סיבוכיות – הרצאה 12

כל העולם הוא חישוב אחד גדול

31.5.11

נמשיך מאיפה שעצרנו. אליס רוצה לשלוח $E(m)$ לבוב בקו פומבי. רוצים שבוב יוכל לקרוא את $D\big(E(m)\big)=m$ ואף אחד אחר לא.
הצעה ראשונה: A ו– B נפגשים קודם, מייצרים מחרוזת אקראית לגמרי r. זה הסוד המשותף שלהם.
כעת A תשלח ל– B את $m'=m\oplus r$. מי שיודע את r יכול לפענח את המסר. מי שלא יודע שום דבר על r, m' נראה אוניפורמי לגמרי. כלומר היריב לא לומד שום דבר על המסר.
הצעה זו טובה במובן הכי חזק שיש לכל יריב, אפילו לא מוגבל בכוחו.
הצעה 2: להחביא את האלגוריתם. אבל הבטיחות של השיטה הזו לא מבוססת על שום דבר.

Public Key Crypthography

כל מי שרוצה שיוכלו לשלוח לו מסרים מוצפנים מכין:

- מפתח פומבי k_{pub} ומפרסם אותו בספר טלפונים ליד שמו.
- מפתח סודי $k_{private}$ שומר בסוד.

מי שרוצה להצפין מסר m, בוחר מחרוזת אקראית r ושולח $m'=E(m,r,k_{pub})$. מי שמקבל את המסר, מפענח אותו ע"י הפעלת $D(m',k_{private})$, וצריך להקיים שלכל r $D\big(E(m,r,k_{pub}),k_{private}\big)=m$.
כלומר, בהינתן הצפנה m ומפתח k_{pub}, יש רק דרך אחת לפתוח את m. כלומר זאת התחייבות על המסר.
בשיטה שנראה, שחקן עם כוח לא מוגבל באמת יכול לגלות את m מתוך m'. אבל, אנחנו מקווים ששחקן עם כוח מוגבל (P או BPP) לא יוכל ללמוד שום דבר מהמסר m'.

El-Gamal

ל– p ראשוני המספרים $F_p^*=\{1,\dots,p-1\}$ זרים ל– p. לכל איבר $a\in F_p^*$ יש הופכי $b\in F_p^*$, כלומר שמקיים $ab\equiv 1 \mod p$.
ל– F_p^* יש יוצר. כלומר יש $g\in F_p^*$ כך ש– $g^0,g^1,g^2,\dots,g^{p-2}$ אלה בדיוק כל איברי F_p^*.
תמיד: אם $a\in F_p^*$ אז $a^{p-1}\cong 1 \mod p$ וכך נוכל למצוא הופכי.
איך נחשב את $a^k \mod p$? אפשר לרוץ $a^1,a^1\cdot a^1,a^1\cdot a^2$ וכו' אבל זה יקח הרבה זמן. אז נעשה fast-exp.
בשביל המפתח הפומבי:

- נבחר ראשוני גדול p (בן 1000 ביטים).
- נמצא יוצר g של F_p^*.
- בוחרים מספר אקראי $S\in\{0,\dots,p-2\}$ שיהיה הסוד שלנו.

- מחשבים $h = g^s \mod p$.

המפתח הפומבי יהיה p, g, h והמפתח הסודי יהיה s.
כדי להצפין בוחרים באקראי $r \in \{0, 1, \ldots, p-2\}$ ושולחים $(g^r, h^r \cdot m)$ (כל החישובים הם מודולו p).
כדי לפענח, מקבלים (α, β) ואז

$$D(\alpha, \beta, s) = (\alpha^s)^{-1} \cdot \beta$$

כי

$$(\alpha^s)^{-1} \cdot \beta = \big((g^r)^s\big)^{-1} \cdot g^{rs} \cdot m = (g^{rs})^{-1} \cdot g^{rs} \cdot m = m$$

רואים שלכל r $D\big(E(m, r, k_{pub}), k_{private}\big) = m$ ואכן ההצפנה היא התחייבות על m.

טענה. g^r קובע r יחיד בין $\{0, 1, \ldots, p-2\}$.

הסבר. יוצר g לכן לכל $r_1, r_2 \in \{0, 1, \ldots, p-2\}$ $g^{r_1} \neq g^{r_2}$.

נסמן $\beta = h^r \cdot m$, $\alpha = g^r$.

$$\alpha = g^r \xrightarrow{\text{קובעת}} r \xrightarrow{\text{קובעת}} h^r \xrightarrow{\text{קובעת}} m$$

כלומר שחקן עם כוח לא מוגבל שרואה את p, g, h, α, β יכול לחשב את m.
צריך להניח שהיריב שרואה $h = g^s$ לא יכול לחשב את s אפילו ש- h קובעת את s. כלומר, יש פה הנחת קושי. נרצה לבדוק מה בדיוק אנחנו מניחים. היינו רוצים שיספיק $P \neq NP$ (ברור שנוכל לפתור ב- NP).
איך בוחרים ראשוני גדול? דוגמים מספר אקראי בן 1000 ביטים. בודקים אם הוא ראשוני. אם כן יופי ואם לא ממשיכים להגריל.
כמה ראשוניים יש בני 1000 ביטים? יש קבוע c קטן כך שמספר הראשוניים בני 1000 ביטים הוא בערך $c \cdot \frac{2^{1000}}{1000}$.
איך נבדוק שהמספר ראשוני? יש אלגוריתם הסתברותי ידוע שעובד לרוב המספרים. לא מזמן מצאו גם אלגוריתם דטרמיניסטי (שהוא דהרנדומיזציה של האלגוריתם ההסתברותי).
עכשיו מה לגבי היוצר? מסתבר שכדאי שהחבורה הכפלית תהיה מסדר ראשוני (לא נסביר מדוע). אז מה שעושים, זה בוחרים q אקראי בן 1000 ביטים, מחשבים את $p = 2q + 1$ ובודקים אם p ראשוני. למה זה מעניין אותנו? נעבוד מודולו p. החבורה הכפלית שלנו תהיה מסדר $p - 1 = 2q$. נמצא יוצר g מודולו p ואז נעבוד עם g^2 שיוצר חבורה מסדר q.
איך נמצא את g? מגרילים ובודקים שהוא יוצר (כלומר שהוא מסדר $2q$. כמובן שהעובדה שקיימים אינסוף p, q כאלה היא השערה מתמטית פתוחה.
נגדיר את הבעיה $DLog$. קלט: p ראשוני, g יוצר של F_p^*. פלט: ה-s היחיד כך ש- $g^s = h \mod p$, $s \in \{0, 1 \ldots, p-2\}$.
$DLog \in NP$, למרות שזאת לא בעיית הכרעה (כלומר או שהיא שייכת למחלקת החיפוש של NP או שנגדיר בעיית הכרעה מתאימה). ל- $DLog$ יש הוכחת 0 מידע (כמו שראינו ל- GI), ולכן אם היא NP קשה אז ההירכיה קורסת. אז $DLog$ כנראה לא NP קשה.
אומרים שבעיה L ניתנת לפתרון במחלקה C אם יש אלגוריתם ב- C שמחשב את L על כל קלט. לפי ההגדרה הזאת L היא קשה אם יש קלט עליו האלגוריתם טועה. אבל זה לא מספיק לנו. זה קושי של worst cast. הקושי שאנחנו צריכים הוא שעל קלט אקראי טיפוסי מההתפלגות הבעיה לא פתירה. במקרה של $DLog$ ההתפלגות שמכינים נראת כך: p, g קבועים, s נבחר באקראי והקלט הוא p, g, g^s. אנחנו צריכים שעל ההתפלגות שאנחנו מכינים הבעיה קשה בממוצע. באופן קצת יותר פורמלי: בהתפלגות שלנו D לכל אלגוריתם A:

משהו זניח $\leq \mathbb{P}_{x\in D}(A(x)$ עונה נכון$)$

עד עכשיו עבדנו עם קושי worst case ועכשיו אנחנו רוצים קושי בממוצע.
נסכם מה אנחנו צריכים:

- $P \neq NP$.
- אי אפשר לפתור את NP ב- P אפילו כשקריטריון ההצלחה נלקח בממוצע על D (כלומר שאי אפשר לפתור את NP ב- P בממוצע).
- ONF כלומר קל לחשב וקשה להכין בממוצע.

וזה הרבה יותר מאשר $P \neq NP$.
נראה עוד שתי אלפיקציות. נתחיל מזה שנתשמש בזה שבעיית אלגמל היא בעיית נחזור לבעיה מהשיעור הקודם. A ו- B יש להם מכונית משותפת. הם נפרדים, לא מוכנים לראות אחד את השני. רוצים להטיל מטבע ולקבוע מי יקח את המכונית. אם הם היו יכולים לדבר הכל היה טוב. אבל הם לא.
המשחק עובד בתורות: A שולחת ל- B. אז B שולח ל- A וכו'. בסוף מחשבים (מעברים) f ולפי התוצאה מחליטים של מי המכונית.
דוגמא לפרוטוקול תקשורת כזה: A מטילה $b \in \{0,1\}$ שולחת ל- B ו- $f(b) = b$. אם A ו- B הוגנים אז $\mathbb{P}(f = 0) = \mathbb{P}(b = 0)$. אם הם לא הוגנים אז יש בעיה.

טענה. לכל פרוטוקול או A או B יכולים לכפות ניצחון.

נחשב עץ משחק. בעלים יש את התוצאות. A מנסה לקבל מינימום, B מנסה לקבל מקסימום, נפתח את העץ ונראה את התוצאה.
מצד שני:

טענה. אם A ו- B שחקנים מוגבלים חישובית אז הם יכולים "להטיל מטבע הוגנת בשיחת טלפון". כלומר קיים פרוטוקול טוב.

ראינו את התשובה לזה. נגיד A תבחר $b \in \{0,1\}$. היא תתחייב עליו בלי לגלות אותו. למשל להצפין אותו. היא תשלח ל-B את $E(b, r, k_{pub}(Alice))$. B יבחר באקראי $c \in \{0,1\}$ ושולח בגלוי ל- A. בשלב שלישי A תשלח את המפתח ל- B. עכשיו שניהם יודעים את b ואת c וחשב את f. A לא יכולה לשלוח מפתח לא נכון כי אפשר לוודא שהמפתח הוא המפתח הנכון.
יותר על העלום הזה אפשר לקרוא בספר של Arona Barak עמודים $369 - 370$.
נראה עכשיו שאם יש commitment schemes אז לכל שפה ב- NP אפשר לתת הוכחת אפס מידע.
מה זה אומר? יש שפה L, קטל x וטענה $x \in L$. אז יש פרוטוקול אינטרקטיבי בין מוכיח כל יכול לבודק הסתברותי V כך ש-

$$x \in L \Rightarrow \mathbb{P}_{\text{מטבעות}}(\text{יקבל } V) = 1$$

ו-

$$x \notin L \Rightarrow \mathbb{P}_{\text{מטבעות}}(\text{יקבל } V) \leq \frac{1}{2}$$

ובנסוף אם $x \in L$ השיחה עם המוכיח לא מגלה שום דבר נוסף (כלומר הבודק היה יכול לסמלץ אותה בעצמו במובן שאין לו דרך להבדיל בין השיחה האמיתית לשיחה שהוא מייצר).

הערה. מספיק להראות את זה עבור בעיה NP קשה אחת.

נבחר לעבוד עם $3Col$. הקלט הוא גרף $G = (V, E)$ והטענה היא שהגרף 3 צביע. במקרה שלנו אפס מידע פירושו שההוכחה לא תגלה את הצביעה.
ניסיון ראשון: במקום להראות לכל קודקוד איך הוא צבוע, המוכיח מוצא 3-צביעה $\pi : V \to \{1, 2, 3\}$ ולכל $v \in V$ שולח $\pi(v)$ לבודק. במקום זה הוא ישלח את $\pi(v)$ בכספת, כלומר שולח התחייבות על $\pi(v)$. כלומר הבודק לא לומד שום דבר על הצביעה בשלב הזה. הבודק בוחר זוג קודקודים שמחוברים בקשת $e = (i, j)$ באקראי ומבקש מהמוכיח לפתוח את הכספות של i ושל j. כעת המוכיח לומד את $\pi(i)$ ואת $\pi(j)$ ועכשיו הוא צריך להחליט אם לקבל או לדחות. אם $\pi(i) = \pi(j)$ הוא דוחה ואחרת מקבל. כעת נקבל במקום

$$x \notin L \Rightarrow \mathbb{P}_{\text{מטבעות}}(\text{יקבל } V) \leq \frac{1}{2}$$

את

$$x \notin L \Rightarrow \mathbb{P}_{\text{מטבעות}}(\text{יקבל } V) \leq 1 - \frac{1}{E}$$

אבל זה בסדר כי נעשה אמפליפיקציה.
האם זה מגלה את הצביעה? נניח $x \in L$, תמיד הבודק מקבל. ז"א לכל קשת $e = (i, j)$ הוא יראה 2 צבעים שונים. אבל זה לא מספיק. נרצה שהצבעים יהיו גם אקראיים. למה? כי אם נחזור על הפרוטוקול נוכל לגלות כבר את כל הצביעה. אם נקבל כל פעם צבעים אקראיים זה לא יעזור אפילו אם נפעיל את הפרוטוקול הרבה פעמים.
הפרוטוקול הסופי נראה כך: המוכיח מוצא צביעה חוקית $\pi : V \to \{1, 2, 3\}$. לכל צביעה, הוא משנה אותה ע"י הפעלה $\sigma \in S_3$ פרמוטציה אקראית. עכשיו באמת נראה שני צבעים אקראיים לכל צלע.
מה ראינו בקורס:

- מה קל ומה קשה לחשב במודלי חישוב שונים.
- מה קל ומה קשה להוכיח במודלי חישוב שונים.

ואז ראינו ש- $PSpace$ היא מצד אחד שייכת לשאלה הראשונה, כלומר מה אפשר לחשב עם זכרון מוגבל ומצד שני, באופן מפתיע, $IP = PSpace$ ואז $PSpace$ היא למעשה המחלקה של כל מה שאפשר להוכיח עם פרוטוקול אינטרקטיבי לבודק הסתברותי יעיל.
ראינו שזה קורה גם ב- $NExp$ ששווה ל- MIP. אם ניקח את הטענה הזאת ונוריד אותה ל- NP (וזה לא טריוויאלי) ואז מקבלים את משפט ה- PCP. הסתבר, שהעובדה שיש מערכות הוכחה מפתיעות וחזקות כלאה גוררת שאם $P \neq NP$ אז גם יש בעיות מקסימיזציה שהקירוב שלהן הוא NP קשה. למעשה זאת עובדה שקולה.
אספקט אחר שעלה הוא השימוש במטבעות אקראיים לפתרון בעיות ובמערכות הוכחה. בכל מערכות ההוכחה שראינו הבודק חייב להיות הסתברותי. אחרת, כל מערכות ההוכחה קורסות ל- NP כי הבודק צפוי. אז הסתלנו על BPP. ראינו את $IT \in BPP$ Identity testing שלא ידועה להיות ב- P. זה אומר שכן יש בה כוח. מצד שני ראינו ש- $BPP \subseteq PSize$ מעגלים בגודל פולינומי וזה אינדיקציה של חולשה. גם ראינו שאם $NP \subseteq PSize$ אז ההירכיה קורסת וזה אומר ש-NP היא קשה ואינדיקציה ש- BPP כנראה לא יפתור את NP.
מצד שני, אם משתחררים מבעיות הכרעה ומסתכלים על דברים כמו דגימה או כמו שימוש במטבעות אקראיים כדי להסתיר דברים, למשל ההוכחות האינטרקטיביות שלנו, שם אין תחליף למטבעות אקראים. ראינו גם שתי דוגמאות יפות לשימוש באקראיות - בIT ובהוכחה ש- $coNP \subseteq IP$.
היום ראינו שאם הקושי הוא חזק מספיק (הרבה יותר מסתם $P \neq NP$) אז הוא שימושי ל- public key cryptography, bit commitence אבל זה מכניס אותנו לעולם חדש של "קשה בממוצע".