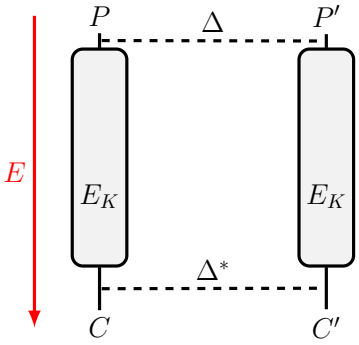
Théorie des codes - TP 2
ZZ3 F5 - Réseaux et Sécurité Informatique
Cryptographie symmétrique : Analyse et Cryptanalyse d'un ciffrement par bloc

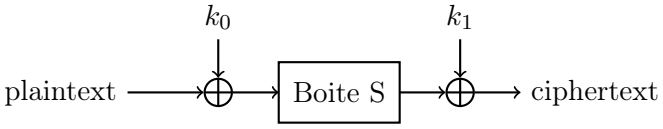
La cryptanalyse est la technique qui consiste à déduire un texte en clair d'un texte chiffré. De nombreuses techniques de cryptanalyses existent pour les chiffrements symmétriques. Une très grande majorité de ces attaques sont dérivées de ce qu'on appelle *cryptanalyse différentielle* et *cryptanalyse linéaire*. On doit la cryptanalyse différentielle à Biham et Shamir en 1990 mais elle était connue des concepteurs du DES dans les années 70. C'est une attaque à messages clairs choisis, sur les constructions itératives de chiffrement par bloc.



Le principe général est le suivant. On considère deux messages clairs m et m^*, et on appelle différence la somme $\alpha = m \oplus m^*$. On s'intéresse à la différence que l'on obtient après chiffrement pour une différence donnée en entrée. Cela va nous permettre de retrouver certaines parties de la clef (et non pas toute la clef).

Un chiffrement de jeu

Voici un petit algorithme de chiffrement qui sera cassé dans la suite de ce TP. Il s'agit d'un SPN simple à deux tours qui n'utilise que des boîtes de substitution (boite S) et deux clefs qui sera xorer (application de $\oplus$) avec les états intermétidaires. L'algorithme accepte un bloc de 4 bits en entrée ainsi qu'une clef de 8 bits ; il produit un bloc de 4 bits en sortie. Le texte en clair est d'abord soumis à une opération XOR avec les 4 premiers bits de la clef (appelée sous-clef 0 ou K0). Ensuite, le résultat est introduit dans une boîte S (déjà implémenté) de 4 bits qui assure la non-linéarité du chiffrement. La sortie de la boîte S est ensuite mélangée par XOR avec les 4 seconds bits de la clef (sous-clef 1 ou K1) et repassée dans la boîte S. La taille des blocs est réduite pour faciliter l'apprentissage, ce qui permet également de réduire la taille de la boîte S. Les paramètres de ce chiffrement-jouet sont tels que suivre le processus "à la main" vous sera possible, ce qui n'est pas le cas habituellement. Le chiffrement :



1. Implémentez le chiffrement décrit et le déchiffrement associé.
2. Quel intérêt y aurait-il à ajouter une deuxième boîte S à la fin du chiffrement ?

Analyse d'un chiffrement symmétrique

Nous allons casser ce code après avoir testé seulement 6 clefs au lieu de 16. Pour ce faire, nous devrons non seulement connaître un texte clair/chiffré, mais aussi le choisir. La cryptanalyse différentielle est une attaque par texte choisi. Dans ce modèle, l'attaquant est capable de faire en sorte qu'un système de cryptographie chiffre des données de son choix en utilisant la clef cible (qui est le secret). En analysant les résultats obtenus (le texte chiffré connu), l'attaquant peut déterminer la clef utilisée. Une fois la clef récupérée de cette manière, les futures transmissions qui l'utilisent peuvent être rapidement décryptées. Les chances que ce modèle d'attaque se produise dans le monde réel sont certes discutables, mais ce scénario est beaucoup plus probable qu'il n'y paraît à première vue.

Nous commencerons par investiguer les différences d'entrée et de sortie des boîtes S afin de déterminer une paire de différences à forte probabilité.

Pour deux entrées X, X' de la boîte S, produisant $Y = S(X)$, $Y' = S(X')$, vous analyserez quelles sont les différences $\Delta Y = Y \oplus Y'$ qui se produisent le plus pour une différence donnée en entrée $\Delta X = X \oplus X'$. Notez que l'opération de XOR ne modifie pas la différentielle.

1. Réalisez une fonction permettant d'afficher l'occurance des différences en sortie en fonction des différences en entrée de la boîte S.

Un peu d'analyse statistique permet de savoir en quoi un différentiel d'entrée est susceptible d'être transformé lorsqu'il passe par la boîte S. Pendant que vous déterminez ce que sera le différentiel de sortie, prenez également note de toutes les valeurs d'entrée possibles (X et X') qui produisent cette transformation. Supposons que nous connaissions une caractéristique différentielle qui a une forte probabilité de se vérifier. Par exemple, disons que p fois sur 16, deux entrées aléatoires dans la boîte S s'opposent pour produire la valeur Δ_X. De plus (6 fois sur 16), lorsque ces valeurs passent par la boîte S et sont ensuite combinées par XOR, elles produisent la valeur différentielle Δ_Y. Nous dirions que la caractéristique différentielle "$\Delta_X \rightarrow \Delta_Y$" se vérifie $p/16$ fois pour cette boîte S. Notez qu'il n'y a que p valeurs réelles qui peuvent donner lieu à cette situation ; ce fait entrera en ligne de compte.

2. En vous basant sur la table obtenu, choissez une paire (Δ_X, Δ_Y) cohérente.

3. Une fois que vous avez trouvé une bonne caractéristique, trouvez toutes les valeurs d'entrée spécifiques qui la produisent et les valeurs de sortie correspondantes. Gardez en mémoire l'ensemble des pairs ayant produits la différentielle choisis.

Nous allons maintenant commencer l'attaque. Rappelons qu'il s'agit d'une attaque par texte choisi. Nous allons générer nos blocs de texte en clair/chiffré connus par paires. Tout d'abord, nous choisissons un bloc de texte en clair au hasard et nous l'appelons P_0. Ensuite, faites un XOR avec le différentiel d'entrée de la caractéristique choisie pour produire P_1. Passez ensuite chacun de ces blocs dans le système de chiffrement (qui utilise les clefs secrètes) pour obtenir C_0 et C_1. Vous venez de produire une paire choisie sur la base de votre caractéristique différentielle. Si vous faites cela et que $C_0 \oplus C_1$ est égal à la différentielle de sortie de notre caractéristique choisie, alors il s'agit d'une bonne paire. Une fois que vous en avez trouvé une, vous pouvez vous arrêter. Conservez les autres mauvaises paires, mais mettez la bonne paire de côté pour l'étape suivante. Les autres paires seront utilisées pour valider les suppositions de clefs.

4. Trouvez une bonne paire et gardez la en mémoire.

A ce point, nous ignorons les entrées et les sorties réelles de la boîte S. Rappelez-vous que seules p paires de valeurs d'entrée peuvent produire cette caractéristique "$\Delta_X \rightarrow \Delta_Y$". Cela signifie que ces valeurs inconnues au milieu ne peuvent avoir que l'une de ces 6 valeurs précalculées.

5. Listez les valeurs de clefs possibles en entrée et en sortie de la boîte S.

6. Testez à présent contre certaines des autres valeurs gardés en mémoire si l'une de ces supposition est correcte. Retournez la clef correcte et vérifiez qu'elle correspond bien à la clef utilisée.

Appendix Cryptographie différentielle

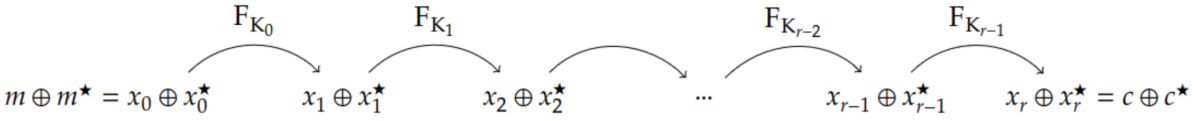
Disclaimer : Cette partie est un condensé introudisant rapidement les principes de la cryptanalyse différentielle, elle est tiré d'un ouvrage plus complet [1]. Un tutoriel plus complet est disponible a

1. `https://www.math.u-bordeaux.fr/~gcastagn/Cryptanalyse/cours-22-23.pdf`

été écris pas Howard M. Heys et est disponible sur internet [2].

On doit la cryptanalyse différentielle à Biham et Shamir en 1990 mais elle était connue des concepteurs du DES dans les années 70. C'est une attaque à messages clairs choisis, sur les constructions itératives de chiffrement par bloc.

Le principe général est le suivant. On considère deux messages clairs m et m^*, et on appelle différence la somme $\alpha = m \oplus m^*$. On s'intéresse à l'évolution des différences à chaque tour (appelée chemin différentiel).



Si la fonction de tour était linéaire, on aurait à chaque tour $x_{i+1} \oplus x^*_{i+1} = F_{K_i}(x_i) \oplus F(x_i)^* = F_{K_i}(x_i \oplus x_i)^*$ Ainsi, quelque soit m, m^* satisfaisant une différence α en entrée, on aurait toujours le même chemin différentiel, α, $F_{K_0}(\alpha)$, $F_{K_1}(F_{K_0}(\alpha))$, ... Comme cette fonction de tour n'est pas linéaire, un tel chemin différentiel ne va pas être suivi avec probabilité 1 (sur les choix de m, m^* tel que $m \oplus m^* = \alpha$) mais va être suivi avec probabilité plus ou moins grande. On s'intéresse en particulier à la probabilité

$$P_{\alpha,\beta} = \Pr[x_{r_1} \oplus x^*_{r_1} = \beta | m \oplus m^* = \alpha]$$

et on suppose connaître (α, β) tel que $P_{\alpha,\beta}$ est élevée (éloignée de l'uniforme, $\frac{1}{2^n}$ où n est la taille des blocs). On se sert de cette probabilité pour tester des candidats pour la clef K_{r-1} du dernier tour comme pour la cryptanalyse linéaire. Lors d'une attaque à clairs choisis, on récupère un grand nombre de couples clairs chiffrés, m, c) et (m^*, c^*) tel que $m \oplus m^* = \alpha$. Pour ces couples, on doit avoir la différence à l'entrée du dernier tour, $x_{r-1} \oplus x^*_{r-1}$ égale à β avec probabilité $P_{\alpha,\beta}$. Pour chaque couple (m, c), (m^*, c^*), on remonte le dernier tour, en calculant pour toutes les valeurs possibles K de la clef de dernier tour K_{r-1},

$$z_K = F_K^{-1}(c) \text{ et } z^*_K = F_K^{-1}(c^*).$$

Si $z_K \oplus z^* = \beta$, on incrémente un compteur pour la clef K. Pour la « bonne » clef, $K = K_{r-1}$, le compteur doit être incrémenté avec probabilité $P_{\alpha,\beta}$. Pour les mauvaises, les valeurs Z_K et z^*_K sont indépendantes de m et de m^*, on s'attend à avoir $z_K \oplus Z^*_K = \beta$ avec probabilité $\frac{1}{2n}$. On ne fait pas une recherche exhaustive sur toute la clef du dernier tour, K_{r-1}. On choisit un β donnant peu de boîtes actives et on se limite à vérifier si $x_{r-1} \oplus x^*_{r-1} = \beta$ au niveau de l'entrée de celles ci. Pour calculer la probabilité $P_{\alpha,\beta}$, toujours dans le cas d'un SPN, on la calcule en observant la propagation des différences sur les étapes élémentaires. Pour les étapes linéaires, de type $v = P(u)$, on sait qu'une différence α devient $\beta = P(\alpha)$ avec probabilité 1. Pour l'ajout de la clef de tour, $x = v \oplus K$, une différence α reste inchangée, donc $\beta = \alpha$ avec probabilité 1. En particulier les probabilités de propagation des différences ne dépendent pas de la clef secrète utilisée, elles ne dépendent que de l'algorithme de chiffrement et peuvent donc être pré calculées. Le calcul des propagations des différences au travers des boîtes S va être le coeur du calcul d'un chemin différentiel. Soit S une boîte S de $F_S^2 \to F_S^2$. On calcule la matrice D à coefficients entiers de taille $2s \times 2s$:

$$D(\alpha, \beta) = \mathsf{Card}\left\{(x, x^*) \in F_S^2 \times F_S^2 | x \oplus x^* = \alpha, S(x) \oplus S(x^*) = \beta\right\}$$

2. `https://ioactive.com/wp-content/uploads/2015/07/ldc_tutorial.pdf`, ce tutoriel introduit aussi la cryptanalyse linéaire, il est parfois utile de se référer à la partie sur celle-ci, en effet les deux principes peuvent parfois avoir des similitudes.

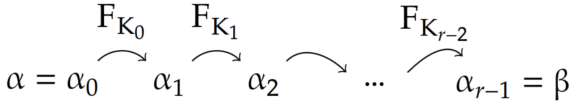
, où α, $\beta \in F_S^2$ sont identifiés avec les entiers de 0 à $2s-1$ pour les indices de positions dans la matrice. De cette matrice, on déduira les probabilités

$$p_{\alpha,\beta} = \Pr[S(x) \oplus S(x^*) = \beta | x \oplus x^* = \alpha] = \frac{D(\alpha,\beta)}{2s}.$$

Une fois calculées les matrices D pour chaque boîte S intervenant dans un SPN, on peut calculer la propagation au travers de l'étape de confusion à chaque tour. Ainsi, si $x = x(1)\| \dots \|x(l)$ est transformé en $u = S_1(x(1))\| \dots \| |S_l(x(l))$ et si $\alpha = \alpha(1)\| \dots \|\alpha(l)$ et $\beta = \beta(1)\| \dots \|\beta(l)$ alors on vérifie facilement, chaque sous bloc évoluant de manière indépendante que

$$\Pr[u \oplus u^* = \beta \| x \oplus x^* = \alpha] = \prod_{j=1}^{l} \Pr[u(j) \oplus u^*(j) = \beta(j) | x(j) \oplus x^*(j) = \alpha(j)].$$

Au final, on est capable de calculer des probabilités de propagation pour chaque tour, du type $P_{\alpha_{i-1},\alpha_i} = Pr[x_i \oplus x_i^* \alpha_i | x_{i-1} \oplus x_{i-1}^* = \alpha_{i-1}]$. Pour calculer $\Pr[x_{r-1} \oplus x_{r-1}^* = \beta | m \oplus m^* = \alpha]$, Biham et Shamir font l'hypothèse que les tours sont indépendants (les clefs de tours rajoutant de l'aléa). Cette hypothèse est assez bien vérifiée en pratique. De plus, plutôt que de considérer toutes les possibilités de chemins différentiels donnant $m \oplus m^* = \alpha$ puis $x_{r-1} \oplus x_{r-1}^* = \beta$, on s'intéresse à un chemin particulier, du type



Sous l'hypothèse d'indépendance des tours, un tel chemin se produit avec probabilité :

$$\prod_{i=1} r - 1 P_{\alpha_{i-1},\alpha_i} = \prod_{i=1} r - 1 \Pr[x_i \oplus x_i^* = \alpha_i | x_{i-1} \oplus x_i^* = \alpha_{i-1}],$$

ce qui minore $\Pr[x_{r-1} \oplus x_{r-1}^* = \beta | m \oplus m^* = \alpha]$