On Cipher-Dependent Related-Key Attacks in the Ideal-Cipher Model

M.R. Albrecht[1], P. Farshim[2], K.G. Paterson[2], and G.J. Watson[3]

[1] INRIA, Paris-Rocquencourt Center, SALSA Project
UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris, France
CNRS, UMR 7606, LIP6, F-75005, Paris, France
malb@lip6.fr
[2] Information Security Group,
Royal Holloway, University of London,
Egham, Surrey, TW20 0EX, United Kingdom.
{pooya.farshim,kenny.paterson}@rhul.ac.uk
[3] Department of Computer Science, University of Calgary,
2500 University Dr. NW, Calgary, Alberta, Canada T2N 1N4
gjwatson@ucalgary.ca

Abstract. Bellare and Kohno introduced a formal framework for the study of related-key attacks against blockciphers. They established sufficient conditions (output-unpredictability and collision-resistance) on the set of related-key-deriving (RKD) functions under which an ideal cipher is secure against related-key attacks, and suggested this could be used to derive security goals for real blockciphers. However, to do so requires the reinterpretation of results proven in the ideal-cipher model for the standard model (in which a blockcipher is modelled as, say, a pseudorandom permutation family). As we show here, this is a fraught activity. In particular, building on a recent idea of Bernstein, we first demonstrate a related-key attack that applies generically to a large class of blockciphers. The attack exploits the existence of a short description of the blockcipher, and so does not apply in the ideal-cipher model. However, the specific RKD functions used in the attack are provably output-unpredictable and collision-resistant. In this sense, the attack can be seen as a separation between the ideal-cipher model and the standard model. Second, we investigate how the related-key attack model of Bellare and Kohno can be extended to include sets of RKD functions that themselves access the ideal cipher. Precisely such related-key functions underlie the generic attack, so our extended modelling allows us to capture a larger universe of related-key attacks in the ideal-cipher model. We establish a new set of conditions on related-key functions that is sufficient to prove a theorem analogous to the main result of Bellare and Kohno, but for our extended model. We then exhibit non-trivial classes of practically relevant RKD functions meeting the new conditions. We go on to discuss standard model interpretations of this theorem, explaining why, although separations between the ideal-cipher model and the standard model still exist for this setting, they can be seen as being much less natural than our previous separation. In this manner, we argue that our extension of the Bellare–Kohno model represents a useful advance in the modelling of related-key attacks. Third, we consider the topic of key-recovering related-key attacks and its relationship to the Bellare–Kohno formalism. In particular, we address the question of whether lowering the security goal by requiring the adversary to perform key-recovery excludes separations of the type exhibited by us in the Bellare–Kohno model.

Keywords. Related-key attack, Ideal-cipher model, Blockcipher.

1 Introduction

BACKGROUND. Related-key attacks were introduced by Biham and Knudsen [6,7,19], and have received considerable attention recently partly due to the discovery of various high-profile key-recovery attacks in this model ([8,9,13]). Some of these new attacks, in particular the family of attacks against AES, do not restrict key-derivation functions to either simple XORs or modular addition of constants. Instead non-linear key-derivation functions are used. This has sparked a debate as to whether these attacks should be considered valid, and in turn whether related-key attacks should be considered valid attacks on blockciphers in general. Part of the debate stems from the question of whether the job of preventing related-key attacks should fall to blockcipher designers or to designers of protocols making use of blockciphers. The latter group could put a stop to such attacks simply by avoiding the use of related keys within their protocols, and this in turn would remove any real incentive for cryptanalysts to consider ever more esoteric key relations. However, taking a pragmatic perspective, there are widely deployed real-world protocols which do make use of such related keys, so the study of related-key attacks holds relevance and interest both from cryptanalytic and theoretical perspectives. For example, key-derivation procedures leading to related-key scenarios seem to be widely used in the financial sector, with a public-domain example being the EMV specifications for card transactions [14, Appendix A1.3.1]. Other examples include the 3GPP confidentiality and integrity algorithms f8,f9 [18].

On the theoretical side, Bellare and Kohno [3] provided a thorough study of related-key attacks. Their main result established a general possibility result concerning security against related-key attacks, for certain classes of related-key-deriving (RKD) functions. Bellare and Kohno have as a thesis that the minimal requirement for a blockcipher security goal to be considered feasible is that it should be provably achievable for an ideal cipher. To this end, they showed that an ideal cipher is secure against related-key attacks involving any set of RKD functions that is both *collision-resistant* and *output-unpredictable*. However, to be usable in studying the security of real blockciphers, we need to be able to interpret such ideal-cipher-model results in the standard model, in which we might model a blockcipher as a pseudorandom permutation family. We note that [3] contains very little in the way of such interpretation.

However, the community's confidence in our ability to translate such results to the standard model has recently received a severe dent. In [17], Harris demonstrated that if the cipher itself is available for use during key derivation, then RKD functions can be constructed using which keys can be recovered for any cipher. Bernstein [5] presented a simple distinguishing attack on AES that also made use of the blockcipher itself in the RKD functions. Moreover, at least heuristically, the sets of RKD functions used in these attacks fulfil the conditions of collision-resistance and output-unpredictability needed to prove Bellare and Kohno's main result about security against related-key attacks. Researchers subsequently argued that, in view of these examples, the model for related-key attacks presented in [3] is broken, in the sense that, since any cipher can be broken in that model, then this model does not tell us anything about ciphers; rather it is simply too strong a model.

CONTRIBUTIONS. We begin by exploring the question of how to interpret the main result of Bellare and Kohno [3], restated here as Theorem 1, in the standard model. We provide two possible interpretations of this result, which vary only in the order in which they invoke certain quantifiers. We then formalise Bernstein's attack as a related-key attack that applies generically to a large class of blockciphers (those having equal-sized keys and messages). Moreover, we formally prove, under the standard assumption that the blockcipher is pseudorandom, that Bernstein's RKD functions meet the sufficient conditions of collision-resistance and output-unpredictability needed for the application of Theorem 1. We then explain how this attack can be seen as a separation between the ideal-cipher model and the standard model in the context of the second of our two interpretations of Theorem 1. We also justify why the first interpretation of Theorem 1 for the standard model is less interesting in practical contexts.

In an attempt to restore confidence in the Bellare–Kohno model, we extend the model to allow RKD functions which access the blockcipher itself. Since we are working with an ideal cipher, we model such access via oracle calls to the ideal cipher and its inverse. This allows us to do several things. Firstly, we can capture attacks like that due to Bernstein in our model (where it shows up as an attack against an ideal cipher for a particular set of RKD functions). Secondly, it allows us to prove the security of an ideal cipher for other sets of RKD functions which make use of the blockcipher during key derivation. Thirdly, it allows us to investigate analogues of Theorem 1 for the new setting. This leads to our main result, Theorem 4, in which we establish that an ideal cipher is secure against related-key attacks for sets of RKD functions that meet certain conditions. More precisely, we introduce oracle versions of collision-resistance and output-unpredictability, along with a new notion called *oracle-independence* of a set of RKD functions; we then show that these three conditions taken together are sufficient to establish the security of an ideal cipher against related-key attacks using that set of RKD functions. We go on to show that our main theorem is not vacuous by exhibiting non-trivial classes of practically relevant RKD functions meeting the new conditions. In particular, we show that RKD function sets like those used in the EMV standard meet the new conditions.

Given the problems we have identified with making standard model interpretations of Theorem 1, we then proceed to a careful discussion of how our main result, Theorem 4, can be translated into the standard model. When restricted to RKD sets which are independent of the blockcipher, our theorem becomes equivalent to that of Bellare and Kohno: its interpretation states that a reasonable blockcipher should resist related-key attacks when restricted to such an RKD set. On the other hand, for RKD sets which depend on the blockcipher, our theorem goes beyond that of Bellare and Kohno (which provides no guarantees) in the following way. Its interpretation asserts that if the dependency of the RKD functions is black box, and furthermore the set satisfies certain conditions, then a good blockcipher is expected to resist related-key attacks when restricted to such an RKD set. In particular, the RKD sets of Bernstein and Harris do not satisfy the required conditions. On the positive side, there exist cipher-dependent RKD sets which satisfy the required conditions.

Our final contribution is to ask whether the problems that arise in translating from the ideal-cipher model to the standard model in the context of related-key attacks can be avoided by lowering our sights. In particular, we consider the topic of related-key attacks that recover keys (rather than breaking the pseudorandomness of a cipher in the sense considered in [3]). This asks more of the adversary and therefore represents a weakening of the model. In turn, this opens up the possibility of excluding separation results like that we have shown. We can in fact show that the particular set of RKD functions used in Bernstein's attack *cannot* be used to mount a key-recovery attack. Unfortunately, we also have a negative result: using a modification of the attack of Harris, we exhibit a specific set of RKD functions that does lead to a full key-recovery attack against real blockciphers, even though the functions satisfy the conditions for Theorem 1 to be applicable. Again, the RKD functions access the blockcipher itself, so the attack can be regarded as another separation between the ideal-cipher model and the standard model, but now for a weaker security notion than was originally considered in [3].

OTHER RELATED WORK. Bellare and Kohno also gave constructions of concrete blockciphers which are secure against adversaries which only partially transform the key. In subsequent work, Lucks [20] investigated RKA-secure blockciphers further, and gave improved security bounds for such partially key-transforming adversaries. In this work, the author also constructed a concrete blockcipher which is RKA-secure with respect to a rich set of related-key-deriving functions. Lucks's construction, however, was based on a non-standard, interactive number-theoretic assumption. The recent work of Goldenberg and Liskov [15] examines whether it is possible to build related-key-secure blockciphers from traditional cryptographic primitives. They show that while a related-key/secret pseudorandom bit is sufficient and necessary to build such a blockcipher, hard-core bits with typical security proofs are not related-secret secure. Very recently, Bellare and Cash [2] managed to construct PRFs and PRPs which are RKA secure with respect to key transformations which involve the action of a group element on the key. Their constructions are based on standard number-theoretic assumptions such as DDH. In yet another recent work, Applebaum, Harnik and Ishai [1] study related-key attacks for randomised symmetric encryption schemes. They also discuss the applications of such RKA-secure primitives to batch and adaptive oblivious transfer protocols.

RELATION TO KDM SECURITY. Key-dependent message (KDM) security [11,16] is a strong notion for primitives such as private/public-key encryption schemes and PRF/PRPs where one requires security in the presence of an adversary which can obtain the outputs of the encryption/function on points which depend, in known (or even chosen) ways, on secret keys. This setting is similar to related-key attacks in the sense that security games involve functions of an unknown key. However, while superficially similar in this sense, the RKA and KDM notions hand different capabilities to the adversary. A fuller discussion of the relations between these notions is beyond the scope of the present paper. In [16], the authors briefly define cipher-dependent KDM security, however their results are about relations which are *independent* of the cipher. We note that analogues of Bernstein's and Harris's attack in the context of KDM security were already noted in [16].

ORGANISATION. In the next section we settle notation and recall a number of definitions from [3]. Section 3 is concerned with the possible interpretations of the main result of [3] in the standard model. In Section 4 we extend the security model of Bellare and Kohno to include RKD sets that access the ideal cipher itself during key derivation. We also discuss some positive and negative results in this new model. We close by discussing the relevance of our results to practice.

2 Notation and related-key attacks

NOTATION. We denote by $s \xleftarrow{\$} S$ the operation of sampling s uniformly at random from set S, and by $x \leftarrow y$ the assignment of value y to x. For a set S, $|S|$ denotes its size. We let $\mathrm{Perm}(\mathcal{D})$ denote the set of all permutations on $\mathcal{D}$. A blockcipher is a family of permutations $E : \mathcal{K} \times \mathcal{D} \to \mathcal{D}$, where $\mathcal{K}$ is the key space and $\mathcal{D}$ is the domain or message space.

We recall a number of definitions from [3].

Definition 1 (Pseudorandomness). *Let $E : \mathcal{K} \times \mathcal{D} \to \mathcal{D}$ be a family of functions. Let A be an adversary. Then*

$$\mathbf{Adv}_{E}^{\mathsf{prp}}(A) := \Pr\left[K \xleftarrow{\$} \mathcal{K} : A^{E(K,\cdot)} = 1\right] - \Pr\left[G \xleftarrow{\$} \mathrm{Perm}(\mathcal{D}) : A^{G(\cdot)} = 1\right]$$

is defined as the prp-advantage *of A against E.*

We let $\mathrm{Perm}(\mathcal{K}, \mathcal{D})$ denote the set of all blockciphers with domain $\mathcal{D}$ and key-space $\mathcal{K}$. Thus the notation $G \xleftarrow{\$} \mathrm{Perm}(\mathcal{K}, \mathcal{D})$ corresponds to selecting a random blockcipher. In more detail, it comes down to defining G via

$$\text{For each } K \in \mathcal{K} : G(K, \cdot) \xleftarrow{\$} \mathrm{Perm}(\mathcal{D}).$$

Given a family of functions $E : \mathcal{K} \times \mathcal{D} \to \mathcal{D}$ and a key $K \in \mathcal{K}$, we define the related-key oracle $E(\mathsf{RK}(\cdot, K), \cdot)$ as an oracle that takes two arguments, a function $\phi : \mathcal{K} \to \mathcal{K}$ and an element $x \in \mathcal{D}$, and returns $E(\phi(K), x)$. We shall refer to ϕ as a related-key-deriving (RKD) function. We let Φ be a set of functions mapping $\mathcal{K}$ to $\mathcal{K}$. We call Φ the set of allowed RKD functions and it will be a parameter of our definitions.

Definition 2 (Pseudorandomness with respect to related-key attacks). *Let $E : \mathcal{K} \times \mathcal{D} \to \mathcal{D}$ be a family of functions and let Φ be a set of RKD functions over $\mathcal{K}$. Let A be an adversary with access to a related-key oracle, and restricted to queries of the form (ϕ, x) in which $\phi \in \Phi$ and $x \in \mathcal{D}$. Then*

$$\mathbf{Adv}_{\Phi,E}^{\mathsf{prp\text{-}rka}}(A) := \Pr\left[K \xleftarrow{\$} \mathcal{K} : A^{E(\mathsf{RK}(\cdot,K),\cdot)} = 1\right] - \Pr\left[K \xleftarrow{\$} \mathcal{K}; G \xleftarrow{\$} \mathrm{Perm}(\mathcal{K}, \mathcal{D}) : A^{G(\mathsf{RK}(\cdot,K),\cdot)} = 1\right]$$

is defined as the prp-rka-advantage *of A in a Φ-restricted related-key attack (RKA) on E.*

Therefore in a related-key attack an adversary's success rate is measured by its ability to distinguish values of the cipher on related-keys from those returned from a random blockcipher.

Definition 3 (RKA pseudorandomness in the ideal-cipher model). *Fix sets $\mathcal{K}$ and $\mathcal{D}$ and let Φ be a set of RKD functions over $\mathcal{K}$. Let A be an adversary with access to three oracles, and restricted to queries of the form (K', x) for the first two oracles and (ϕ, x) for the last, where $K' \in \mathcal{K}$, $\phi \in \Phi$ and $x \in \mathcal{D}$. Then*

$$\begin{aligned}\mathbf{Adv}_{\Phi,\mathcal{K},\mathcal{D}}^{\mathsf{prp\text{-}rka}}(A) := {} & \Pr\left[K \xleftarrow{\$} \mathcal{K}; E \xleftarrow{\$} \mathrm{Perm}(\mathcal{K}, \mathcal{D}) : A^{E, E^{-1}, E(\mathsf{RK}(\cdot,K),\cdot)} = 1\right] \\ & - \Pr\left[K \xleftarrow{\$} \mathcal{K}; E \xleftarrow{\$} \mathrm{Perm}(\mathcal{K}, \mathcal{D}); G \xleftarrow{\$} \mathrm{Perm}(\mathcal{K}, \mathcal{D}) : A^{E, E^{-1}, G(\mathsf{RK}(\cdot,K),\cdot)} = 1\right]\end{aligned}$$

is defined as the prp-rka-advantage *of A in a Φ-restricted related-key attack on an ideal cipher with keys $\mathcal{K}$ and domain $\mathcal{D}$.*

This definition is simply an adaptation of Definition 2 to the ideal-cipher model by allowing oracle access to E and E^{-1}. To de-clutter the notation, we use E and E^{-1} as shorthand for $E(\cdot,\cdot)$ and $E^{-1}(\cdot,\cdot)$, respectively. An $E^{-1}(\mathsf{RK}(\cdot,K),\cdot)$ oracle can be added to the above definition to get *strong RKA pseudorandomness*. In this paper, however, we will work with the standard (i.e. non-strong) pseudorandomness. Our results can be extended to the strong setting.

Definition 4 (Output-unpredictability-2). *Let Φ be a set of RKD functions on the key-space $\mathcal{K}$. Let $\mathcal{P}_K(\cdot)$ and $\mathcal{X}(\cdot)$ be a pair of oracles. The oracle $\mathcal{P}_K(\cdot)$ takes as input an element $\phi \in \Phi$ and the oracle $\mathcal{X}(\cdot)$ takes as input an element $K' \in \mathcal{K}$. Neither oracle returns a value. An adversary wins if it queries its $\mathcal{X}(\cdot)$ oracle with a key K' and if it queries its $\mathcal{P}_K(\cdot)$ oracle with a function ϕ such that $\phi(K) = K'$. We define the* up2-advantage *of an adversary A as*

$$\mathbf{Adv}_{\Phi}^{\mathsf{up2}}(A) := \Pr\left[K \xleftarrow{\$} \mathcal{K} : A^{\mathcal{P}_K(\cdot),\mathcal{X}(\cdot)} \textit{ wins}\right].$$

The above definition captures the intuition that no adversary is able to predict the value $\phi(K)$, for a random K, with a high probability.

Definition 5 (Collision-resistance-2). *Let Φ be a set of functions on the key-space $\mathcal{K}$. Let $\mathcal{C}_K(\cdot)$ be an oracle that takes as input a function $\phi \in \Phi$ and that returns no value. An adversary wins if it queries its oracle with two distinct functions $\phi_1, \phi_2 \in \Phi$ such that $\phi_1(K) = \phi_2(K)$. We define the* cr2-advantage *of an adversary A as*

$$\mathbf{Adv}_{\Phi}^{\mathsf{cr2}}(A) := \Pr\left[K \xleftarrow{\$} \mathcal{K} : A^{\mathcal{C}_K(\cdot)} \textit{ wins}\right].$$

The intuition here is that no adversary can trigger a collision between two different ϕ's with high probability. Note also that output-unpredictability is simply collision-resistance between a non-constant and a constant function. Throughout the paper we call an RKD set Φ output-unpredictable-2 or collision-resistant-2 if the corresponding advantage is "small" for efficient any adversary.

REMARK. Alternative and stronger notions of output-unpredictability and collision-resistance are also presented in [3]. However, the above definitions are enough for the main result there. We note that an attractive feature of the above definitions is their *non-interactiveness*. In fact, it is possible to simplify these definitions further by requiring an adversary which returns a single pair (K,ϕ) in the output-unpredictability-2 game, and two distinct RKD functions (ϕ_1,ϕ_2) in the collision-resistance-2 game. Using a standard reduction one can show that the simplified definitions are equivalent to the above definitions (respectively). In the first case a (multiplicative) security loss of $qq'/2$, where q and q' are, respectively, the number of queries to the $\mathcal{X}$ and $\mathcal{P}_K$ oracles, is introduced. In the second case, a loss of $q(q-1)/2$ is introduced, where q is the number of queries to the $\mathcal{C}_K$ oracle.

3 A generic cipher-dependent attack

Bellare and Kohno established the following theorem as their main result in [3].

Theorem 1 (Bellare and Kohno [3]). *Fix a key space $\mathcal{K}$ and domain $\mathcal{D}$. Let Φ be a set of RKD functions over $\mathcal{K}$. Let A be an ideal-cipher-model adversary that queries its first two oracles with a total of at most q' different keys and that queries its last oracle with a total of at most q different RKD functions from Φ. Then there are output-unpredictability-2 and collision-resistance-2 adversaries B and C such that*

$$\mathbf{Adv}_{\Phi,\mathcal{K},\mathcal{D}}^{\mathsf{prp\text{-}rka}}(A) \le \mathbf{Adv}_{\Phi}^{\mathsf{cr2}}(B) + \mathbf{Adv}_{\Phi}^{\mathsf{up2}}(C),$$

where B queries its $\mathcal{C}_K$ oracle q times, and C queries its $\mathcal{P}_K$ and $\mathcal{X}$ oracles q and q' times (respectively).

The above theorem states that for all Φ satisfying appropriate properties, an ideal cipher is secure against Φ-restricted related-key attacks. It is tempting to try to translate this ideal-cipher-model result to the standard model. Indeed, it is conceivable that a real blockcipher might also resist such Φ-restricted attacks under the same conditions. This statement can be interpreted in (at least) two ways.

1. For any Φ which is collision-resistant and output-unpredictable, there is a standard model blockcipher which resists Φ-restricted attacks; and
2. There is a standard model blockcipher E which resists all Φ-restricted attacks, as long as Φ is collision-resistant and output-unpredictable.

The essential difference between these two interpretations is their *order of quantifiers*. In the first interpretation, there may be no dependencies of Φ on E, whereas the blockcipher in the second interpretation should resist all Φ-restricted attacks, including those which depend on E. The theorem of Bellare and Kohno, on the other hand, does not allow the functions in Φ to depend on the ideal cipher itself, as the latter is chosen uniformly at random and independently from Φ. Therefore, the first interpretation is, in our opinion, a more accurate translation of the theorem to the standard model. In fact no natural counterexamples to this interpretation are yet known.[4] On the other hand, based on the aforementioned recent example of Bernstein, we show in the next theorem that the second interpretation is *in*valid. This result utilises RKD sets Δ_E, which depend on E, for each blockcipher E. The proof is given in Appendix A.

Theorem 2. *Let $E : \mathcal{K} \times \mathcal{D} \to \mathcal{D}$ be a family of functions with $\mathcal{K} = \mathcal{D}$. Let $0, 1 \in \mathcal{D}$ be any two distinct elements of $\mathcal{D}$ and consider the set of RKD functions*

$$\Delta_E := \{K \mapsto K, K \mapsto E(K, 0)\}.$$

Then there are a Δ_E-restricted related-key adversary A against E, and a prp adversary B against E such that

$$\mathbf{Adv}^{\mathsf{prp\text{-}rka}}_{\Delta_E,E}(A) \geq 1 - \mathbf{Adv}^{\mathsf{prp}}_{E}(B) - 2/|\mathcal{K}|.$$

Furthermore, for any collision-resistant-2 or output-unpredictable-2 adversary A, there is a prp adversary B such that

$$\mathbf{Adv}^{\mathsf{up2}}_{\Delta_E}(A) \leq \mathbf{Adv}^{\mathsf{prp}}_{E}(B) + 2q'/|\mathcal{K}| \quad \textit{and} \quad \mathbf{Adv}^{\mathsf{cr2}}_{\Delta_E}(A) \leq \mathbf{Adv}^{\mathsf{prp}}_{E}(B) + 1/|\mathcal{K}|,$$

where q' is the number of queries that an output-unpredictability-2 adversary A makes to its $\mathcal{X}$ oracle.

Hence if the blockcipher E is prp secure, then Δ_E is both collision-resistant-2 and output-unpredictable-2. This theorem therefore exhibits a class of ciphers E for which the second standard model interpretation of Theorem 1 does not hold. Note that we have in fact established a strong falsification of the second interpretation as it is enough to show that the inequality in the statement of Theorem 1 does not hold in the standard model. Note also that the inequalities in the above theorem can be somewhat simplified by observing that prp-rka security with respect to Δ_E-restricted adversaries implies prp security.

Note that in the Δ_E set, one can replace 0 with any $x \in \mathcal{D}$. Furthermore, there is no special role played by the identity function as a similar attack applies if the set was defined to be $\{K \mapsto E(K, 0), K \mapsto E(E(K, 0), 0)\}$. Note also that no efficiency requirements on RKD functions are made in Theorem 1, and the result holds even for ϕ that are infeasible to compute (in the ideal-cipher model). This allows us to define an RKD set containing a single function which allows an attacker to recover the key of a concrete cipher: $K \mapsto K'$ with K' such that $E(K', 0) = K$. We stress that this failure of the model, although technically allowed in the model of [3], is only of theoretical interest: the ability to compute this function would immediately break the prp-security of the cipher.

Harris [17] presents another cipher-dependent related-key attack which breaks every cipher in the standard model. In Appendix C we formalise this attack and study its implications. In particular, the description

[4] Although artificial code-based separation results akin to that in [10] might be constructible.

of the RKD set is unclear in the original work, and depending on the interpretation, the set might or might not satisfy the collision-resistance-2 property. We clarify this issue by deriving accurate bounds for the advantage of a related-key adversary.

Theorem 2 and Theorem 9 (see Appendix C) can be seen as a *weak* separation between the standard model and the ideal-cipher model as they (only) rule out the second interpretation. It remains an open problem to prove or disprove the first interpretation. Disproving it would demonstrate a *strong* separation result as it also implies the weak separation.[5] We, however, do not consider this to be an important issue since, in a real-world attack, the attacker can choose its set of RKD functions after seeing the cipher, which relates more closely to the second interpretation. Put differently, at the core of the above attacks lies the dependence of the RKD set on E, which cannot be replicated in the Bellare–Kohno (BK) model.

Most concrete related-key attacks [8,9,13] lead to the recovery of a blockcipher's key. Hence one way to restore confidence in the BK model would be to raise the security bar for an attacker, and require it to recover keys. We formalise key-recovery in Appendix B and show that this approach cannot succeed.

In the next section, we investigate how the model can be modified so as to capture such cipher-dependent related-key attacks.

4 RKD functions with access to E and E^{-1}

As discussed above, one weakness of the BK model lies in its inability to model related-key functions which depend on the blockcipher. In this section, we extend the BK model to address this issue and prove a result akin to Theorem 1 for this extended setting. In doing so, we treat the RKD functions as being oracle Turing machines and write each RKD function as $\phi^{\mathcal{O}_1,\mathcal{O}_2}$, where $\mathcal{O}_i$ are oracles. These oracles will be instantiated with a random blockcipher E and its inverse E^{-1} during security games. We denote a set of such oracle RKD functions by[6] $\Phi^{E,E^{-1}}$. We note that such oracle RKD functions are of interest in the ideal-cipher model only: a concrete blockcipher has a compact description and there is no need to grant access to the oracle.

We are now ready to define a refined notion of RKA pseudorandomness in the ideal-cipher model.

Definition 6 (Oracle RKA pseudorandomness in the ideal-cipher model). *Fix sets $\mathcal{K}$ and $\mathcal{D}$, and let $\Phi^{E,E^{-1}}$ be a set of oracle RKD functions over $\mathcal{K}$. Let A be an adversary with access to three oracles, and restricted to queries of the form (K',x) for the first two oracles and $(\phi^{E,E^{-1}},x)$ for the last, where $K' \in \mathcal{K}$, $\phi^{E,E^{-1}} \in \Phi^{E,E^{-1}}$, and $x \in \mathcal{D}$. Then*

$$\mathbf{Adv}^{\mathsf{prp\text{-}orka}}_{\Phi^{E,E^{-1}},\mathcal{K},\mathcal{D}}(A) := \Pr\left[K \xleftarrow{\$} \mathcal{K}; E \xleftarrow{\$} \mathrm{Perm}(\mathcal{K},\mathcal{D}) : A^{E,E^{-1},E(\mathsf{RK}(\cdot,K),\cdot)} = 1\right]$$
$$- \Pr\left[K \xleftarrow{\$} \mathcal{K}; E \xleftarrow{\$} \mathrm{Perm}(\mathcal{K},\mathcal{D}); G \xleftarrow{\$} \mathrm{Perm}(\mathcal{K},\mathcal{D}) : A^{E,E^{-1},G(\mathsf{RK}(\cdot,K),\cdot)} = 1\right]$$

is defined as the prp-orka-advantage *of A in a $\Phi^{E,E^{-1}}$-restricted related-key attack on an ideal cipher with keys $\mathcal{K}$ and domain $\mathcal{D}$.*

Informally, we say that an ideal cipher is secure against $\Phi^{E,E^{-1}}$-restricted related-key attacks if the above advantage is "small" for any efficient A.

We now define a set of oracle RKD functions, which is the ideal-cipher model counterpart[7] of Δ_E defined in Theorem 2, as follows:

$$\Delta^E := \{K \mapsto K, K \mapsto E(K,0)\}.$$

The next theorem shows that this set can be used to break an ideal cipher in the sense of Definition 6.

[5] We remark that Proposition 9.1 of [3] demonstrates an intermediate result: it considers a restricted set of RKD functions which only alter the last few bits of the key, but may depend on the cipher.

[6] Although we use E and E^{-1} in the exponent, in the following security definitions they are chosen during each game.

[7] Note that we use superscripts to denote an oracle access whereas subscripts are used to denote dependence. The former is of interest in the ideal-cipher model, and the latter in the standard model.

Theorem 3. *Fix a set $\mathcal{K}$ and let $\mathcal{D} = \mathcal{K}$. Then there exists the ideal-cipher model Δ^E-restricted adversary A such that*

$$\mathbf{Adv}^{\mathsf{prp\text{-}orka}}_{\Delta^E,\mathcal{K},\mathcal{D}}(A) \geq 1 - 2/|\mathcal{K}|.$$

Proof. The proof of this theorem is similar to that of Theorem 2. Adversary $A^{E,E^{-1},f(\mathsf{RK}(\cdot,K),\cdot)}$, whose goal is to decide whether $f = E$ or $f = G$, operates as shown in Figure 1.

```
Algorithm $A^{E,E^{-1},f}$:
Query RK oracle on $(K \mapsto K, 0)$ to get $x = f(K,0)$
Query RK oracle on $(K \mapsto E(K,0), 0)$ to get $y = f(E(K,0),0)$
Query oracle $E$ on $(x,0)$ to get $z = E(x,0)$
Return $(z = y)$
```

Fig. 1: Δ^E-restricted adversary breaking an ideal cipher with $\mathcal{K} = \mathcal{D}$.

When $f = E$, we have that $x = E(K,0)$, $y = E(E(K,0),0)$, and $z = E(E(K,0),0)$. Hence $z = y$ with probability 1. On the other hand, if $f = G$ then $x = G(K,0)$, $y = G(E(K,0),0)$, and $z = E(G(K,0),0)$. Similarly to the proof of Theorem 2 we have that $\Pr[E(G(K,0),0) = G(E(K,0),0)] \leq 2/|\mathcal{K}|$. The theorem follows. □

As it can be seen, the proof is analogous to that of Theorem 2, reflecting the ability of our extended model to capture such cipher-dependent attacks. Another way to look at this result is to interpret it in terms of the event whose analysis underpins the proof of Theorem 1 in [3]. In that proof one needs to upper-bound the probability of:

Event D: A queries its related-key oracle with a function ϕ and queries its ideal cipher (in either the forward or backward directions) with a key K' such that $\phi(K) = K'$.

Looking at the code of A in Figure 1, it is easy to check that event D happens with probability 1: In A's attack $K' = E(K,0)$ is a key queried to E and the RK oracle will have a key equal to K' when queried with $K \mapsto E(K,0)$. This observation motivates the introduction of appropriately modified notions of output-predictability-2 and collision-resistance-2, as well as additional definitions which might enable a proof of oracle RKA pseudorandomness in the ideal-cipher model for $\Phi^{E,E^{-1}}$-restricted adversaries to be constructed.

Our first two conditions are modified versions of the collision-resistance-2 and output-unpredictability-2 notions of Bellare and Kohno as recalled in Section 2.

Definition 7 (Oracle-output-unpredictability-2). *Fix a key space $\mathcal{K}$ and domain $\mathcal{D}$ and let $\Phi^{E,E^{-1}}$ be a set of RKD functions on the key-space $\mathcal{K}$. Let $\mathcal{P}_K(\cdot)$ and $\mathcal{X}(\cdot)$ be a pair of oracles for each E. The oracle $\mathcal{P}_K(\cdot)$ takes as input an element $\phi^{E,E^{-1}} \in \Phi^{E,E^{-1}}$ and the oracle $\mathcal{X}(\cdot)$ takes as input an element $K' \in \mathcal{K}$. Neither oracle returns a value. An adversary wins if it queries its $\mathcal{X}(\cdot)$ oracle with a key K' and if it queries its $\mathcal{P}_K(\cdot)$ oracle with a function $\phi^{E,E^{-1}}$ such that $\phi^{E,E^{-1}}(K) = K'$. We define the* oup2-advantage *of an adversary A as*

$$\mathbf{Adv}^{\mathsf{oup2}}_{\Phi^{E,E^{-1}},\mathcal{K},\mathcal{D}}(A) := \Pr\left[K \xleftarrow{\$} \mathcal{K}; E \xleftarrow{\$} \mathrm{Perm}(\mathcal{K},\mathcal{D}) : A^{\mathcal{P}_K(\cdot),\mathcal{X}(\cdot)} \textit{ wins}\right].$$

Definition 8 (Oracle-collision-resistance-2). *Fix a key space $\mathcal{K}$ and domain $\mathcal{D}$ and let $\Phi^{E,E^{-1}}$ be a set of functions on the key-space $\mathcal{K}$. Let $\mathcal{C}_K(\cdot)$ be an oracle for each E that takes as input a function $\phi^{E,E^{-1}} \in \Phi^{E,E^{-1}}$ and that returns no value. An adversary wins if it queries its oracle with two distinct functions $\phi_1^{E,E^{-1}}, \phi_2^{E,E^{-1}} \in \Phi^{E,E^{-1}}$ such that $\phi_1^{E,E^{-1}}(K) = \phi_2^{E,E^{-1}}(K)$. We define the* ocr2-advantage *of an adversary A as*

$$\mathbf{Adv}^{\mathsf{ocr2}}_{\Phi^{E,E^{-1}},\mathcal{K},\mathcal{D}}(A) := \Pr\left[K \xleftarrow{\$} \mathcal{K}; E \xleftarrow{\$} \mathrm{Perm}(\mathcal{K},\mathcal{D}) : A^{\mathcal{C}_K(\cdot)} \textit{ wins}\right].$$

Once again we note that oracle-output-unpredictability can be seen as oracle-collision-resistance between a constant and a non-constant function.

Our third definition provides a sufficient condition on a set of oracle RKD functions $\Phi^{E,E^{-1}}$ to enable an analogue of Theorem 1 to be proved. Intuitively speaking, if an oracle RKD set is *oracle-independent*, then no collisions can take place between the explicit queries made by an adversary to one of its three oracles, and those made implicitly through the oracle RKD function, during its attack.

Definition 9 (Oracle-independence). *Fix a key space $\mathcal{K}$ and domain $\mathcal{D}$ and let $\Phi^{E,E^{-1}}$ be a set of functions on the key-space $\mathcal{K}$. Let $\mathcal{Q}_K(\cdot,\cdot)$ be an oracle for each E that takes as input a function $\phi^{E,E^{-1}} \in \Phi^{E,E^{-1}} \cup \mathrm{K}$, where K is the set of constant functions, and an $x \in \mathcal{D}$ and returns no value. An adversary wins if it queries its oracle with two (not necessarily distinct) oracle RKD functions $\phi_1^{E,E^{-1}}$ and $\phi_2^{E,E^{-1}}$, and a point $x_1 \in \mathcal{D}$ such that*

$$(\phi_1^{E,E^{-1}}(K), x_1) \in \{(K', x') : \phi_2^{E,E^{-1}}(K) \textit{ queries } (K', x') \textit{ to } E \textit{ or } E^{-1}\}.$$

We define the oind-advantage *of an adversary A as*

$$\mathbf{Adv}^{\mathsf{oind}}_{\Phi^{E,E^{-1}},\mathcal{K},\mathcal{D}}(A) := \Pr\left[K \xleftarrow{\$} \mathcal{K}; E \xleftarrow{\$} \mathrm{Perm}(\mathcal{K},\mathcal{D}) : A^{\mathcal{Q}_K(\cdot,\cdot)} \textit{ wins}\right].$$

Similarly to the remark at the end of Section 2, simpler alternatives of the above definitions can be formulated. For convenience, we call an oracle RKD set which satisfies the above three requirements *valid*.

Let us now state and prove the main result of this section.

Theorem 4. *Fix a key space $\mathcal{K}$ and a domain $\mathcal{D}$, and let $\Phi^{E,E^{-1}}$ be a set of RKD functions over $\mathcal{K}$. Let A be an ideal-cipher-model adversary that queries its first two oracles with a total of at most q' different keys and that queries its last oracle with a total of at most q different RKD functions from $\Phi^{E,E^{-1}}$. Then there exists an oracle-output-unpredictability-2 adversary B, an oracle-collision-resistance-2 adversary C, and an oracle-independence adversary D such that*

$$\mathbf{Adv}^{\mathsf{prp\text{-}orka}}_{\Phi^{E,E^{-1}},\mathcal{K},\mathcal{D}}(A) \le \mathbf{Adv}^{\mathsf{ocr2}}_{\Phi^{E,E^{-1}},\mathcal{K},\mathcal{D}}(B) + \mathbf{Adv}^{\mathsf{oup2}}_{\Phi^{E,E^{-1}},\mathcal{K},\mathcal{D}}(C) + \mathbf{Adv}^{\mathsf{oind}}_{\Phi^{E,E^{-1}},\mathcal{K},\mathcal{D}}(D),$$

where B queries its $\mathcal{C}_K$ oracle q times, C queries its $\mathcal{P}_K$ and $\mathcal{X}$ oracles q and q' times (respectively), and D queries its $\mathcal{Q}_K$ oracle $q + q'$ times.

The intuition behind the proof of this theorem is similar to that for the proof of Theorem 1. The three conditions allow us to *separate* various oracle queries enabling us to simulate them by returning independently chosen random values. The output-unpredictability property is used to separate the ideal-cipher oracles from the related-key oracle. Collision-resistance is used to separate different ϕ's queried to the related-key oracle. The third condition is used in separating ϕ's oracles from those directly given to A; this was not necessary in the previous model when ϕ did not have access to any oracles. We now present the formal proof of Theorem 4.

Proof. Let A^{Game_L} denote the game in the left hand side of the oracle RKA advantage, namely,

$$K \xleftarrow{\$} \mathcal{K}; E \xleftarrow{\$} \mathrm{Perm}(\mathcal{K},\mathcal{D}) : A^{E(\cdot,\cdot),E^{-1}(\cdot,\cdot),E(\mathsf{RK}(\cdot,K),\cdot)},$$

and A^{Game_R} the game on the right hand side, i.e.

$$K \xleftarrow{\$} \mathcal{K}; E \xleftarrow{\$} \mathrm{Perm}(\mathcal{K},\mathcal{D}); G \xleftarrow{\$} \mathrm{Perm}(\mathcal{K},\mathcal{D}) : A^{E(\cdot,\cdot),E^{-1}(\cdot,\cdot),G(\mathsf{RK}(\cdot,K),\cdot)}.$$

Our goal is to show that the difference $\Pr\left[A^{\mathsf{Game}_L} = 1\right] - \Pr\left[A^{\mathsf{Game}_R} = 1\right]$ is small. Let us start by defining the following events:

- Event E_1: A triggers an oracle-collision-resistance-2 event, i.e. it submits two oracle RKD functions to its related-key oracle which are equal when evaluated on the key K.
- Event E_2: A triggers an oracle-output-unpredictability-2 event, i.e. it submits to one of its ideal-cipher oracles a key K' equal to the output of an oracle RKD function submitted to its related-key oracle.
- Event E_3: A triggers an oracle-independence event, i.e. it queries its related-key oracle with a first pair, and its ideal-cipher oracle or its related-key oracle with a second pair such that the first pair causes an oracle RKD function to query the ideal cipher on a point equal to that queried by the second pair.
- Event Bad: A triggers an oracle-collision-resistance-2, an oracle-output-unpredictability-2, or an oracle-independence event, i.e. $\mathsf{Bad} = \mathsf{E}_1 \vee \mathsf{E}_2 \vee \mathsf{E}_3$.

Before event Bad happens, Game_L and Game_R are identical as all oracles queries made by the adversary are independent of each other. More precisely, there are no collisions between two different ϕ's, nor between a key and a ϕ; additionally there are no dependencies between A's answers received from its oracle queries and those made implicitly by the submitted oracle RKD functions. Hence by the fundamental lemma of game-playing [4] we have that

$$\mathbf{Adv}^{\mathsf{prp\text{-}orka}}_{\Phi^{E,E^{-1}},\mathcal{K},\mathcal{D}}(A) \le \Pr[\mathsf{Bad}] .$$

In order to upper-bound this probability, we suppose that event Bad happens. This means that exactly one of the following mutually exclusive events takes place:

Event F_i for $i = 1, 2, 3$: A triggers event E_i before it does this for the other two remaining E-events.

This means that

$$\Pr[\mathsf{Bad}] = \Pr[\mathsf{F}_1] + \Pr[\mathsf{F}_2] + \Pr[\mathsf{F}_3] .$$

We upper-bound this sum by relating each of its terms to the $\mathsf{ocr2}$, $\mathsf{oup2}$, and oind advantages of the following adversaries, respectively.

- An oracle-collision-resistance-2 adversary B which submits A's related-key queries to its $\mathcal{C}_K(\cdot)$ oracle.
- An oracle-output-unpredictability-2 adversary C which submits A's ideal-cipher queries to its $\mathcal{X}(\cdot)$ oracle, and A's related-key queries to its $\mathcal{P}_K(\cdot)$ oracle.
- An oracle-independence adversary D which submits A's ideal-cipher queries as well as related-key queries to its $\mathcal{Q}_K(\cdot, \cdot)$ oracle.

Furthermore, the adversaries B, C and D, for each of the keys K' with which A queries its ideal-cipher oracle (in either the forward or backward directions), reply to A's queries using an independently selected random permutation (or its inverse as appropriate). They do so by picking and returning random points in $\mathcal{D}$ subject to the constraint that they simulate a permutation. This is easily done by maintaining lists of inputs and outputs for each K' in the usual way.

It is clear that if event F_1 happens, then adversary B wins the oracle-collision-resistance-2 game (and similarly for events F_2 and F_3, and adversaries C and D). Furthermore, before event F_1 happens (respectively event F_2 and event F_3), none of the events E_1, E_2, or E_3 happen, and B's (respectively C's and D's) simulation above is consistent with the view of A in Game_L (and Game_R). This is because of the same reasons as those discussed above: there are no collisions between two different ϕ's, nor between a key and a ϕ; additionally there are no dependencies between A's answers received from its oracle queries and those made implicitly by the submitted oracle RKD functions. We conclude that

$$\Pr[\mathsf{F}_1] = \mathbf{Adv}^{\mathsf{ocr2}}_{\Phi^{E,E^{-1}},\mathcal{K},\mathcal{D}}(B) , \Pr[\mathsf{F}_2] = \mathbf{Adv}^{\mathsf{oup2}}_{\Phi^{E,E^{-1}},\mathcal{K},\mathcal{D}}(C) , \text{ and } \Pr[\mathsf{F}_3] = \mathbf{Adv}^{\mathsf{oind}}_{\Phi^{E,E^{-1}},\mathcal{K},\mathcal{D}}(D) .$$

The theorem follows. □

Note that if a set of RKD functions does not make any oracle calls to E or E^{-1} then the set automatically satisfies the oracle-independence criterion (with advantage 0). The oracle-collision-resistance-2 and oracle-output-unpredictability-2 conditions are identical to collision-resistance-2 and output-unpredictability-2 conditions of Bellare and Kohno respectively, and we also recover Theorem 1 in this case.

Let us now check why the attacks of Bernstein and Harris fail to satisfy the conditions required in Theorem 4 for an ideal cipher to be resistant to oracle related-key attacks. For convenience, we have included a slightly modified and improved version of Harris's attack in Appendix C.

Theorem 5. *Let Δ^E and Ψ_i^E denote Bernstein's and Harris's set of oracle RKD functions, respectively. Then the oracle RKD sets Δ^E and Ψ_i^E do not satisfy the oracle-independence property.*

Proof. In Bernstein's attack, the function $K \mapsto E(K, 0)$ queries E on K, as does the related-key oracle when queried on the identity function. For Harris's attack, note that $H_{i,p}^E(K)$ queries E on $K \oplus p$ and $K \oplus p \oplus [1^k]_i$ and then, to compute the actual value of the related-key oracle when queried with this function, once again E is queried on one of these values. □

So far we concentrated on ruling out attacks, and have not demonstrated how our choice of modelling can be used in a positive way. In other words, could it be the case that any non-trivial access to E or E^{-1} violates one of the three needed properties, rendering Theorem 4 meaningless. Fortunately, this is not the case. Our next two results demonstrate how one can model *new* cipher-dependent RKD functions which do not compromise security. The next theorem considers an RKD set from the EMV specification [14], and is proved in Appendix D.

Theorem 6. *Fix a key space $\mathcal{K}$ and let $\mathcal{D} = \mathcal{K}$. Define*

$$\Omega^E := \{K \mapsto E(K, x) : x \in \mathcal{D}\}.$$

Then for any adversary ocr2 *adversary A, any* oup2 *adversary B making at most q and q' queries to its $\mathcal{P}_K$ and $\mathcal{X}$ oracles (respectively), and any* oind *adversary C making at most q queries to its $\mathcal{Q}_K$ oracle, we have that*

$$\mathbf{Adv}^{\mathsf{ocr2}}_{\Omega^E,\mathcal{K},\mathcal{D}}(A) = 0, \mathbf{Adv}^{\mathsf{oup2}}_{\Omega^E,\mathcal{K},\mathcal{D}}(B) \leq qq'/(2|\mathcal{K}|), \text{ and } \mathbf{Adv}^{\mathsf{oind}}_{\Omega^E,\mathcal{K},\mathcal{D}}(C) \leq q^2/(2|\mathcal{K}|).$$

The next theorem provides a possibility result in a scenario where the adversary has access to the identity function as well as other RKD functions.

Theorem 7. *Fix a key space $\mathcal{K}$ and let $\mathcal{D} = \mathcal{K}$. Define*

$$\Theta^E := \{K \mapsto K, K \mapsto E(K', K) : K' \in \mathcal{K}\}.$$

Then for any adversary ocr2 *adversary A making at most q queries to its $\mathcal{C}_K$ oracle, any* oup2 *adversary B making at most q and q' queries to its $\mathcal{P}_K$ and $\mathcal{X}$ oracles (respectively), and any* oind *adversary C making at most q queries to its $\mathcal{Q}_K$ oracle, we have that*

$$\mathbf{Adv}^{\mathsf{ocr2}}_{\Theta^E,\mathcal{K},\mathcal{D}}(A) \leq q^2/(2|\mathcal{K}|), \mathbf{Adv}^{\mathsf{oup2}}_{\Theta^E,\mathcal{K},\mathcal{D}}(B) \leq qq'/(2|\mathcal{K}|), \text{ and } \mathbf{Adv}^{\mathsf{oind}}_{\Theta^E,\mathcal{K},\mathcal{D}}(C) \leq q^2/(2|\mathcal{K}|).$$

The proof of this theorem is presented in Appendix E.

5 Interpretations in the standard model

Standard model interpretations of cryptographic results in an idealised model have always existed in the research community. The random oracle model and its real-world interpretations [12] provide a good example of the difficulties involved in attempting such translations. Another example is a result of Black [10], which gives a hash function construction provably secure in the ideal-cipher model, but insecure if the ideal cipher is instantiated with any concrete blockcipher. The result of [10] holds under related-key attacks as long as the RKD set under consideration contains the identity function (as in this case prp-rka security is at least as strong as the standard notion of prp security). This separation result, although theoretically valid, is *unnatural* as it is unlikely that a real-world hash function depends on the code of a blockcipher in the same artificial way as that used to derive the result in [10].

On the other hand, as shown in Theorem 2, in the related-key attack model of Bellare and Kohno, a *natural* (weak) separation result exists. This possibility seems to have been over-looked by the authors of [3], who did not discuss interpretations of their main result, Theorem 1, in the standard model. As pointed out in Section 3, this theorem can be interpreted in two different ways: the first interpretation, which was argued to be a more accurate translation, lacked a natural separation result, but was of smaller relevance to practice; the second interpretation, on the other hand, although relevant to practice, was shown to be invalid in Theorem 2.

Theorem 4 is an attempt to overcome the limitations of Theorem 1: oracle RKD functions enable modelling RKD functions which might depend on the cipher. For consistency and completeness, we should also investigate possible interpretations of this result in the standard model.

The first issue in interpreting this result arises when one attempts to relate an oracle RKD set $\Phi^{E,E^{-1}}$ to a concrete RKD set in the standard model. Our theorem concerns oracle RKD functions, that is RKD functions which use a blockcipher in a black-box way. Its relevance to the standard model is therefore restricted to RKD functions which use the cipher in a black-box (or a symbolic) way. Hence, given such an RKD set $\Phi_{E^\star}$ making subroutine calls to $E^\star$ and $E^{\star-1}$, one can rewrite it in the form of a natural oracle RKD set $\Phi^{E,E^{-1}}$, such that if it is instantiated at $E^\star$ (i.e. the oracle calls to E and E^{-1} are replaced with subroutine calls to $E^\star$ and $E^{\star-1}$), one recovers the original RKD set $\Phi_{E^\star}$.

Next, the validity of $\Phi_{E^\star}$ should be interpreted in terms of validity of $\Phi^{E,E^{-1}}$, as these sets are no longer the same (Note that this issue did not exist in interpreting Theorem 1 as the sets were identical). The minimum requirement on the set $\Phi_{E^\star}$ is that the associated oracle RKD set is valid, i.e. it satisfies the oracle-collision-resistant-2, oracle-output-unpredictable-2, and oracle-independent conditions. We additionally require that $\Phi^{E,E^{-1}}$ satisfies these conditions if E is no longer sampled uniformly at random in the games but is fixed to be $E^\star$. This latter condition is due to the fact that validity of the set for a random E is not enough to guarantee that at a specific $E^\star$ the set is also "reasonable".

We are now ready to present interpretations of our theorem that are analogous to those of Theorem 1. In the following, we let $\Phi^{E,E^{-1}}$ and $\Phi_{E^\star}$ be a pair of associated sets as discussed above. Our two interpretations, as before, concern the choice of order of quantifiers:

1. For all valid $\Phi^{E,E^{-1}}$, there exists a concrete blockcipher $E^\star$ which resists $\Phi_{E^\star}$-restricted attacks if $\Phi^{E,E^{-1}}$ is also valid at $E = E^\star$.
2. There is a concrete blockcipher $E^\star$, such that for all valid $\Phi^{E,E^{-1}}$, $E^\star$ resists $\Phi_{E^\star}$-restricted attacks if $\Phi^{E,E^{-1}}$ is also valid at $E = E^\star$.

In attempting to derive counterexamples to the above two interpretations a similar, but higher level, line of argument to that given for Theorem 1 applies. In Theorem 4, the strategy of dependence on E is fixed as E is chosen randomly and independently of $\Phi^{E,E^{-1}}$. In other words, for each E, each RKD function depends on E in the same way. This is exactly what is expressed by the first interpretation, and hence as in Theorem 1, we take this choice of order of quantifiers to be a more accurate interpretation. What is important here is that *unlike the first interpretation of the theorem of Bellare and Kohno, the RKD functions here may depend on* E, and hence this theorem has a greater relevance to practice than that provided by Theorem 1. As in Theorem 1, we do not expect there to be natural counterexamples to this interpretation.

Let us turn to the second interpretation. Due to the reversed order of quantifiers the strategy of dependence in a counterexample may itself depend on each cipher $E^\star$. In fact, Bernstein's attack still constitutes a counterexample to the second interpretation of Theorem 4, if one chooses the oracle RKD set to be identical to the concrete RKD set for each $E^\star$, i.e.

$$\Delta_{E^\star} := \{K \mapsto K, K \mapsto E^\star(K, 0)\} \text{ and } \Delta^E_{E^\star} := \{K \mapsto K, K \mapsto E^\star(K, 0)\}.$$

Note however that, as pointed out before, the dependency of $\Delta^{E}_{E^\star}$ on $E^\star$ is black box. This oracle RKD set may then be rewritten symbolically as

$$\Delta^E := \{K \mapsto K, K \mapsto E(K, 0)\},$$

and as we saw in Theorem 5, this set is not valid. A similar observation applies to Harris's RKD set. In general this dependency in a (natural) counterexample is likely to be black box, and the functions can be rewritten as an oracle RKD set with a *fixed* dependence strategy. This in turn would either constitute a counterexample to the first interpretation, which we have assumed to be unlikely, or the resulting new oracle RKD set will be invalid. On the other hand, a non-black-box dependency seems difficult to achieve. In conclusion, the second interpretation, for practical purposes, is the same as the first one.

Turning to positive results, Theorems 6 and 7 can be interpreted in the standard model in the following way. It is a "reasonable" goal to design a blockcipher which resist Ω_E- and Θ_E-restricted related-key attacks where

$$\Omega_E := \{K \mapsto E(K, x) : x \in \mathcal{D}\} \text{ and } \Theta_E := \{K \mapsto K, K \mapsto E(K', K) : K' \in \mathcal{K}\}$$

are respectively the RKD sets associated to Ω^E and Θ^E as defined in Theorems 6 and 7. These results may have applications in establishing the security of key hierarchies which use the cipher to derive new keys.

Let us look at the first standard model interpretations of Theorem 1 and Theorem 4 from a cryptanalytic perspective. Theorem 1 classifies a blockcipher E as broken if there exists a collision-resistance-2 and output-unpredictable-2 RKD set which can be used to break E in the related-key attack model *and furthermore* this set does not depend on E. This theorem provides no answers for RKD sets, such as Bernstein's set or that given in Theorem 7, which depend on E. Theorem 4, on the other hand, allowed dependency on E at the expense of an extra condition. This theorem classified a blockcipher E as broken in two cases: 1) the attack is independent of E and we are back at the conditions of Theorem 1; or 2) the attack is dependent on E in a black-box way, and the associated oracle RKD set is valid for a random E and also at $E^\star$.

According to the above cryptanalytic perspective, Bernstein's and Harris's attacks should not be seen as harmful. Attacks using RKD sets which involve a cipher's building blocks demonstrated by Biryukov et al. [8,9] on AES raise the following question: can Biryukov et al.'s set of RKD functions be simulated using calls to the full encryption and decryption routines of AES? If this is not the case, or if this is the case and the resulting oracle RKD set is valid, then the related-key attack against AES should be seen as interesting. Formalising such a natural dependency remains an open problem, and hence, in our opinion, Biryukov et al.'s attack should then be seen as a threat against AES in the related-key attack model (assuming the relevant RKD functions are available to the cryptanalyst). We note that our model might be further extended to consider RKD sets with oracle access to round functions so as to model Biryukov et al.'s results which exploit relations of this type.

REMARK. The requirements of Theorem 4 constitute a set of sufficient conditions for an ideal cipher to be secure in the prp-orka sense. These conditions, however, are quite strong and one might alternatively directly prove that an ideal cipher is prp-orka secure.[8] Validity at $E^\star$ now means resilience of an ideal cipher to related-key attacks when the oracle RKD set is instantiated with $E^\star$. Such proofs can then be used to conjecture the existence of a blockcipher resisting related-key attacks under the associated RKD set. From a cryptanalytic perspective, if these proofs exists, and the associated RKD set breaks a specific cipher, then this cipher should be seen as broken. Conversely, if an oracle RKD set can be used to break an ideal cipher, then the associated RKD set should not be seen as valid. This in turn means that one should neither expect there to be a blockcipher which resists such attacks, nor should such an attack be seen as harmful. These observations also apply to the two conditions used in Theorem 1.

IMPLICATIONS FOR PRACTICE. As well as considering standard-model interpretations of our main result, we also wish to reflect on what our results might mean for practice. Suppose we have an RKD set that is

[8] This is the case for functions expressed in a contrived way such as $K \mapsto K \oplus 1 = K \oplus E(K, E^{-1}(K, 1))$ or $K \mapsto K \oplus E(0, 0)$.

*in*valid for a blockcipher, so that the conditions of our main theorem are not met. Does this mean that there must be a related-key attack against the blockcipher? The answer is clearly no, since the possibility of a related-key attack depends on exactly how the blockcipher is used as a component in an overall system or protocol: if that environment does not make available to the attacker the relevant RK oracles, then the related-key attack will not be mountable in practice. A recent example of an interesting related-key attack which is not mountable in practice as is would be the attack of Dunkelman et al. [13] against KASUMI when used in the 3G network. On the other hand, even if we have an RKD set that is valid, then we can still not rule out related-key attacks altogether, because of the gap that exists between the ideal-cipher model and the standard model. Finally, we ask: what should a protocol designer do? The simple answer is to avoid the use of related keys in protocols altogether. If this is not possible, then our best advice is to only use related-keys in such a way that the relevant RKD set satisfies the conditions of Theorem 4. For example, the sets of functions exhibited in Theorems 6 and 7 would be suitable in this case.

Acknowledgements

The authors would like to thank Thomas Shrimpton and Martijn Stam for their detailed and insightful comments on earlier drafts of this paper. The work described in this paper has in part been supported by the Commission of the European Communities through the ICT program under contract ICT-2007-216676 (ECRYPT-II). M.R. Albrecht was supported by the Royal Holloway Valerie Myerscough Scholarship. The work of K.G. Paterson was supported by an EPSRC Leadership Fellowship, EP/H005455/1.

References

1. B. Applebaum, D. Harnik, and Y. Ishai: Semantic Security Under Related-Key Attacks and Applications. In *Cryptology ePrint Archive*, Report 2010/544, 2010.
2. M. Bellare and D. Cash: Pseudorandom Functions and Permutations Provably Secure Against Related-Key Attacks. In *Advances in Cryptology – CRYPTO 2010*, LNCS 6223:666–684, Springer, 2010.
3. M. Bellare and T. Kohno: A Theoretical Treatment of Related-Key Attacks: RKA-PRPs, RKA-PRFs, and Applications. In *Advances in Cryptology – EUROCRYPT 2003*, LNCS 2656:491–506, Springer, 2003.
4. M. Bellare and P. Rogaway: The Security of Triple Encryption and a Framework for Code-Based Game-Playing Proofs. In *Advances in Cryptology – EUROCRYPT 2004*, LNCS 4004:409–426, Springer, 2006.
5. D.J. Bernstein: E-mail Discussion among the Participants of the Early Symmetric Crypto Seminar 2010.
6. E. Biham: New Types of Cryptoanalytic Attacks Using Related Keys (Extended Abstract). In *Advances in Cryptology – EUROCRYPT '93*, LNCS 765:398–409, Springer, 1993.
7. E. Biham: New Types of Cryptoanalytic Attacks Using Related Keys. *Journal of Cryptology*, 7(4):229–246, 1994.
8. A. Biryukov and D. Khovratovich: Related-Key Cryptanalysis of the Full AES-192 and AES-256. In *Advances in Cryptology – ASIACRYPT 2009*, LNCS 5912:1–18, Springer, 2009.
9. A. Biryukov, D. Khovratovich, and I. Nikolic: Distinguisher and Related-Key Attack on the Full AES-256. In *Advances in Cryptology – CRYPTO 2009*, LNCS 5677:231–249, Springer, 2009.
10. J. Black: The Ideal-Cipher Model, Revisited: An Uninstantiable Blockcipher-Based Hash Function. In *Fast Software Encryption*, LNCS 4047:328–340, Springer, 2006.
11. J. Black, P. Rogaway, and T. Shrimpton: Encryption-Scheme Security in the Presence of Key-Dependent Messages. In *Selected Areas in Cryptography*, LNCS 2595:62–75, Springer, 2002.
12. R. Canetti, O. Goldreich, and S. Halevi: The Random Oracle Methodology, Revisited. *JACM*, 51(4): 557–594, 2004.
13. O. Dunkelman, N. Keller, and A. Shamir: A Practical-Time Attack on the A5/3 Cryptosystem Used in Third Generation GSM Telephony. In *Cryptology ePrint Archive*, Report 2010/013, 2010.
14. EMV Integrated Circuit Card Specifications for Payment Systems, Book 2 Security and Key Management, Version 4.2, June 2008.
15. D. Goldenberg and M. Liskov: On Related-Secret Pseudorandomness. In *Theory of Cryptography*, LNCS 5978:255–272, Springer, 2010.
16. S. Halevi and H. Krawczyk: Security under Key-Dependent Inputs. In *ACM Conference on Computer and Communications Security, CCS 2007*, pages 466–475, ACM, 2007.
17. D.G. Harris: Generic Ciphers are More Vulnerable to Related-Key Attacks than Previously Thought. In *WCC 2009*, 2009.
18. T. Iwata and T. Kohno: New Security Proofs for the 3GPP Confidentiality and Integrity Algorithms. In *Fast Software Encryption*, LNCS 3017:427–445, Springer, 2004.
19. L.R. Knudsen: Cryptanalysis of LOKI91. In *Advances in Cryptology – AUSCRYPT '92*, LNCS 718:196–208, Springer, 1992.
20. S. Lucks: Ciphers Secure against Related-Key Attacks. In *Fast Software Encryption*, LNCS 3017:359–370, Springer, 2004.
21. E. Razali, R.C.-W. Phan, and M. Joye: On the Notions of PRP-RKA, KR and KR-RKA for Block Ciphers. In *Provable Security*, LNCS 4784:188–197, Springer, 2007.

A Proof of Theorem 2

Proof. Adversary $A^{f(\mathsf{RK}(\cdot,K),\cdot)}$, where $f = E$ or $f = G$, operates as shown in Figure 2.

```
Algorithm $A^f$:
Query RK oracle on $(K \mapsto K, 0)$ to get $x = f(K, 0)$
Query RK oracle on $(K \mapsto E(K, 0), 0)$ to get $y = f(E(K, 0), 0)$
Calculate $z = E(x, 0)$
Return $(z = y)$
```

Fig. 2: Δ_E-restricted adversary breaking a blockcipher with $\mathcal{K} = \mathcal{D}$.

When $f = E$, we have that $x = E(K, 0)$, $y = E(E(K, 0), 0)$, and $z = E(E(K, 0), 0)$. Hence $z = y$ with probability 1. On the other hand, if $f = G$, then $x = G(K, 0)$, $y = G(E(K, 0), 0)$, and $z = E(G(K, 0), 0)$. Let

$$\mathsf{Event} := E(G(K,0),0) = G(E(K,0),0) \text{ and } \delta := \Pr\left[K \overset{\$}{\leftarrow} \mathcal{K} : E(K,0) = K\right].$$

From the proof of the second part of the theorem we have

$$\delta \leq \mathbf{Adv}_E^{\mathsf{prp}}(B) + 1/|\mathcal{K}|,$$

where algorithm B is described in Figure 3. Hence

$$\begin{aligned}
\Pr[\mathsf{Event}] &= \Pr[\mathsf{Event}|E(K,0) = K] \cdot \delta + \Pr[\mathsf{Event}|E(K,0) \neq K] \cdot (1-\delta) \\
&= \Pr\left[t \overset{\$}{\leftarrow} \mathcal{D} : E(t,0) = t\right] \cdot \delta + \Pr\left[z \overset{\$}{\leftarrow} \mathcal{D}; x \overset{\$}{\leftarrow} \mathcal{D} \setminus z : E(x,0) = z\right] \cdot (1-\delta) \\
&= \delta^2 + (1-\delta)/|\mathcal{K}| \leq \delta + 1/|\mathcal{K}| \leq \mathbf{Adv}_E^{\mathsf{prp}}(B) + 2/|\mathcal{K}|.
\end{aligned}$$

This proves the first part of the theorem.

We now prove the second part of the theorem. Let us first consider output-unpredictability-2, and let X be the set of all queries made to the $\mathcal{X}$ oracle, where $|X| = q'$. We consider two cases: 1) the identity function results in the adversary winning the output-unpredictability-2 game; and 2) the RKD function $K \mapsto E(K, 0)$ results in the adversary winning the output-unpredictability-2 game. In the first case the up2-advantage is at most $q'/|\mathcal{K}|$ as the identity map is a permutation [3]. For the second case, we note that the map is no longer a permutation. In this case, however, we bound the advantage of A by relating it to the prp-advantage of adversary B^f shown in Figure 3, where f is either $E(K, \cdot)$ or $G(\cdot)$.

```
Algorithm $B^f$:
Construct $X$, a list of $A$'s $\mathcal{X}$ queries
Query $f$ on 0 to get $x = f(0)$
Return $(x \in X)$
```

Fig. 3: Adversary distinguishing a blockcipher from an ideal cipher.

It is easily seen that

$$\begin{aligned}
\Pr\left[B^{E(K,\cdot)} = 1\right] &= \Pr\left[K \overset{\$}{\leftarrow} \mathcal{K} : E(K,0) \in X\right] \quad \text{and} \\
\Pr\left[B^{G(\cdot)} = 1\right] &= \Pr\left[G \overset{\$}{\leftarrow} \mathrm{Perm}(\mathcal{D}) : G(0) \in X\right] = |X|/|\mathcal{K}|.
\end{aligned}$$

Therefore

$$\Pr\left[K \overset{\$}{\leftarrow} \mathcal{K} : E(K,0) \in X\right] = \mathbf{Adv}_E^{\mathsf{prp}}(B) + |X|/|\mathcal{K}|.$$

The result follows by applying the union bound to the probabilities in the above two cases.

We now consider collision-resistance-2. Since Δ_E contains only two functions, we need only bound the probability that for a random K we have that $E(K, 0) = K$. This in turn will provide a bound on $\mathbf{Adv}^{\mathsf{cr2}}_{\Delta_E}(A)$ for any adversary A. To do so, we consider the prp adversary B as shown in Figure 4. Intuitively, this adversary first attempts to recover the key K by asking for the encryption of 0, and then checks its validity through a test encryption.

```
Algorithm $B^f$:
Query $f$ on 0 to get $x = f(0)$
Query $f$ on 1 to get $z = f(1)$
Calculate $y = E(x, 1)$
Return $(y = z)$
```

Fig. 4: Adversary distinguishing a concrete blockcipher from an ideal cipher using when $K = E(K, 0)$.

Note that whenever $E(K, 0) = K$, it is also necessarily the case that $E(E(K, 0), 1) = E(K, 1)$. Therefore by a conditional probability argument we get that

$$\Pr\left[B^{E(K,\cdot)}=1\right] \geq \Pr\left[K \xleftarrow{\$} \mathcal{K} : E(K,0)=K\right].$$

On the other hand

$$\Pr\left[B^{G(\cdot)} = 1\right] = \Pr\left[G \xleftarrow{\$} \mathrm{Perm}(\mathcal{D}) : E(G(0), 1) = G(1)\right] = 1/|\mathcal{K}|.$$

It follows that

$$\Pr\left[K \xleftarrow{\$} \mathcal{K} : E(K, 0) = K\right] \leq \mathbf{Adv}^{\mathsf{prp}}_{E}(B) + 1/|\mathcal{K}|.$$

□

B Key recovery under related-key attacks

Most concrete related-key attacks [8,9,13] lead to recovery of a blockcipher's key. Hence an alternative way to restore confidence in the BK model would be to raise the security bar for an attacker, and require it to recover keys. In this section, we investigate this notion.

Let us start with key-recovery without related-key attacks.

Definition 10 (Key recovery). *Let $E : \mathcal{K} \times \mathcal{D} \to \mathcal{D}$ be a family of functions. Let A be an adversary. Then*

$$\mathbf{Adv}^{\mathsf{kr}}_{E}(A) := \Pr\left[K \xleftarrow{\$} \mathcal{K} : A^{E(K,\cdot)} = K\right]$$

is defined as the kr-advantage *of A against E.*

It is easily seen that prp security implies security in the above sense: the recovered key can be used to distinguish the cipher from a random blockcipher through a test encryption (only in the unlikely event that the two outputs match does this test fail).

We extend the above definition to related-key attacks following the approach of [3]. Key-recovery attacks in this setting were first introduced in [21]. Here we present a different, more natural variant.

Definition 11 (Key recovery with respect to related-key attacks). *Let $E : \mathcal{K} \times \mathcal{D} \to \mathcal{D}$ be a family of functions and let Φ be a set of RKD functions over $\mathcal{K}$. Let A be an adversary with access to a related-key oracle, and restricted to queries of the form (ϕ, x) in which $\phi \in \Phi$ and $x \in \mathcal{D}$. Then*

$$\mathbf{Adv}^{\mathsf{kr\text{-}rka}}_{\Phi,E}(A) := \Pr\left[K \xleftarrow{\$} \mathcal{K} : A^{E(\mathsf{RK}(\cdot,K),\cdot)} = K\right]$$

is defined as the kr-rka-advantage *of A in a Φ-restricted related-key attack on E.*

It is easy to see that prp-rka security (in the sense of Definition 2) implies kr-rka security in the above sense. This implication still holds in the ideal-cipher model (once an ICM version of Definition 11 is formulated), and hence collision-resistance-2 and output-unpredictability-2 conditions constitute a set of sufficient conditions for resistance against key-recovery in the ideal-cipher model. These conditions may well be stronger than needed. In fact, note that the set of constant functions should not help a key-recovery attack to succeed. We therefore provide the following simple, but useful, tool which allows us to extend the set of allowed related-key-deriving functions in key-recovery attacks.

Theorem 8. *Let $E : \mathcal{K} \times \mathcal{D} \to \mathcal{D}$ be a family of functions and let Φ and Φ' be two disjoint sets of RKD functions over $\mathcal{K}$. Suppose that E is secure against Φ-restricted kr-rka attacks. Suppose further that for any $\phi' \in \Phi'$ there is an oracle Turing machine $F_{\phi'}$ such that for any $K \in \mathcal{K}$ we have*[9]

$$F_{\phi'}^{E(\mathsf{RK}(\cdot,K),\cdot)} = \phi'(K),$$

where $F_{\phi'}$ queries its related-key oracle on pairs (ϕ, x) with $\phi \in \Phi$ and $x \in \mathcal{D}$. Then for any $\Phi \cup \Phi'$-restricted adversary A, there is a Φ-restricted adversary B such that

$$\mathbf{Adv}^{\mathsf{kr\text{-}rka}}_{\Phi\cup\Phi',E}(A) \leq \mathbf{Adv}^{\mathsf{kr\text{-}rka}}_{\Phi,E}(B)\,.$$

An analogous result holds in the ideal-cipher model (with $F_{\phi'}$, A, and B having access to E and E^{-1} oracles).

Proof (Sketch). Adversary B is able to perfectly simulate the environment for A: related key oracles with functions in Φ are answered using B's own oracle, whereas those in Φ' are simulated using $F_{\phi'}$. Figure 5 describes this procedure in more detail.

```
Algorithm $B^{E(\mathsf{RK}(\cdot,K),\cdot)}$:
Run $A$ answering its queries as follows.
    On RK query $(\phi, M)$, with $\phi \in \Phi$,
        Query the RK oracle on $(\phi, M)$ and return the answer
    On RK query $(\phi', M)$, with $\phi' \in \Phi'$,
        Run $F_{\phi'}$
            Answer $F_{\phi'}$'s RK queries using provided RK oracle
        Return the output of $F_{\phi'}$
Eventually $A$ returns $K'$
Return $K'$
```

Fig. 5: Φ-restricted adversary based on a $\Phi \cup \Phi'$-restricted adversary.

It is clear from the above code that B perfectly simulates the environment for A. Furthermore, whenever A is successful, so is B. The first part of the theorem follows.

The proof in the ideal-cipher model is analogous and is omitted. □

As an example, note that such an F_{const} for each constant function trivially exists.

The above theorem provides a simple proof that Bernstein's RKD set Δ_E cannot be used to mount a key-recovery attack against a prp-secure blockcipher. This follows from the observation that the RKD function $K \mapsto E(K, x)$ for any $x \in \mathcal{D}$ can be simulated using oracle access to $E(K, \cdot)$ available in the prp security game.

One might hope that if a key-recovery analogue of Theorem 1 were to be formulated (in the ideal-cipher model), an analogous second interpretation of it (see Section 4) might hold in moving to the standard model. In Appendix C we formalise Harris's attack [17] to show that such an RKD set invalidating the second

[9] The Turing machines $F_{\phi'}$ are similar to key-transformers introduced recently in [2].

interpretation does indeed exist. However, Harris's original key-recovery attack uses a set which is not quite collision-resistant-2. By rigorously analysing this attack, we are able to slightly tweak Harris's RKD set such that the resulting set has the collision-resistant-2 property.[10] We therefore conclude that the second interpretation in this context also fails, and the BK model cannot be rescued by raising the adversarial goal to key-recovery.

Note also that security against key-recovery related-key attacks can be formulated for oracle RKD sets analogously to Definition 6. An oracle-equipped version of Theorem 8 can also be proved. We omit the details.

C Harris's attack

In this section we formally analyse Harris's attack [17]. Fix a key space $\mathcal{K}$ and let $\mathcal{D} = \mathcal{K}$. We define a set of oracle RKD functions as follows:

$$H^E_{i,p}(K) := \begin{cases} K \oplus p, & \text{if } [E(K \oplus p, p)]_i = [E(K \oplus p \oplus [1^k]_i, p)]_i \,; \\ K \oplus p, & \text{if } [E(K \oplus p, p)]_i = [K]_i \,; \\ K \oplus p \oplus [1^k]_i, & \text{if } [E(K \oplus p \oplus [1^k]_i, p)]_i = [K]_i \,. \end{cases}$$

Here 1^k denotes a string of k 1's, and $[x]_i$ denotes the all-zero k-bit string with the i-th bit set to the i-th bit of x. We now define the following oracle RKD sets:

$$\Psi^E_i := \{H^E_{i,p} : p \in \mathcal{D} \wedge [p]_i = [1^k]_i\} \text{ and } \Psi^E := \bigcup_{i=1}^{k} \Psi^E_i.$$

Note that each function as defined above has an inverse, and hence [3] is oracle-output-unpredictable-2:

$$H^{-1}{}^E_{i,p}(K) := \begin{cases} K \oplus p, \text{ if } [E(K \oplus p, p)]_i = [E(K \oplus p \oplus [1^k]_i, p)]_i \,; \\ K \oplus p \text{ with } i\text{-th bit set to } i\text{-th bit of } E(K,p), \text{ otherwise.} \end{cases}$$

We now check oracle-collision-resistance-2 for the set Ψ^E_i. If $H^E_{i,p}(K) = H^E_{i,p'}(K)$ then either we have that (without loss of generality) $K \oplus p = K \oplus p'$, in which case $p = p'$ and the maps are the same, or that $K \oplus p = K \oplus p' \oplus [1^k]_i$, in which case p and p' are identical except for the i-th bit. But since the i-th bit of p is 1 for all allowed p, this is not possible.

The set Ψ^E is also oracle-output-unpredictable-2 as it consists of permutations. The set, however, is not oracle-collision-resistance-2. To see this, take $i \neq j$, and define a p such that i-th and j-th bits of p are both 1. Then with probability at least $1/4$ the first if-statements for both $H^E_{i,p}$ and $H^E_{j,p}$ are satisfied, and we have that $H^E_{i,p}(K) = K \oplus p = H^E_{j,p}(K)$.

To solve this problem, we slightly modify Ψ^E to a new oracle RKD set $\tilde{\Psi}^E$ as follows. Suppose the cipher has a k-bit key-space and an n-bit domain with $k|2^{n-1}$. We partition the domain into k components $\mathcal{D}_i$ such that if $p \in \mathcal{D}_i$ then so is $p \oplus [1^k]_i$. Note that each component has size $2^{n-1}/k$. We now allow p in $H^E_{i,p}$ to be chosen only from $\mathcal{D}_i$. This would remove collisions as if $i \neq j$ then there is no way to choose a p which ensures $H^E_{i,p}(K) = K \oplus p = H^E_{j,p}(K)$, nor is there a way to choose p and q such that $K \oplus p = K \oplus q \oplus [1^k]_j$, as $q \oplus [1^k]_j$ belongs to the same partition as that which q belongs to.

The following theorem formalises Harris's attack [17] for oracle RKD sets in the ideal-cipher model.

Theorem 9. *Fix a key space $\mathcal{K}$ and let $\mathcal{D} = \mathcal{K}$. Then for any i, there exists a Ψ^E_i-restricted adversary A such that*

$$\mathbf{Adv}^{\mathsf{prp\text{-}orka}}_{\Psi^E_i, \mathcal{K}, \mathcal{D}}(A) = 1/8.$$

[10] Note that this example provides an even stronger failure of the second interpretation: an ideal cipher is prp-rka secure against Harris's RKD set, whereas a concrete blockcipher is not even kr-rka secure.

Furthermore, there is a $\tilde{\Psi}^E$-restricted adversary A such that

$$\mathbf{Adv}^{\mathsf{kr\text{-}orka}}_{\tilde{\Psi}^E,\mathcal{K},\mathcal{D}}(A) \geq (1 - \exp(-\ell/32))^k$$

for any $\ell < 2^{n-1}/k$.

Proof. We construct an adversary A as shown in Figure 6.

```
Algorithm $A^{E,E^{-1},f(\mathsf{RK},\cdot)}$:
$p \leftarrow 1^k$; $q \leftarrow [1^k]_i$ // distinct plaintexts with $i$-th bit equal to 1
Query $f(\mathsf{RK}(H_{i,p}, K), p)$ to get $x$
Query $f(\mathsf{RK}(H_{i,q}, K), q)$ to get $y$
Return $([x]_i = [y]_i)$
```

Fig. 6: Ψ_i^E-restricted adversary in the ideal-cipher model.

In the following, we drop $[\cdot]_i$ for readability and assume all variables are considered as their i-th bit.

$$\begin{aligned}
\Pr\left[A^{E,E^{-1},E(\mathsf{RK},\cdot)} = 1\right] &= \Pr[x = y] \\
&= \Pr[x = K | y = K] \Pr[y = K] + \Pr\left[x = 1^k \oplus K | y = 1^k \oplus K\right] \Pr\left[y = 1^k \oplus K\right] \\
&= 3/4 \cdot 3/4 + 1/4 \cdot 1/4 = 10/16.
\end{aligned}$$

Here we note that $\Pr[y = K] = 3/4$ as the first if-statement in definition of H holds with probability $1/2$, in which case the bits match with probability $1/2$. In the two remaining cases (which happens with probability $1/2$), the bits always match. Therefore $\Pr[y = K] = 1/2 \cdot 1/2 + 1 \cdot 1/2 = 3/4$.

On the other hand, since G is a random permutation and p and q are distinct we have

$$\Pr\left[A^{E,E^{-1},G(\mathsf{RK},\cdot)} = 1\right] = 1/2.$$

The first part of the theorem follows.

To prove the second part, we construct a key-recovery adversary A as shown in Figure 7. Intuitively, this adversary uses different indices i to recover the i-th bit of the key. In doing so, it collects a sample of size ℓ for each i, by randomly choosing different p's from $\mathcal{D}_i$. It then takes the majority answer to guess the i-bit. The probability that A recovers the bit correctly in each step is $3/4$ (and hence it errs with probability $1/4$). Using the Chernoff bound we deduce that the probability that the majority answer is an incorrect guess is at most

$$\exp(-(1/2 - 1/4)^2 \ell/2).$$

Therefore the probability that all the computed bits are correct is given by the expression in the theorem statement.

By taking ℓ large enough this probability can be made arbitrarily close to 1. However, we need to ensure that each segment $\mathcal{D}_i$ contains at least ℓ different plaintexts. This introduces the requirement that $2^{n-1} \geq k\ell$, which is easily achievable as k and n are of the same order. □

REMARK. In order to simplify the analysis, we have presented Harris's attack in the ideal-cipher model using oracle RKD sets. Harris's attack for each concrete cipher $E^\star$ is obtained by replacing the oracle with a subroutine computing $E^\star$. The derivation of various probabilities above should then be adapted to $E^\star$, as the cipher no longer returns outputs which are perfectly uniformly distributed.

```
Algorithm $A^{E,E^{-1},E(\mathsf{RK}(\cdot,K),\cdot)}$:
For $i = 1$ to $k$ do
    For $j = 1$ to $\ell$ do
        $p_{ij} \xleftarrow{\$} \mathcal{D}_i$
        Query $E(\mathsf{RK}(H_{i,p_{ij}}, K), p_{ij})$ to get $x_{ij}$
    Set $b_i$ to be the majority of $[[x_{i1}]_i, \ldots, [x_{i\ell}]_i]$
Return $(b_1 \oplus b_2 \oplus \ldots \oplus b_k)$
```

Fig. 7: $\tilde{\Psi}^E$-restricted adversary recovering the key of an ideal cipher.

D Proof of Theorem 6

Proof. For oracle-collision-resistance-2, the advantage is zero as E is a permutation. For oracle-output-unpredictability-2 note that for any K'

$$\Pr\left[K \xleftarrow{\$} \mathcal{K}; E \xleftarrow{\$} \mathrm{Perm}(\mathcal{K}, \mathcal{D}) : E(K, x) = K'\right] = 1/|\mathcal{K}|.$$

The advantage term follows by considering all possible collisions between the q queries made to the $\mathcal{P}_K$ oracle, and the q' queries made to the $\mathcal{X}$ oracle. For oracle-independence, note that for any K', x, x_i, and x_j we have

$$\Pr\left[K \xleftarrow{\$} \mathcal{K}; E \xleftarrow{\$} \mathrm{Perm}(\mathcal{K}, \mathcal{D}) : (K', x) \in \{(K, x_i)\}\right] \leq 1/|\mathcal{K}| \quad \text{and}$$

$$\Pr\left[K \xleftarrow{\$} \mathcal{K}; E \xleftarrow{\$} \mathrm{Perm}(\mathcal{K}, \mathcal{D}) : (E(K, x_j), x) \in \{(K, x_i)\}\right] \leq 1/|\mathcal{K}|.$$

The advantage term follows by considering the q queries made to the $\mathcal{Q}_K$ oracle. □

REMARK. Proof of validity at $E = E^\star$ is similar to that given for the second part of Theorem 2.

E Proof of Theorem 7

Proof. For oracle-collision-resistance-2 note that since E's output is randomly and independently distributed, for any K' and K'' we have

$$\Pr\left[K \xleftarrow{\$} \mathcal{K}; E \xleftarrow{\$} \mathrm{Perm}(\mathcal{K}, \mathcal{D}) : K = E(K', K)\right] = 1/|\mathcal{K}| \quad \text{and}$$

$$\Pr\left[K \xleftarrow{\$} \mathcal{K}; E \xleftarrow{\$} \mathrm{Perm}(\mathcal{K}, \mathcal{D}) : E(K', K) = E(K'', K)\right] = 1/|\mathcal{K}|.$$

The advantage term follows as A makes at most q queries to its $\mathcal{C}_K$ oracle.

The argument for oracle-output-unpredictability-2 is analogous. For any K' and K'' we have

$$\Pr\left[K \xleftarrow{\$} \mathcal{K}; E \xleftarrow{\$} \mathrm{Perm}(\mathcal{K}, \mathcal{D}) : K'' = E(K', K)\right] = 1/|\mathcal{K}|.$$

For any oracle-output-unpredictability-2 adversary A making at most q queries to its $\mathcal{X}$ oracle and at most q' queries to its $\mathcal{P}_K$ oracle, we get the advantage term.

For oracle-independence, note that for any K', K_1, K_2 and x_1 we have

$$\Pr\left[K \xleftarrow{\$} \mathcal{K}; E \xleftarrow{\$} \mathrm{Perm}(\mathcal{K}, \mathcal{D}) : (K', x_1) \in \{(K_2, K)\}\right] \leq 1/|\mathcal{K}| \quad \text{and}$$

$$\Pr\left[K \xleftarrow{\$} \mathcal{K}; E \xleftarrow{\$} \mathrm{Perm}(\mathcal{K}, \mathcal{D}) : (E(K_1, K), x_1) \in \{(K_2, K)\}\right] \leq 1/|\mathcal{K}|.$$

Therefore for any oracle-independence adversary A making at most q queries to its $\mathcal{Q}_K$ oracle, the advantage is at most as given in the statement of the theorem. □

REMARK. Proof of validity at $E = E^\star$ assumes $F^\star(K, x) := E^\star(x, K)$ is a pseudorandom function.