and n=7, predicted to produce a 29% improvement for DES attacks that use an r-1 round approximation.**

## I. Introduction

Attacks that exploit the non-random behavior of symmetric-key ciphers (such as linear or differential cryptanalysis) typically require a large number of ciphertext values to successfully estimate the key. Because of this, it is generally assumed that changing the key often offers good protection against such attacks. This is clearly true, of course, if the different keys are independent. On the other hand, relationships among keys can arise in a number of situations: when the random number generators used in key generation are weak, or when the adversary is powerful enough to control the relationship. While formal models of block cipher cryptanalysis [15], [13], [9] and of related-key attacks (RKAs) [1] exist, there is, however, no model of the combination. In particular, it is not known how the relationship among keys affects the success probability of a statistical attack.

A more formal statement of the problem for the specific case of the linear cryptanalytic attack is as follows. Consider a single linear cryptanalytic equation of bias $b$, using $N$ plaintext-ciphertext (P/C) pairs to determine $d$ bits of a single key. Denote by $\nu(N)$ its amortized cost, in P/C pairs required per key bit determined – i.e. $\nu(N) = \frac{N}{d}$, and by $\epsilon(N)$ the corresponding probability of error. It is well-known that error decreases indefinitely only if $N$ increases correspondingly. Because $d$ is fixed, error decreases indefinitely only if amortized cost also increases indefinitely.

$$\epsilon(N) \to 0 \Rightarrow \nu(N) = \frac{N}{d} \to \infty \tag{1}$$

Now consider a set of linear cryptanalytic attacks using $n$ related keys, constructed from $k$ independent ones. If $n = k$ and all $k$ keys are independent, the best the adversary can do is to launch $k$ independent linear cryptanalytic attacks, and (1) represents the behaviour of $\nu$ with $N$ and $\epsilon(N)$ for each independent key. When $n \neq k$, however, and the adversary uses $N$ P/C pairs for each of the $n$ related keys, is $\nu = \frac{Nn}{kd}$ lower or higher or the same for a fixed value of $\epsilon$? Are there relationships among the keys for which it behaves one way or another? $\nu$ measures the communication complexity of the attack – that part of it that is online and requires an interaction with the sender of the message. A significant change in it could be of considerable importance.

This paper's contributions are threefold.

- It presents a *formal model* for RKAs on block ciphers that are already vulnerable to linear cryptanalysis. The model focuses on attacks where $r - 1$ of the $r$ rounds are linearly approximated, and may be easily extended to other types of statistical cryptanalysis and to stream ciphers.
- It shows that the general RKA provides an asymptotically lower value of $\nu(N)$ than do $k$ independent linear cryptanalytic attacks. In fact, it shows that $\nu(N)$ *can be maintained at a constant, finite value while decreasing* $\epsilon(N)$, i.e. that

  $$\epsilon(N) \to 0 \; and \; \nu(N) \simeq \Lambda \tag{2}$$

  are simultaneously possible, for some constant finite $\Lambda$. While the values of $k$ and $N$ for which practical improvements are seen depend on the particular cipher, the fact that asymptotic values of $\nu$ can be finite does not.
- It *describes an RKA* that provides a modest improvement (a decrease of 29%) for DES with only a small redundancy in keys ($k = 5$, $n = 7$). A larger improvement is expected for a larger number of keys. It appears that this RKA is general enough to be useful for other ciphers as well.

Thus the results provide a means of designing new attacks on block ciphers that are vulnerable to linear cryptanalysis. The results also imply that, not only is changing the keys often not sufficient to prevent against a statistical attack, but that, with a particularly strong adversary or a particularly weak key generator, it can be worse than using the same key, and prove beneficial to the adversary. In coding theory terms, a single-key attack is similar to a repetition code, but an RKA is similar to a channel code and provides the associated improvement in

1-4244-0504-1/06/$20.00 ©2006 IEEE

communication efficiency to the adversary. With this general premise, even in the absence of a strong relationship among the keys, the techniques described here should be useful in various other settings where key relationships are examined, including ciphertext-only as well as non-linear cryptanalysis, and stream cipher cryptanalysis.

## II. The Approach

The paper treats linear cryptanalysis as communication over a very noisy channel, using a model of [3] extended to address known-plaintext attacks and RKAs[1]. The message consists of the $d$ key bits determined using a single linear approximation of the cipher. The cipher provides the encoding of the message bits, and the randomness of the channel is provided by the plaintexts used. A property of each of the $N$ known plaintext-ciphertext (P/C) pairs provides an $N$-bit received codeword. The rate of the transmission is $\frac{d}{N}$. Because $d$ is fixed by the linear approximation used, it is not possible to maintain a constant rate while increasing $N$. Thus it is not possible to achieve the limits of the channel coding theorem [12] with a single key attack, and, beyond a certain point, using the same key repeatedly has the disadvantage of a repetition code. This paper hence examines RKAs.

RKAs provide improvements in error performance similar to those of channel codes. The paper shows that RKAs correspond to concatenated codes, where the inner code is defined by the linear cryptanalytic attack, and the outer code by the relationship among the keys. It translates the *wrong key hypothesis* [7] to an assumption in the model, which affects the error-correcting properties of the inner code, and hence the error performance of the single key linear cryptanalytic attack.

Using Forney's constructions [4], the paper applies the channel coding theorem [12] to the super-channel, consisting of the single-key attack and its estimate, to obtain (2). Note that $\Lambda$ does not correspond to the channel capacity of the linear cryptanalytic channel. That capacity cannot be achieved because $d$ cannot be increased indefinitely.

The theoretical result (2) obtained is asymptotic. To determine whether there would be sufficient decrease in $\nu$ for a small enough value of $k$ and $n$ to make the attack practical, a careful calibration of the error of the single-key attack, and the error of the RKA, is needed. The paper uses a few values from Matsui's [10] first theoretical and experimental results to represent the error of a single key attack as $N$ is increased. This value represents the probability of error for the super-channel. Motivated by [4], the paper uses an outer Reed Solomon code for the RKA, and obtains estimates of improvement in amortized cost for DES over that of $k$ single key attacks.

## III. Related Work

Filiol [3] first suggested that a known probabilistic relationship – between $C$ and a single binary property of $K$ – be modeled as a communication channel. In his model, for ciphertext-only attacks, the input to the channel is a single binary property of fixed key $K$, denoted $\mathcal{I}(K)$ (the parity of a few bits, say), see Figure 1. Its output is a bit of $C$, or the parity

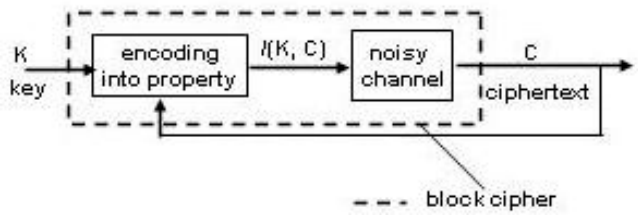


Fig. 1. Filiol's Channel Model [3]

of a few bits of $C$. The channel output is equal to the channel input with a probability slightly greater than half. The property may depend on the output, i.e. it could be an encoded bit of the key with feedback. Each use of the cipher transmits the same property over the channel, and corresponds to a repetition code on the property. Our model shows that known-plaintext cryptanalysis, approximating the cipher for $r-1$ rounds, can be used to generate several, distinct, encoded bits of the key.

[3] also describes how the same set of $N$ received bits may be decoded as a single repetition code of length $N$, or as $n$ codes of length $\frac{N}{n}$. While not explicitly described thus, this is the decoding technique for a concatenated code, with an inner repetition code of length $\frac{N}{n}$ (over the property of the key, $\mathcal{I}(K)$), and an outer repetition code of length $n$ (over the key). [3] correctly indicates that, in this case, concatenation provides no advantage, and that the most efficient decoding is one where the received bits are treated as consisting of a single codeword. [3] does not treat RKAs, and uses concatenation only for decoding, not for the purpose of increasing the efficiency of transmission across the cipher channel.

In other related work, Jakobsen [9] treats attacks on ciphers whose properties can be modeled as polynomials of small degree, and uses recent work in computational coding theory to efficiently decode attacks. In particular, he proposes the list decoding model, where the key estimate consists of a small set of possibilities, as opposed to a unique estimate. RKAs are not addressed in this model. The framework of Wagner [15] describes more formally the techniques for obtaining the probabilistic relationships among $P$, $C$ and $K$. It models the relationships as Markov chains, in the manner of [13], [14]. Our work models a relationship as a channel, which allows us to address RKAs, and provides access to a rich literature in coding theory. At the same time, our work allows, in a very natural way, the use of Wagner's model to determine the communication channel, and the properties transmitted across it.

Biham examines RKAs on block ciphers tracing the relationships among the keys to the key scheduling algorithm [2]. After demonstrating how RKAs lower the complexity for specific block ciphers, he stresses the need for a careful design of the key scheduling algorithm. Kelsey, Schneier and Wagner

[1]Though it was later shown that the attack of [3] was not proven to be valid for the AES as claimed, the core idea is very useful in characterizing attacks.

[5], [6] further present more RKAs on various other block ciphers and demonstrate how real protocols can be exploited to mount such attacks on them.

## IV. Our Framework: Single Key Attacks

### A. Preliminaries

We use notation very similar to that of Harpes, Kramer and Massey [7], from where we also draw our description of linear cryptanalysis. An $r$-round block cipher of block size $q$ consists of $r$ rounds of application of a keyed round function $\mathcal{F}_K$, a bijection, using a different *round key* $K^{(i)}$ for each round. The key to the cipher consists of all the round keys: $K^{(1,2,..r)} = (K^{(1)}, K^{(2)}, ...K^{(r)})$, where $K^{(i)} \in \mathcal{K}$, the round-key space. Plaintext $P$ and ciphertext $C$ belong to $\Sigma^q$, the set of all binary $q$-tuples. We consider the attack described by Matsui [10] that uses approximations of $r-1$ rounds of the cipher, and then uses the round function itself for the $r^{th}$ round.

In linear cryptanalysis, [10], [8], a single round of the cipher may be approximated using a linear expression of the form:

$$Pr[h_1(X) \oplus h_2(Y) \oplus h_3(K) = 0] = \frac{1}{2} \pm \gamma \qquad (3)$$

where $\gamma$ is positive, $X$ is the round input, $Y$ the output, and $K$ the round key, i.e. $Y = \mathcal{F}_K(X)$, and $h_1$, $h_2$ and $h_3$ are linear or affine. The randomness is across plaintexts, and not necessarily across keys. Through the repeated use of approximations like (3) for $r-1$ rounds, and the exact round function $\mathcal{F}$ for the last round, one may obtain an expression of the following form:

$$Pr[f(P) = g(\mathcal{F}^{-1}_{K^{(r)}}(C))] = \frac{1}{2} \pm b \qquad (4)$$

for some non-zero bias $b$ and last round key $K^{(r)}$. The exact value of the first-round function may also be used instead of an approximation, and it is straightforward to see how our model translates to this and other similar attacks.

To determine $K^{(r)}$, all possible values are tried, and the right and left-hand sides of equation (4) computed for all $N$ P/C pairs. The sub-key chosen is the one that satisfies the equation most often (or least often, to allow for the probability of (4) being $\frac{1}{2} - b$). In addition, one bit of the rest of the cipher key is also revealed through whether the sub-key chosen satisfied the equation most or least often. The other bits of the key may either be similarly determined, or determined by brute force. Thus linear cryptanalysis reduces, by one-$r^{th}$ and one bit, the length of the key that is to be determined by brute force.

### B. The Model

This paper views linear cryptanalysis as communication across a very noisy channel. $K^{(r)}$, denoted $K$ in Figure 2, forms the message. The value of a binary property of $K$ and $C_j$, the $j^{th}$ ciphertext:

$$\mathcal{I}_j(K) = g(\mathcal{F}^{-1}_K(C_j)) \qquad (5)$$

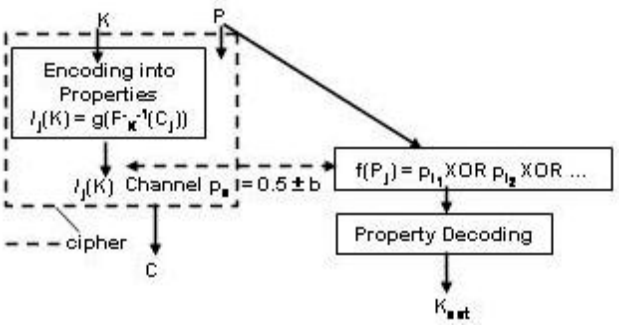


Fig. 2. Linear cryptanalysis as channel communication

is the $j^{th}$ codebit, and the codeword is:

$$\alpha(K) = (\mathcal{I}_1(K), \mathcal{I}_2(K), ...\mathcal{I}_N(K)) \qquad (6)$$

The codeword itself is not accessible to the adversary. However, the set $(f(P_1), f(P_2), ...f(P_N))$, of the binary property $f$ of the plaintexts, is, and, from (4), is a very noisy value of the codeword (6). It hence provides the output of the channel.

The randomness of the channel is provided by the different values of plaintext encrypted, and the channel flips each bit of the encoding almost as often as not, i.e. with a probability $0.5 \pm b$. The channel is a binary symmetric channel with probability of error $0.5 \pm b$, whose capacity may be estimated using the second order term in the Taylor series expansion (zeroth and first order terms are zero). We assume, as in [7], that $b$ is not significantly dependent on $K$, i.e. the channel is identical for all keys. The process of determining the target subkey from the values $f(P_j)$ can easily be shown to be maximum-likelihood decoding [11]. We further assume that the small dependence between $K$ and $b$ does not affect the maximum-likelihood estimation procedure [7]. This gives us the following observation.

**Observation 1:** *Linear cryptanalysis, using $N$ P/C pairs and a single linear cryptanalytic relationship, corresponds to the transmission of $K^{(r)}$ across a communication channel of capacity $\mathcal{C} \simeq \frac{b^2}{0.34}$, using the encoding $\alpha(K^r)$ of length $N$, and maximum-likelihood decoding.*

### C. The Encoding

Experimental reports of linear cryptanalysis imply that $\alpha(K)$ typically contains enough information to accurately determine $K^{(r)}$. A formal statement of the assumption that this is true is frequently made in the form of the *wrong key hypothesis* [7] which states that, for an incorrect key, $\mathcal{I}_j(K)$ is close to as likely to be represented by $f(P_j)$ as not. In our model, we incorporate an assumption that we show is equivalent to wrong key randomization, examine the error-correcting behavior of $\alpha(K)$ under this assumption, and relate it to a property of the cipher.

Recall that the randomness of the channel, whether representing the entire cipher or a single round, is provided by the plaintexts used. Hence, given a fixed key, there is a partition of the plaintext space, into those plaintexts for which the last

round approximation, of the form (3), is true, and those for which it is not, denote these by $\mathcal{P}_K$ and $\overline{\mathcal{P}_K}$ respectively. Note that the noise in the channel is "0", when $P \in \mathcal{P}_K$, and "1" otherwise.

**Definition 1:** *For round key $K$, $\mathcal{P}_K$ is the set of all $r^{th}$-round input for which $h_1(X) \oplus h_2(Y) = h_3(K)$ is true.*

We denote by $(P_1, K) \leftrightarrow (P_2, K')$ that $P_1$, with key $K$, and $P_2$, with key $K'$, represent the same value of the noise bit for the channel, i.e. the last round approximation (3) is either true for both or untrue for both. The following lemma holds.

**Lemma 1:**
$h_3(K) = h_3(K')$ and $(\mathcal{F}_K^{-1}(C_j), K) \leftrightarrow (\mathcal{F}_{K'}^{-1}(C_j), K')$
$\Rightarrow \mathcal{I}_j(K) = \mathcal{I}_j(K')$

*Proof:* $(\mathcal{F}_K^{-1}(C_j), K) \leftrightarrow (\mathcal{F}_{K'}^{-1}(C_j), K')$ implies that the value of $h_1(X) \oplus h_2(Y) \oplus h_3(K)$ is the same for both values of $K$. Further, because $h_3(K) = h_3(K')$, and $h_2(C)$ is the same for both $K$ and $K'$, this implies that $h_1(X)$ is also the same for both values of $K$. The value of $X$ itself is different, but $h_1(X)$ is not. Further, note that $h_1(X) = \mathcal{I}(C, K)$, hence $\mathcal{I}_j(K) = \mathcal{I}_j(K')$. ⊔

Similarly,

**Lemma 2:**
$h_3(K) \neq h_3(K')$ and $(\mathcal{F}_K^{-1}(C_j), K)$ !$(\leftrightarrow)$ $(\mathcal{F}_{K'}^{-1}(C_j), K')$
$\Rightarrow \mathcal{I}_j(K) = \mathcal{I}_j(K')$

*Proof:* Clear. ⊔

If $\mathcal{I}_j(K) = \mathcal{I}_j(K')$ too often, $\alpha(K)$ will not differentiate well between $K$ and $K'$. This motivates the following assumption and Theorem:

**Assumption 1:** *Given any two $K$, $K'$, such that $K \neq K'$, and $C$ chosen uniformly at random,*

$$|Pr[(\mathcal{F}_K^{-1}(C_j), K) \leftrightarrow (\mathcal{F}_{K'}^{-1}(C_j), K')] - \frac{1}{2}| \leq \delta$$

*where $\delta$ is small.*

**Theorem 1:** *Assumption 1 is equivalent to the wrong key randomization hypothesis.*

*Proof Sketch:* For distinct $K$, $K'$, let $Pr[(\mathcal{F}_K^{-1}(C_j), K) \leftrightarrow (\mathcal{F}_{K'}^{-1}(C_j), K')] = \frac{1}{2} + c$. Further, let $K$ be the right key, and $K'$ a wrong key. Wrong key randomization is equivalent to:

$$\frac{|Pr[f(P_j) = \mathcal{I}_j(K')] - \frac{1}{2}|}{|Pr[f(P_j) = \mathcal{I}_j(K)] - \frac{1}{2}|} << 1$$

$$\Leftrightarrow \frac{|(\frac{1}{2} \pm c)(\frac{1}{2} + b) + (\frac{1}{2} \mp c)(\frac{1}{2} - b) - \frac{1}{2}|}{b} << 1$$

$$\Leftrightarrow \frac{b|c|}{b} = |c| << 1$$

⊔

For a "good" round function, $\delta = 0$, so that the channel is completely independent of the key. As one might expect from Theorem 1, the error-correcting behavior of $\alpha$ depends on $\delta$, and, in particular, a "good" round function results in a more efficient attack.

**Theorem 2:** $\lim_{N\to\infty} \frac{min\_dist(\alpha)}{N} = \frac{1}{2} - \delta$, *where $min\_dist(\alpha)$ is the minimum distance of the code $\alpha$. Hence, a smaller value of $\delta$ results in a lower attack error.*

*Proof Sketch:* Straighforward. ⊔

## V. Related Keys and Concatenation

Consider an RKA, where $k$ independent keys are used to generate $n$ related keys. Suppose the function used is

$$\mathcal{H} : \mathcal{K}^k \rightarrow \mathcal{K}^n$$

$$\mathcal{H}(\mathcal{L}_1, \mathcal{L}_2, ...\mathcal{L}_k) = (K_1, K_2, ..K_n)$$

Each related key $K_i$ can be used for a linear cryptanalytic attack, to produce a key estimate, $K_{est,i}$. The relation among the keys may then be inverted in some manner.

### A. Related-Key Attacks as Concatenated Codes

**Theorem 3:** *The RKA described above is a concatenated code over the cipher channel.*

*Proof Sketch:* $\mathcal{H}$ is the outer code, the inversion procedure is its decoding. $\alpha(K)$ with maximum-likelihood decoding forms the inner code. The key estimates of a single linear cryptanalytic attack, $K_{est,i}$, form the output of the super-channel, whose input consists of the related keys $K_i$, and probability of error is that of the single-key attack. ⊔

Figure 3 shows such an attack.

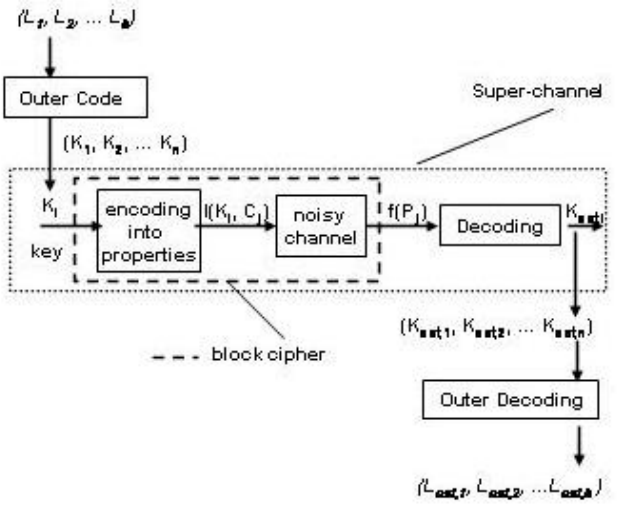


Fig. 3. Related-key Attacks as Concatenated Codes

### B. The Existence of an Efficient RKA

Consider any error value, $e$, reasonably small, corresponding to an amortized cost of $\frac{N}{d}$ in a single key attack. Assuming that the super-channel is symmetric, let its capacity be $\mathcal{C}_S(e)$. (For small values of $e$, $\mathcal{C}_S(e)$ is close to unity). This gives us:

**Theorem 4:** $\epsilon(N) \rightarrow 0$ *and* $\nu(N) \simeq \Lambda$, *for all* $\Lambda \geq \frac{N}{d\mathcal{C}_S(e)}$ *is possible, where* $\frac{N}{d} \geq \frac{0.34}{b^2}$.

*Proof:* Follows from the application of the channel coding theorem to the outer code and the super-channel with capacity $\mathcal{C}_S(e)$. ⊔

Theorem 4 implies that, while it is possible to transmit efficiently, one may not be able to transmit at the capacity of the inner channel if the inner code is bad, even if the outer code is good. Because the inner code has a finite number of message bits, it is not, in general, a good code.

### C. Construction of a Good RKA

Forney's constructions of concatenated codes motivate the use of RS codes as the relationship among the keys. Forney uses an RS outer code with a good inner code to reduce error indefinitely while maintaining any rate smaller than channel capacity ($\frac{b^2}{0.34}$ in this case). A similar attack, with $N \simeq \frac{0.34d}{b^2}$ P/C pairs in each of the $n$ single-key attacks, and $n$ only slightly larger than $k$ (as large as required by super-channel capacity, $\mathcal{C}_S(e)$), cannot be used to indefinitely decrease linear cryptanalytic error while maintaining rate at inner channel capacity, because the inner code is unable to maintain rate while decreasing error. It should provide some improvement, however, and we show that it can provide a reasonable improvement over single key attacks on DES.

### D. The Good RKA is Practical

To examine the improvement in amortized cost provided by the RKA that uses an RS code for the relationship among keys, one needs an expression for the error of the single-key attack in terms of $N$, $d$ and $b$, and of the RS code in terms of $d$, $k$, $n$, and super-channel error, i.e. the error of the single-key attack. Expressions for outer code error in terms of super-channel error are well-known [4]. The expressions for single-key attack error would generally depend on the attack itself. We use Matsui's theoretical and experimental values for the linear cryptanalytic attack on DES that uses approximations for $r-1$ rounds [10]. In this case, $d = 7$. [10, Table 3] predicts estimation accuracies of $48.6\%$, $78.5\%$, $96.7\%$, and $99.9\%$ for $N = \frac{2}{b^2}, \frac{4}{b^2}, \frac{8}{b^2}$ and $\frac{16}{b^2}$ respectively.

We construct various RKA attacks using RS outer codes and the above single-key attacks as inner codes. For these attacks, we obtain amortized cost estimates, requiring each of the $k$ keys to be determined at an accuracy level of 99.9%. We then observe the factor by which $\frac{16}{b^2}$, the cost for $k$ single key attacks each achieving a 99.9% accuracy, is larger than the amortized cost of the RKA, call this ratio $\tau$. Larger values of $\tau$ imply a greater improvement in amortized cost. We observe the following:

- An upper bound on $\tau$ corresponds to the RKA achieving inner channel capacity. The upper bound is about $6.75$.
- $N = \frac{8}{b^2}$ and the $(7, 5)$ RS code over $GF(2^3)$ (three groups of three bits each are encoded with the outer code) gives $\tau = 1.43$.
- $N = \frac{4}{b^2}$, and the $(127, 65)$ RS code over $GF(2^7)$ gives $\tau = 2$.

Matsui also reports experimental accuracies of $0.88$ and $0.99$ for $N = \frac{4}{b^2}$ and $\frac{8}{b^2}$ respectively. Using these values, we observe that an attack with $N = \frac{4}{b^2}$ and the $(127, 99)$ RS code over $GF(2^7)$ gives $\tau = 1.6$, and that the upper bound on $\tau$, corresponding to inner code channel capacity, is about $3.4$.

Larger values of $\tau$ would be obtained if larger values of $n$ were acceptable, or if larger accuracies were desired.

## VI. Conclusions and Future Directions

We have presented a model for RKAs that treats the RKA as a concatenated code. We have shown that RKAs can asymptotically achieve lower amortized cost than an equivalent set of many single-key attacks, and that this result does not depend on specific properties of the cipher, but simply on the fact that it is vulnerable to linear cryptanalysis. We have described an RKA expected to increase the efficiency of the linear cryptanalytic attack on DES for a small number of related-keys.

A number of future directions present themselves. First and foremost, an implementation of the RKA on specific block ciphers would indicate whether it is practical, and, if so, on what types of ciphers. Second, an implementation of similar attacks on stream ciphers would be interesting. Third, an examination of RKAs within the list decoding framework [9] might result in more efficient attacks, and could also provide insights into what types of round functions are resilient to such attacks. Fourth, an examination of key scheduling algorithms in this framework could be very interesting. Finally, other attacks, such as ciphertext-only and higher-order approximation attacks are also expected to lend themselves well to study in this framework.

## Acknowledgment

Poorvi L. Vora would like to thank the Imaging Systems Laboratory, Hewlett-Packard Laboratories, for partial support of the research.

## References

[1] Mihir Bellare and Tadayoshi Kohno. A Theoretical Treatment of Related-Key Attacks: RKA-PRPs, RKA-PRFs, and Applications. *Eurocrypt '03*.
[2] Eli Biham. New types of cryptanalytic attacks using related keys. *Eurocrypt '93*.
[3] Eric Filiol. Plaintext-dependent Repetition Codes Cryptanalysis of Block Ciphers - The AES Case. IACR eprint archive, http://eprint.iacr.org/2003/003/, 8th January 2003.
[4] David G. Forney. *Concatenated Codes*, MIT Press, 1966.
[5] John Kelsey, Bruce Schneier and David Wagner. Key-Schedule Cryptoanalysis of IDEA, G-DES, GOST, SAFER, and Triple-DES. *Crypto '96*.
[6] John Kelsey, Bruce Schneier and David Wagner. Related-key cryptanalysis of 3-WAY, Biham-DES, CAST, DES-X, NewDES, RC2, and TEA. *ICICS '97*.
[7] Carlo Harpes, Gerhard G. Kramer, and James L. Massey. A Generalization of Linear Cryptanalysis and the Applicability of Matsui's Piling-up Lemma. *Eurocrypt '95*.
[8] M. Heys. A Tutorial on Linear and Differential Cryptanalysis. *Technical Report CORR 2001-17*, Centre for Applied Cryptographic Research, Department of Combinatorics and Optimization, University of Waterloo, Mar. 2001. (Also appears in Cryptologia, vol. XXVI, no. 3, pp. 189-221, 2002.)
[9] Thomas Jakobsen. Ph.D. Dissertation, Dept. of Mathematics, Technical University of Denmark.
[10] M. Matsui. Linear Cryptanalysis Method for DES Cipher. *Eurocrypt '93*.
[11] S. Murphy, F.Piper, M.Walker and P.Wild. *Maximum Likelihood Estimation for Block Cipher Keys*. Research Report 1994 (Original version 1992).
[12] Claude Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27: 379-423, July 1948.
[13] S. Vaudenay. An Experiment on DES: Statistical Cryptanalysis. *ACM CCS '96*: 139-147, ACM Press.
[14] S. Vaudenay. Decorrelation: A Theory for Block Cipher Security. *Journal of Cryptology*, 16(4):249-286, Sept. 2003.
[15] David Wagner. Towards a unifying view of block cipher cryptanalysis. *Fast Software Encryption 2004*.