PRF Domain Extension using DAGs

Charanjit S. Jutla

IBM T. J. Watson

Abstract. We prove a general domain extension theorem for pseudo-random functions (PRFs). Given a PRF F from n bits to n bits, it is well known that employing F in a chaining mode (CBC-MAC) yields a PRF on the bigger domain of mn bits. One can view each application of F in this chaining mode to be a node in a graph, and the chaining as the edges between the node. The resulting graph is just a line graph. In this paper, we show that the underlying graph can be an arbitrary directed acyclic graph (DAG), and the resulting function on the larger domain is still a PRF. The only requirement on the graph is that it have unique source and sink nodes, and no two nodes have the same set of incident nodes. A new highly parallelizable MAC construction follows which has a critical path of only $3 + \log^* m$ applications of F.

If we allow Galois field arithmetic, we can consider edge-colored DAGs, where the colors represent multiplication in the field by the color. We prove an even more general theorem, where the only restriction on the colored DAGs is that if two nodes (u and v) have the same set of incident nodes W, then at least one w in W is incident on u and v with a different colored edge. PMAC (parallelizable message authentication [5]) is a simple example of such graphs. Finally, to handle variable length domain extension, we extend our theorem to a collection of DAGs. The general theorem allows one to have further optimizations over PMAC, and many modes which deal with variable lengths.

All the results proven are under the adaptive adversary model.

Keywords: PRF, MAC, DAG, partial order, Galois field

1 Introduction

There is often a need to extend the domain of a given pseudo-random function (PRF). One of the most popular and well-known such schemes is the CBC-MAC [1]. In [3] it was shown that if F is a secure pseudo-random function from n bits to n bits, then the CBC (cipher block chaining) construction yields a secure PRF from mn bits to n bits. Although the construction is called a MAC (message authentication code), which is a strictly weaker notion than PRF ([8]), the above shows that it is indeed a more general PRF domain extension method. Other domain extension schemes are known as well, for example, the cascade construction [2] and the protected counter sum construction [4]. Recently, a scheme PMAC (or Parallelizable Message Authentication)[5] (also see XECB [10]) was also shown to be a domain extension scheme.

Despite all these results, there is no unifying theme in these results. In this paper, we attempt to remedy this situation, by proving a general theorem for domain extension. In essence, we show that arbitrary acyclic networks of the same pseudo-random function can be used to build a pseudo-random function on a larger domain. To illustrate this paradigm, consider the CBC-MAC scheme. Let F be a PRF from n bits to n bits (and which takes k bits of secret key). For example, DES[9] is usually assumed to be such a PRF on 64 bits, with 56 bits of secret key. A PRF $\tilde{F}$ from mn bits to n bits is defined as follows. The mn bit input is divided into m blocks $P_1, P_2, ..., P_m$. The function F_K (i.e. F with key K) is applied to the first block P_1 to yield an intermediate value C_1. The function F_K is then invoked on the xor of the next block P_2 and previous intermediate value C_1, to yield C_2. This chaining process is continued, and the output of $\tilde{F}$ is just C_m. The chaining process defines an underlying directed graph of m nodes $V_1, V_2, ..., V_m$, with an edge from V_i to V_{i+1}.

Now, consider an arbitrary directed acyclic graph (DAG) $G = (V, E)$, with m nodes V, and edges E. Assume that G has only one source node V_1, and only one sink node V_m. Given a function F from n bits to n bits, a composite function $\tilde{F}$ from mn bits to n bits is defined as follows. As before, assume that the input is a sequence $P_1, ..., P_m$. The first intermediate value is just $C_1 = F(P_1)$. Inductively assume that we have computed the intermediate values of all predecessors of a node V_i. Then, the intermediate value C_i for the node V_i is the result of applying F to the xor sum of P_i and all the C_j, such that (V_j, V_i) is a directed edge in the graph. The output of the composite function $\tilde{F}$ is just C_m. See figure 1 for an example.

Of course, not all DAGS are expected to yield a PRF. However, consider DAGs with the restriction that no two nodes have the same set of incident nodes (u is said to be incident on v if there is an edge from u to v), and that they have unique source and sink nodes. In this paper we show that given a PRF F from n bits to n bits, the composite function $\tilde{F}$, defined using such DAGs as above, is a PRF from mn bits to n bits.

An immediate application is that if a party has access to parallel hardware, then instead of simple chaining as in CBC-MAC, it can compute the PRF in parallel. For instance, if it has four processors, then it can employ the method given by the graph in figure 2 . A parallel mode with critical path of length only $3 + \log^* m$ also follows (see appendix C). Unlike PMAC[5], this mode does not use any Galois arithmetic.

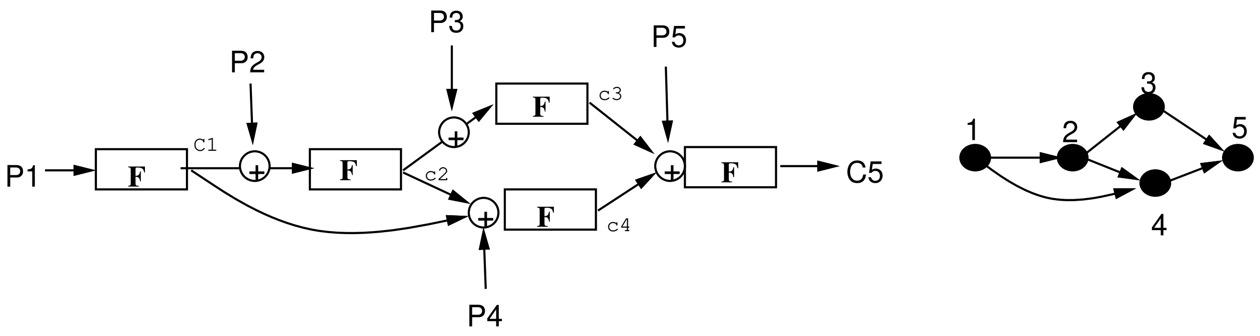

Fig. 1. A PRF Domain Extension Mode and its DAG

If we allow Galois Field arithmetic (in particular, fields $GF(2^n)$), we can consider edge-colored DAGs. The colors on the edges represent multiplication in the field by the color (assume that each color is mapped to a unique element in the field). For example, going back to figure 1, suppose we employ three colors, col1, col2 , and col3. Let w be a primitive element in the field. We map col1 to unity in the field, col2 to w, and col3 to w^2. Then, if we color the edge $(1,4)$ by col2, then in the definition of the composite function, we multiply the intermediate result C_1 with w in the field, before xoring it with the plaintext P_4 and C_2, and applying F (see fig 6 in Appendix C).

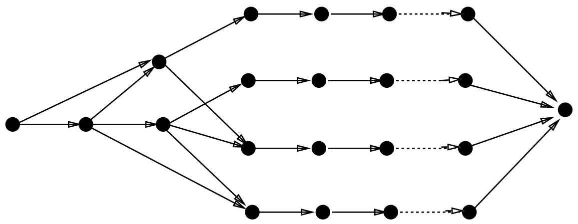

Fig. 2. A Parallel Mode for four processors

The main result of the paper can be stated as follows. Consider an edge colored DAG G with unique source and sink nodes and m total nodes, and with the condition that if two nodes (say u and v) have the same set of incident nodes (say W), then for at least one node w in W, the color on the edge (w,u) is different from the color on the edge (w,v). Given a PRF F from n bits to n bits, the composite function $\tilde{F}$ built using the graph G as above, is a PRF from mn bits to n bits. The result is proven under the adaptive adversary model, which is of course the difficult case.

The mode in fig 2 can now be parallelized further as in fig 5 (see appendix C). The additional cost is a few GF(2^n) operations. Security of PMAC follows (see fig 7 in appendix C), as it is a simple example of such a colored DAG. Further, we obtain the additional optimization over PMAC, because unlike PMAC, we do not even need to compute F on the all zero word (i.e. $F(0^n)$).

We now address the issue of *variable length domain extension.* The previous constructions were devoted to extending the domain of a function from n bits to mn bits, for a fixed m. In other words, the plaintext queries of the adversary were restricted to be exactly mn bits. We could fix m to be large enough, say $m = 2^n$, and use a canonical encoding of smaller sized plaintexts into length mn bit strings. Such an encoding exists for all plaintexts of size less than mn by appending plaintexts of size q bits, by 10^i, where $i = mn - q - 1$. In other words, 10^i acts as an end marker. However, smaller sized plaintexts have to undergo $m = 2^n$ applications of F, which is very inefficient. This problem of a really long end marker was resolved by [16] (also see [6]) by noting that the end marker can actually be of length zero, if it can be authenticated.

So, given a function F on n bits, consider a collection of graphs, one graph G_q in the family for each plaintext length q. Then if we define $\tilde{F}^{G_q}$ similarly to as before, we have a composite function from all strings to n bits. We know that individually each $\tilde{F}^{G_q}$ is a PRF given F is a PRF. We need to assure that these different functions are almost independent. We prove that if the family of graphs satisfy certain constraints then this is indeed the case. We defer the details to section 6. Also, this general theorem leads to interesting new applications which are discussed in section 7.

2 Definitions

Definition 1. For positive integers n, m, let $\mathcal{F}(n \to m)$ be the set of all functions from n bits to m bits.

Definition 2. (PRF) A *pseudo-random function* has signature

$$F : \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^l.$$

Define $\mathrm{Sec}_F(q, T)$ to be the maximum advantage an adaptive adversary can obtain when trying to distinguish between $F_K(\cdot)$ (with K chosen uniformly at random) and a function chosen uniformly at random from $\mathcal{F}(n \to l)$, when given q queries and time T.

3 Domain Extension using arbitrary acyclic graphs

Definition 3. Let $G = (V, E)$, be a directed acyclic graph (DAG) [11] with a finite vertex set V and edges E. A node u is said to be **incident** on a node v, if there is an edge from u to v, i.e $E(u, v)$. Such an edge will sometimes be denoted $\langle u, v \rangle$. Define a DAG to be **non-redundant** if for every pair of nodes, the set of their incident nodes is different. For two vertices u and v, we say that $u \prec v$ if there is a directed path from u to v. Since G is a finite DAG, the relation $\prec$ is a finite partial order.

Definition 4. Given a function F from n bits to n bits, and a non-redundant DAG $G = (V, E)$ with only one source node and only one sink node, and a total of m nodes, define $F^G : \{0,1\}^{nm} \to \{0,1\}^n$ as follows:

- Let the input to F^G be mn bit string P, which is divided into m n-bit strings $P_1, P_2, ..., P_m$.
- Since $|V| = m$, let V_1,....,V_m be an enumeration of the nodes. When it is clear from context, we will identify the index of a vertex with the vertex itself. Let the unique source node be V_1, and the unique sink node be V_m.
- For the unique source node, define $M_1 = P_1$.
- For every non-source node V_j, $j > 1$, inductively define $M_j = P_j \oplus_{u:E(u,j)} F(M_u)$
- For notational convenience, for every node V_j, let C_j denote $F(M_j)$.
- The output of the function F^G is just C_m.

It is clear that the restriction of one sink node is crucial, for if there was another sink node other than V_m, then the plaintext fed into this other sink node has no influence on C_m. It is possible that there are instances of DAGs G with *two source nodes* such that F^G is a PRF; however, a more stringent requirement than non-redundancy will definitely be required. Consider a DAG G, with two source nodes V_1 and V_2, both with only one outgoing edge and that too to the same vertex. Then, the resulting function is clearly not a PRF. A similar situation motivates the requirement of non-redundancy.

One may be tempted to weaken the non-redundancy requirement. For instance, one idea is to have the condition on the DAG that it have no non-trivial automorphism. However, such a DAG may not yield a secure PRF, as illustrated in Figure 8. The two queries $\langle p1, p2, p2, p4, p5, p6 \rangle$ and $\langle p1, p2, p2, p5, p4, p6 \rangle$ yield the same result.

Theorem 1. *For a non-redundant DAG $G = (V, E)$ with unique source and sink nodes, and m total nodes, let F^G be as above. Then, no adaptive adversary, with q queries, can distinguish between (a) F^G where F is chosen uniformly at random from $\mathcal{F}(n \rightarrow n)$, (b) and a function chosen uniformly at random from $\mathcal{F}(nm \rightarrow n)$, with probability more than $(mq)^2 2^{-(n+1)}$.*

In the next section, we state and prove a more general theorem.

4 Domain Extension using colored DAGs and GF(2^n)

If we allow Galois field arithmetic, we get an even more general construction, and a corresponding PRF domain extension theorem. Assuming that the underlying function F has an n-bit output, we will use the Galois field GF(2^n). Such fields have the property that they have exactly 2^n elements. Moreover, each element can be represented as a n bit vector, with addition in the field being just the bitwise xor ($\oplus$). Since multiplication distributes over addition in a field, it follows that if a, b and c are three elements in the field then $a * (b \oplus c) = a * (b + c) = (a * b) + (a * c) = (a * b) \oplus (a * c)$. A further useful property of finite fields is that for a fixed non-zero a in the field, if b is picked uniformly at random from the field, then $a * b$ is also uniformly distributed in the field.

Definition 5. Let $G = (V, E)$, be a directed acyclic graph (DAG). Let $|V| = m$. A coloring χ of the edges of the graph is a map $\chi : E \rightarrow [1..m]$. The triple (V, E, χ) will be called an **edge-colored** DAG. Define an edge-colored DAG to be **non-singular** if for every pair of nodes u, v, if the set of their incident nodes is same (say W), then at least for one $w \in W$, $\chi(\langle w, u \rangle) \neq \chi(\langle w, v \rangle)$. For two vertices u and v, we say that $u \prec v$ if there is a directed path from u to v. Since G is a finite DAG, the relation $\prec$ is a finite partial order.

Definition 6. Given a function F from n bits to n bits, and a non-singular edge-colored DAG $G = (V, E, \chi)$ with only one source node and only one sink node and a total of $m < 2^n$ nodes, define $F^G : \{0,1\}^{nm} \rightarrow \{0,1\}^n$ as follows:

- Since $m < 2^n$, we can view χ as a map from E to GF(2^n)*, i.e. the non- zero elements of the field.
- Let the input to F^G be mn bit string P, which is divided into m $n$bit strings $P_1, P_2, ..., P_m$.
- Since $|V| = m$, let V_1,.....,V_m be an enumeration of the nodes When it is clear from context, we will identify the index of a vertex with the vertex itself. Let the unique source node be V_1, and the unique sink node be V_m.
- For the unique source node, define $M_1 = P_1$.
- For every non-source node V_j, $j > 1$, inductively define $M_j = P_j + \sum_{u:E(u,j)} \chi(\langle u, j \rangle) * F(M_u)$, where $F(M_u)$, which is an n-bit quantity, is viewed as an element of GF(2^n). The summation is addition in the field, which is the same as n-bit xor.
- For notational convenience, for every j, we denote $F(M_j)$ by C_j.
- The output of the function F^G is just C_m.

Theorem 2. *: (Main Theorem) For a non-singular edge-colored DAG $G = (V, E, \chi)$ with unique source and sink nodes, and $m < 2^n$ total nodes, let F^G be as above. Then, no adaptive adversary, with q queries, can distinguish between (a) F^G where F is chosen uniformly at random from $\mathcal{F}(n\to n)$, (b) and a function chosen uniformly at random from $\mathcal{F}(nm\to n)$, with probability more than $(mq)^2 2^{-(n+1)}$.*

Theorem 3. *Given a PRF $F : \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^n$, and a non-singular edge-colored DAG $G = (V, E, \chi)$ with unique source and sink nodes, and $m < 2^n$ total nodes, a function $F^G : \{0,1\}^k \times \{0,1\}^{mn} \to \{0,1\}^n$ can be defined be letting for each K, $(F^G)_K$ to be $(F_K)^G$ (as in definition 6). Then,*

$$Sec_{F^G}(q,T) \leq Sec_F(q,T) + (mq)^2 2^{-(n+1)}$$

The proof follows from Theorem 2 by standard techniques.

Before we give the proof of theorem 2, we need to fix more notation and give a general idea of the proof. We first note that we allow arbitrary functions as adversaries and not just computable functions. Then without loss of generality, we can assume that the adversary is deterministic, as every probabilistic adversary is just a probability distribution over all deterministic adversaries[14].

Since we are going to show the no adaptive adversary can distinguish, fix an adaptive adversary. Since the adversary is deterministic, the first query's plaintext (say $P^1 = \langle P_1^1, ..., P_m^1 \rangle$) is fixed for that adversary. Thus, the first query's output, say C_m^1 is only a function of F. The adversary being adaptive, its second query is a function of C_m^1. But, since C_m^1 is only a function of F, the second query's plaintext can also be written just as a function of F. Thus, C_m^2 is only a function of F, and so forth.

Notation. We will denote probabilities under the first scenario, i.e. (a) in the theorem 2 statement, as Pr, and the probabilities in the second scenario, i.e. (b) in the theorem 2 statement, as $\Pr_{(b)}$. Most of the analysis will be devoted to the first scenario. So, unless otherwise mentioned, all random variables from now on are in the first scenario.

All random variables will be denoted by *upper case letters*. A constant value which a random variable can take will be denoted by the corresponding *small case letter*. For all random variables corresponding to a query, we will use superscripts to denote the query number. Subscripts will be used to denote blocks within a query. The random variables will be as in Definition 6, i.e. P standing for plaintext input, M standing for the variable on which the F function is applied, and C standing for the output of the F function.

Thus, for i in $[1..q]$, we will use C^i to denote the sequence C_1^i,...,C_m^i, where C_j^i is $F(M_j^i)$ as in definition 6. We will use C_j^* to denote the sequence C_j^1,...,C_j^q, i.e. the jth blocks from all the queries. We will use C to denote the sequence $C^1, ..., C^q$. Thus, C denotes the whole transcript of F outputs. More precisely, this random variable and other such random variables should be written $C(F)$, as it is a function of F and only F as argued above. However, we will drop the arguments when it is clear from context. For a fixed f, we will write it as $C(f)$. Small case c, by the convention above, denotes a fixed transcript.
(*end of notation*)

Since the adversary is adaptive, the variables P_j^i are a function of C (more precisely C_m^*). Although as argued above, P_j^i is ultimately a function of F, it will be convenient to write P_j^i as functions of only C. Thus,

M_j^i **can be viewed as a function of only** C. and we will write it as $M_j^i(C)$. For a fixed c, we will write it as $M_j^i(c)$.

Having fixed the notation, lets try to get a sense of what we are trying to prove, and how we may get there. Since the adversary decides 0 or 1 based on the oracle replies, i.e. C_m^* (in scenario (b), C_m^* is just a uniformly random string of length qm), the adversary's output is a random variable, say $A(C_m^*)$. We want to show that

$$|\,\Pr_F[A(C_m^*) = 0] - \Pr_{(b)}[A(C_m^*) = 0]\,|$$

is small. For any qn length constant string r_m^*, on the condition that all the M_m^i are distinct (call this condition D) in scenario (a), we would like to prove that $\Pr_F[C_m^* = r_m^* \;\wedge\; D]$ is same as $\Pr_{(b)}[C_m^* = r_m^*] * \Pr_F[D]$. Even if we prove this, estimating $\Pr_F[D]$ is not easy as the adversary is adaptive. There are many ways one could try to prove this, but they all are erroneous. For instance, one would like to argue that $\Pr_F[C_m^* = r_m^* \mid D]$ is same as $\Pr_{(b)}[C_m^* = r_m^*]$. But this would only hold if one can show that the condition D has not put additional constraints on F. The condition D depends on the whole of C, and the condition that it holds puts additional constraints on C and hence on F.

Since D depends on whole of C, it maybe fruitful to estimate for each mqn bit constant c, $\Pr_F[C(F) = c \;\wedge\; D]$. Let us also *generalize* D to *all* M_j^i being distinct now. Consider a fixed transcript c. Since c is the whole transcript, it contains c_m^*, and hence the plaintext is fixed (as the adaptive adversary's plaintext choice depends on c_m^*). Since M_j^i is a function of only the plaintext and c, each $M_j^i(c)$ is also fixed. Thus, D is either true or false, independent of F. Thus the predicate D can be written as a function of c: $D(c)$. For each c such that $D(c)$, we then have that $\Pr_F[C(F) = c \;\wedge\; D]$ is same as 2^{-mqn}. Thus, we have made some progress.

However, what we really want is an average over all c, as $\Pr_F[C_m^* = r_m^* \;\wedge\; D]$ is same as sum over all c: $\Pr_F[C(F) = c \;\wedge\; C_m^* = r_m^* \;\wedge\; D]$. Thus, we will need to determine for how many c, the predicate $D(c)$ holds. Recall that the plaintext is a function of c_m^*, and hence for a fixed c_m^* (the plaintext being fixed) the adversary cannot force all c to fail $D(c)$. In fact, the intuition is that for a fixed c_m^* it can only force a few c to fail $D(c)$. There is a caveat though; if the adversary retains the same plaintext (except for the last few blocks) over two queries, then it is **forcing** $D(c)$ to fail, regardless of c. This suggests that our definition of $D(c)$ may be too strong. We will weaken the definition of $D(c)$ by allowing M_j^i and $M_j^{i'}$ to agree if the adversary is forcing these values to be same (call this predicate PD; see the precise definition of PD in definition 8 in the next section). However, now as opposed to the previous para, for each c such that $\mathrm{PD}(c)$, we *do not* have that $\Pr_F[C(F) = c \wedge \mathrm{PD}]$ is same as 2^{-mqn}. This is because even though $\mathrm{PD}(c)$ requires all "unforced" M to be distinct, the c values must be consistent at the blocks where M values are forced to be equal (since $c = C(F)$).

Instead of calculating $\Pr_F[C(F) = c \;\wedge\; \mathrm{PD}]$ for each c such that $\mathrm{PD}(c)$ holds, and then estimating the number of c for which $\mathrm{PD}(c)$ holds, we will calculate the *average* probability (over c) as required above, i.e. $\Pr_{c,F}[C(F) = c \wedge C_m^* = r_m^* \;\wedge\; \mathrm{PD}]$ directly. Before we can proceed we need to precisely define the notion of a consistent transcript.

Given a constant c, let the plaintext chosen by the adversary be p. For any vertices j, j', and query indices

i, i' we say that $(i,j) \equiv_c (i',j')$ if

$$(j = j') \text{ and } \forall k \preceq j : p^i_k = p^{i'}_k$$

That this is an equivalence relation is easy to see. Also define

$$\mu_c(i,j) = \min\{i' | (i',j) \equiv_c (i,j)\}.$$

Thus, for each vertex j, $\mu_c(i,j)$ maps query i to the smallest query number i' s.t. all plaintext blocks till j are same in i and i'. We call c **consistent** (con(c)) if

$$\forall j \in [1..m], \forall i, i' \in [1..q],\ (i,j) \equiv_c (i',j)\ : c^i_j = c^{i'}_j$$

Let $I = \{(i,j) | \mu_c(i,j) = i\}$ be the *core* index set of c. Let $l = |I|$. Going back to the average probability above, the key technical idea (lemma 4 of the next section) for the rest of the proof as follows. Firstly for any c, for $C(F) = c$ to hold c must be consistent, and hence there are exactly $(mq - l)$ n-bit linear constraints on c. We also show that for every consistent c, a function f_c can be defined using the M and the c values at core indices I such that $C(f_c) = c$. For this, it is important that there are no collisions among the M values at indices I. Moreover, such an f_c is *unique* on these l input values M. Thus, there are exactly l n-bit constraints on F such that $C(F) = c$. Thus, for a uniformly chosen c, on the condition that there were no collisions, there are a total of mq n-bit constraints, and hence the above average probability is 2^{-mqn}. All that remains is to estimate the probability of collisions in the M values at indices I of a uniformly chosen c (lemma 5 in the next section). This is a much easier problem, as there is no adversary involved in this. We do this argument rigorously in the next section.

5 Proof of Main Theorem

We first collect all the key definitions from the end of the previous section.

Definition 7. For any vertices j, j', and query indices i, i' we say that $(i,j) \equiv_c (i',j')$ if

$$(j = j') \text{ and } \forall k \preceq j : p^i_k = p^{i'}_k$$

Define

$$\mu_c(i,j) = \min\{i' | (i',j) \equiv_c (i,j)\}.$$

We call c **consistent** (con(c)) if

$$\forall j \in [1..m], \forall i, i' \in [1..q],\ (i,j) \equiv_c (i',j)\ : c^i_j = c^{i'}_j$$

Define the following "correcting" function ρ from mq n-bit blocks to mq n-bit blocks:

$$\rho(c) = \overline{c},\ \text{ where } \overline{c}^i_j = c^{\mu_c(i,j)}_j.$$

Fact 1(e) below shows that for each c there is a consistent b, whereas fact 1(f) shows that for each b there is only one consistent c, such that b agrees with c at core indices.

Fact 1: For all $i, i' \in [1..q]$, $i \neq i'$, for all $j \in [1..m]$ and mqn bit constant transcript c:
(a) $(i,m) \not\equiv_c (i',m)$, i.e. $\mu_c(i,m) = i$,
(b) $\equiv_c$ is an equivalence relation,
(c) $\mu_c(\mu_c(i,j),j) = \mu_c(i,j)$,
(d) $\mu_c = \mu_{\rho(c)}$,
(e) $\rho(c)$ is consistent,
(f) Let c be consistent, and let b be such that for all i : $\mu_c(i,j) = i$: $b^i_j = c^i_j$. Then $\rho(b) = c$,
(g) For $u \preceq j$, $\mu_c(i,u) = \mu_c(\mu_c(i,j),u)$.
(h) For consistent c, $M^i_j = M^{\mu_c(i,j)}_j$
(i) $C(F)$ is consistent.
Proof: see Appendix A.

The condition D from the previous section needs to be extended. Consider the following event **PD** (*pairwise different*).

Definition 8. For any constant c, define PD(c) to be

$$\forall i, i' \in [1..q], \forall j, j' \in [1..m],\ j \neq j'\ :\ M^i_j(c) \neq M^{i'}_{j'}(c),$$

$$\text{and}\quad \forall i, i' \in [1..q], \forall j \in [1..m]\ :\ (i,j) \not\equiv_c (i',j) \Rightarrow M^i_j(c) \neq M^{i'}_j(c).$$

Note that Ms are required to be distinct only if $(i,j) \not\equiv_c (i',j)$. If we did not have this condition then (see lemma 6) we will not be able to prove that PD happens with high probability.

Any c such that PD(c) can be used to define a function, denoted f_c, such that C(f_c) is actually same as $\rho(c)$. Thus, for consistent c it will turn out to be same as c. We will directly prove this for a consistent c.

Definition 9. For each c, such that PD(c) holds, define f_c as follows. Let $I = \{(i,j) \mid \mu_c(i,j) = i\}$ be the *core* index set. For $(i,j) \in I$, define $f_c(M^i_j(c)) = c^i_j$. This is well defined as PD(c) holds. We will not need to define f_c on other values.

Fact 2: For any consistent c such that PD(c) holds:

$$C(f_c) = c$$

Proof: see Appendix A.

Lemma 4. *(PRF Technical Lemma) For every qn bit constant r^*_m*

$$Pr_{c \in_U \{0,1\}^{mqn}, F}[C(F) = c\ \wedge\ PD(c) \mid c^*_m = r^*_m]\ =\ 2^{-mqn} * Pr_{c \in_U \{0,1\}^{mqn}}[PD(\rho(c)) \mid c^*_m = r^*_m]$$

Proof: We first show that the LHS above is same as

$$\Gamma = \Pr_{c,b \in_U \{0,1\}^{mqn}}[b^i_j = c^i_j|_{(i,j):\mu_c(i,j)=i}\ \wedge\ \mathrm{con}(c)\ \wedge\ \ \mathrm{PD}(c) \mid c^*_m = r^*_m]$$

By fact 1(i), the conjunct con(c) can be added to the LHS of the lemma. We show that the two probabilities are same for every constant c. So, fix a c. As before, let $I = \{(i,j) \mid \mu_c(i,j) = i\}$ be the core index set of c.

Let $S = \{M_j^i(c) \mid (i,j) \in I\}$. Since PD(c) holds, $|S| = |I|$. Let S' be an arbitrary set of n bit strings, disjoint from S, and $|S'| = mq - |I|$. Thus, $|S \cup S'| = mq$.

By fact 2, $C(f_c) = c$. Thus, for each b agreeing with c on I, we have a function f_c defined on $|I|$ inputs S, such that $C(f_c) = c$. We can use the remaining $mq - |I|$ values of b (i.e. from indices which are not in I) to extend f_c to be defined on $S \cup S'$. This map from b to the extended f_c is 1-1.

Similarly, for any function f defined on $S \cup S'$, such that $C(f) = c$ (note that f need only be defined on S for $C(f)$ to be well defined), we can define a mqn-bit long b which agrees with c on I. For indices in $(i,j) \in I$, use $f(M_j^i(c))$ to define b_j^i, and use $f(s)$, $s \in S'$, to define the remaining part of b. This map from f to b is also 1-1. This shows that the LHS of the statement of the lemma is same as Γ.

We next show that, the RHS of the statement of the lemma is same as Γ. To this end, we show that the following two sets are equinumerous, i.e. we show a bijection between the two sets. The first set is

$$\mathcal{C} = \{c | c \in \{0,1\}^{mqn}, \mathrm{PD}(\rho(c)), \text{ and } c_m^* = r_m^*\}$$

The second set is

$$\mathcal{D} = \{(c,b) | c,b \in \{0,1\}^{mqn}, b_j^i = c_j^i|_{(i,j):\mu_{c,j}(i,j)=i}, \ \mathrm{con}(c), \ PD(c), \text{ and } c_m^* = r_m^*\}$$

That they are equinumerous follows easily from facts 1(e), 1(f), 1(a) and 1(d), but to be rigorous consider the following extension of ρ to a function $\hat{\rho}$ from $\mathcal{C}$ to $\mathcal{D}$.

$$\hat{\rho}(c) = (\rho(c), c)$$

It needs to be shown that the function has $\mathcal{D}$ as its range, is 1-1 and onto. The function is obviously 1-1. To prove that its range is $\mathcal{D}$, we need to prove three things:
(1) $\rho(c)$ is consistent: follows by fact 1(e).
(2) $c_j^i = \rho(c)_j^i|_{\mu_{\rho(c)}(i,j)=i}$: follows directly from definition of ρ and fact 1(d).
(3) $\forall i$, $\rho(c)_m^i = r_m^i$: by fact 1(a) and definition of ρ we have $\rho(c)_m^i = c_m^i$; and hence $c_m^i = r_m^i$ implies $\rho(c)_m^i = r_m^i$.

To prove that it is onto, for any (c,b) in $\mathcal{D}$, we show that b is in $\mathcal{C}$ and $\hat{\rho}(b) = (c,b)$. But for any (c,b) in $\mathcal{D}$, by fact 1(f), $\rho(b) = c$. Thus, $\hat{\rho}(b) = (c,b)$. It also follows that PD($\rho(b)$) holds. Moreover, by fact 1(a), $b_m^* = c_m^*$. Thus b is in $\mathcal{C}$. □

Fact 3: For any mqn bit constant c let p be its corresponding plaintext.

If for all u s.t. $E(u,j)$, $\mu_c(i,u) = \mu_c(i',u)$, and $p_j^i = p_j^{i'}$, then $\mu_c(i,j) = \mu_c(i',j)$.
Proof: see Appendix A.

We will denote by Δ the quantity $(mq)^2 2^{-(n+1)}$.

Lemma 5. *For every qn bit constant r_m^*,*

$$Pr_{c \in_{\mathcal{U}} \{0,1\}^{mqn}}[\ PD(\rho(c)) \mid c_m^* = r_m^*] \ \geq 1 - \Delta$$

Proof: First note that for all i, $c_m^i = \rho(c)_m^i$, by fact 1 (a) and definition of ρ. Thus, once c_m^* is fixed (and hence $\rho(c)_m^*$) is fixed to r_m^*, the plaintext p is fixed, independent of other c_i ($i < m$). We will prove the lemma by upper bounding the probability of ¬PD by union bound.

For each vertex j, let V_j be its set of incident vertices, i.e. $V_j = \{u | E(u,j)\}$. Recall,

$$M^i_j(\rho(c)) = p^i_j + \sum_{u:E(u,j)} \chi(\langle u,j\rangle) * c_u^{\mu_c(i,u)}$$

If $j \neq j'$, and $V_j \neq V_{j'}$, wlog let $w \in V_j$ and $w \notin V_{j'}$. Then $M^i_j(\rho(c)) = M^{i'}_{j'}(\rho(c))$ iff

$$\chi(\langle w,j\rangle) * c_w^{\mu_c(i,w)} \;=\; p^i_j + p^{i'}_{j'} + \sum_{u:E(u,j),u\neq w} \chi(\langle u,j\rangle) * c_u^{\mu_c(i,u)} + \sum_{u:E(u,j')} \chi(\langle u,j'\rangle) * c_u^{\mu_c(i',u)}$$

Since, $c_w^{\mu_c(i,w)}$ does not appear on the RHS, and $w < m$, and $\chi(\langle w,j\rangle) \neq 0$, the probability of above is 2^{-n}.

If $j \neq j'$, and $V_j = V_{j'}$, then for some $w \in V_j$, $\chi(\langle w,j\rangle) \neq \chi(\langle w,j'\rangle)$, as the underlying graph G is non-singular. Thus, similarly to the argument above, $M^i_j = M^{i'}_{j'}$ happens with probability 2^{-n}.

When j equals j' (and $i \neq i'$), we have three cases. If for some u incident on j ($E(u,j)$), $\mu_c(i,u) \neq \mu_c(i',u)$, then the probability of the two Ms being equal is at most 2^{-n}. Otherwise, if $p^i_j \neq p^{i'}_j$, then the probability is zero. If $p^i_j = p^{i'}_j$, we have $\mu_c(i,j) = \mu_c(i',j)$ by fact 3, and hence the corresponding disjunct in ¬PD is false.

Since all the probabilities are 2^{-n} or zero, the bound in the lemma follows. □

Lemma 6.

$$Pr_F[\; PD(C(F))] \;\geq\; 1 - \Delta$$

Proof: see Appendix A.

Since the adversary A decides 0 or 1 based on the oracle replies, i.e. C^*_m, we can write its output as $\mathrm{A}(C^*_m)$. Recall, $\mathrm{Pr}_{(b)}$ is the probability under oracle (b), i.e. when the oracle is a random function with range n bits.

Lemma 7.

$$Pr_{(b)}[A(C^*_m) = 0] \geq Pr_F[A(C^*_m) = 0 \;\wedge\; PD(C(F))] \geq (1-\Delta) Pr_{(b)}[A(C^*_m) = 0]$$

Proof: see Appendix A.

Proof of Theorem 2 (Main Theorem): By lemma 7 and lemma 6 it follows that

$$|\,\mathrm{Pr}_{F,H}[A(C^*_m) = 0] - \mathrm{Pr}_{(b)}[A(C^*_m) = 0]\,| \;\leq\; \Delta$$

□

6 Variable Length Domain Extension and Family of Graphs

We consider a fixed n throughout the rest of this section. We will assume that we are only interested in domain extension up to length $2^n * n$ bits, as theorem 2 is ineffective beyond that length (this restriction is only for sake of simplicity). Each query of the adversary will be a string p of length q bits, $(0 < q < 2^n * n)$. We let the composite function answers the query as follows: If q is a multiple of n, then it returns $F^{G_q}(p)$. Otherwise, let p' be p appended with 10^i, where i is the smallest positive number to make $|p'|$ a multiple of n. The composite function then returns $F^{G_q}(p')$.

For every $0 \leq l < 2^n$, since strings of length $ln + 1$ to $ln + n - 1$ bits get canonically encoded in the above method, we can use the same graph for all these lengths. Thus, for each l, we really need only two graphs ([6]), one for lengths $ln + 1$ to $ln + n - 1$, and one for length $ln + n$. From now on, we will assume that all plaintexts are of bit length multiples of n. Each adversarial query will be a pair: (p, z), where p is a bit string of length multiple of n, and z is in $\{0, 1\}$ (we can generalize z to be in an arbitrary finite set, but for our application this suffices). Note that we can no longer assume that p does not repeat, though we can assume that (p, z) does not repeat.

Definition 10. Let S be the set of all binary strings of length non-zero multiples of n, but less than $2^n * n$. Let $\mathcal{F}$ be the set of all functions:

$$S \times \{0,1\} \rightarrow \{0,1\}^n$$

Let $\tilde{F}$ be a function with signature:

$$\{0,1\}^k \times S \ \times \{0,1\} \rightarrow \{0,1\}^n$$

Given a PRF F from n bits to n bits, we need to define $\tilde{F}$ such that no adaptive adversary can distinguish between $\tilde{F}_K$, with K chosen randomly, and a function chosen uniformly at random from $\mathcal{F}$. As in the previous sections, given a function F from n bits to n bits, and given a collection of graphs $\mathcal{G}$, we first define a function $F^{\mathcal{G}}$ in $\mathcal{F}$.

Definition 11. Let $\mathcal{G}$ be a collection of edge-colored DAGs $G(l)$ (see definition 5), $l \leq (2^n - 1) * 2$. Each $G(l)$ is required to have unique source and sink nodes. Each $G(l)$ must have at least $\lceil \frac{l}{2} \rceil$ nodes. Define a function $F^{\mathcal{G}}$ as follows:

$$F^{\mathcal{G}}(p, z) = F^{G(2*|p|-z)}(p)$$

where F^G is as in definition 6. If the graph has more nodes than the length of the plaintext, then append enough zeroes to the plaintext. Usually, graphs will have exactly the required number of nodes. However, at the base cases, i.e. small length plaintexts, it may be necessary to have extra nodes.

Thus, if p repeats we definitely use a different graph. For a theorem similar to theorem 2 to hold, we need further restrictions on $\mathcal{G}$. In particular, it will not be enough that individual graphs in $\mathcal{G}$ be non-singular. Since, we will need to extend the notion of non-singularity to the whole collection of graphs, it is best to fix a set of vertices V, and just define the edges and colorings for the individual graphs. Thus, we will define $E(l)$, and $\chi(l)$. The partial order $\prec_l$ is, as before, just the transitive closure of $E(l)$.

To motivate the generalized definition of non-singularity, we first consider an example (see fig 4 in appendix C) where it is not enough for individual graphs to be non-singular. We have $V = [1..4]$. Ignore the colorings for now. The graphs are identical, except that the second graph $G(2)$ has an extra edge from 3 to 4. The first graph $G(1)$ is used to answer queries of length 3 blocks, and the second to answer queries of length 4. Clearly, both graphs are individually non-redundant. Consider two queries, one of length three, and another of length four, the latter being just an extension of the first. However, the first graph's output is C_3, and is accessible to the adversary. Thus, during the second query the internal state C_3 is available to the adversary, and it can force M_4 to be any value of its choice.

This suggests that for each graph $G(i)$, it cannot be allowed to be an induced subgraph of another graph $G(i')$. We prove that this condition is sufficient for the composite function to be a PRF.

Because of lack of space, we formalize this condition, state the theorem, and prove the theorem in the appendix (see Appendix B).

7 Applications to Variable Length Domain Extension

As an application of Theorem 8 (appendix B), we get the variable length domain extension scheme as described in figure 3 (see appendix C). In the figure, for each plaintext block length two graphs are given as required in definition 11. The number on the left of the graphs denotes the block length applicable to those graphs. We have only illustrated graphs up to length five, as for larger lengths, we follow similar methods as for length four and five. This mode has an advantage over XCBC [6], and OMAC [12] that it does not even need to employ the initial F on a constant like 0^n. Moreover,the scheme shows that if the plaintexts are restricted to be more than 3 blocks in length, then no Galois field arithmetic is required.

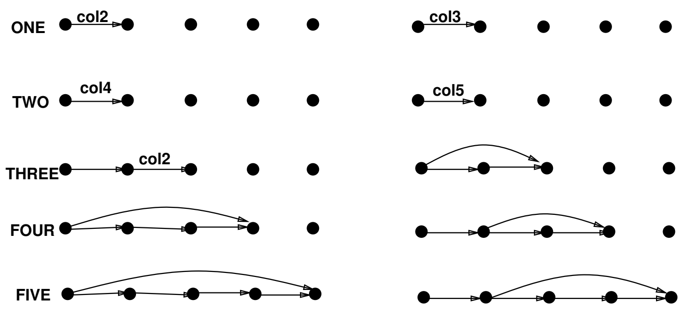

Fig. 3. A Variable Length Mode

References

1. ANSI X3.106, "American National Standard for Information Systems - Data Encryption Algorithm - Modes of Operation", *American National Standards Institute, 1983.*

2. M . Bellare, R. Canetti, H. Krawczyk, " Pseudorandom Functions Revisited: The Cascade Construction and its Concrete Security", Proc. IEEE FOCS 1996.
3. M. Bellare, J. Kilian, P. Rogaway, "The Security of Cipher Block Chaining", *JCSS*, Vol. 61, No. 3, Dec 2000, pp. 362-399
4. D. Bernstein, " How to Stretch Random Functions: The security of Protected Counter Sums", J. of Cryptology, Vol 12,No. 3, (1999).
5. J. Black, P. Rogaway, " A Block Cipher Mode of Operation for Parallelizable Message Authentication", Proc. Eurocrypt 2002.
6. J. Black, P. Rogaway, "CBC MACs for arbitrary length messages: The three key constructions". CRYPTO 2000, LNCS 1880.
7. J. Carter, M. Wegman, "Universal Classes of Hash Functions", *JCSS*, Vol. 18, 1979, pp 143-154.
8. O. Goldreich, S. Goldwasser, and S. Micali, " How to construct random functions", J. ACM, vol. 33, no. 4, 1986.
9. National Bureau of Standards, Data Encryption Standard, U.S. Department of Commerce, FIPS 46 (1977)
10. V.D. Gligor, P. Donescu, "Fast Encryption Authentication: XCBC Encryption and XECB Authentication Modes",
`http://csrc.nist.gov/encryption/modes/workshop1`
11. F. Harary, *Graph Theory*, Addison-Wesley 1969.
12. T. Iwata, K. Kurosawa, " OMAC: One -key CBC-MAC", FSE 2003, LNCS 2887.
13. Hugo Krawczyk, "LFSR-based Hashing and Authentication", *Proc. Crypto 94*, LNCS 839, 1994
14. H.W. Kuhn, "Extensive games and the problem of information" in *Contributions to the Theory of Games II*, H.W. Kuhn and A. W. Tucker eds., Annals of Mathematical Studies No. 28, Princeton Univ. Press, 1950.
15. M. Luby, "Pseudorandomness and Cryptographic Applications", *Princeton Computer Science Notes*, Princeton Univ. Press, 1996
16. E. Petrank, C. Rackoff, "CBC-MAC for real-time data sources", J. of Cryptology, vol 13, no. 3, nov 2000.

Appendix A

Fact 1: For all $i, i' \in [1..q]$, $i \neq i'$, for all $j \in [1..m]$ and mqn bit constant transcript c:
(a) $(i,m) \not\equiv_c (i',m)$, i.e. $\mu_c(i,m) = i$,
(b) $\equiv_c$ is an equivalence relation,
(c) $\mu_c(\mu_c(i,j),j) = \mu_c(i,j)$,
(d) $\mu_c = \mu_{\rho(c)}$,
(e) $\rho(c)$ is consistent,
(f) Let c be consistent, and let b be such that for all $i:\ \mu_c(i,j) = i$: $b^i_j = c^i_j$. Then $\rho(b) = c$,
(g) For $u \preceq j$, $\mu_c(i,u) = \mu_c(\mu_c(i,j),u)$.
(h) For consistent c, $M^i_j = M^{\mu_c(i,j)}_j$
(i) $C(F)$ is consistent.

Proof: (a) As we have assumed, wlog, that the adversary does not repeat queries, it follows that i and i' ($i \neq i'$) can never be equivalent over all vertices V. In particular, it is not the case that $(i,m) \equiv_c (i',m)$. To see this, note that we have assumed that the graph has only one sink node, i.e. V_m. It follows that for every node j, $j \preceq m$, hence the claim.
(b) & (c) straightforward.
(d) Note that the adversary's choice of p depends only on c^*_m. So we first show that for all i, $\rho(c)^i_m = c^i_m$. This follows as $\mu_c(i,m) = i$ by (a). Thus p remains same for $\rho(c)$.

(e) We just note that for all i, i', $(i,j) \equiv_c (i',j)$ implies $\mu_c(i,j) = \mu_c(i',j)$. Thus, by definition of ρ, we have $\rho(c)^i_j = \rho(c)^{i'}_j$.
(f) We first note that, since by (a), $\mu_c(i,m) = i$, we have $b^i_m = c^i_m$. Thus, as in proof of (d) above, $\mu_b = \mu_c$. Now, $\rho(b)^i_j = b_j^{\mu_b(i,j)} = b_j^{\mu_{c,j}(i)} = c_j^{\mu_{c,j}(i)}$, the last equality following from (c) and condition on b. For consistent c, this is same as c^i_j.
(g) For $u \preceq j$, $(i,j) \equiv_c (i',j)$ implies $(i,u) \equiv_c (i',u)$. So, let $i' = \mu_c(i,j)$. Then, $(i,u) \equiv_c (\mu_c(i,j),u)$.
(h) $M^i_j = p^i_j + \sum_{u: E(u,j)} \chi(\langle u,j\rangle) * c^i_u$. First note that $p^i_j = p_j^{\mu_c(i,j)}$. Also, for consistent c and $u \preceq j$, $c^i_u = c_u^{\mu_c(i,u)} = c_u^{\mu_c(\mu_c(i,j),u)}$ by (g). Again by consistency of c, the latter is same as $c_u^{\mu_c(i,j)}$. This shows that $M^i_j = M_j^{\mu_c(i,j)}$.
(i) by induction on the finite partial order $\prec$. □

Fact 2: For any consistent c such that PD(c) holds:

$$C(f_c) = c$$

Proof: For clarity sake, the M values, the plaintext values, and the C values in the evaluation of $C(f_c)$ will be denoted by a bar above them.
Base Case: Since the adversary is fixed, the first plaintext message is the same, i.e. $\bar{p}^1 = p^1$. Since $\bar{M}^1_1 = p^1_1$, $\bar{c}^1_1 = f_c(\bar{M}^1_1) = f_c(M^1_1) = c^1_1$, as $(1,1)$ is trivially in I. For $j > 1$, $\bar{M}^1_j = p^1_j + \sum_{u: E(u,j)} \chi(\langle u,j\rangle) * \bar{c}^1_u$ But, by induction over the partial order $\prec$, $\bar{c}^1_u = c^1_u$, hence $\bar{M}^1_j = M^1_j$. Moreover, $(1,j)$ is trivially in I, and hence $\bar{c}^1_j = c^1_j$.

So, assume that for all $i' < i$, and all j, $\bar{c}^{i'}_j = c^{i'}_j$. Thus, $\bar{p}^i = p^i$. Again, $\bar{M}^i_1 = p^i_1 = M^i_1$. Thus, $\bar{c}^i_1 = f_c(M^i_1) = f_c(M_1^{\mu_c(i,1)})$ by fact 1(h). By definition of f_c, this is same as $c_1^{\mu_c(i,1)} = c^i_1$. For $j > 1$, $\bar{M}^i_j = p^i_j + \sum_{u: E(u,j)} \chi(\langle u,j\rangle) * \bar{c}^i_u$. But, by induction over the partial order $\prec$, $\bar{c}^i_u = c^i_u$, thus $\bar{M}^i_j = M^i_j$. As before, using fact 1(h), we are done. □

Fact 3: For any mqn bit constant c let p be its corresponding plaintext.
If for all u s.t. $E(u,j)$, $\mu_c(i,u) = \mu_c(i',u)$, and $p^i_j = p^{i'}_j$, then $\mu_c(i,j) = \mu_c(i',j)$.
Proof: We just need to show that $(i,j) \equiv_c (i',j)$, from which the claim follows by fact 1(b). But again, $\mu_c(i,u) = \mu_c(i',u)$ implies $(i,u) \equiv_c (i',u)$ by fact 1(b). This along with $p^i_j = p^{i'}_j$ shows that p agrees in queries i and i' over all blocks $j' \preceq j$. □

Proof of Lemma 6:

$$\begin{aligned}
&\Pr_F[\ \mathrm{PD}(C(F))] \\
&= \sum_{r^*_m} \sum_{c} \Pr_F[C(F) = c \wedge \mathrm{PD}(c) \wedge c^*_m = r^*_m] \\
&= \sum_{r^*_m} \Pr_{c,F}[C(F) = c \wedge \mathrm{PD}(c) \wedge c^*_m = r^*_m] * 2^{mqn} \\
&= \sum_{r^*_m} \Pr_{c,F}[C(F) = c \wedge \mathrm{PD}(c) \mid c^*_m = r^*_m] * 2^{-qn} * 2^{mqn} \\
&= \sum_{r^*_m} 2^{-qn} * \Pr_c[\mathrm{PD}(\rho(c)) \mid c^*_m = r^*_m] \qquad \text{(by lemma 4)} \\
&\geq 1 - \Delta \quad \text{(by lemma 5)}
\end{aligned}$$

$\square$

Proof of Lemma 7: To begin with, we have

$$\begin{aligned}
&\mathrm{Pr}_F[A(C_m^*) = 0 \wedge \mathrm{PD}(C(F))\,] = \sum_c \mathrm{Pr}_F[A(c_m^*) = 0 \wedge C(F) = c \wedge \mathrm{PD}(c)] \\
&= 2^{mqn} * \mathrm{Pr}_{c \in_{\mathcal{U}} \{0,1\}^{mqn}, F}[A(c_m^*) = 0 \wedge C(F) = c \wedge \mathrm{PD}(c)] \\
&= 2^{mqn} * \mathrm{Pr}_{c \in_{\mathcal{U}} \{0,1\}^{mqn}, F}[C(F) = c \wedge \mathrm{PD}(c) \mid A(c_m^*) = 0] * \mathrm{Pr}_{c \in_{\mathcal{U}} \{0,1\}^{mqn}}[A(c_m^*) = 0]
\end{aligned}$$

The above is at least $(1 - \Delta)\mathrm{Pr}_{(b)}[A(C_m^*) = 0]$ by lemma 4 and 5, and at most $\mathrm{Pr}_{(b)}[A(C_m^*) = 0]$. $\square$

APPENDIX B

For sake of completeness, we repeat definitions 10 and 11 here.

Definition 10. Let S be the set of all binary strings of length multiples of n, but less than $2^n * n$. Let $\mathcal{F}$ be the set of all functions:

$$S \times \{0,1\} \rightarrow \{0,1\}^n$$

Definition 11. Let $\mathcal{G}$ be a collection of edge-colored DAGs $G(l)$, $l \leq (2^n - 1) * 2$. Each $G(l)$ is required to have unique source and sink nodes. Each $G(l)$ must have at least $\lceil \frac{l}{2} \rceil$ nodes. Define a function $F^{\mathcal{G}}$ as follows:

$$F^{\mathcal{G}}(p, z) = F^{G(|p|*(z+1))}(p)$$

where F^G is as in definition 6.

Definition 12. For any vertex j in V, let V_j^l be the set of incident vertices of j in $G(l)$.
For any vertex j in V, we say $(l, j) \cong (l', j)$ if either $(j = 1)$ or
- $V_j^l = V_j^{l'}$, and
- for all $u \in V_j^l$: $\chi_l(\langle u, j\rangle) = \chi_{l'}(\langle u, j\rangle)$, and inductively $(l, u) \cong (l', u)$.

Essentially, (l, j) is congruent to (l', j) if the two graphs $G(l)$ and $G(l')$ are identical till j.

Definition 13. Let $\mathcal{G} = \langle G(l)\rangle$, where each $G(l) = (V, E(l), \chi(l))$ is an edge-colored DAG, be a collection of graphs.

- With each $G(l)$ we associate its size $m(l)$ to be the largest numbered node in V such that there is an edge directed to it in $G(l)$.
- For each $G(l)$ we define the graph $\tilde{G}(l) = ([1..m(l)], E(l), \chi(l))$, to be the induced subgraph of $G(l)$ on vertices $[1..m(l)]$.

The collection $\mathcal{G}$ is called **PRF-preserving** if

- each $\tilde{G}(l)$ has only one source node, one sink node, has at least $\lceil \frac{l}{2} \rceil$ nodes, and

– if for any pair of nodes u, v ($u \neq v$), and graphs $G(l)$ and $G(l')$, the set of incident nodes of u in $G(l)$, and the set of incident nodes of v in $G(l')$ are same (say W), then for at least one $w \in W$, $\chi_l(\langle w, u\rangle) \neq \chi_{l'}(\langle w, v\rangle)$.
– for each graph $G(l)$, it is not the case that there is another graph $G(l')$, $l' \neq l$, s.t. $(l, m(l')) \cong (l', m(l'))$

Basically, the second condition above has extended the non-singularity requirement to be over all graphs.

Theorem 8. *: For a PRF-preserving collection of* $2 * (2^n - 1)$ *DAGs* $\mathcal{G}$*, let* $F^{\mathcal{G}}$ *be as in definition 11. Then, no adaptive adversary, with* q *adaptive queries* $\langle (p^i, z^i)\rangle$ *(*$i \in [1..q]$*, and* $|p^i| \leq 2^n - 1$*), can distinguish between (a)* $F^{\mathcal{G}}$ *where* F *is chosen uniformly at random from* $\mathcal{F}(n \rightarrow n)$*, (b) and a function chosen uniformly at random from* $\mathcal{F}$*, with probability more than* $(\sum_{i \in [1..q]} |p^i|)^2 2^{-(n+1)}$*.*

Proof: To adapt the proof of theorem 2, we first need to redefine the notion of consistent transcripts c. First note that, on a fixed transcript c, the queries of the adversary are fixed. Recall, by definition of $F^{\mathcal{G}}$, on input p^i, z^i the graph $G(2 * |p^i| - z^i)$ is used. We just denote this graph by G^i. The corresponding edge relation, coloring and partial order will be denoted E^i, χ^i, and $\prec^i$ resp. Also, for the graph G^i, its induced subgraph as per definition 13, will be denoted $\tilde{G}^i$. Similarly, the size of the graph $\tilde{G}^i$ will be denoted by m^i. Note that $m^i = |c^i| \geq |p^i|$.

Definition 14. For any vertex j in V, let V^i_j be the set of incident vertices of j in G^i.
For any vertex j in V, we say $(i, j) \cong_c (i', j)$ if either $(j = 1)$ or
- $V^i_j = V^{i'}_j$, and
- for all $u \in V^i_j$: $\chi^i(\langle u, j\rangle) = \chi^{i'}(\langle u, j\rangle)$, and inductively $(i, u) \cong_c (i', u)$.
Essentially, (i, j) is congruent (wrt c) to (i', j) if the two graphs G^i and $G^{i'}$ are identical till j.

Once we generalize the definition of $\equiv_c$, rest of the definitions and proofs remain almost same.

Definition 15. For any vertices j, j', and query indices i, i' we say that $(i, j) \equiv_c (i', j')$ if

$$(j = j') \text{ and } (i, j) \cong_c (i', j) \text{ and } \forall k \preceq^i j : p^i_k = p^{i'}_k$$

As before, define

$$\mu_c(i, j) = \min\{i' | (i', j) \equiv_c (i, j)\}.$$

We call c **consistent** (con(c)) if

$$\forall j \in [1..2^n - 1], \forall i, i' \in [1..q],\ (i, j) \equiv_c (i', j)\ : c^i_j = c^{i'}_j$$

Define the following "correcting" function ρ:

$$\rho(c) = \overline{c}, \text{ where } \overline{c}^i_j = c^{\mu_c(i,j)}_j, \text{ for} j \in [1..m^i]$$

We will denote all facts and lemmas corresponding to theorem 5 by the prime symbol. In the proof of fact 1(a)$'$, if $m^i \neq m^{i'}$, then $(i, m) \ncong_c (i', m')$. Otherwise, if the plaintexts p^i and $p^{i'}$ are different, then again $(i, m^i) \not\equiv_c (i', m^i)$. If the plaintexts are also same, then as the adversary does not repeat queries, wlog let

$G^i = G(2*m^i - 1)$, and $G^{i'} = G(2*m^i)$. But $(i, m^i) \cong_c (i', m^i)$ is not allowed in $\mathcal{G}$ which is PRF-preserving. That proves fact $1(a)'$.

Proof of rest of fact $1'$ is similar to proof of fact 1. In the statement and proof of fact $1(f)'$, j must be restricted to be $[1..m^i]$. Similar restrictions apply in the definition of PD (definition 8) and definition of f_c (definition 9). Proof of fact $2'$ is similar to proof of fact 2.

Lemma 4 is now restated as (recall S from definition 10):

Lemma 9. *For every qn bit constant $\langle r^i \rangle$ $(i \in [1..q])$*

$$Pr_{c \in_U S^q, F}[C(F) = c \ \wedge \ \ PD(c) \mid c^i_{m^i} = r^i] \ = \ 2^{-mqn} * Pr_{c \in_U S^q}[PD(\rho(c)) \mid c^i_{m^i} = r^i]$$

Proof Sketch: The proof is similar to proof of lemma 4, if we notice that we fix c in the first part of the proof. For a fixed c, let $I = \{(i,j) \mid \mu_c(i,j) = (i,j), j \in [1..m^i]\}$. Let $T = \{M^i_j(c) \mid (i,j) \in I\}$. Since PD(c) holds, $|T| = |I|$. Let T' be an arbitrary set of n bit strings, disjoint from T, and $|T'| = \sum_{i \in [1..q]} m^i - |I|$. Thus, $|T \cup T'| = \sum_{i \in [1..q]} m^i$.

By, fact $2'$, $C(f_c) = c$. Thus, for each b agreeing with c on I, we have a function f_c defined on $|I|$ inputs T, such that $C(f_c) = c$. We can use the remaining $\sum_{i \in [1..q]} m^i - |I|$ values of b (i.e. from indices which are not in I) to extend f_c to be defined on $T \cup T'$. This map from b to the extended f_c is 1-1.

The reverse direction is done as in lemma 4.

Rest of the proof is also as in proof of lemma 4. □

Let Δ denote $(\sum_{i \in [1..q]} m^i)^2 * 2^{-(n+1)}$.

Lemma 10. *For every qn bit constant $\langle r^i \rangle$ $(i \in [1..q])$,*

$$Pr_{c \in_U S^q}[\ PD(\rho(c)) \mid c^i_{m^i} = r^i] \ \geq 1 - \Delta$$

Proof: First note that for all i, $c^i_{m^i} = \rho(c)^i_{m^i}$, by fact $1(a)'$ and definition of ρ. As opposed to lemma 5, we need to show that it is not the case that a $\rho(c)^i_j$, with $j \neq m^i$, can be defined to be a $c^{i'}_j$, such that $j = m^{i'}$. Suppose, there is indeed an $(i', j) \equiv_c (i, j)$, such that $j = m^{i'}$. Since, $(i', j) \equiv_c (i, j)$, we have $(i', j) \cong_c (i, j)$. Thus the graphs G^i and $G^{i'}$ are identical till $j = m^{i'}$. Thus, unless they are the same graph, this is not allowed by the condition on PRF-preserving $\mathcal{G}$. If they are the same graph, then $j = m^i$, a contradiction.

Rest of the proof is similar to proof of lemma 5. □

Rest of the proof of theorem 8 is identical to that of theorem 2.

Appendix C

Consider a layered graph G with vertex sets $V_1, V_2, ..., V_t$. V_t has only one vertex m, and all vertices in V_{t-1} have an edge to m. For every s, $0 < s \leq t-1$, each vertex in V_s has edges from vertices in V_{s-1}, such that no two vertices in V_s have the same set of incident nodes. Thus, we can take all non-empty subset of vertices in V_{s-1}, and let each such subset be the incident set of a vertex in V_s. Thus, $|V_s|$ can be as large as $2^{|V_{s-1}|} - 1$. We can add another $2^{|V_{s-1}|} - 1$ nodes to V_s by using an edge from V_{s-2}, if $s > 2$. Thus, we can assume that if $s > 2$, $|V_s| \geq 2^{|V_{s-1}|}$.

Define $\mathrm{tower}(2,1) = 2$. For each $n > 1$, let $\mathrm{tower}(2,n) = 2^{\mathrm{tower}(2,n-1)}$. Let $\log^* n$ be the smallest number m such that $\mathrm{tower}(2,m) = n$.

It follows that G has $\mathrm{tower}(2,t-3)$ vertices. Thus if we need m vertices, $t = 3 + log^* m$.

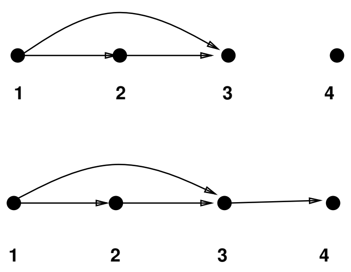

Fig. 4. An Incorrect Construction

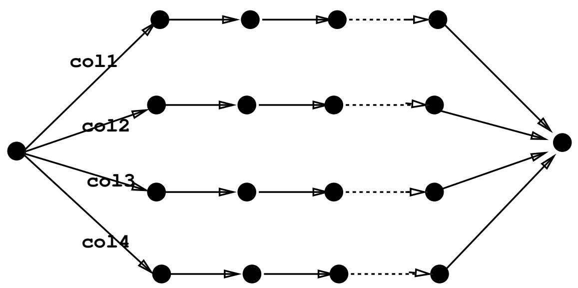

Fig. 5. A Parallel Mode using $\mathrm{GF}(2^n)$

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

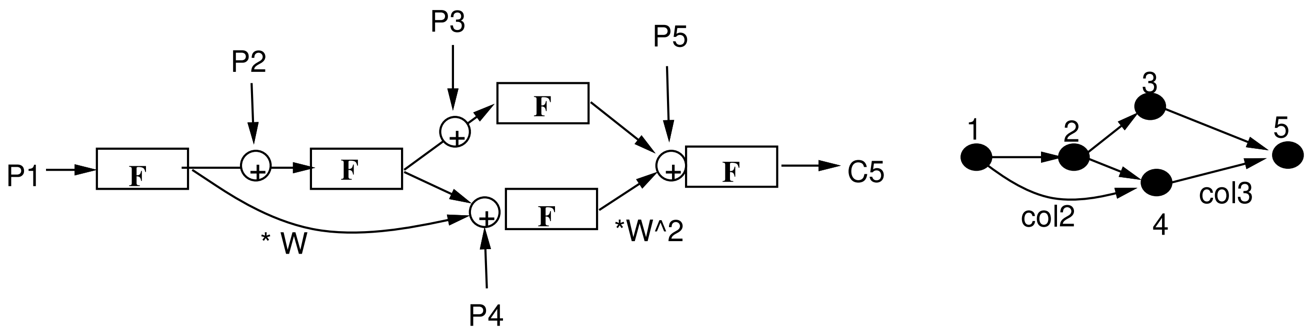

Fig. 6. A PRF Domain Extension using GF(2^n) and its colored DAG

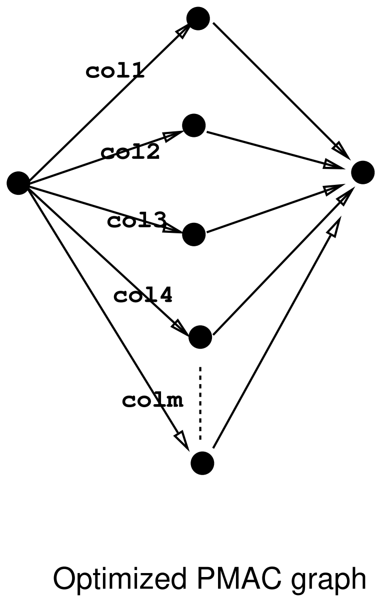

Fig. 7. PMAC without $F(0^n)$

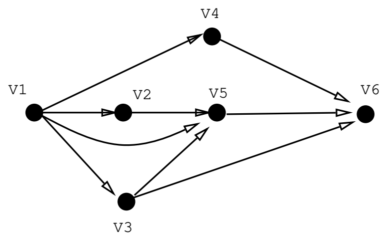

Fig. 8. A non-automorphic DAG