קריפטוגרפיה ואבטחת תוכנה 67392

מקורות

מקורות שעליהן מבוסס הקורס:

1. הספר Introduction to Modern Cryptography (J. Katz and Y. Lindell).
2. קורס בקורסרה Software Security (Michael Hicks).

1 הרצאה 1: קריפטוגרפיה קלאסית מול מודרנית

מהי קריפטוגרפיה?

החלה בתור אומנות עתיקה, מטרתה היא למצוא שיטות המאפשרת לשני אנשים לתקשר ביניהם בצורה סודית, כאשר אדם שלישי מנסה להאזין להם.

הנפשות הפועלות: אליס (**A**lice) מנסה לדבר עם בוב (**B**ob), כאשר (**E**ve) מנסה להאזין להם.

במשך תקופה ארוכה, הקריפטוגרפיה הייתה נחלתה של ארגונים גדולים, והתבססה יותר על יצירות וכשרון אישי מאשר עקרונות מדעיים.

מקור מעולה לסיפורים מרתקים על ההיסטוריה של הקריפטוגרפיה הוא הספר The Code Breakers.

השינוי הגדול בקריפטוגרפיה שהוביל ללידתה של הקריפטוגרפיה המודרנית התרחש לקראת סוף שנות ה-70 של המאה שעברה. השינוי הזה הפך את הקריפטוגרפיה מאומנות למדע, ובראייה לאחור השפיע רבות על המדע בכלל. אנחנו נדון בנקודת המפנה החשובה הזאת רבות במהלך הקורס. עם התפתחות הטכנולוגיה בעשורים האחרונים הקריפטוגרפיה הפכה להיות זמינה לכולם, ובשימוש נרחב מאוד על ידי כולנו. כיום היא מתבוססת על מודלים מוגדרים היטב, יסודות מוצקים והוכחות.

> **הגדרה.** קריפטוגרפיה
>
> העיסוק המדעי במערכות העמידות בפני התנהגות עוינת.

> **הגדרה.** התנהגות עוינת
>
> כל התנהגות שמתכנן המערכת לא הגדיר כהתנהגות הרגילה והמתוכננת של המערכת.

> **הגדרה.** מערכת
>
> כל אוביקט עבורו הגדרנו התנהגות קלט פלט מסוימת, או אופן פעולה מסוים לאורך. (למשל: מערכות הצפנה, מערכות לחתימה דיגיטלית, מערכות לאבטחת מידע ברשת, מערכות לזיהוי ביומטרי, מערכות הצבעה מקוונות, מערכות בנקאות מקוונות ועוד)

קווי השיעור

1. מערכת הצפנה סימטרית
2. סקירת מערכות הצפנה היסטוריות(לא פורמלי)
3. העקרונות הבסיסיים של הקריפטוגרפיה המודרנית

4. הגדרת בטיחות עבור מערכת הצפנה סימטרית ומגבלותיה (סודיות מושלמת).

1.1 מערכת הצפנה סימטרית

מערכת מסוג זה מאפשרת לשני אנשים לתקשר בצורה סודית, אפילו אם אדם שלישי מנסה להאזין לערוץ התקשורת המחבר אותם. בשלב זה של הקורס, אנו נניח כי איב מסוגלת אך ורק להאזין לערוץ - בפרט כל אחת מההודעות שנמצאות על הערוץ מגיעות גם לאיב, אך לאיב אין שליטה על הערוץ (מחיקה / הזרקה של הודעות חדשות).

הנחה בסיסית בקריפטוגרפיה סימטרית - יש לאליס ולבוב מידע סודי משותף שידוע אך ורק להם, ולא ידוע לאיב. מידע זה נקרא מפתח ההצפנה. זו הנחה די רצינית שלא מתאימה לכל השימושים של מערכות הצפנה.

הערה. סימון פלט אלגוריתם בקורס.

בקורס ישנם שני סימונים שונים לפלט של אלגוריתם:

1. $=$ עבור אלגוריתם דטרמיניסטי (קיים רק ערך אפשרי אחד המתקבל מהרצה זו).
2. $\leftarrow$ עבור אלגוריתם הסתברותי (קיימים יותר מערך אפשרי אחד).

למשל: $k \leftarrow KeyGen()$, $c \leftarrow Enc_k(m)$, $m = Dec_k(c)$.

הגדרה. מערכת הצפנה סימטרית

מורכבת משלושה אלגוריתמים:

1. אלגוריתם יצירת המפתחות ($KeyGen$). אלגוריתם הסתברותי שבשלב זה של הקורס אין לא כל קלט. מחזיר מפתח הנדגם מהתפלגות כלשהי $k \in \mathcal{K}$.
2. אלגוריתם ההצפנה (Enc). מקבל בתור קלט מפתח $k \in \mathcal{K}$ והודעה $m \in \mathcal{M}$, ומחזיר בתור פלט הצפנה $c \in \mathcal{C}$. ייתכן כי אלגוריתם Enc הוא דטרמיניסטי (ואז לכל מפתח והודעה מותאמת הצפנה יחידה אפשרית). ייתכן גם כי אלגוריתם Enc הוא הסתברותי, ואז לכל מפתח והודעה יש אוסף של הצפנות אפשריות שדוגמים מהן באקראי.
3. אלגוריתם הפענוח (Dec). מקבל בתור קלט מפתח $k \in \mathcal{K}$ והצפנה $c \in \mathcal{C}$ ומחזיר בתור פלט הודעה $m \in \mathcal{M}$.

מערכות אלה נקראות סימטריות כי המפתח הסודי המשותף משמש גם כדי להצפין הודעות וגם כדי לפענח אותן. (עם זאת אופן ההצפנה ואופן הפענוח יכולים להיות שונים לגמרי בשימוש במפתח הסודי).

יש שתי דרישות:

1. נכונות. לכל מפתח $k \in \mathcal{K}$ ולכל $m \in \mathcal{M}$ מתקיים $Dec_k(Enc_k(m)) = m$.
2. בטיחות. ישנן מספר הגדרות אפשריות.

(עקרון קרקהוף) כאשר מתכננים מערכות הצפנה, יש להניח כי שלושת האלגוריתמים של המערכת ידועים לחלוטין לתוקף, ולא רק למשתמשים הלגיטימיים. הסוד היחידי הוא מפתח ההצפנה. זה שיקוף של המציאות שכן מערכות ההצפנה שכיום בשימוש הן מוכרות לציבור הרחב.

1.2 מערכות הצפנה היסטוריות

1.2.1 צופן הסתה (צופן קיסר)

- $KeyGen$ דוגם באופן אחיד $k \leftarrow \{0, ..., 25\}$
- $\mathcal{M} = \{a, ..., z\}^\ell$, ו- $\mathcal{C} = \{A, ..., Z\}^\ell$

- Enc מסיט כל אות k תווים קדימה.
- Dec מסיט כל אות k תווים אחורה.

דוגמא עם $k=1$:

$$Enc_k(welcome) = XFMDPNF$$

האם הצופן בטוח?

לא. יש רק 26 מפתחות, כלומר $|\mathcal{K}|$ קטן, וניתן לפענח עם כל המפתחות עד שמקבלים משפט הגיוני. (לא נעשה כל שימוש במבנה הספציפי של הצופן, לכן תמיד נרצה שמספר המפתחות יהיה גדול).

1.2.2 צופן החלפה ($Substitution$)

- $KeyGen$ דוגם באופן אחיד k מקבוצת כל הפרמוטציות של $\{a, ..., z\}$. (גודל קבוצה זו הוא $26!$)
- $\mathcal{M} = \{a, ..., z\}^\ell$, ו- $\mathcal{C} = \{A, ..., Z\}^\ell$
- Enc מפעיל את הפרמוטציה k לכל אות.
- Dec עושה את הפרמוטציה ההפוכה.

דוגמא :

$$k = \begin{array}{cccccccccccccccccccccccccc} a & b & c & d & e & f & g & h & i & j & k & l & m & n & o & p & q & r & s & t & u & v & w & x & y & z \\ X & E & U & A & D & N & B & K & V & M & R & O & C & Q & F & S & Y & H & W & G & L & Z & I & J & P & T \end{array}$$

$$Enc_k(tellhim) = GDOOKVC$$

האם הצופן בטוח?

יש הרבה מפתחות ($26!$). עם זאת, ההחלפה של כל אות באות אחרת היא קבועה. לכן מערכת הצפנה זו משמרת באופן מושלם את השכיחות היחסית של אותיות בשפה. ולכן ניתן למפות כל אות בעזרת נתוני השכיחות (לפחות כאשר ההודעה לא קצרה מדי).

https://en.wikipedia.org/wiki/Letter_frequency

(לכן כמות מפתחות גדולה היא תנאי הכרחי אך **לא** מספיק)

1.2.3 צופן ויגנר ($Vigenere$)

- $KeyGen$ דוגם וקטור באופן אחיד : $k = (k_0, ..., k_{t-1}) \leftarrow \{0, ..., 25\}^t$. ($|\mathcal{K}| = 26^t$)

- $\mathcal{C} = \{A, ..., Z\}^\ell$ ו- $\mathcal{M} = \{a, ..., z\}^\ell$,
- Enc מסיט את האות ה-i $k_{i \bmod t}$ מקומות קדימה.
- Dec מסיט את האות ה-i $k_{i \bmod t}$ מקומות אחורה.

דוגמא עם $t = 3$, $k = (1, 2, 3)$:

$$Enc_k(hello) = IGOMQ$$

האם הצופן בטוח?

אם מתבוננים אך ורק באותיות המופיעות במיקומים $i \bmod t$, לכל $i \in \{0, t-1\}$ שוב ניתן להשתמש בתבניות סטטיסטיקה של השפה. להסבר מעמיק ניתן לפנות לספר של קץ ולינדל.

1.2.4 סיכום מערכות היסטוריות

1. מרחב המפתחות חייב להיות גדול.
2. אסור למפות כל אות לאות אחרת באופן קבוע.
3. אלו אינם תנאים מספיקים.

כמעט כל המערכות הפותחו עד לתחילתה של הקריפטוגרפיה המודרנית הן לגמרי לא בטוחות (למרות שעדיין לא הגדרנו את זה).

1.3 העקרונות הבסיסיים של הקריפטוגרפיה המודרנית

אחת התרומות הראשיות של הקריפטוגרפיה המודרנית היא ההבחנה שהגדרות בטיחות מדויקות הן נחוצות. אחרת לא נוכל להבין מתי ואם בכלל הצלחנו להשיג בטיחות. העקרונות:

1. הגדרת הבטיחות חייבת לכלול שני רכיבים מדויקים: 1. התיאור של המתקפות מפניהן אנו רוצים להגן על המערכת (כולל בפרט את אופן הגישה של התוקפים למערכת ואת הכוח החישובי שעומד ברשותם). 2. תיאור של מה נחשב בעינינו כשבירה של הבטיחות (למשל מצד אחד תוקף שמצליח לגמרי לשחזר את המפתח הסודי, ומצד שני תוקף שמצליח לקבל מידע לא טריוויאלי כלשהו על ההודעה - למשל הביט השביעי). זה גם מאפשר לנו להשוות בין מערכות שונות ולנסות להביא לאיזון בין הבטיחות שלהן ליעילות שלהן כתלות במטרה שלה נעשה שימוש במערכת.
2. הצורך בתיאור של ההנחות עליהן הבטיחות של המערכת מתבססת. זה נובע מהעובדה כי עבור כמעט כל המערכות הקריפטוגרפיות לא ניתן להוכיח את בטיחותן באופן בלתי מותנה. אנו שואפים לבסס את הקריפטוגרפיה על הנחות וותיקות יחסית שבחנו אותן במשך מספר שנים ועדיין משערים שהן נכונות. (למשל ההנחה שקשה לפרק מספר גדול מספיק לגורמיו הראשוניים).
3. ניתן להשתמש בשני העקרונות הראשונים כדי להוכיח באופן מתמטי שמערכת היא בטוחה, ולא להשתמש רק באינטואיציה.

העקרונות האלו לא מבטיחים שמערכת תהיה בלתי שבירה, אבל בזכות העקרונות האלו אנחנו יכולים להבין את נקודת הכשל - להבין למה מערכת נשברה. ייתכנו שני מקרים: 1. הגדרת הבטיחות (כחלק מהעקרון הראשון) הייתה חלשה מדי, ולא הצליחה למדל איזהשהו תוקף אמיתי ששבר את המערכת. 2. אחת ההנחות (כחלק מהעקרון השני) התבררה כהנחה שגויה. (למשל אם יום אחד יגלו אלגוריתם פולינומיאלי לפירוק מספר לגורמיו הראשוניים). 3. ההגדרות טובות, ההנחות נכונות, אבל המערכת לא מומשה בצורה נכונה (נדבר על כך בחלק אבטחת התוכנה).

1.4 הגדרת בטיחות עבור מערכת הצפנה סימטרית ומגבלותיה (סודיות מושלמת)

הגדרה. סודיות מושלמת

עבור מערכת הצפנה סימטרית הנתונה על ידי $(KeyGen, Enc, Dec)$ ומפתח משותף לאליס ובוב $k \leftarrow KeyGen()$. והתפלגות K מעל $\mathcal{K}$ ממנה המפתח נגדם.

נניח כי איב יודעת התפלגות אפריורית M על מרחב ההודעות ($\mathcal{M}$), למשל, איב עשויה לדעת: $Pr[M = \text{"}hi\text{"}] = 0.75$ ו-$Pr[M = \text{"}hello\text{"}] = 0.25$.

$C = Enc_K(M)$ היא התפלגות ההצפנות (מעל המרחב $\mathcal{C}$). בסך הכול יש לנו שלושה משתנים מקריים K, M, C.

הגדרת סודיות מושלמת (לא פורמלי): ההודעה המוצפנת c לא מגלה לאיב שום מידע נוסף על m. (גם אם היא לא חסומה מבחינת זמן חישוב ומשאבי חישוב).

הגדרה פורמלית: מערכת הצפנה סימטרית $\Pi = (KeyGen, Enc, Dec)$ מספקת סודיות מושלמת אם לכל התפלגות M מעל $\mathcal{M}$, לכל הודעה $m \in \mathcal{M}$, ולכל $c \in \mathcal{C}$ כך ש-$Pr[\mathcal{C} = c] > 0$ מתקיים כי:

$$Pr[M = m|\mathcal{C} = c] = Pr[M = m]$$

כלומר יש אי תלות בין המשתנים המקריים M ו-C.

טענה.

צופן ההסתה וצופן ההחלפה אינם מספקים סודיות מושלמת עבור הודעה באורך $\ell > 1$.

הוכחה.

נתמקד בצופן ההסתה. נזכיר כי $k \leftarrow \{0, ..., 25\}$ ואלגוריתם ההצפנה עובר אות אחרי אות ומזיז אותה k מקומות קדימה. נשלול סודיות מושלמת. נניח כי M מוגדרת על ידי $Pr[M = \text{"}aa\text{"}] = Pr[M = \text{"}ab\text{"}] = \frac{1}{2}$.

(לא פורמלי) עבור ההצפנה של "aa" נקבל בהודעה המוצפנת שני תווים זהים. עבור ההצפנה של "ab" נקבל בהודעה המוצפנת שני תווים עוקבים. ולכן ניתן לדלות מידע מההודעה המוצפנת.

(פורמלי):

$$Pr[M = \text{"}aa\text{"}|C = \text{"}AB\text{"}] = 0$$
$$Pr[M = \text{"}aa\text{"}] = \frac{1}{2}$$

□

טענה.

מערכת OTP מספקת סודיות מושלמת.

הוכחה.

יהיו התפלגות M מעל $\mathcal{M}$, הודעה $m \in \mathcal{M}$, והצפנה $c \in \mathcal{C}$.

$$\begin{aligned}
Pr[C=c] &\overset{*}{=} \sum_{w\in\mathcal{M}} Pr[M=w]\cdot Pr[C=c|M=w] \\
&\overset{*_2}{=} \sum_{w\in\mathcal{M}} Pr[M=w]\cdot Pr[K=c\oplus w|M=w] \\
&\overset{*_3}{=} \sum_{w\in\mathcal{M}} Pr[M=w]\cdot Pr[K=c\oplus w] \\
&\overset{*_4}{=} \sum_{w\in\mathcal{M}} Pr[M=w]\cdot \left(\frac{1}{2}\right)^{\ell} \\
&= \left(\frac{1}{2}\right)^{\ell}
\end{aligned}$$

$*$ נובע מנוסחאת ההסתברות השלמה.

$*_2$ נובע מכך ש-$C=c$ אם ורק אם $c=k\oplus w$ אם ורק אם $K=c\oplus w$. (כי $k\oplus w=c\oplus w\oplus w=c$).

$*_3$ נובע מאי תלות בין המשתנה המקרי של המפתח K למשתנה המקרי של ההודעה M.

$*_4$ נובע מכך ש-K מתפלג אחיד.

$$\begin{aligned}
Pr[M=m|C=c] &\overset{*}{=} \frac{Pr[C=c|M=m]\cdot Pr[M=m]}{Pr[C=c]} \\
&\overset{*_2}{=} \frac{Pr[K=c\oplus w]\cdot Pr[M=m]}{Pr[C=c]} \\
&= \frac{\left(\frac{1}{2}\right)^{\ell}\cdot Pr[M=m]}{\left(\frac{1}{2}\right)^{\ell}} \\
&= Pr[M=m]
\end{aligned}$$

$*$ נובע מהגדרת הסתברות מותנית (!).

$*_2$ כמו מקודם.

1.4.1 חסרונות סכימת ה-OTP

כיום היא לא בשימוש מעשי, לפחות לא בתור מערכת הצפנה עצמאית.

1. אורך המפתח צריך להיות לאורך ההודעה. בלתי שמיש עבור הודעות ארוכות (למשל קובץ וידאו גדול).

2. מספק בטיחות אך ורק עבור הצפנה של הודעה אחת. אם נתונות שתי הודעות $c = Enc_k(m)$ ו-$c' = Enc_k(m')$ אז מתקיים $c \oplus c' = m \oplus m'$ וזו בעיה.

3. אם ידועה הודעה אחת וההצפנה המתאימה לה, ניתן לשחזר את מפתח ההצפנה כי $k = m \oplus c$.

שאלה: האם ניתן לתכנן סכימה אחרת שעונה להגדרת הסודיות המושלמת, אבל עושה שימוש במפתחות שאורכן קצר מאורך האורך?

התשובה: סכימת ה-OTP היא הכי טובה מבין כל הסכימות שמקיימות את הגדרת הסודיות המושלמת. אזי לכל סכימה אחרת שמקיימת את ההגדרה יש את אותן החסרונות.

משפט. תהי $\Pi = (Keygen, Enc, Dec)$ מערכת הצפנה סימטרית עם מרחב מפתחות $\mathcal{K}$ ומרחב הודעות $\mathcal{M}$. אם Π מקיימת את הגדרת הסודיות המושלמת, אז $|\mathcal{K}| \geq |\mathcal{M}|$ (עבור OTP מתקיים השוויון).

הוכחה.
נניח כי $|\mathcal{K}| < |\mathcal{M}|$ ונראה כי הסכימה אינה מקיימת את הגדרת הסודיות המושלמת. נסמן ב-M את ההתפלגות האחידה על מרחב ההודעות $\mathcal{M}$. תהא $m \in \mathcal{M}$ ותהא $c \in \mathcal{C}$ הצפנה אפשרית של m. נגדיר:

$$\mathcal{M}(c) \stackrel{def}{=} \{\hat{m} | \hat{m} = Dec_{\hat{k}}(c) \text{ some for } \hat{k} \in \mathcal{K}\}$$

כלומר כל ההודעות שניתן לפענח מ-c בעזרת מפתח כלשהו. אז $m \in \mathcal{M}(c)$ מדרישת הנכונות של מערכת ההצפנה.
אז מתקיים $|\mathcal{M}(c)| \leq |K|$ כי כל אחד מהמפתחות מפענח את c להודעה אחת בלבד.
ההנחה $|\mathcal{M}| > |\mathcal{K}|$ גוררת ש-$|\mathcal{M}| > |\mathcal{M}(c)|$. בפרט, קיימת הודעה $m^* \in \mathcal{M}$ כך ש-$m^* \notin \mathcal{M}(c)$. אז אין אף מפתח המפענח את ההצפנה c להודעה m^*. מנכונותה של מערכת ההצפנה, אין אף מפתח המצפין את m^* ל-c. לכן:

$$Pr[M = m^* | C = c] = 0 \neq \frac{1}{|\mathcal{M}|} = Pr[M = m^*]$$

לכן הסכימה לא מקיימת סודיות מושלמת. □

2 הרצאה 2 - Private Key Encription I

קווי השיעור

1. גישות בטיחות חישוביות

2. הצפנות בלתי ניתנות להבחנה

3. פרימיטב בסיסי: יצרן פסאודו אקראי (PRG)

4. $One\ Time\ Pad$ מבוסס PRG

5. חזרה להצפנות בלתי ניתנות להבחנה

2.1 גישות בטיחות חישוביות

השבוע נעבור מבטיחות מושלת ל"בטיחות חישובית" על מנת להתגבר על הבעיות של סודיות מושלמת. נחליש את המערכת בכך שנדרוש בטיחות רק כנגד יריבים חסומים חישובית (למשל, 2000 שנים בעזרת מחשבי העל הטובים ביותר כיום). בנוסף נאפשר ליריבים להצליח בהסתברות מאוד נמוכה.

2.1.1 גישת הבטיחות הקונקרטית

(לא אוהבים אותה) נגיד שהמערכת $(t, \epsilon) - Secure$ אם יריב שרץ ל-t זמן, יכול לשבור את הסכימה בהסתברות לכל היותר t. וניתן למשל לקבוע $t = 2^{60}$, $\epsilon = 2^{-60}$.

נקודת חוזקה של הגישה היא שניתן להתאים את t ו-ϵ למערכות החישוב הקיימות קיימות. עם זאת, לא בטוח שנוכל להתמודד עם שיפורים אפילו קטנים ביכולות החישוב.

2.1.2 גישת הבטיחות האסימפטוטית

מטרתה של הגישה היא להיפטר מרגישות לשינויים קטנים. ההגדרה המילולית היא שכל יריב עם אלגוריתם הסתברותי פולינומיאלי יכול לשבור אותה בהסתברות זניחה.

הגדרה. נאמר כי אלגוריתם A רץ ב**זמן הסתברותי פולינומיאלי (Probabilistic Polynomial-Time - PPT)** אם קיים פולינום $p(\cdot)$ כך שלכל קלט $x \in \{0,1\}^*$ ולכל מחרוזת רנדומלית של הטלות מטבע $r \in \{0,1\}^*$ (שהיא המימד ההסתברותי - שהוא כאילו מראש) החישוב של $A(x, r)$ מסתיים תוך $p(|x|)$ צעדים.

נוסיף למשחק פרמטר נוסף: פרמטר הבטיחות. משלב זה של הקורס, כל האלגוריתמים שלנו יקבלו את פרמטר הבטיחות בתוך קלט, הן של המערכות עצמן, והן של היריבים שתוקפים אותם. נסמן אותו 1^n (ייצוג אונארי של המספר n). והמפתחות גם יהיו באורך n $k \in \mathcal{K}_n$.

הגדרה. פונקציה $f : \mathbb{N} \to \mathbb{R}^+$ היא **זניחה** אם לכל פולינום $p(\cdot)$ קיים N כך לכל $n > N$ מתקיים:

$$f(n) < \frac{1}{p(n)}$$

דוגמה. הפונקציות $2^{-n}, 2^{-\sqrt{n}}, 2^{-log^2(n)}$ הן זניחות.הפונקציות $\frac{1}{2}, \frac{1}{log^2(n)}, \frac{1}{n^5}$ אינן זניחות.

טענה. יהיו $v_1(n)$ ו-$v_2(n)$ פונקציות זניחות. אזי, לכל פולינום חיובי $p(n)$, הפונקציה $p(n) \cdot (v_1(n) + v_2(n))$ היא זניחה.

נשים לב לתכונות המועילות הבאות:

1. $poly(n) \times poly(n) = poly(n)$
2. $poly(n) \times negligible(n) = negligible(n)$

2.2 הצפנות בלתי ניתנות להבחנה (Indistinguishable Encryptions)

זו הגדרת הבטיחות הבסיסית ביותר מבחינת חוזק עבור מערכות הצפנה הצפנה סימטריות. לעומת סודיות מושלמת, שבה ההצפנה לא מגלה שום מידע, כאן אנחנו דורשים חוסר יכולת להבדיל בין הצפנות של הודעות שונות.

הגדרה. בהינתן מערכת הצפנה סימטרית Π, רמת בטיחות 1^n, $k \leftarrow KeyGen(1^n)$. $b \leftarrow \{0,1\}$ ביט שנבחר באקראיות. ויריב $\mathcal{A}$. היריב בוחר זוג הודעות m_0, m_1. היריב מקבל את ההצפנה של m_b הנבחרה באקראיות והוא צריך לקבוע האם זו m_1 או m_2. אז **ניסוי ה-Indinstinguishability** מוגדר כך:

$$IND_{\Pi,\mathcal{A}}(n) := \begin{cases} 1 & \mathcal{A}\text{Correct Is} \\ 0 & \mathcal{A}\text{Wrong Is} \end{cases}$$

הגדרה. מערכת הצפנה סימטרית Π היא בעלת הצפנות **בלתי ניתנות להבחנה** אם לכל יריב התסברותי פולינומיאלי $\mathcal{A}$, קיימת פונקציה זניחה $v(\cdot)$ כך ש:

$$Pr(IND_{\Pi,\mathcal{A}}(n) = 1) \leq \frac{1}{2} + v(n)$$

כאשר ההסברות נלקחת גם על הטלות המטבע הדרושות לניסוי, וגם על הטלות המטבע של $\mathcal{A}$.

2.3 יצרן פסאודו אקראי ($Pseudorandom\ Generator - PRG$)

הגדרה. תהא $G : \{0,1\}^* \to \{0,1\}^*$ פונקציה הניתנת לחישוב בזמן פולינומיאלי, ויהי $\ell(\cdot)$ פולינום הקובע את אורך הפלט של G, כך שלכל קלט $s \in \{0,1\}^n$ מתקיים $G(s) \in \{0,1\}^{\ell(n)}$. אז הפונקציה G היא **יצרן פסאודו-אקראי** אם שתי הדרישות הבאות מתקיימים:

1. הרחבה: $\ell(n) > n$ (במילים אחרות: $|G(s)| > |s|$)
2. פסאודו-אקראיות: לכל "מבחין" הסתברותי פולינומיאלי (PPT) $\mathcal{D}$, קיימת פונקציה זניחה $\nu(\cdot)$ כך ש:

$$\left|\Pr_{s\leftarrow\{0,1\}^n}\left[\mathcal{D}\left(G(s)\right)=1\right] - \Pr_{r\leftarrow\{0,1\}^{\ell(n)}}\left[\mathcal{D}\left(r\right)=1\right]\right| \leq \nu(n)$$

הערה: הסימון $x \leftarrow \{0,1\}^m$ מסמל ש-x נדגם מהתפלגות אחידה מעל $\{0,1\}^m$, (כל ערך בהסתברות $\frac{1}{2^m}$)

במילים: ההתפלגות $G(s)$ הנוצרת כתוצאה מהפעלת G על $seed$ קצר הנדגם באופן אחיד, למעשה זהה להתפלגות האחידה על פני $\ell(n)$ ביטים מנקודת המבט של כל אלגוריתם הסתברותי פולינומיאלי.

2.3.1 האם קיימים $PRGs$? ונסיונות ייצור כושלים

אנחנו מאמינים שכן, ניתקל באחד בהרצאה הבאה. אבל די קשה לבנות אחד במסגרת ההבנה המדעית הנוכחית של קושי חישובי.

ננסה לבנות יצרן באורך הפלט שלו גדול ב-1 מאורך הפלט. נסמן את הביטים של הקלט: $s = s_1, ..., s_n \in \{0,1\}^n$.

הצעה ראשונה:

$$G(s) = s_1, ..., s_n, 0$$

האם ניתן להבחין ביעילות וביתרון לא זניח בין ההתפלגות $G(s)$ לבין מחרוזת $r_1, ..., r_{n+1}$ הנדגמת באופן אחיד באמת? התשובה היא בוודאי שכן. הגדרה פשוטה למבחין: להחזיר את הביט ה-$n+1$. בעולם האחיד $Pr(1) = 0.5$, אך ב-PRG מתקיים $Pr(1) = 0$.

הצעה שנייה:

$$G(s) = s_1, ..., s_n, s_1$$

גם כאן ניתן להבחין לאור הקשר בין הביט הראשון לאחרון. הגדרה למבחין: להחזיר 1 אמ"מ הביט הראשון שווה לביט האחרון. בעולם האחיד $Pr(1) = 0.5$, אך ב-PRG מתקיים $Pr(1) = 1$.

הצעה שלישית:

$$G(s) = s_1, ..., s_n, z$$
$$\text{where} z = s_1 \oplus ... \oplus s_n$$

גם כאן ניתן להבחין. כי $s_1 \oplus ... \oplus s_n \oplus z = 0$ תמיד. הגדרה למבחין: המציין של האם $s_1 \oplus ... \oplus s_n \oplus z = 0$.בעולם האחיד $Pr(1) = 0.5$, אך ב-PRG מתקיים $Pr(1) = 1$.

לסיכום, לא ממש פשוט לבנות יצרן פסאודו אקראי. למעשה קיום של PRG גורר $P \neq NP$. אבל יש מספר לא קטן של מועמדים.

2.3.2 עובדה שימושית

כל תכונה של ההתפלגות האחידה שניתנת לזיהוי על ידי אלגוריתם הסתברותי פולינומיאלי, צריכה להתקיים גם על ידי הסתברות הפלט של כל יצרן פסאודו אקראי. למעשה ראינו 3 מהתכונות האלו כשניסינו לייצר $PRGs$.

דוגמה. אם G הוא PRG אז קיימת פונקציה זניחה $v(\cdot)$ כך ש:

$$\Pr_{s \leftarrow \{0,1\}^n}\left[\text{in 1's of fraction} G(s) < \frac{1}{4}\right] \leq v(n)$$

2.4 One Time Pad מבוסס PRG

הגדרה. מערכת הצפנה סימטרית One-Time Pad מבוססת PRG

- יהא n פרמטר הבטיחות.
- יהא G יצרן פסאודו אקראי שלכל seed באורך n, מרחיב אותו לאורך $\ell(n)$.
- $\mathcal{K}_n = \{0,1\}^n$ אבל $\mathcal{M}_n = \mathcal{C}_n = \{0,1\}^{\ell(n)}$ (מרחב המפתחות קטן ממרחב ההודעות)
- $KeyGen(1^n)$ דוגם $k \leftarrow \{0,1\}^n$
- $Enc_k(m) = m \oplus G(k)$
- $Dec_k(m) = c \oplus G(k)$

ניתן לחשוב על המערכת בתור OTP כאשר המפתח האפקטיבי שלה מתקבל בתור ערך פלט של יצרן פסאודו אקראי ולא צריך להישמר בצורה מפורשת.

משפט. אם G הוא PRG אז לסכימה יש הצפנות בלתי ניתנות להבחנה.

הוכחה. נוכיח באמצעות שיטת ההוכחה הנפוצה והחשובה בקורס. הוכחת בטיחות בעזרת רדוקציה.

נניח בשלילה שההצפנות ניתנות להבחנה, כלומר, קיים יריב PPT, $\mathcal{A}$, וקיים פולינום $p(n)$ כך ש:

$$Pr(IND_{\Pi,\mathcal{A}}(n)=1) \geq \frac{1}{2}+\frac{1}{p(n)}$$

עבור מספר אינסופי של ערכי n.

נגדיר מבחין $\mathcal{D}$ באופן הבא:

1. בהינתן קלט z, המבחין מריץ את $\mathcal{A}$ ומקבל ממנו זוג הודעות.
2. המבחין ידגום ביט $b \leftarrow \{0,1\}$ המפולג באופן אחיד, ויחזיר ל-$\mathcal{A}$ את ההצפנה $c^* = z \oplus m_b$. הוא יקבל מ-$\mathcal{A}$ את הפלט b' (הניחוש שלו להודעה שממנה ההצפנה הגיעה).
3. המבחין מחזיר 1 אמ"ם $b' = b$. (כלומר אם $\mathcal{A}$ הצליח בניסוי המסומלץ)

זמן הריצה של $\mathcal{D}$ הוא פולינומיאלי (הרצה של $\mathcal{A}$, פעולות xor) לכן הוא PPT כנדרש.

כעת ננתח את המקרה בו $z \leftarrow \{0,1\}^{\ell(n)}$. נשים לב שמהגדרה, c^* מפולג באופן אחיד כי z מפולג באופן אחיד. מכאן שאין ל- $\mathcal{A}$ כל אינפורמציה על b, ולכן ההסתברות שהוא שווה לפלט b' היא בדיוק חצי. כלומר:

$$\Pr_{z\leftarrow\{0,1\}^{\ell(n)}}\left[\mathcal{D}\left(G(s)\right)=1\right]=\frac{1}{2}$$

במקרה השני, בו $z = G(k)$ עבור $k \leftarrow \{0,1\}^n$. כעת מנקודת המבט של $\mathcal{A}$ זה זהה לניסוי $IND_{\Pi,\mathcal{A}}$. לכן מההנחה:

$$\Pr_{k\leftarrow\{0,1\}^{n}}\left[\mathcal{D}\left(G(s)\right)=1\right]=Pr(IND_{\Pi,\mathcal{A}}(n)=1) \geq \frac{1}{2}+\frac{1}{p(n)}$$

נאסוף יחד את שני המקרים, נקבל:

$$\left|\Pr_{k\leftarrow\{0,1\}^{n}}\left[\mathcal{D}\left(G(s)\right)=1\right]-\Pr_{z\leftarrow\{0,1\}^{\ell(n)}}\left[\mathcal{D}\left(G(s)\right)=1\right]\right| \geq \frac{1}{p(n)}$$

בסתירה לתכונת פסאודו האקראיות של G. $\square$

2.5 חזרה להצפנות בלתי ניתנות להבחנה

אנחנו מעוניינים להבין את הבטיחות המתקבלת מהצפנות בלתי ניתנות להבחנה. אי אפשר להבחין בין $Enc_k(m_0)$ ו-$Enc_k(m_1)$. אבל האם אפשר ללמוד משהו על m מ-$Enc_k(m)$?

טענה. תהא Π מערכת הצפנה בעלת הצפנות בלתי ניתנות להבחנה, ויהי $\mathcal{B}$ אלגוריתם הסתברותי פולינומיאלי המקבל בתור קלט את הייצוג האונארי של פרמטר הבטיחות n, והצפנה של הודעה $m \leftarrow \{0,1\}^{\ell(n)}$. אז קיימת פונקציה זניחה $v(\cdot)$ כך ש-

$$Pr(\mathcal{B}(1^n, Enc_k(m)) = LSB(m)) \leq \frac{1}{2} + v(n)$$

($LSB =$ Bit Significant Least)

(במילים אחרות: אם המערכת הצפנה בעלת הצפנות בלתי ניתנות להבחנה, אז לאף אלגוריתם אין ייתרון לא זניח מעבר לחצי, אפילו בהינתן הצפנה של m (מקבלים את הסיכוי של חצי גם אם סתם מנחשים).

הוכחה. נוכיח ברדוקציה. נניח בשלילה שקיים אלגוריתם הסתברותי פולינומיאלי $\mathcal{B}$, וקיים פולינום $p(n)$ כך ש-

$$Pr(\mathcal{B}(1^n, Enc_k(m)) = LSB(m)) > \frac{1}{2} + \frac{1}{p(n)}$$

עבור מספר אינסופי של ערכי n.

לכל $\sigma \in \{0,1\}$ נסמן $I_\sigma \subseteq \{0,1\}^\ell$ את אוסף כל המחרוזות שה-LSB שלהן הוא σ. נשים לב ששתי הקבוצות הללו מחלקות את מרחב ההסתברות באופן שווה.

נגדיר יריב $\mathcal{A}$ כך:

1. $\mathcal{A}$ידגום הודעה $m_0 \leftarrow I_0$ ו-$m_1 \leftarrow I_1$ באופן אחיד ובלתי תלוי. ויפלו אותן.
2. $\mathcal{A}$יקבל בחזרה הצפנה $c^* = Enc_k(m_b)$ עבור $k \leftarrow KeyGen(1^n)$ $b \leftarrow \{0,1\}$ (b נדגם אחיד). (הוא לא יודע את k ו-b.
3. $\mathcal{A}$ מחזיר את $b' = \mathcal{B}(1^n, c^*)$

$\mathcal{A}$ הוא אלגוריתם הסתברותי פולינומיאלי. בנוסף:

$$\begin{aligned}
Pr(IND_{\Pi,\mathcal{A}}(n) = 1) &= Pr(\mathcal{B}(1^n, c^*) = b) \\
&= \frac{1}{2}Pr_{m_0 \leftarrow I_0}(\mathcal{B}(1^n, c^*) = 0) + \frac{1}{2}Pr_{m_1 \leftarrow I_1}(\mathcal{B}(1^n, c^*) = 1) \\
&= Pr_{n \leftarrow \{0,1\}^\ell}(\mathcal{B}(1^n, c^*) = LSB(m)) \\
&\geq \frac{1}{2} + \frac{1}{p(n)}
\end{aligned}$$

בסתירה לכך שמערכת ההצפנה Π מספקת הצפנות בלתי ניתנות להוכחה. $\square$

הטענה הזו מרמזת לנו שהגדרת ההצפנות הבלתי ניתנות להבחנה חזקה יותר מהרושם הראשוני שלה.

2.5.1 בטיחות סמנטית

באופן אינטואיטיבי ההגדרה דורשת שכל מה שאפשר לחשב ביעילות בהינתן הצפנה של הודעה כלשהו m הנדגמת מהתפלגות ידועה, אפשר לחשב גם בלי לקבל את ההצפנה עצמה.

הגדרה. מערכת הצפנה Π היא בעלת **בטיחות סמנטית**, אם לכל יריב הסתברותי פולינומיאלי $\mathcal{A}$, קיים אלגוריתם הסתברותי פולינומיאלי $\mathcal{S}$ (נקרא לו סימולטור) כך שלכל התפלגות $\mathcal{M}$ על M הניתנת לדגימה ביעילות ולכל פונקציות f ו-h הניתנות לחישוב בזמן פולינומיאלי, ישנה פונקציה זניחה $v(\cdot)$ כך ש-

$$|Pr\left(\mathcal{A}(1^n, Enc_k(m), h(m)) = f(m)\right) - Pr\left(\mathcal{S}(1^n, h(m)) = f(m)\right)| \leq v(n)$$

כאשר $k \leftarrow KeyGen(1^n)$ ו-$m \leftarrow \mathcal{M}_n$.
(ההפרש מייצג את ההבדל בין היכולת של היריב $\mathcal{A}$ ללמוד מידע מההצפנה של ההודעה m, לבין היכולת של הסימולטור $\mathcal{S}$ ללמוד בדיוק את אותו המידע מבלי שבכלל קיבל את ההצפנה).
(המידע שאותו $\mathcal{A}$ ו-$\mathcal{S}$ רוצים ללמוד הוא הפונקציה $f(m)$ - לדוגמא, הביט השמאלי).
(המידע המוקדם שלהם על ההודעה m נתון על ידי $h(m)$).

משפט. Π היא בטוחה סמנטית אמ"ם ההצפנות שלה בלתי ניתנות להבחנה.

3 הרצאה 3 - Private Key Encription II

קווי השיעור

- בלתי ניתנות להבחנה חישובית (Indistinguishability Computational)
  - טכניקת הוכחה הוכחה חדשה : ארגומנט היברידי
- בטיחות נגד מתקפות Chosen-plaintext (CPA)
- אובייקט בסיסי שני - פונקציה פסאודו-אקראית PRF
- מערכת הצפנה בטוחה כנגד CPA באמצעות PRF
- $Block\ Cipher$ - היוריסטיקה פרקטית לבניית פונקציות פסאודו אקראיות.

3.1 בלתי ניתנות להבחנה חישובית

אפורמלית, שתי התפלגויות בלתי ניתנות להבחנה חישובית, אם אף אלגוריתם יעיל לא יכול להבחין ביניהן. אזי מבחינת בטיחות חישובית, זה לא משנה אם נשתמש באחת או בשנייה.

באופן כללי אנו מעוניינים בהתפלגויות שניתן לדגום בזמן פולינומיאלי, בפרט בסדרות $X = \{X_n\}_{n\in\mathbb{N}}$, $Y = \{Y_n\}_{n\in\mathbb{N}}$ כאשר n הוא פרמטר הבטיחות.

נתקלנו במקרים פרטיים של זה -

- בהגדרה של יצרן פסאודו אקראי. $X_n = G(s)$ עבור $s \leftarrow \{0,1\}^n$ ו-Y היא ההתפלגות האחידה על $\{0,1\}^{\ell(n)}$.
- בהגדרה של הצפנות בלתי ניתנות להבחנה. $X_n = Enc_k(m_1)$, $Y_n = Enc_k(m_2)$ כאשר ההסתברות היא על $k \leftarrow KeyGen(1^n)$.

הגדרה. התפלגויות $X = \{X_n\}_{n\in\mathbb{N}}$, $Y = \{Y_n\}_{n\in\mathbb{N}}$ **בלתי ניתנות להבחנה חישובית** אם לכל אלגוריתם PPT, $\mathcal{D}$ (המבחין) קיימת פונקציה זניחה $\nu(\cdot)$ כך ש:

$$|Pr_{x\leftarrow X_n}(\mathcal{D}(1^n,x)=1) - Pr_{y\leftarrow Y_n}(\mathcal{D}(1^n,y)=1)| \le \nu(\cdot)$$

נסמן זאת על ידי $X \approx^c Y$.

נשים לב כי הגדרה זו מכלילה פסאודו-אקראיות, והתפלגות היא פסאודו אקראית אם היא בלתי ניתנת להבחנה עם ההתפלגות האחידה.

3.1.1 טכניקת הוכחה: ארגומנט היברידי

נכיר את הטכניקה על ידי הוכחת המשפט הבא:

משפט. יהי $G:\{0,1\}^n \to \{0,1\}^{4n}$ יצרן פסאודו אקראי. אזי $H(s_1,s_2) = G(s_1)||G(s_2)$ הוא יצרן פסאודו אקראי.

הוכחה. נניח בשלילה ש-H הוא לא יצרן פסאודו אקראי, ונוכיח כי במצב זה גם G אינה פסאודו אקראית. אז קיים מבחין $\mathcal{D}$ כך שהוא מבדיל בין ההתפלגות $(G(s_1),G(s_2))$ ובין (r_1,r_2) (כאשר $s_1,s_2 \leftarrow \{0,1\}^n$ ו-$r_1,r_2 \leftarrow \{0,1\}^{4n}$).

הרעיון הבסיסי של <u>ארגומנט היברידי</u> הוא להסתכל על התפלגות היברידית. במקרה שלנו: $(G(s_1),r_2)$.

מה ההסתברות ש-$\mathcal{D}$ מחזיר 1 בהינתן ההתפלגות ההיברידית? קשה לדעת במדויק. אך אם אנחנו יודעים של-$\mathcal{D}$ יש ייתרון ϵ בהבדלה בין $(G(s_1),G(s_2))$ ו-(r_1,r_2), זה אומר שיש לו <u>או</u> ייתרון $\frac{\epsilon}{2}$ בהבדלה בין $(G(s_1),G(s_2))$ ו-$(G(s_1),r_2)$ <u>או</u> ייתרון $\frac{\epsilon}{2}$ בהבדלה בין (r_1,r_2) ו- $(G(s_1),r_2)$. ההוכחה לכך נובעת מאי שוויון המשולש:

$$\begin{aligned}\epsilon &\le |Pr(\mathcal{D}(G(s_1)||G(s_2))=1) - Pr(\mathcal{D}(r_1||r_2)=1)| \\ &\le |Pr(\mathcal{D}(G(s_1)||G(s_2))=1) - Pr(\mathcal{D}(G(s_1)||r_2)=1)| + |Pr(\mathcal{D}(G(s_1)||r_2)=1) - Pr(\mathcal{D}(r_1||r_2)=1)|\end{aligned}$$

עבור כל אחד משני המקרים הללו, נבנה מבחין $\mathcal{A}$ על קלטים באורך $4n$ שייסתור את פסאודו האקראיות של G.

נתחיל במקרה הראשון בו $\mathcal{D}$ מבחין בין $(G(s_1),G(s_2))$ ו-$(G(s_1),r_2)$. נשים לב כי ההתפלגויות זהות בחציין השמאלי. ננצל את העובדה שההתפלגויות שונות בחציין הימני. $\mathcal{A}$ יוגדר כך עבור קלט $z \in \{0,1\}^{4n}$.

- דוגם סיד $s_1 \leftarrow \{0,1\}^n$ ומחזיר את $\mathcal{D}(G(s_1)||z)$.

נשים לב כי אם $z = G(s_2)$ אז המבחין $\mathcal{A}$ יריץ את $\mathcal{D}$ על $(G(s_1),G(s_2))$ ואם $z = r_2$ אז המבחין $\mathcal{A}$ יריץ את $\mathcal{D}$ על $(G(s_1),r_2)$. מאחר ו-$\mathcal{D}$ מבחין, גם $\mathcal{A}$ מבחין בייתרון לא זניח. והוא גם PPT. בסתירה.

במקרה השני בו $\mathcal{D}$ מבחין בין (r_1,r_2) ו-$(G(s_1),r_2)$. נגדיר את A כך עבור קלט $z \in \{0,1\}^{4n}$:

- דוגם סיד $s_1 \leftarrow \{0,1\}^n$ ומחזיר את $\mathcal{D}(z||r_2)$.

העקרון זהה. □

3.2 בטיחות נגד מתקפות Chosen Plaintext (CPA)

נשים לב לבעיה בהגדרת הצפנות בלתי ניתנות להבחנה - היא לא נותנת הגנה במצב בו היריב נחשף ליותר מהצפנה אחת. זו כמובן ציפייה לא ריאליסטית. נרצה למדל התקפות מציאותיות יותר מסוג זה. נכליל את הניסוי IND.

הגדרה. ניסוי **CPA** הוא כמו ניסוי **IND**, אך בכל נקודת זמן היריב יכול לקבל הצפנה c של כל הודעה m שיירצה (בעזרת המפתח k שנבחר בראשית הניסוי). נסמן אותו כך:

$$IND_{\Pi,\mathcal{A}}^{CPA} = \begin{cases} 1 & \text{if} b = b' \\ 0 & \text{otherwise} \end{cases}$$

ננסח זאת במילים אחרות "גישת אורקל לאלגוריתם ההצפנה". נסמן זאת בסימון $\mathcal{A}^{Enc_k(\cdot)}$.

נשים לב כי בגלל שהיריב $\mathcal{A}$ הוא PPT, גם מספר ההצפנות שהוא יכול לבקש הוא פולינומיאלי.

הגדרה. מערכת הצפנה סימטרית Π היא בעלת הצפנות בלתי ניתנות להבחנה תחת מתקפת CPA (או בקיצור - בטוחה CPA) אם לכל יריב PPT, $\mathcal{A}$, קיימת פונקציה זניחה $\nu(\cdot)$ כך ש-

$$Pr(IND_{\Pi,\mathcal{A}}^{CPA}(n) = 1) \leq \frac{1}{2} + v(n)$$

ההסתברות פה היא הן על ידי הטלות המטבע של היריב $\mathcal{A}$ והן על פני הטלות המטבע של האלגוריתמים $KeyGen$ ו-Enc, והן על בחירת הביט b.

- על פניו מתקבל הרושם שקל ליריב לנצח בניסוי CPA אם $\mathcal{A}$ בודק מה ההצפנות של m_1 ו-m_2. אך Enc היא לא בהכרח דטרמיניסטית ולכן זה לא כך. אם כך תמיד נרצה להשתמש ב-Enc הסתברותית כך שאם מצפינים אותה הודעה יותר מפעם אחת ההסתברות שנקבל את אותה הצפנה הוא זניח.

- אינטואיטיבית - נותן בטיחות כנגד ידיעת ההצפה של מספר הודעות.

- האם ההגדרה "חזקה מדי"? איך נבנה מערכות הצפנה שמספקות אותה?

3.3 פונקציה פסאודו-אקראית PRF

זו פונקציה ש"נראית" כמו פונקציה אקראית לחלוטין. ננסה להבין מה זו פונקציה אקראית לחלוטין. נסמן ב-$Func_{n\to\ell}$ את קבוצה כל הפונקציות מ-$\{0,1\}^n$ ל-$\{0,1\}^\ell$. מספר הפונקציות האלו הוא $2^{\ell\cdot 2^n}$. פונקציה אקראית $h(x)$ לחלוטין היא פונקציה הנדגמת באופן אחיד מ-$Func_{n\to\ell}$.

בפועל, זה אומר שלכל $x \in \{0,1\}^n$, הערך $h(x) \in \{0,1\}^\ell$ נבחר באופן אחיד ובלתי תלוי לכל x אחר.

הגדרה. תהי $F:\{0,1\}^n\times\{0,1\}^n\to\{0,1\}^{\ell(n)}$ (הקלט הראשון הוא המפתח $k\in\{0,1\}^n$, והקלט השני הוא הקלט עצמו, $x\in\{0,1\}^n$) פונקציה הניתנת לחישוב בזמן פולינומיאלי. הפונקציה F **פונקציה פסאודו-אקראית** אם לכל מבחין PPT, $\mathcal{D}$, קיימת פונקציה זניחה $\nu(\cdot)$ כך ש:

$$\left|Pr\left(\mathcal{D}^{F_k(\cdot)}(1^n)=1\right)-Pr\left(\mathcal{D}^{h(\cdot)}(1^n)=1\right)\right|\leq\nu(n)$$

כאשר $k\leftarrow\{0,1\}^n$ ו-$h\leftarrow Func_{n\to\ell}$.

המשמעות של $\mathcal{D}^{F_k(\cdot)},\mathcal{D}^{h(\cdot)}$ היא שיש ל-$\mathcal{D}$ גישת אורקל לכל אחת מהפונקציות, כלומר הוא יכול לשלוח לה קלטים ולקבל חזרה את ערכי הפונקציה (לכל היותר מספר פולינומי פעמים).

- אנו מניחים לשם פשטות כי המפתח k נדגם באופן אחיד $k\leftarrow\{0,1\}^n$ כאשר n הוא פרמטר הבטיחות. באופן כללי אפשר להגדיר ביחד עם הפונקציה F גם אלגוריתם $KeyGen$ המייצר התפלגות כלשהי על מפתחות. במשך לרוב לא נעשה את זה.

3.3.1 הוכחת בטיחות עם פונקציות פסאודו-אקראיות

לרוב נשתמש בפונקציות פסאודו-אקראיות כדי לבנות מערכות הצפנה. הוכחת הבטיחות של מערכות של פונקציות פסאודו אקראיות לרוב הולכת ככה:

1. ננסה לחשוב מה היה קורה בתקיפה של היריב $\mathcal{A}$ אילו היינו מחליפים את הפונקציה הפסאודו-אקראית לפונקציה אקראית לחלוטין.
2. נוכיח כי בניסוי המחשבתי הזה, היריב לא היה מצליח לשבור את המערכת, אלא בהסתברות זניחה.
3. נוכיח כי היריב לא מסוגל להבחין בין המערכת עם הפונקציה האקראית לחלוטין לבין המערכת המקורית. נעשה זאת בשלילה - נראה כי אם היריב מסוגל לשבור את המערכת המקורית, שעושה שימוש בפונקציה פסאודו-אקראית, אז ניתן לעשות בו שימוש כדי להוכיח שהפונקציה לא פסאודו אקראית.
4. נסיק כי המערכת המקורית בטוחה.

3.4 מערכת הצפנה בטוחה כנגד CPA באמצעות PRF

תהא $F:\{0,1\}^n\times\{0,1\}^n\to\{0,1\}^\ell$ פונקציה פסאודו אקראית. נגדיר מערכת הצפנה Π_F באופן הבא:

- $KeyGen$: נדגום באופן אחיד $k\leftarrow\{0,1\}^n$.
- Enc: עבור מפתח $k\in\{0,1\}^n$ והודעה-$m\in\{0,1\}^\ell$, האלגוריתם דוגם באופן אחיד $r\leftarrow\{0,1\}^n$ ומחזיר את

$$c=(r,F_k(r)\oplus m)$$

(נשים לב כי האלגוריתם הסתברותי, ומשתמש ב-$One-Time-Pad$ עם הפלט של F).

- Dec: עבור מפתח $k\in\{0,1\}^n$ והצפנה $c=(r,s)$ האלגוריתם מחזיר $m=F_k(r)\oplus m$.

משפט. אם F היא PRF (פונקציה פסאודו אקראית) אזי הסכימה Π_F לעיל היא בטוחה כנגד מתקפות CPA.

הוכחה. נסמן ב-Π_h את המערכת המתקבלת מ-Π_F כאשר מחליפים את הפונקציה הפסאודו-אקראית F בפונקציה אקראית לחלוטין h.
בשלב הראשון נוכיח Π_h בטוחה כנגד מתקפות CPA ובשלב השני נוכיח שאף יריב לא יכול להבחין בין שתי המערכות.
יהי $\mathcal{A}$ יריב PPT.

טענת עזר 1: קיימת פונקציה זניחה $\nu(n)$ כך ש:

$$\left|Pr(IND^{CPA}_{\Pi_F,\mathcal{A}}(n)=1)-Pr(IND^{CPA}_{\Pi_h,\mathcal{A}}(n)=1)\right|\leq\nu(n)$$

טענת עזר 2: אם $q(n)$ הוא מספר השאלות של $\mathcal{A}$ לאורקל ההצפנה, אז:

$$Pr(IND^{CPA}_{\Pi_h,\mathcal{A}}(n)=1)\leq\frac{1}{2}+\frac{q(n)}{2^n}$$

כנשחבר את שתי הטענות נקבל:

$$\begin{aligned}Pr(IND^{CPA}_{\Pi_F,\mathcal{A}}(n)=1)&\leq\left|Pr(IND^{CPA}_{\Pi_F,\mathcal{A}}(n)=1)-Pr(IND^{CPA}_{\Pi_h,\mathcal{A}}(n)=1)\right|+Pr(IND^{CPA}_{\Pi_h,\mathcal{A}}(n)=1)\\&\leq\frac{1}{2}+\left(\frac{q(n)}{2^n}+\nu(n)\right)\end{aligned}$$

נשים לב כי מספר השאלות של $\mathcal{A}$ לאורקל ההצפנה הוא פולינומיאלי ולכן $\frac{q(n)}{2^n}$ היא פונקציה זניחה. כלומר זה מה שרצינו להוכיח.

הוכחת טענת עזר 1: נניח בשלילה שקיים יריב $\mathcal{A}$,PPT המקיים

$$\left|Pr(IND^{CPA}_{\Pi_F,\mathcal{A}}(n)=1)-Pr(IND^{CPA}_{\Pi_h,\mathcal{A}}(n)=1)\right|\geq\frac{1}{p(n)}$$

עבור מספר אינסופי של ערכי n. אז נגדיר מבחין $\mathcal{D}^{\mathcal{O}}$ כך:
$\mathcal{D}$ יגריל $b\leftarrow\{0,1\}$ וייקרא ל-$\mathcal{A}$.
$\mathcal{A}$ יבקש הצפנות לאורך הניסוי. $\mathcal{D}$ יחקה את אלגוריתם ההצפנה של המערכת, כאשר הוא משתמש ב-$\mathcal{O}$ בתור הפונקציה שפונקציית ההצפנה משתמשת בה. ויחזיר ל-$\mathcal{A}$ את ההצפנות.
$\mathcal{A}$ יחזיר שתי הודעות m_0,m_1. $\mathcal{D}$ יחזיר ל-$\mathcal{A}$ את ההצפנה של m_b כמו שהוגדרה בשלב 2.
$\mathcal{A}$ יחזיר ביט b'. $\mathcal{D}$ יחזיר 1 אם $b'=b$ ו-0 אחרת.
נשים לב זמן הריצה של המבחין הוא פולינומיאלי. כעת ננתח את ההסתברות שהמבחין מחזיר אחת בהינתן ש-$\mathcal{O}$ היא פונקציה פסאודו אקראית אל מול אקראית.
אם $\mathcal{O}=F_k$ עבור $k\leftarrow\{0,1\}^n$:

$$Pr(\mathcal{D}^{F_k(\cdot)}(1^n)=1)=Pr(IND^{CPA}_{\Pi_F,\mathcal{A}}(n)=1)$$

אם $\mathcal{O}=h$ כאשר $h\leftarrow Func_{n\to\ell}$:

$$Pr(\mathcal{D}^{h(\cdot)}(1^n)=1)=Pr(IND^{CPA}_{\Pi_h,\mathcal{A}}(n)=1)$$

3.5 $Block\ Cipher$ - היוריסטיקה פרקטית לבניית פונקציות פסאודו אקראיות.

Block Cipher הוא מימוש שאמור להיחשב בטוח היוריסטי (לא מוכח) לפונקציה פסאודו אקראית ואף לפרמוטציה פסאודו אקראית. השוני המרכזי בין זה לבין פונקציה פסאודו אקראית אמיתית היא שכאן אנחנו מעוניים בבטיחות קונקרטית ולא אסימפטוטית, כלומר עבור פרמטרים ספציפיים של n ו-ℓ. לכן לא ניתן להגדיר בטיחות כמו שעשינו עד כה. אז נגיד שבלוק סייפר בטוח (באופן היוריסטי) אם לא ניתן לתקוף אותו (באופ היוריסטי) על ידי ביצוע מספר פעולות הקטן משמעותית מ-2^n.

AES היא דוגמא יחסית מודרנית לפונקציה כזאת. צריך לחלק את ההודעה לבלוקים ולהצפין כל אחד בנפרד. כשדוגמים r לכל בלוק בנפרד זה אומנם עדיין בטוח, אבל הופך את ההודעה המוצפנת להיות פי 2 מאורך ההודעה המקורית. לכן קיים מצב $Counter(CTR)$:

$$Enc_k(m_1, ..., m_\ell : r) = (r, F_k(r+1) \oplus m_1, F_k(r+2) \oplus m_2, ..., F_k(r+\ell) \oplus m_\ell)$$

משפט. אם F היא פונקציה פסאודו אקראית אז מצב $Counter$ הוא בטוח כנגד CPA.

4 הרצאה 4 - Message Authentication and Hash Functions

קווי השיעור

- מערכת לאימות הודעות - Message Authentication Code (MAC)
- אובייקט בסיסי שלישי - פונקציות האש עמידות בפני התנגשויות - Collision-Resistant hash functions
- חזרה להצפנה

4.1 מערכות לאימות הודעות - Message Authentication Code (MAC)

4.1.1 הגדרות

מטרתן של מערכות לעימוד הודעות היא להבטיח, שכאשר אליס מעבירה הודעה לבוב, איב לא משנה את ההודעה. כלומר לבדוק האם ההודעה שהתקבלה באמת נשלחה. כאן איב יכולה לא רק לצפות בהודעות, אלא גם לשנות אותן ולהכניס הודעות חדשות.

בניגוד להצפנה, אנחנו לא רוצים להבטיח את סודיות המידע.

הגדרה. **<u>מערכת לאימות הודעות</u>**, בקיצור MAC, מורכבת משלושה אלגוריתם.

- אלגוריתם יצירת המפתחות Gen, מקבל בתור קלט את הייצוג האונארי של פרמטר הבטיחות 1^n, ומחזיר מפתח סודי k הנדגם מהתפלגות כלשהי.
- אלגוריתם יצירת ערכי האימות Mac, מקבל בתור קלט מפתח סודי k והודעה $m \in \{0,1\}^*$ ומחזיר ערך אימות (טאג) $t \in \{0,1\}^*$. (יכול להיות הסתברותי)
- אלגוריתם הווידוא $Vrfy$, מקבל בתור קלט מפתח סודי k, הודעה m, וערך אימות (טאג) פוטנציאלי t, ומחזיר ביט b המסמל קבלה או דחייה של ערך האימות t עבור ההודעה m ביחס למפתח הסודי k.

ומקיימת:

- <u>נכונות:</u> לכל k, m מתקיים $Vrfy_k(m, Mac_k(m)) = 1$.
- <u>בטיחות:</u> לכל יריב PPT, $\mathcal{A}$ ישנה פונקציה זניחה $\nu(\cdot)$ כך שהחל מ-n מסוים:

$$Pr\left[MacForge_{\Pi,\mathcal{A}}(n) = 1\right] \leq \nu(\cdot)$$

בכל פעם שאליס תרצה לשלוח לבוב הודעה m, היא תשלח ביחד עם ההודעה ערך אימות t, שהיא הפיקה בעזרת המפתח הסודי המשותף שלהם k. בצד של בוב, הוא יוודא בעזרת k את ערך האימות עבור ההודעה ובכך בודק אם איב שינתה אותה בדרך.

הגדרה. <u>ניסוי $MacForge$:</u>

1. מפעילים את אלגוריתם יצירת המפתחות $k \leftarrow Gen(1^n)$.
2. מריצים את היריב $\mathcal{A}$ על הייצוג האונארי של פרמטר הבטיחות בתור קלט, ונותנים לו גישה לאורקל Mac_k. היריב יכול לבקש ערכי אימות עבור כל מספר פולינומיאלי של הודעות. נסמן ב-Q את אוסף כל ההודעות ש-$\mathcal{A}$ ביקל מהאורקל.
3. $\mathcal{A}$ מחזיר זוג (m^*, t^*).
4. תוצאת הניסוי:

$$MacForge_{\Pi,\mathcal{A}}(n) = \begin{cases} 1 & \text{if} Vrfy(m^*, t^*) = 1 \text{ and } m^* \notin Q \\ 0 & \text{otherwise} \end{cases}$$

נשים לב שזה לא מונע מתקפות "Replay" כלומר הגדרת הבטיחות לא מונעת מיריב לשלוח אותה הודעה שכבר אומתה שוב.

4.1.2 בניית מערכת Mac להודעות באורך קבוע

תהא $F:\{0,1\}^n\times\{0,1\}^n\to\{0,1\}^n$ פונקציה פסאודו-אקראית. נגדיר

- Gen דוגם באופן אחיד $k\leftarrow\{0,1\}^n$.
- Mac מקבל $k\in\{0,1\}^n$, $m\in\{0,1\}^n$ ומחזיר את $t=F_k(m)$.
- $Vrfy_{..}$ מקבל $k\in\{0,1\}^n$, $m\in\{0,1\}^n$, $t\in\{0,1\}^n$ ומחזיר 1 אם $t=F_k(m)$ ו-0 אחרת.

משפט. אם F היא פונקציה פסאודו-אקראית, אז מערכת ה-Mac לעיל בטוחה.

הוכחה. נניח בשלילה כי המערכת לא בטוחה, ונוכיח כי קיים מבחין $\mathcal{D}$ ל-F בסתירה.

נניח בשלילה שקיים יריב PPT, $\mathcal{A}$, ושקיים פולינום $p(n)$ כך ש:

$$Pr\left[MacForge_{\Pi,\mathcal{A}}(n)=1\right]\geq\frac{1}{p(n)}$$

עבור אינסוף ערכי n. נגדיר מבחין $\mathcal{D}^{\mathcal{O}}$ כך.

1. מקבל את הייצוג האונארי של פרמטר 1^n וגישת אורקל לפונקציית $\mathcal{O}$.
2. מריץ את היריב $\mathcal{A}$ על 1^n. $\mathcal{A}$ חושב שהוא בניסוי $MacForge$ ולכן מבקש ערכי אימות כרצונו אותם $\mathcal{D}$ צריך להחזיר לו. בהינתן כל הודעה m עבורה $\mathcal{A}$ מבקש ערך אימות, $\mathcal{D}$ מבקש מהאורקל שלו את $t=\mathcal{O}(m)$ ומחזיר את t ל-$\mathcal{A}$.
3. כאשר $\mathcal{A}$ מסיים את ריצתו ומחזיר (m^*,t^*), $\mathcal{D}$ מחזיר 1 אמ"ם $m^*\notin Q$ (כאשר Q היא קבוצת ההודעות ש-$\mathcal{A}$ ביקש) וגם $t^*=\mathcal{O}(m^*)$ ואחרת הוא מחזיר 0.

נשים לב כי $\mathcal{D}$ הוא אלג' PPT. ננתח את ההסתברות ש-$\mathcal{D}$ מחזיר 1 עבור כל אחד מהמקרים. כאשר $\mathcal{O}=F_k$:

$\mathcal{D}$ מסמלץ ל-$\mathcal{A}$ בדיוק את הניסוי $MacForge_{\Pi,\mathcal{A}}(n)$. על כן:

$$Pr\left[D^{F_k}(1^n)=1\right]=Pr\left[MacForge_{\Pi,\mathcal{A}}(n)=1\right]$$

כאשר $\mathcal{O}=h$ פונקציה אקראית לחלוטין:

כאשר $Q\neq m^*$, מנקודת המבט של היריב $\mathcal{A}$, הערך $h(m^*)$ מפולג באופן אחיד ולא תלוי בשום ערך ערך אחר שהיריב $\mathcal{A}$ צפה בו במהלך הניסוי. על כן:

$$Pr\left[D^{F_k}(1^n)=1\right]=2^{-n}$$

על כן:

$$\left|Pr\left[D^{F_k}(1^n)=1\right]-Pr\left[D^h(1^n)=1\right]\right| = \left|Pr\left[MacForge_{\Pi,\mathcal{A}}(n)=1\right]-2^{-n}\right|$$
$$\geq \frac{1}{p(n)}-2^{-n}$$

עבוא אינסוף ערכי n, וזו אינה פונקציה זניחה בסתירה לכך ש-F היא פונקציה פסאודו-אקראית. $\square$

4.1.3 בניית מערכת Mac להודעות באורך משתנה

נניח שנתונה לנו מערכת Mac: $\hat{\Pi}=(\hat{Gen},\hat{Mac},\hat{Vrfy})$ לאימות הודעות באורך קבוע כלשהו. נבנה מערכת Mac להודעות באורך משתנה $\Pi=(Gen,Mac,Vrfy)$ כך:

- <u>ניסיון ראשון:</u> נחלק את ההודעה m לבלוקים $m_1,...,m_d$ באורך הקבוע. נפעיל את$\hat{Gen}$פעם אחת כדי לקבל מפתח k. נפעיל את $\hat{Mac}$ על כל בלוק m_i בנפרד כדי לקבל ערכי אימות $t_1,...,t_d$. נחזיר 1 ב-$Vrfy$ אם $\hat{Vrfy}$ מחזיר 1 לכל (m_i,t_i).

אבל זה לא טוב. אם $t=(t_1,t_2)$ זה ערך אימות עבור $m=(m_1,m_2)$, היריב יכול לבקש את ערך האימות עבור m, ואז להחזיר (t_1,m_1), והסתברותו לנצח בניסוי $MacForge$ היא 1.

- <u>ניסיון שני:</u> נחלק את ההודעה m לבלוקים $m_1,...,m_d$ באורך הקבוע. נפעיל את$\hat{Gen}$פעם אחת כדי לקבל מפתח k. נפעיל את $\hat{Mac}$ על כל שרשור $d|m_i$ בנפרד כדי לקבל ערכי אימות $t_1,...,t_d$. כאשר d הוא מספר הבלוקים. נחזיר 1 ב-$Vrfy$ אם $\hat{Vrfy}$ מחזיר 1 לכל $(d|m_i,t_i)$.

גם ניסיון זה אינו בטוח. אם $t=(t_1,t_2)$ זה ערך אימות עבור $m=(m_1,m_2)$, היריב יכול לבקש את ערך האימות עבור m, ואז להחזיר $((t_2,t_1),(m_2,m_1))$.

- <u>ניסיון שלישי:</u> נחלק את ההודעה m לבלוקים $m_1,...,m_d$ באורך הקבוע. נפעיל את$\hat{Gen}$פעם אחת כדי לקבל מפתח k. נפעיל את $\hat{Mac}$ על כל שרשור $d|i|m_i$ בנפרד כדי לקבל ערכי אימות $t_1,...,t_d$. כאשר d הוא מספר הבלוקים. נחזיר 1 ב-$Vrfy$ אם $\hat{Vrfy}$ מחזיר 1 לכל $(d|i|m_i,t_i)$.

גם ניסיון זה אינו בטוח. אם $t=(t_1,t_2)$ זה ערך אימות עבור $m=(m_1,m_2)$, וגם $t'=(t'_1,t'_2)$ זה ערך אימות עבור $m'=(m'_1,m'_2)$, היריב יכול לבקש את ערך האימות עבור m,m', ואז להחזיר $((t_1,t'_2),(m_1,m'_2))$.

- <u>**פתרון 1:**</u> נחלק את ההודעה m לבלוקים $m_1,...,m_d$ באורך הקבוע. נפעיל את$\hat{Gen}$פעם אחת כדי לקבל מפתח k. נדגום באופן אחיד ערך r ונפעיל את $\hat{Mac}$ על כל שרשור $r|d|i|m_i$ בנפרד כדי לקבל ערכי אימות $(r,t_1,...,t_d)$.נחזיר 1 ב-$Vrfy$ אם $\hat{Vrfy}$ מחזיר 1 לכל $(r|d|i|m_i,t_i)$.

נשים לב כי אם הערך r לא קצר מדי, ההסתברות שלשתי הודעות יש את אותו ערך r היא זניחה. דרך זו אכן עובדת! אבל חיסרון משמעותי של הפתרון היא העובדה כי ערכי האימות שלו מאוד ארוכים! נרצה ערכי אימות קצרים שאינם תלויים באורך ההודעה.

- <u>**פתרון 2:**</u> $CBC-MAC$: $Mac_k(m)=t_d$ כאשר: $t_0=0^n$ ו-$t_i=F_k(t_{i-1}\oplus m_i)$ לכל $1\leq i\leq d$. F היא פונקציה פסאודו אקראית ו-d נבחר מראש.

זה פתרון מאוד נפוץ.

- **פתרון 3:** Hash and Authenticate נכווץ את ההודעה m ל"טביעת אצבע"-$H(m)$ ואז נאמת את $H(m)$.

כאן צריך שיהיה קשה למצוא $m \neq m'$ כך ש-$H(m) = H(m')$.

4.2 פונקציות האש העמידות בפני התנגשויות - Collision-Resistant hash functions

זה האובייקט הבסיסי השלישי שלנו בקורס. לפונקציות האש יש שימושים רבים ומגוונים בקריפטוגרפיה ואנו נראה כיצד ניתן לעשות בהן שימוש לעיבוד הודעות בכל אורך פולינומיאלי ולאו דווקא קבוע אלא משתנה.

פונקציות האש עמידות בפני התנגשויות צריכות מצד אחד, לכווץ כל קלט לפלט באורך קבוע כלשהו ומצד שני, צריך שיהיה קשה למצוא $x \neq x'$ כך ש-$H(x) = H(x')$.

הגדרה. **פונקציית האש עמידה בפני התנגשויות** היא זוג $\Phi = (Gen, H)$ כך ש:

- אלגוריתם ייצור המפתחות Gen הוא אלג' PPT המקבל בתור קלט את פרמטר הבטיחות 1^n ומחזיר מפתח s.
- אלגוריתם החישוב H מקבל בתור קלט מפתח s וקלט $x \in \{0,1\}^*$ ומחזיר פלט $H_s(x) \in \{0,1\}^{\ell(n)}$.
- בטיחות: לכל יריב PPT ,$\mathcal{A}$ קיימת פונקציה זניחה $\nu(\cdot)$ כך ש-
-

$$Pr\left[HashColl_{\Phi,\mathcal{A}}(n)\right] \leq \nu(\cdot)$$

כלומר הפונקציה מכווצת כל קלט שאורכו לפחות $\ell(n)+1$ ביטים.

הגדרה. ניסוי_ $HashColl_{\Phi,\mathcal{A}}(n)$:

1. מריצים את אלגוריתם יצירת המפתחות כדי לקבל $s \leftarrow Gen(1^n)$.
2. היריב $\mathcal{A}$ מקבל את המפתח s, והוא מחזיר שני ערכים (x, x')
3. תוצאת הניסוי:

$$HashColl_{\Phi,\mathcal{A}}(n) = \begin{cases} 1 & \text{if} H_s(x) = H_s(x') \text{ and } x \neq x' \\ 0 & \text{otherwise} \end{cases}$$

נשים לב כי בניגוד לניסויים אחרים, כאן היריב יודע את המפתח.

4.2.1 מתקפת יום ההולדת

נדגום $q \approx 2^{\frac{\ell}{2}}$ קלטים באקראי. נחשב את ערך הפונקציה H עבור כל אחד ונקווה להתנגשות. מתקפה זו תמצא התנגשות לא טריוואלית בהתסברות לא רעה - אפילו אם H הייתה פונקציה אקראית לחלוטין, ההסתברות של כל זוג לעבור התנגשות לא טריוויאלית היא $2^{-\ell}$, לכן התוחלת עבור 2^{ℓ} זוגות היא 1. לאור זאת לפונקציות האש הנמצאות כיום בשימוש, $128 \leq \ell$ כדי להבטיח בטיחות.

דוגמאות כיום בשימוש - $MD5, SHA-1, ..., SHA-3$

4.2.2 גישת ה-Hash and Authenticate במערכות MAC

נניח שנתונה לנו מערכת Mac : $\hat{\Pi} = (\hat{Gen}, \hat{Mac}, \hat{Vrfy})$ לאימות הודעות באורך קבוע כלשהו. נתונה לנו גם פונקציית $\Phi = Hash$ (Gen_H, H). אנחנו מניחים שאורך הפלט של Φ היא ℓ ביטים וכי המערכת $\hat{\Pi}$ פועלת על הודעות באורך ℓ ביטים. נבנה מערכת Mac להודעות באורך פולינומיאלי משתנה $\Pi = (Gen, Mac, Vrfy)$ כך :

- Gen מקבל בתור פלט את הייצוג האונארי של מערכת הבטיחות 1^n ודוגם שתי מפתחות $k \leftarrow \hat{Gen}(1^n)$ ו-$s \leftarrow Gen_H(1^n)$, ומחזיר את (k,s)
- Mac מקבל בתור קלט (k,s) ו-$m \in \{0,1\}^*$ ומחזיר $\hat{Mac}_k(H_s(m))$
- $Vrfy$ מקבל בתור קלט (k,s) ו-$m \in \{0,1\}^*$ ו-$t \in \{0,1\}^*$ ומחזיר את $\hat{Vrfy}_k(H_s(m), t)$

משפט. אם $\hat{\Pi}$ מערכת MAC בטוחה, ו-Φ חסינה בפני התנגשויות, אז Π היא מערכת MAC בטוחה.

הוכחה. בהינתן יריב $\mathcal{A}$ ששובר את הבטיחות של Π נוכל או לשבור את הבטיחות של $\hat{\Pi}$ או לשבור את החסינות להתנגשויות של Φ. בדומה לארגומנט היברידי, גם כאן לא יהיה לנו נחוץ לדעת איזה משני המקרים מתקיים.

יהא יריב $\mathcal{A}$ כנגד המערכת החדשה Π. נגדיר את המאורע $Collision$ המציין אם בניסוי $MacForge$ $\mathcal{A}$ ביקש מהאורקל אימות עבור $m_i \neq m^*$ כך ש-$H_s(m_i) = H_s(m^*)$ (כאשר m^* זו ההודעה שהוא מחזיר בניסוי ה-$MacForge$).

נשים לב כי בכל הרצה של הניסוי בה המאורע $Collision$ מתרחש, היריב $\mathcal{A}$ מציע עבורנו התנגשות לא טריוויאלית.

נשים לב כי בכל הרצה של הניסוי בה המאורע $Collision$ לא מתרחש (ו-$\mathcal{A}$ מנצח) נוכל להשתמש ב-$\mathcal{A}$ כדי לזייף אימות של ההודעה $H_s(m^*)$. כי $H_s(m^*) \notin \{H_s(m_1), ..., H_s(m_q)\}$.

עד כה זו הייתה רק אינטואציה. פורמלית, אם $\mathcal{A}$ יריב הסתברותי פולינומיאלי אז :

$$Pr\left[MacForge_{\Pi,\mathcal{A}}(n) = 1\right] \leq Pr\left[Collision\right] + Pr\left[MacForge_{\Pi,\mathcal{A}}(n) = 1 \wedge \overline{Collision}\right]$$

<u>תת-טענה 1 :</u> קיימת פונקציה זניחה $\nu_1(n)$ כך ש-$Pr\left[Collision\right] \leq \nu_1(n)$.

<u>תת-טענה 2 :</u> קיימת פונקציה זניחה $\nu_2(n)$ כך ש-$Pr\left[MacForge_{\Pi,\mathcal{A}}(n) = 1 \wedge \overline{Collision}\right] \leq \nu_2(n)$.

אם נשתמש בשתי הטענות נקבל :

$$Pr\left[MacForge_{\Pi,\mathcal{A}}(n) = 1\right] \leq \nu_1(n) + \nu_2(n)$$

החל מ-n מסוים כנדרש.

<u>הוכחת תת-טענה 1</u>: נניח בשלילה שקיים יריב PPT , $\mathcal{A}$, ושקיים פולינום $p(n)$ כך ש-$Pr\left[Collision\right] \geq \frac{1}{p(n)}$ עבור מספר אינסופי של ערכי n.

נבנה אלגוריתם $\mathcal{C}$ כך:

1. בהינתן מפתח s בתור קלט, דוגם מפתח $k \leftarrow \hat{Gen}(1^n)$ ומריץ את $\mathcal{A}$.
2. $\mathcal{A}$ חושב שהוא בניסוי $MacForge$ והוא מבקש את האימות עבור הודעות $m_1, ..., m_q$.
3. $\mathcal{C}$ יחשב את ערכי האימות $\hat{Mac}_k(H_s(m_i))$ ויחזיר אותם ל-$\mathcal{A}$.
4. $\mathcal{A}$ יחזיר (m^*, t^*).
5. אם קיים $m \in \{m_1, ..., m_q\}$ כך ש-$m \neq m^*$ ו-$H_s(m) = H_s(m^*)$, $\mathcal{C}$ יחזיר (m, m^*). אחרת יחזיר "לא".

מבחינת זמן ריצה $\mathcal{C}$ פולינומיאלי. נשים לב כי $\mathcal{C}$ מסמלץ ל-$\mathcal{A}$ באופן מושלם את הניסוי $MacForge_{\Pi,\mathcal{A}}$. לכן מהגדרתם של המאורע $Collision$ ושל האלגוריתם $\mathcal{C}$ נקבל:

$$Pr\left[HashColl_{\Phi,\mathcal{C}}(n) = 1\right] = Pr\left[Collsion\right] \geq \frac{1}{p(n)}$$

בסתירה להנחה כי Φ עמידה בפני מציאת התנגשויות.

<u>הוכחת תת-טענה 2</u>: נניח בשלילה שקיים יריב PPT , $\mathcal{A}$, ושקיים פולינום $p(n)$ כך ש-$Pr\left[MacForge_{\Pi,\mathcal{A}}(n) = 1 \wedge \overline{Collision}\right] \geq \frac{1}{p(n)}$ עבור מספר אינסופי של ערכי n.

נבנה יריב $\hat{\mathcal{A}}$ כנגד מערכת הMAC $\hat{\Pi}$ כך:

1. בהינתן 1^n דוגם מפתח $s \leftarrow Gen_H(1^n)$ ומריץ את $\mathcal{A}$.
2. $\mathcal{A}$ חושב שהוא בניסוי $MacForge$ והוא מבקש את האימות עבור הודעות $m_1, ..., m_q$.
3. $\hat{\mathcal{A}}$ יחשב את ההודעות $H_s(m_i)$ ויבקש עבורם את ערכי האורקל $\hat{Mac}_k(H_s(m_i))$ ויחזיר אותם ל-$\mathcal{A}$.
4. $\mathcal{A}$ יחזיר (m^*, t^*).
5. $\hat{\mathcal{A}}$ יחזיר $(H_S(m^*), t^*)$.

מבחינת זמן ריצה $\hat{\mathcal{A}}$ פולינומיאלי (מס פולינומיאלי של פולינומיאלי). שוב אנחנו מסמלצים ל-$\mathcal{A}$ באופן מושלם את הניסוי $MacForge_{\Pi,\mathcal{A}}$.

לכן:

$$Pr\left[MacForge_{\hat{\Pi},\mathcal{A}}(n) = 1\right] = Pr\left[MacForge_{\Pi,\mathcal{A}}(n) = 1 \wedge \overline{Collision}\right] \geq \frac{1}{p(n)}$$

בסתירה להנחה $\hat{\Pi}$ מערכת MAC בטוחה. $\square$

4.3 שילוב מערכות הצפנה ומערכות לאימות הודעות

הצפנה ואימות הן שתי מטרות נפרדות לגמרי, סודיות מול אימות שההודעות לא השתנו. באופן כללי אף אחת משתי המטרות לא מבטיחה את האחרת.

נזכור את מערכת ההצפנה הבטוחה כנגד CPA המוגדרת כך בהינתן פונקציה פסאודו אקראית F:

$$Enc_k(m,r) = (r, F_k(r) \oplus m)$$

נשים לב כי יריב שיושב על הערוץ ומקבל את ההצפנה הזאת, יכול בקלות לשנות את ההודעה באופן שעדיין משאיר את ההודעה המקורית התיאורטית כמשהו פוטנציאלית חוקי:

$$Enc_k(m \oplus 1^n, r) = (r, F_k(r) \oplus m \oplus 1^n)$$

כשרוצים להשיג את שתי המטרות בו זמנית, זה נקרא Authenticated Encription. אלו כנראה סוג מערכות ההצפנה הנמצאות ביותר בשימוש כיום. אם כך נשאלת השאלה כיצד בונים אותן. בקורס הזה נוותר על הגדרה פורמלית ורק ננסה להבין אותו מבחינה איטואיטיבית.

- ניסיון 1: (Encrypt-and-Authenticate) לאליס ולבוב יהיו את המפתח $k = (k_E, k_M)$. אליס תחשב $c \leftarrow Enc_{k_E}(m)$ וגם $t \leftarrow Mac_{k_M}(m)$ ותשלח לבוב את (c,t). בוב יחשב $m \leftarrow Dec_{k_E}(c)$ ואת $Vrfy_{k_M}(m,t)$.

זה לא בטוח כי אין שום דבר שמבטיח שלא ניתן ללמוד את ההודעה m מהערך t.

- ניסיון 2: (Encrypt-then-Authenticate) לאליס ולבוב יהיו את המפתח $k = (k_E, k_M)$. אליס תחשב $c \leftarrow Enc_{k_E}(m)$ וגם $t \leftarrow Mac_{k_M}(c)$ ותשלח לבוב את (c,t). בוב יחשב $m \leftarrow Dec_{k_E}(c)$ ואת $Vrfy_{k_M}(c,t)$.

אינטואיטיבית הפעם t לא חושף כלום על ההודעה m אלא רק על ההצפנה c. באופן כללי זו גישה טובה!

עולם הפרימטיבים שלנו בינתיים:

The World of Crypto Primitives
(so far)