Hawkes-Rose variant of IAPM and message integrity

This note should be read in conjunction with
http://csrc.nist.gov/encryption/modes/proposedmodes/iapm/integrityproofs.pdf

In this note we show that the variant of IAPM proposed by Hawkes and Rose is secure for message integrity. In this variant, not all blocks need be sent encrypted. Without loss of generality, let all the plaintexts which need to be sent unencrypted be at the beginning of the message. We allow the adversary to pick for each message how many blocks are to be sent unencrypted. Let this random variable be called U^i, i.e. U^i of the L^i blocks are to be sent unencrypted, and this includes the first block (0th block) which is just the IV.

The HR scheme works as follows. The blocks 1 to $U^i - 1$ are encrypted anyway, to generate ciphertext blocks as usual (i.e. as in IAPM). However, only the plaintext is sent to the receiving party. In other words, for j in $1..U^i - 1$

$$C^i_j = P^i_j$$

But now a new checksum A^i is computed as follows

$$A^i = \sum_{j=1}^{U^i-1} N^i_j \oplus S^i_j$$

In other words, this checksum is the xor sum of what used to be the ciphertext blocks upto block $U^i - 1$. This is then xored to what used to be $C^i_{L^i-1}$. Thus,

$$C^i_{L^i-1} = N^i_{L^i-1} \oplus S^i_0 \oplus A^i$$

Proof (Message Integrity): We follow the proof idea of Johan Håstad as given in section 6.1. Note that section 6.1 inturn refers to section 6 (Theorem 1) for details. At some point, we hope to write this proof complete in itself.

There are two different ways of proving this result. The proof of message integrity goes by first building the tree of computation paths, and labelling each leaf with $C = c$. Now we could include in C and c the blocks of "ciphertext" which the adversary does not see, but then we have to restrict the paths by forcing the adversary to take the same "next choice" when the part of C visible to it is same. In another proof, which we prefer, we leave C and c to be exactly what it is , i.e. it doesn't include the $N^i_j \oplus S^i_j$ and its fixed values, for $0 < j \leq U^i - 1$. However, the definition of event E6, and hence proof of lemma 4 needs to be modified.

Each constant ciphertext c, now has an auxillary information u correspnding to U. Given $C = c$, and $G = g$, where c is a constant sequence of ciphertexts and g is a constant permutation, the M values are fixed, because $M^i_j = P^i_j \oplus S^i_j$. The variable P^i_j is completely fixed by c, and S^i_j is fixed by $g(c^i_0)$'s. We will write $M^i_j(c, g)$ for this value of M^i_j. Similarly, for N^i_j, $j > U^i - 1$. So, for any c and g, and $y \leq z$, define E6(y, c, g) to be

$$\forall i, i' \in [1..y], \forall j, j', j \in [1..l^i - 1], j' \in [1..l^{i'} - 1], (i, j) \neq (i', j') :$$

$$(M^i_j(c,g) \neq M^{i'}_{j'}(c,g)) \wedge$$
$$\forall i, i' \in [1..y], \forall j, j', j \in [u^i..l^i - 1], j' \in [u^{i'}..l^{i'} - 1], (i,j) \neq (i',j') :$$
$$(N^i_j(c,g) \neq N^{i'}_{j'}(c,g))$$

We add a pseudo round to the first z rounds of queries made by the adversary. This pseudo rounds is built out of the query made by the adversary in the second stage, i.e. using C'. In the pseudo round $z+1$, for $j \in [U'..L'-2]$, if N'_j is not equal to any of the previous N values, then it is decrypted to yield the corresponding M value. Note that we do not decrypt the last block.

Now we generalize event E6 above to round $z+1$, where for round $z+1$ we only consider those M and N blocks as in the previous paragraph (i.e. new Ns and their corresponding Ms). Recall that c' is completely determined by c.
Now we can generalize lemma 4 of section 6.1 as follows:
For every constant c, and for any permutation g such that $E6(z+1,c,g)$,

$$\Pr[G = g | C = c \wedge E6(z+1,c,G)] = \frac{\Pr[G=g]}{\Pr[E6(z+1,c,G)]}$$

Proof: Let U be the universe of G. Under the condition $C = c$ and $\mathrm{E6}(z+1,c,G)$ we show that every g such that $\mathrm{E6}(z+1,c,g)$ holds, is equally likely to be G. Since c is fixed, fixing G to g, fixes the N variables to a single value for all $i,j : u^i \leq j \leq l^i - 1$ (with all N's different, for otherwise $\mathrm{E6}(z+1,c,g)$ wouldn't hold). This value of the N variables is not ruled out as all the M variables are different (by $\mathrm{E6}(z,c,G)$), and F is a random permutation.

For the other N values we have the following condition (for each i)

$$\sum_{j=u^i}^{l^i-1} N^i_j = S^i_0 \oplus c^i_{l^i-1} \oplus \sum_{j=u^i}^{l^i-2} S^i_j$$

Again, for each value of g (which fixes S), given $\mathrm{E6}(z+1,c,g)$, the number of possibilites for the remaining N variables is the same (one could actually calculate this expression, but that is not required), as F is a random permutation.

Thus,

$$\Pr[G = g | C = c \wedge E6(z,c,G)] = \frac{1}{\#g : E6(z,c,g)}$$

Thus,

$$\Pr[G = g | C = c \wedge E6(z,c,G)] = \frac{1}{|U| * \Pr[E6(z,c,G)]} = \frac{\Pr[G=g]}{\Pr[E6(z,c,G)]}$$

□

Now for the main proof. Let's assume that c' has the same C_0 as some first stage message (query) C^k (case (b) claim 4), because the other case is routine– needs to be written though! There are three cases.
(a) If in the second stage, the first U^k blocks are the same in C' as in C^k (regardless of whether $U' = U^k$), and there is a ciphertext later than $U^k - 1$ which is different, then the analysis is similar to original IAPM.

In the other cases we can assume that $U' = U^k$, for it is unlikely for the adversary to figure out unknown portions of ciphertexts. This needs to be written too.
(b) If, the ciphertext is same in the latter half, i.e blocks U' to $L' - 1$, but different in the first half, then we show that $N'_{L'-1}$ is different from all previous N (i.e. upto round $z+1$). First note,

$$\begin{aligned}&\Pr[N'_{L'-1} = N^k_{L'-1} \mid C = c \wedge E6(z+1, c, G) \ \wedge \ E3(\overrightarrow{r})] \\ &= \Pr[\sum_{x=1}^{U'-1} N'_x = \sum_{x=1}^{U'-1} N^k_x \mid C = c \wedge E6(z+1, c, G) \ \wedge \ E3(\overrightarrow{r})]\end{aligned}$$

By E6$(z+1, c, G)$, the M's are all different and hence the probability of this event is (about) 2^{-n}, given that there is at least one x such that $M'_x \neq M^k_x$.

Next for 0 U', $< s$

$$\begin{aligned}&\Pr[N'_{L'-1} = N'_s \mid C = c \wedge E6(z+1, c, G) \ \wedge \ E3(\overrightarrow{r})] \\ &= \Pr[\sum_{x=1}^{U'-1} (N'_x \oplus S^k_x) \oplus c'_{l'-1} \oplus S^k_0 = N'_s \mid C = c \wedge E6(z+1, c, G) \ \wedge \ E3(\overrightarrow{r})]\end{aligned}$$

If $x = s$ is the only x such that M'_x is different from M^k_x, then by uniformity of $S^k_s \oplus S^k_0$ (by lemma 4) the probability above is 2^{-n}. Otherwise given E6, N'_x can be chosen from a set of size almost 2^n.

If s U', the proof is similar. Finally, a similar proof shows that $N'_{L'-1}$ is different from all N^i_s, for all other i.
(c) If the ciphertext is different in both halves, the proof is same as in case (b), except that it is possible that $c'_{L'-1} \neq c^k_{L'-1}$, in which case for the case

$$\Pr[N'_{L'-1} = N^k_{L'-1}]$$

we get another additive term $c'_{L'-1} \oplus c^k_{L'-1}$, which doesn't make any difference to the analysis.

Thus, by an equation similar to equation (1) we get that in all cases (a), (b) and (c), the checksum validates with negligible probability $\square$

This article was processed using the LaTeX macro package with LLNCS style