Avoid Mask Re-use in Masked Galois Multipliers

D. Canright
Applied Math., Naval Postgraduate School, Monterey CA 93943, USA
dcanright@nps.edu

26 November 2008

Abstract

This work examines a weakness in re-using masks for masked Galois inversion, specifically in the masked Galois multipliers. Here we show that the mask re-use scheme included in our work[1] cannot result in "perfect masking," regardless of the order in which the terms are added; explicit distributions are derived for each step. The same problem requires new masks in the subfield calculations, not included in [1]. Hence, for resistance to first-order differential attacks, the masked S-box must use distinct, independent masks for input and output bytes of the masked inverter, and new masks in the subfields, resulting in a larger size.

keywords: AES, S-box, masking, DPA, composite Galois field

1 Introduction

Implementations of cryptographic algorithms, e.g. the Advanced Encryption Standard (AES), may be vulnerable to "side-channel attacks" such as differential power analysis (DPA). One countermeasure against such attacks is adding a random mask to the data; this randomizes the statistics of the calculation at the cost of computing "mask corrections." Oswald et al.[2] showed how the "tower field" representation of the Galois field allows maintaining an additive mask throughout the Galois inverse calculation. For a hardware implementation of AES, they suggested re-using masks from input to output, which allows re-use of some previously computed products to save circuitry. We incorrectly applied a similar masking strategy in developing a compact masked S-box[1]. Both of these works claimed "perfect masking" (by the definition of Blömer[3]) giving suitable implementations immunity to first-order differential side-channel attacks. However, only [2] was indeed secure; as shown below, the incorrect mask re-use employed in [1] cannot give perfect masking.

Note that, even though the masking scheme of [2] was provably secure, nonetheless Mangard et al.[4] successfully attacked an ASIC implementing this scheme. The weakness was attributed to CMOS "glitches," and indeed Mangard and Schramm[5] showed specifically how timing glitches could actually leak information even in a "perfectly masked" CMOS implementation.

1.1 Security Model

We adopt the notion of "perfect masking" given by Blömer[3], that is, assuming a source of truly random uniformly distributed masks, then the distribution of each intermediate result is independent of both the plaintext data and the key. This gives protection from first-order differential side-channel attacks (neglecting specific hardware problems such as CMOS glitches).

For reference, here we paraphrase Lemmas 1 and 2 of [3].

Lemma 1 *Given x uniformly distributed over a finite field $\mathbb{F}$, and any $y \in \mathbb{F}$ independent of x, then $z = x+y$ is also uniformly distributed and independent of y.[3]*

Lemma 2 *Given x and y independent and both uniformly distributed over a finite field $\mathbb{F}_q$ of size q, then $z = x\,y$ is distributed according to*

$$Pr(z = i) = \begin{cases} (2q-1)/q^2 & , \ i = 0 \\ (q-1)/q^2 & , \ i \neq 0 \end{cases}$$

here called the random product distribution.[3]

2 Masked Multipliers

The re-use of masks in the Galois inverter part of an AES S-box in [1], turns out *not* to give "perfect masking," where each intermediate operand has a distribution independent of the masked data. The problem is in masking the multiplications inherent in evaluating the inverse over the subfield. We adopt the notation of Mangard and Schramm[5], who pinpointed the CMOS glitch problem to the adding of terms in the masked multiplier.

To mask the Galois field multiplication

$$a\,b = c \tag{1}$$

we need masks m_a, m_b, m_c for a, b, c respectively. (We assume a source of uniformly distributed random masks.) Define the masked variables by

$$a_m = a + m_a\,,\ b_m = b + m_b\,,\ c_m = c + m_c \tag{2}$$

Then the masked multiplier of [5] is

$$c_m = a_m\,b_m + (m_a\,b_m + (a_m\,m_b + (m_a\,m_b + m_c))) \tag{3}$$

where + indicates Galois field addition (bitwise XOR, assuming an extension of the binary field), multiplications are in the Galois field, and the parens indicate the order of operations to avoid any intermediate result with a distribution dependent on the data a and/or b. If the masks are indeed random and independent, this is provably secure, and can be made immune to CMOS glitches by enforcing some timing constraints[5].

Oswald et al.[2] suggested re-using masks in the Galois inverter, to be able to replace some subfield multiplications by previously known products. The approach shown in, e.g., equation [2, (15)], is to replace the independent mask m_c above with m_a, one of the input masks. After all the additions are complete, the result would be $c + m_a$, so the product (output data) is correctly masked by an independent uniform mask.

But before [2, (15)] and related formulas, Oswald specifically warns:

> It needs to be pointed out that the formulae, which we derive in this section, do not lead to a secure implementation when directly implemented. The secure implementation of these formulae requires the addition of an independent value to the first intermediate value that is computed.

Unfortunately, in [1] we failed to heed this warning.

The problem is, you can't get there $(c + m_a)$ from here:

$$a_m\,b_m + m_a\,b_m + a_m\,m_b + m_a\,m_b + m_a \tag{4}$$

No matter what order you choose to perform the additions, some intermediate result will have a distribution that depends on the data. The source of the trouble is that the mask, though uniform, is *not independent* of the other terms.

Here is an exhaustive examination of the possible addition strategies. First, it is well known that adding any two of the products

$$P_1 = a_m\,b_m\,,\ P_2 = m_a\,b_m\,,\ P_3 = a_m\,m_b\,,\ P_4 = m_a\,m_b \tag{5}$$

gives a data-dependent distribution, even though each product has the "random product" distribution (hereafter denoted as P). But for completeness, we will consider what distributions result from adding two products. $P_1 + P_2 = a\,b_m$ with a distribution of the form $a\,X$ (where X is independent and uniform); this distribution depends on a since $a = 0$ gives the constant zero distribution while $a \neq 0$ gives a uniform distribution. Similarly, $P_3 + P_4 = a\,m_b$ (of the form $a\,X$) and $P_1 + P_3 = a_m\,b$ and $P_2 + P_4 = m_a\,b$ (of the form $b\,X$) give data-dependent distributions. Both $P_1 + P_4 = a\,b_m + m_a\,b$ and $P_2 + P_3 = a\,m_b + m_a\,b$ have distributions of the form $a\,X + b\,Y$ (with both X and Y uniform and independent); if $a = b = 0$ the zero distribution results, otherwise the uniform distribution: data-dependent.

So the addition must start with the mask. Consider first adding

$$P_1 + m_a = a\,b_m + m_a\,(b_m + 1) \tag{6}$$

The result is of the form $aX + P$ where X is uniform and P is a random product and both are independent of a. Then if $a = 0$ we get the product distribution P, but if $a \neq 0$ we get a uniform distribution (by Lemma 1 since then aX is uniform); the distribution depends on a. Similarly,

$$P_3 + m_a = a\,m_b + m_a\,(m_b + 1) \tag{7}$$

is of the form $aX + P$.

So we must start with either

$$P_2 + m_a = m_a\,(b_m + 1) \tag{8}$$

or

$$P_4 + m_a = m_a\,(m_b + 1) \tag{9}$$

each having distribution P.

Say we take the first choice (8). Consider adding P_4 (or equivalently, take the second choice and add P_2):

$$(P_2 + m_a) + P_4 = m_a\,(b + 1) \tag{10}$$

which has the form $(b + 1)\,X$. So if $b = 1$ we get the constant zero distribution; if not we get a uniform distribution; no go. Or try adding P_3:

$$(P_2 + m_a) + P_3 = a\,m_b + m_a\,(b + 1) \tag{11}$$

of the form $a\,X + (b + 1)\,Y$. So if a and $b + 1$ are both zero, we get the zero distribution; otherwise uniform: no go. Similarly if we start with the second choice (9) and add P_1

$$(P_4 + m_a) + P_1 = a\,b_m + m_a\,(b + 1) \tag{12}$$

no go.

So the first choice (8) leads to

$$(P_2 + m_a) + P_1 = a\,b_m + m_a \tag{13}$$

of the form $a\,X + Y$, giving a uniform distribution. Similarly, the second choice (9) gives

$$(P_4 + m_a) + P_3 = a\,m_b + m_a \tag{14}$$

again $a\,X + Y$.

From (13), if we add P_4

$$((P_2 + m_a) + P_1) + P_4 = a\,b_m + m_a\,(m_b + 1) \tag{15}$$

we again get the form $a\,X + P$; no go. Similarly if we take (14) and add P_2

$$((P_4 + m_a) + P_3) + P_2 = a\,m_b + m_a\,(b_m + 1) \tag{16}$$

also of form $a\,X + P$.

But from (13), if we add P_3

$$((P_2 + m_a) + P_1) + P_3 = a\,b + m_a\,(m_b + 1) \tag{17}$$

we get the form $c + P$, which is the product distribution if $c = 0$ and not otherwise (each different value of c gives a different distribution, which could be called an “offset” product distribution). And similarly if we add P_1 to (14)

$$((P_4 + m_a) + P_3) + P_1 = a\,b + m_a\,(b_m + 1) \tag{18}$$

again of form $c + P$. (Note that this shows that all four possible third steps, independent of the earlier steps, give data-dependent distributions, so consideration of third steps alone would have sufficed.)

So even though each individual term in (4) has a data-independent distribution, and so does their sum, there is *no way* to add them up without revealing a data-dependent distribution! Hence, the direct re-use of masks in masked products, as in [2, (15)], cannot give perfect masking. ([2] does not use m_a to protect the additions, but rather points out that “every summation of variables must start with the addition of an independent mask M.”) Therefore the statement before eq. [1, (29)], that the output mask **S** could be the original input mask **M** (without adding an independent “fresh mask” first), is *incorrect.*

2.1 Intermediate Masks in Subfields

What about the re-use of masks in the subfield calculations of [1, (23–28)]?

There is a big difference there, in that the mask re-used on each output product is *neither* of the masks of the factors. The output mask is re-used from a different part of the calculation.

For example, in [1, (23)] the two factors $\tilde{\mathbf{b}}_0$ and $\tilde{\mathbf{c}}^{-1}$ are respectively masked by $\mathbf{m}_0$ and $\mathbf{m}_1$, which are the two halves of the uniform mask $M_2 = N\ (M_1 + M_0)^2$ [1, (13)]. But the output $\tilde{\mathbf{b}}_1^{-1}$ is masked by $\mathbf{m}_{11}$, the upper half of M_1, which is uniform and independent of the two input masks (due to the M_0). Similarly with [1, (25)].

2.2 Other Dependent Masks

How about other combinations of inputs to mask products, such as in [1, (12)]? There we asserted that the addition must start with the uniformly distributed term $N\left(\tilde{A}_1 + \tilde{A}_0\right)^2$. One might suspect that, because this term does not seem to be independent of the others, that there may be a problem.

Unfortunately, there is a problem. Again, even though the final result would be correctly masked, no order of additions maintains data-independent distributions.

Rewrite [1, (12)] in the notation used above:

$$N(a_m + b_m)^2 + a_m\, b_m + m_a\, b_m + a_m\, m_b + m_a\, m_b \tag{19}$$

where N is a known nonzero constant. Use the following respective labels for these terms:

$$M = N(a_m + b_m)^2\ ,\ P_1 = a_m\, b_m\ ,\ P_2 = m_a\, b_m\ ,\ P_3 = a_m\, m_b\ ,\ P_4 = m_a\, m_b \tag{20}$$

Again, since no two products can be added, the sum must begin with the mask M. (In the algebra that follows, recall that, since the field has characteristic 2, subtraction is addition, and a sum squared is the sum of the squares.)

Consider beginning with

$$M + P_1 = N(a_m + b_m)^2 + a_m\, b_m \tag{21}$$

This is clearly independent of the data (a & b), being solely a function of the uniform, masked quantities a_m & b_m. The specific distribution that results depends on the value of N[1].

The N in [1] comes from earlier work[6]: N is the norm of an element of the normal basis for the larger field, where the basis element was chosen to have a trace of 1. So in $GF(2^4)$, N is a root of one of the two polynomials[2] $x^4 + x^3 + 1$ or $x^4 + x^3 + x^2 + x + 1$; in $GF(2^2)$ it is a root of $x^2 + x + 1$. For these values of N, the polynomial

$$x^2 + N^{-1}\, x + 1 \tag{22}$$

is irreducible and so has no roots in the field. (It turns out that, for all *other* values of N [the other half of the field], then (22) factors and $M + P_1$ gives the product distribution; this case is discussed in the appendix A.)

Then the distribution of $M + P_1$ is uniform over the *nonzero* elements of the field $\mathbb{F}_q$, each with probability $(q+1)/q^2$, except zero occurs with probability $1/q^2$. This distribution can be understood as follows. If $a_m = b_m = 0$, then clearly (21) is 0. But if a_m or b_m is nonzero, we can "divide it out". For example, suppose $b_m \neq 0$, that is, consider the nonzero portion of the uniform distribution of b_m. Then (21) = $N\, b_m\, (X^2 + N^{-1}X + 1)$ where $X = a_m b_m^{-1}$. But, for any value of X, the polynomial (22) gives a nonzero result, and $N\, b_m$ is uniformly distributed over the nonzero multiplicative group $\mathbb{F}_q^*$, so for each value of X then (21) is uniformly distributed over the nonzero group. The same reasoning applies for $a_m \neq 0$. Hence, only the case $a_m = b_m = 0$ leads to (22) = 0; any other case leads to a uniform distribution over the nonzero values of the field. Let us call this distribution the "unproduct" distribution, since the paucity of zeros is the opposite of the product distribution.

The other three possible beginnings all give data-dependent distributions. For

$$M + P_2 = N a^2 + [N(m_a + b_m)^2 + m_a\, b_m] \tag{23}$$

[1] These results were guided and/or checked by calculations using the *Maple* mathematics software.

[2] The specific value used in [1] is a root of the first minimal polynomial.

the terms in brackets give the unproduct distribution, and adding the first term gives a different "offset unproduct" distribution (where one value occurs with probability $1/q^2$, the rest uniformly) for each a. Similarly,

$$M + P_3 = N b^2 + [N(a_m + m_b)^2 + a_m\, m_b] \tag{24}$$

gives a different offset unproduct distribution for each b. And

$$M + P_4 = N(a+b)^2 + [N(m_a + m_b)^2 + m_a\, m_b] \tag{25}$$

gives a different offset unproduct distribution for each $a + b$.

For the second addition, one possibility is

$$(M + P_1) + P_2 = N m_a^2 + N(a + b_m)^2 + a\, b_m \tag{26}$$

where the first term is uniform and independent of the rest, so the sum gives a uniform distribution. Similarly

$$(M + P_1) + P_3 = N m_b^2 + N(a_m + b)^2 + a_m\, b \tag{27}$$

also gives a uniform distribution. The remaining case is

$$(M + P_1) + P_4 = N(m_a + m_b)^2 + N(a+b)^2 + a\, b + a\, m_b + b\, m_a \tag{28}$$

For most choices of a and b the first (uniform) term is independent of the rest and so adds to give a uniform distribution. But when $a = b \neq 0$ then the result has the form $N\, X^2 + a\, X + a^2$, where $X = m_a + m_b$ is uniform, but the range of this quadratic polynomial only gives *half* of the field; which half depends on a: data-dependent.

But now all three feasible third steps give data-dependent distributions (offset unproduct distributions with the offset given by the data):

$$((M + P_1) + P_2) + P_3 = N(a+b)^2 + a\, b + [N(m_a + m_b)^2 + m_a\, m_b] \tag{29}$$

$$((M + P_1) + P_2) + P_4 = N b^2 + a\, b + [N(a_m + m_b)^2 + a_m\, m_b] \tag{30}$$

$$((M + P_1) + P_3) + P_4 = N a^2 + a\, b + [N(m_a + b_m)^2 + m_a\, b_m] \tag{31}$$

(In fact, skipping over first and second steps, the fourth possible third step $((M + P_2) + P_3) + P_4 = a\, b + [N(a_m + b_m)^2 + a_m\, b_m]$ also gives a data-dependent distribution, so consideration of third steps suffices.)

Again, you can't get there from here. Similarly for [1, (14)].

So the masking scheme as given in [1] is *incorrect.* Dang. In order to achieve perfect masking, not only must top level input and output masks be different ($\mathbf{S} \neq \mathbf{M}$ in [1, (29)]), but also new independent uniform masks must be introduced in the subfield calculations (similar to the approach in Blömer et al.[3]).

2.3 Re-using Masks Between Rounds

What about re-using masks between rounds, a possible implementation choice discussed in [1]? That is a completely different sort of mask re-use, where the masks used on a block for one round are used again in a later round. Note that for each Galois inverter, the input mask is different from the output mask in this proposed masking scheme. This sort of re-use does not introduce any data-dependent operand distributions, and so (assuming the rest of the masking scheme is secure) is resistant to first-order DPA (although such re-use probably decreases resistance to higher-order attacks).

3 Conclusion

Masked multipliers that re-use an input mask (or a mask dependent on the input) for output are *insecure* (unless an independent "fresh mask" is added first to protect the summations).

The "perfectly masked" compact AES S-box of [1] was *not* perfectly masked, because of mask re-use and input-dependent masks. (A corrected version [7] is available.) Because re-using input masks as output masks in the Galois inverter is precluded, and additional masks must be included in the subfield calculations, then the size of a compact masked S-box is significantly larger.

References

[1] Canright, D., Batina, L.: A very compact "perfectly masked" S-box for AES. In Bellovin, S.M., et al., eds.: Proceedings of 6th International Workshop on Applied Cryptography and Network Security (ACNS). Volume 5037 of Lecture Notes in Computer Science., Springer-Verlag (2008) 446–459

[2] Oswald, E., Mangard, S., Pramstaller, N., Rijmen, V.: A side-channel analysis resistant description of the AES S-box. In Gilbert, H., Handschuh, H., eds.: Fast Software Encryption: 12th International Workshop, FSE 2005. Volume 3557 of Lecture Notes in Computer Science., Springer-Verlag (2005) 413–23

[3] Blömer, J., Guajardo, J., Krummel, V.: Provably secure masking of AES. In Handschuh, H., Hasan, M.A., eds.: Selected Areas in Cryptography, 11th International Workshop, Springer-Verlag (2004) 69–83

[4] Mangard, S., Pramstaller, N., Oswald, E.: Successfully attacking masked AES hardware implementations. In Rao, J.R., Sunar, B., eds.: Proceedings of 7th International Workshop on Cryptograpic Hardware and Embedded Systems (CHES). Volume 3659 of Lecture Notes in Computer Science., Springer-Verlag (2005) 157–171

[5] Mangard, S., Schramm, K.: Pinpointing the side-channel leakage of masked AES hardware implementations. In Goubin, L., Matsui, M., eds.: Proceedings of 8th International Workshop on Cryptograpic Hardware and Embedded Systems (CHES). Volume 4249 of Lecture Notes in Computer Science., Springer-Verlag (2006) 76–90

[6] Canright, D.: A very compact S-box for AES. In Rao, J.R., Sunar, B., eds.: Proceedings of 7th International Workshop on Cryptograpic Hardware and Embedded Systems (CHES). Volume 3659 of Lecture Notes in Computer Science., Springer-Verlag (2005) 441–455

[7] Canright, D., Batina, L.: A very compact "perfectly masked" S-box for AES (corrected). `http://web.nps.navy.mil/~dcanrig/pub/acns2008corr.pdf` (November 2008)

A Other Values of N

For values of N other than the norms discussed above (i.e., *excluding* roots of x^4+x^3+1 and $x^4+x^3+x^2+x+1$ in $GF(2^4)$, and roots of x^2+x+1 in $GF(2^2)$), then some of the distributions are different from those above, though the conclusions about data dependency remain unchanged. Of course, if $N=0$ there is no mask, so we only consider nonzero N here.

For these values of N, the polynomial (22) $x^2+N^{-1}\,x+1$ has two distinct roots in the field, which we can call k and k^{-1} (since their product is 1). Then $k+k^{-1}=N^{-1}$, and

$$M+P_1 = N(a_m+b_m)^2 + a_m\,b_m = N(a_m+k\,b_m)(a_m+k^{-1}b_m) \tag{32}$$

gives a distribution of the form $N\,X\,Y$ with X and Y uniform and independent and N nonzero, i.e., we get the product distribution P, independent of the data.

Again, the other three possible beginnings all give data-dependent distributions.

$$M+P_2 = N(a+m_a+b_m)^2 + m_a\,b_m = Na^2 + N(m_a+k\,b_m)(m_a+k^{-1}b_m) \tag{33}$$

of the form Na^2+P, gives a different offset product distribution for each a. Similarly,

$$M+P_3 = N(a_m+b+m_b)^2 + a_m\,m_b = Nb^2 + N(a_m+k\,m_b)(a_m+k^{-1}m_b) \tag{34}$$

is of the form Nb^2+P, a different offset product distribution for each b. And

$$M+P_4 = N(a+m_a+b+m_b)^2 + m_a\,m_b = N(a+b)^2 + N(m_a+k\,m_b)(m_a+k^{-1}m_b) \tag{35}$$

is of the form $N(a+b)^2+P$, a different offset product distribution for each $a+b$.

For the second addition, the distributions remain as given above; the specific value of N is not important there.

And again all three feasible third steps give data-dependent distributions (offset product distributions with the offset given by the data):

$$((M+P_1)+P_2)+P_3 = N(a+m_a+b+m_b)^2+a\,b+m_a\,m_b = N(a+b)^2+a\,b+N(m_a+k\,m_b)(m_a+k^{-1}m_b) \quad (36)$$

$$((M+P_1)+P_2)+P_4 = N(a_m+b+m_b)^2+a\,b+a_m\,m_b = Nb^2+a\,b+N(a_m+k\,m_b)(a_m+k^{-1}m_b) \quad (37)$$

$$((M+P_1)+P_3)+P_4 = N(a+m_a+b_m)^2+a\,b+m_a\,b_m = Na^2+a\,b+N(m_a+k\,b_m)(m_a+k^{-1}b_m) \quad (38)$$

And so does the fourth possible third step, without considering previous steps:

$$((M+P_2)+P_3)+P_4 = N(a_m+b_m)^2+a\,b+a_m\,b_m = a\,b+N(a_m+k\,b_m)(a_m+k^{-1}b_m) \quad (39)$$

So masks using these other values of N also reveal data-dependent distributions: imperfect masking.