Problem 1.

Part A For any p.p.t. distinguisher $\mathcal{D}$ that tries to distinguish $F_k^\$$ and $f^\$$, we can construct $\mathcal{D}'$ who emulates $\mathcal{D}$. Given input 1^λ and oracle access to $\mathcal{O} \in \{F_k, f\}$, the distinguisher $\mathcal{D}'$ emulates the execution of $\mathcal{D}(1^\lambda)$, upon each query from $\mathcal{D}$, samples $r \leftarrow \{0,1\}^\lambda$ and feed $(r, \mathcal{O}(r))$ to $\mathcal{D}$.

$$\begin{aligned} &\left|\Pr[\mathcal{D}^{F_k^\$}(1^\lambda) = 1] - \Pr[\mathcal{D}^{f^\$}(1^\lambda) = 1]\right| \\ =& \left|\Pr[\mathcal{D}'^{F_k}(1^\lambda) = 1] - \Pr[\mathcal{D}'^{f}(1^\lambda) = 1]\right| \leq \text{negl}(\lambda) \end{aligned}$$

Hence F is also a *weak* PRF.

Part B Let $f' : \{0,1\}^\lambda \to \{0,1\}^\lambda$ be a random function and define

$$f(x) := \begin{cases} f'(x), & \text{if } x \text{ is even} \\ f'(x+1), & \text{if } x \text{ is odd} \end{cases}$$

By the same argument as **Part A**, for any p.p.t. distinguisher $\mathcal{D}$

$$\left|\Pr[\mathcal{D}^{F_k^\$}(1^\lambda) = 1] - \Pr[\mathcal{D}^{f^\$}(1^\lambda) = 1]\right| \leq \text{negl}(\lambda).$$

Let BAD denote the event that in the execution of $\mathcal{D}$, $|r_i - r_j| = 1$ for some two random inputs r_i, r_j. Conditioning on BAD doesn't happen, f will act identically to a random function.

$$\begin{aligned} &\left|\Pr[\mathcal{D}^{f^\$}(1^\lambda) = 1] - \Pr[\mathcal{D}^{f'^\$}(1^\lambda) = 1]\right| \\ =& \Big|\big(\Pr[\mathcal{D}^{f^\$}(1^\lambda) = 1 \mid \mathsf{BAD}] - \Pr[\mathcal{D}^{f'^\$}(1^\lambda) = 1 \mid \mathsf{BAD}]\big) \Pr[\mathsf{BAD}] \\ &\quad - \big(\Pr[\mathcal{D}^{f^\$}(1^\lambda) = 1 \mid \neg\mathsf{BAD}] - \Pr[\mathcal{D}^{f'^\$}(1^\lambda) = 1 \mid \neg\mathsf{BAD}]\big) \Pr[\neg\mathsf{BAD}]\Big| \\ =& \Big|\big(\Pr[\mathcal{D}^{f^\$}(1^\lambda) = 1 \mid \mathsf{BAD}] - \Pr[\mathcal{D}^{f'^\$}(1^\lambda) = 1 \mid \mathsf{BAD}]\big) \Pr[\mathsf{BAD}]\Big| \\ \leq& \Pr[\mathsf{BAD}] \leq \text{negl}(\lambda). \end{aligned}$$

Then by the triangular inequality, $\mathcal{D}$ can not distinguish between $F_k^\$$ and $f'^\$$, hence F is a *weak* PRF.

However, F is not a PRF since $F_k(2x+1) = F_k(2x+2)$ holds for all x.

Part C The scheme is not secure even in the presence of an eavesdropper.

Assume the *weak* PRF we use is constructed as in **Part B**, choose $m_0 = x\|y\|x, m_1 = x\|x\|x$ where $x \neq y$ and output 1 if any two adjacent blocks of ciphertext are identical.

For the ciphertext of m_1 always has two identical adjacent blocks. While for m_0, such event happens with probability $\Pr[F'_k(r) \oplus F'_k(r+1) = x \oplus y]$, which is negligible.

Part D Recall how we prove the CPA security of Π when the function F is a PRF. For any adversary $\mathcal{A}$ targeting the CPA security of Π, we construct a distinguisher $\mathcal{D}$, which is essentially the CPA security game $\mathsf{PrivK}^{\text{CPA}}_{\Pi,\mathcal{A}}$. The only difference is that, in $\mathcal{D}$'s emulation, the challenger does not sample k, the computation of F_k is delegated to the oracle $\mathcal{O}$.

The proof of Part D is very similar. For any adversary $\mathcal{A}$ targeting the CPA security of Π, we construct a distinguisher $\mathcal{D}_{\text{new}}$, which is essentially the CPA security game $\mathsf{PrivK}^{\text{CPA}}_{\Pi,\mathcal{A}}$. The only difference is that, in $\mathcal{D}_{\text{new}}$'s emulation, the challenger does not sample k and whenever the challenger need to sample a random r and computes $F_k(r)$, the task is delegated to the probabilistic oracle. Since F is a weak PRF,

$$\left|\Pr[\mathcal{D}_{\text{new}}^{F_k^{\$}}(1^\lambda) \to 1] - \Pr[\mathcal{D}_{\text{new}}^{f^{\$}}(1^\lambda) \to 1]\right| \leq \text{negl}(\lambda).$$

Since the challenger in $\mathsf{PrivK}^{\text{CPA}}_{\Pi,\mathcal{A}}$ only evaluates F_k on fresh random points, the behavior of $\mathcal{D}^f$ and $\mathcal{D}_{\text{new}}^{f^{\$}}$ are identical for any f,

$$\begin{aligned}
\Pr[\mathcal{D}_{\text{new}}^{F_k^{\$}}(1^\lambda) \to 1] &= \Pr[\mathcal{D}^{F_k}(1^\lambda) \to 1] = \Pr[\mathsf{PrivK}^{\text{CPA}}_{\Pi,\mathcal{A}}(\lambda) \to 1], \\
\Pr[\mathcal{D}_{\text{new}}^{f^{\$}}(1^\lambda) \to 1] &= \Pr[\mathcal{D}^{f}(1^\lambda) \to 1] = \frac{1}{2} \pm \text{negl}(\lambda).
\end{aligned}$$

(Both can be directly verified. But relying the equivalence between $\mathcal{D}^f$ and $\mathcal{D}_{\text{new}}^{f^{\$}}$ simplifies the proof.) Thus $\Pr[\mathsf{PrivK}^{\text{CPA}}_{\Pi,\mathcal{A}}(\lambda) \to 1] = \frac{1}{2} \pm \text{negl}(\lambda)$, the scheme Π is CPA-secure.

Problem 2.

Part A P' is not a PRP.

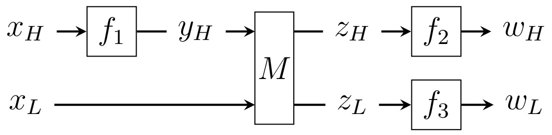
Given oracle $\mathcal{O}$, the distinguisher picks $x_L^0 \neq x_L^1, x_H^0 \neq x_H^1$ and checks if the lower parts of $\mathcal{O}(x_L^0, x_H^0) + \mathcal{O}(x_L^1, x_H^1)$ and $\mathcal{O}(x_L^0, x_H^1) + \mathcal{O}(x_L^1, x_H^0)$ are equal.

If $\mathcal{O}$ is P', their parts are the same, which equals to

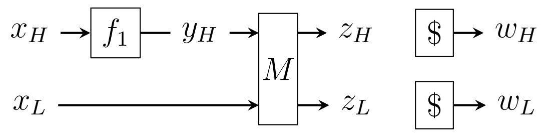
$$M \begin{pmatrix} y_H^0 + y_H^1 \\ y_L^0 + y_L^1 \end{pmatrix}.$$

If $\mathcal{O}$ is a random permutation, such probability is negligible.

Part B Let $x_L^i, x_H^i, y_H^i, z_L^i, z_H^i, w_L^i, w_H^i$ denote the input, output and intermediate values of the i-th query. W.l.o.g., we assume the queries (x_L^i, x_H^i)r are distinct.



World 1



World 2

Consider *World 1*, where PRP $F_{k_1}, F_{k_2}, F_{k_3}$ are replaced by random functions f_1, f_2, f_3 respectively. Due to the security of PRF, the distinguisher cannot distinguish the real world from World 1 with a non-negligible margin.

Consider *World 2*, where f_2, f_3 are further replaced by "random boxes". Upon a query, a random box ignores the input and samples a fresh random value. The distinguisher cannot distinguish the ideal world from World 2 with a non-negligible margin.

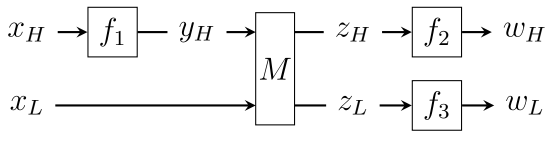
It remains to show that the distinguisher cannot distinguish World 1 and World 2.

Define event $\mathsf{Repeat} = \{\exists i < j \text{ s.t. } z_L^i = z_L^j \vee z_H^i = z_H^j\}$. When Repeat does not happen, World 1 and World 2 perform identically. So $\Pr[\mathsf{Repeat}]$ in World 1 equals $\Pr[\mathsf{Repeat}]$ in World 2, and is an upper bound of distinguishing margin.

It is easier to bound the probably of Repeat in World 2. In World 2, the adversary receives no information of y_H^i, z_H^i, z_L^i, so it has to *non-adaptively* choose $(x_L^i, x_H^i)_i$. For each $i < j$, the probability $\Pr[z_L^i = z_L^j]$ and $\Pr[z_H^i = z_H^j]$ are bounded by $2^{-\lambda}$.

Part B alternative proof P'' is a PRP.

Since F is a PRP, we can replace $F_{k_1}, F_{k_2}, F_{k_3}$ by i.i.d. uniform f_1, f_2, f_3 respectively.



Let $x_L^i, x_H^i, y_H^i, z_L^i, z_H^i, w_L^i, w_H^i$ denote the input, output and intermediate values of the i-th query. W.l.o.g., all (x_L^i, x_H^i) are distinct.

Due to the randomness of f_1, with overwhelming probability, $z_L^i \neq z_L^j \wedge z_H^i \neq z_H^j$ for any $i \neq j$. In such case, every output (w_L^i, w_H^i) is fresh random, and thus cannot be distinguished from a random permutation.

The intuition can be formalized. Define the following statements:

- A_t: with overwhelming probability, for all $i < j \leq t$, $z_L^i \neq z_L^j \wedge z_H^i \neq z_H^j$.
- B_t: the joint distribution of the first t outputs $(w_L^i, w_H^i)_{i=1}^t$ is close to uniform.
- C_t: the distribution of f_1 conditioning on the first t outputs $(w_L^i, w_H^i)_{i=1}^t$ is close to uniform

$A_t \implies B_t$ follows directly from the randomness of f_2, f_3.

$A_t \implies C_t$ also follows from the randomness of f_2, f_3. Due to the effect of f_2, f_3, the only leaked information of f_1 is whether z_L^i equals z_L^j and whether z_H^i equals z_H^j. As claimed by A_t, such leakage is negligible.

$C_{t-1} \implies A_t$ follows from the randomness of f_1. For each $j < t$, if $x_H^j = x_H^t$ then there is definitely no collision; if $x_H^j \neq x_H^t$, the randomness of y_H^t ensures $z_L^t \neq z_L^j \wedge z_H^t \neq z_H^j$ with overwhelming probability.

Problem 3.

Part A A counterexample is 3-round Feistel network.

Part B F' is a PRF.

When k is hidden, $F(k, \cdot)$ is indistinguishable from a random function $f(\cdot)$ under oracle access. Therefore, as a standard trick, it suffices to show that $f'(x) = x \oplus f(x)$ is indistinguishable from a random function under oracle access. Note that the distribution of f' is identical to a random function, thus it is indistinguishable from a random function.

More formally, consider the following three oracles:

$F'(k, \cdot)$ Given x, output $F'(k, x) = x \oplus F(k, x)$. ($k$ is a random key.)

f' Given x, output $f'(x) = x \oplus f(x)$. (f is a random function.)

f Given x, output $f(x)$. (f is a random function.)

The first two are indistinguishable because F being a PRF. The last two are indistinguishable because they are identical.

Part C F' is a PRP.

Consider the following three oracles:

$F'(k, \cdot)$ Given x, output $F'(k, x) = F(k_2, F(k_1, x))$. ($k = k_1 \| k_2$ is a random key.)

f' Given x, output $f'(x) = f_2(f_1(x))$. (f_1, f_2 are random permutations.)

f Given x, output $f(x)$. (f is a random permutation.)

The first two are indistinguishable because F being a PRP. The last two are indistinguishable because they are identical.

Part D F' is a PRP.

Consider the following three oracles:

$F'(k, \cdot)$ Given x, output $F'(k, x) = F(k_2, F(k_1, x))$. ($k = k_1 \| k_2$ is a random key.)

f' Given x, output $f'(x) = f(f(x))$. (f is a random function.)

f Given x, output $f(x)$. (f is a random function.)

The first two are indistinguishable because F being a PRF. Thus it suffices to show the indistinguishability between the last two. (We also rely on the fact that a random function and a random permutation are indistinguishable.)

Without loss of generality, we assume the distinguisher never query the same input twice. The oracle f always returns a fresh random output upon a new query.

When the distinguisher is interacting with the oracle f', let x_i denote the i-th query, let y_i, z_i denote the corresponding intermediate value and output. For each t, $x_t \notin \{x_1, \ldots, x_{t-1}, y_1, \ldots, y_{t-1}\}$ with overwhelming probability, thus $y_t = f(x_t)$ is a fresh random value and $y_t \notin \{x_1, \ldots, x_t, y_1, \ldots, y_{t-1}\}$ with overwhelming probability, then $z_t = f(y_t)$ is a fresh random value.

The intuition can be formalized.

Formalization 1. Define the following statements:

- A_t: with overwhelming probability, $x_t \notin \{x_1, \ldots, x_{t-1}, y_1, \ldots, y_{t-1}\}$.
- B_t: with overwhelming probability, $y_t \notin \{x_1, \ldots, x_t, y_1, \ldots, y_{t-1}\}$.
- C_t: the joint distribution of the first t outputs $z_1, \ldots, z_t$ is close to uniform.
- D_t: the distribution of $y_1, \ldots, y_t$ conditioning on $x_1, \ldots, x_t, z_1, \ldots, z_t$ is close to uniform

$D_{t-1} \implies A_t$: $x_t \notin \{x_1, \ldots, x_{t-1}\}$ comes from the assumption of no duplicated queries. $y_t \notin \{x_1, \ldots, x_{t-1}\}$ w.h.p. follows from D_{t-1}.

$A_t \implies B_t$ follows directly from the randomness of f.

$B_t + C_{t-1} \implies C_t$ also follows from the randomness of f.

$D_{t-1} + B_t \implies D_t$ because x_t is determined by $x_1, \ldots, x_{t-1}, z_1, \ldots, z_{t-1}$, thus revealing no information; and z_t is just fresh randomness.

Formalization 2. The oracle f' can be implemented by the following program, if parameter threshold is set as 0.

Initialize an empty table f in the setup phase.

Upon receiving input x_i,

if $f(x_i)$ is defined and $i \geq$ threshold

let $y_i \leftarrow f(x_i)$

otherwise, sample y_i randomly and set $f(x_i) = y_i$

if $f(y_i)$ is defined and $i \geq$ threshold

let $y_i \leftarrow f(z_i)$

otherwise, sample z_i randomly and set $f(y_i) = z_i$

output z_i

If threshold is set to be the number of queries, then the program always return i.i.d. random outputs.

Comparing the program when threshold $= t$ and threshold $= t + 1$, the only difference is in the t-th query. When x_t is received, $f(x_t)$ is undefined with overwhelming probability because $y_1, \ldots, y_{t-1}$ are completely hidden from the distinguisher. Then as a consequence, y_t is random and $f(y_t)$ is undefined with overwhelming probability. In short, the program parameterized by threshold $= t$ and the program parameterized by threshold $= t + 1$ perform exactly the same with overwhelming probability. By a hybrid argument, the program parameterized by threshold $= 0$ and the program parameterized by threshold $= \text{poly}(\lambda)$ perform exactly the same with overwhelming probability.

TODO: update proof

Problem 4.

Part A. It is a PRP.

The proof is almost the same as the proof for independent-key 3-round Feistel.

The first step is to replace the PRFs by random functions. No PPT adversary can distinguish

$$\text{Feistel}_{f(k_1,\cdot),f(k_2,\cdot),f(k_2,\cdot)} \text{ from } \text{Feistel}_{F_1(\cdot),F_2(\cdot),F_2(\cdot)},$$

by having oracle access to them. (As we have shown in problem 10.) So it suffices to show no PPT adversary can distinguish

$$\text{Feistel}_{F_1(\cdot),F_2(\cdot),F_2(\cdot)}$$

from a random permutation over $\{0,1\}^{2n}$, by having oracle access.

W.l.o.g., we can assume the adversary never makes duplicate queries. Under such assumption, when the oracle is a random permutation, it gets a random $2n$-bit-string whenever it queries the oracle. We need to show that the same happens when the oracle is $\text{Feistel}_{F_1(\cdot),F_2(\cdot),F_2(\cdot)}$.

Let (x_0^i, x_1^i) denotes the adversary's i-th query, let (x_3^i, x_4^i) denotes the corresponding output, and let x_2^i denote the corresponding intermediate value.

Consider statement P_t: with probability $1 - \text{negl}(n)$, all of $(P_t.1)$, $(P_t.2)$, $(P_t.3)$ hold.

(P_t.1) "There is no collision on x_2, x_3." That is, $x_2^1, x_2^1, \ldots, x_2^t, x_3^t$ are all distinct values.

(P_t.2) "The i-th output is uniform." That is, conditioning on $(x_0^i, x_1^i, x_3^i, x_4^i)_{i<t}$, the conditional distribution of (x_3^t, x_4^t) is close to uniform.

(P_t.3) "F_1 is hidden." That is, conditioning on $(x_0^i, x_1^i, x_3^i, x_4^i)_{i\le t}$, the conditional distribution of F_1 is close to uniform.

We prove that statement P_t holds for all $t \le \text{poly}(n)$ inductively.

Assume P_{t-1} holds. Due to $(P_{t-1}.3)$, x_2^t collides with a previous value with at most negligible probability. Since x_2^t does not collides with any previous x_2^i or x_3^i, thus $F_2(x_2^t)$ is a fresh random value. Since x_3^t is one-time padded by $F_2(x_2^t)$, the value of x_3^t does not collides with any previous x_2^i or x_3^i with overwhelming probability. (So $(P_t.1)$ holds.) Then $F_2(x_3^t)$ is also a fresh random value.

Since $F_2(x_2^t), F_2(x_3^t)$ are fresh random values, the distribution of (x_3^t, x_4^t) is uniform, even conditioning on previous information. (So $(P_t.2)$ holds.)

In the t-th query, the only information about F_1 is $F_1(x_1^t)$. But the information is perfectly hidden, because the output (x_3^t, x_4^t) is one-time padded by $F_2(x_2^t), F_2(x_3^t)$. (So $(P_t.3)$ holds.)

Part A alternative proof. Here we present a more formal proof. W.l.o.g., we assume the distinguisher never makes duplicate queries. Let $x_0^i, x_1^i, x_2^i, x_3^i, x_4^i$ denotes the input, intermediate value, output corresponding to the i-th query.

- Real world: The distinguisher has oracle access to $\text{Feistel}_{f(k_1,\cdot),f(k_2,\cdot),f(k_2,\cdot)}$.
- World 1: PRFs are replaced by random functions. The distinguisher has oracle access to $\text{Feistel}_{F_1(\cdot),F_2(\cdot),F_2(\cdot)}$.

Problem 4 continued on next page...

- World 2: F_2 is further replaced by a "random box". Upon a query, it will always sample fresh random output.
- World 3: x_2^i is computed, and is ignored. x_3^i, x_4^i are freshly uniformly sampled.
- Ideal world: The distinguisher has oracle access to a random function $\{0,1\}^{2\lambda} \to \{0,1\}^{2\lambda}$.

It is easy to argue that real world and World 1 are indistinguishable, World 2 and World 3 are identical, ideal world and World 3 are indistinguishable.

If F_2 is never evaluated upon same input twice, it behaves exactly the same as a random box. Let Repeat denote the event that F_2 is evaluated on some input twice. Then the advantage distinguishing World 1 and World 2 is no more than $\Pr[\mathsf{Repeat}]$ (in World 1 or World 2 or World 3).

$$\mathsf{Repeat} = \{x_r^i = x_s^j \text{ for some } i, j \in [T], r, s \in \{2,3\} \text{ s.t. } (i,r) \neq (j,s)\}$$

It is easier to bound $\Pr[\mathsf{Repeat}]$ in World 3. In World 3, fresh random x_3^i is unlikely to collide with other values. In World 3, the distinguisher learns no information about F_1, so the distinguisher can only make non-adaptive queries $\{x_0^i, x_1^i\}_i$. The randomness of F_1 ensures x_2^i will not collides with other values with overwhelming probability.

Part B. It is not even a PRP, because

$$f_4((k_1,k_2),(x_0,x_1)) = (x_4,x_5) \implies f_4((k_1,k_2),(x_5,x_4)) = (x_1,x_0).$$