תורת הסיבוכיות — תרגול 7
משחקי Arthur–Merlin

ארתור–מרלין: הגדרה

מערכת ההוכחה האינטראקטיבית ארתור–מרלין כוללת שני שחקנים - ארתור (A) שמשמש כמוודא, ומרלין (M) שמשמש כמוכיח. ארתור הוא מכונה פולינומית הסתברותית בעוד שמרלין הוא מכונה כל־יכולה, בדומה למצב במערכות הוכחה אינטראקטיביות רגילות. ההבדל הוא במגבלה הנוספת שמושתתת על ארתור: ההודעות שהוא יכול לשלוח למרלין כוללות ביטים אקראיים בלבד, והחישוב הסופי שהוא מבצע כדי לקבוע אם לקבל או לדחות הוא דטרמיניסטי בהינתן הפרוטוקול. בניסוח לא פורמלי, כל הטלות המטבע של ארתור הן פומביות; ארתור לא יכול להשתמש בהטלות מטבע שמרלין לא ראה בשום שלב בפרוטוקול (אם כי ארתור לא מחוייב לשלוח את כל הביטים האקראיים שלו בסיבוב הראשון בפרוטוקול, כך שתשובות הביניים של מרלין עדיין מוחזרות תוך כדי חוסר ידע כלשהו).

כרגיל, פרוטוקול ארתור–מרלין עבור שפה L חייב לקיים שלמות ונאותות: ההסתברות לטעות של המוודא, לכל קלט, לא יכולה לעלות על $\frac{1}{3}$.

עבור מספר k נגדיר $\mathsf{AM}[k]$ להיות מחלקת השפות L שקיים עבורן פרוטוקול ארתור–מרלין בן k סיבובים. כמו כן נסמן $\mathsf{AM} \triangleq \mathsf{AM}[2]$.

לפעמים מדברים גם על פרוטוקול מרלין–ארתור שבו מרלין מתחיל, עם מחלקות אנלוגיות MA ו־ $\mathsf{MA}[k]$. לא נעסוק בפרוטוקולים הללו כאן.

תוצאה מפתיעה של Goldwasser ו־ Sipser מ־ 1986 היא שהמעבר להטלות מטבע פומביות לא מחלישה באופן מהותי את מערכת ההוכחה; ספציפית, $\mathsf{IP}[k] \subseteq \mathsf{AM}[k+2]$ לכל פונקציה $k(n)$ החשיבה בזמן פולינומי ב־ n. מקרה פרטי מייצג של המשפט הוא התוצאה $\mathrm{GNI} \in \mathsf{AM}[2]$, שמראה כי קיים פרוטוקול הוכחה אינטראקטיבית לבעיית הגרפים הלא איזומורפיים שבו כל הטלות המטבע פומביות למרות שהפרוטוקול הקלאסי לבעיה זו דורש הטלות מטבע שאינן גלויות למוכיח.

הפרוטוקול ל־ GNI (נלמד בהרצאה)

הרעיון בפרוטוקול הקלאסי ל־ GNI הוא פשוט: בהינתן (G_0, G_1), המוודא בוחר ביט אקראי b, מגריל גרף H כך ש־ $H \cong G_b$ (על ידי הגרלת פרמוטציה אקראית של צמתי G_b) ושולח למוכיח את H. המוכיח שולח בחזרה ביט β והמוודא מקבל אם ורק אם $b = \beta$. אם $G_0 \ncong G_1$ אז H מזהה באופן יחיד את b ולכן המוכיח תמיד יצליח לשכנע את המוודא; אם לעומת זאת $G_0 \cong G_1$ אז ההתפלגות של H שהמוודא שולח למוכיח אינה תלויה ב־ b, ולכן ההסתברות של המוכיח לשלוח את b בחזרה היא $\frac{1}{2}$. כמובן שהפרוטוקול קורס אם הטלות המטבע של המוודא היו פומביות כי אז המוכיח יודע את b ויכול פשוט לשלוח אותו בחזרה תמיד.

הפרוטוקול הפומבי ל־ GNI שונה למדי באופיו (אם כי כפי שנראה, במובן מסויים הוא פשוט הכללה של הפרוטוקול הקלאסי). האבחנה הבסיסית היא שאם $G_0 \ncong G_1$ אז קיימים "הרבה" גרפים שאיזומורפיים או ל־ G_0 או ל־ G_1, בעוד שאם $G_1 \cong G_2$ קיימים הרבה פחות גרפים כאלו.

פורמלית, כל גרף H שאיזומורפי ל־ G מתקבל על ידי הגרלת פרמוטציה $\pi \in S_n$ וחישוב $\pi(G)$. לעתים עלול להתקיים $\pi(G) = G$ כך שלא נקבל גרף שונה. במקרה כזה, אומרים ש־ π הוא **אוטומורפיזם** של G_b (תמיד יש לפחות אחד כזה — תמורת הזהות). לא קשה לבדוק שקבוצת כל האוטומורפיזמים של גרף נתון היא תת־חבורה נורמלית בחבורה S_n והקוסטים שלה מתאימים בדיוק לפרמוטציות שונות ששולחות את G לאותו גרף H. לכן מספר הגרפים שאיזומורפיים ל־ G הוא בדיוק $\frac{n!}{t}$ כש־ t הוא מספר האוטומורפיזמים של G.

כדי להותיר את המספרים שלנו "נקיים", נתחשב בתוצאה זו: בהינתן (G_0, G_1) נגדיר את הקבוצה הבאה:

$$S := \left\{(H, \pi) \;\middle|\; (H \cong G_0 \vee H \cong G_1) \wedge \pi \in \mathrm{Aut}(G)\right\}$$

בבירור אם $G_0 \ncong G_1$ אז $|S| = 2n!$ ואם $G_0 \cong G_1$ אז $|S| = n!$.

הבעיה שלנו הוחלפה בבעיה אחרת — כיצד מרלין יכול לשכנע את ארתור ש־ S היא קבוצה "גדולה"? לשם כך משתמשים ברעיון של Goldwasser ו־ Sipser — פרוטוקול Set Lower Bound.

פרוטוקול Set Lower Bound (נלמד בהרצאה)

כלי מרכזי שבו משתמשים בפרוטוקול הוא הכלי הבא.

<u>הגדרה:</u> **פונקציות תמצות בלתי תלויות בזוגות**, באנגלית **pairwise independent hash function**, עם פרמטרים n,k טבעיים, הן אוסף $\mathcal{H}_{n,k}$ של פונקציות $h:\{0,1\}^n\to\{0,1\}^k$ כך שלכל $x\neq x'\in\{0,1\}^n$ ו־ $y,y'\in\{0,1\}^k$ מתקיים $\underset{h\in_R\mathcal{H}_{n,k}}{\mathbf{Pr}}\left[h(x)=y\wedge h(x')=y'\right]=2^{-2k}$.

נשים לב שבפרט מתקיים $\underset{h\in_R\mathcal{H}_{n,k}}{\mathbf{Pr}}\left[h(x)=y\right]=2^{-k}$ לכל $x\in\{0,1\}^n$ ו־ $y\in\{0,1\}^k$.

<u>טענה:</u> לכל n,k טבעיים קיים אוסף של פונקציות תמצות בלתי תלויות בזוגות. מעבר לכך, ניתן ביעילות לדגום את הפונקציות.

הוכחה באמצעות בנייה. נניח תחילה $k=n$, ונגדיר $\mathcal{H}_{n,n}:=\left\{h_{a,b}(x)=a\cdot x+b\,\middle|\,a,b\in\mathrm{GF}(2^n)\right\}$, כאשר $\mathrm{GF}(2^n)$ זה שדה גלואה עם 2^n איברים. כעת, אם $n>k$ אז נרחיב קלטים מאורך k ל־ n על ידי ריפוד באפסים, ואם $k<n$ אז נקצר פלטים מאורך n ל־ k על ידי קטימת $n-k$ הביטים האחרונים. קל לראות שדרישת היעלות שבהגדרה מתקיימת; את דרישת הנכונות נשאיר כתרגיל.

כעת נוכל להציג את הפרוטוקול. ננסח את הבעיה כך: נתונה קבוצה $S\subseteq\{0,1\}^n$ כך שלכל $x\in S$ קיימת הוכחה בגודל פולינומי ב־ n לכך ש־ $x\in S$ הניתנת לווידוא בזמן פולינומי. נתון מספר טבעי K, ומטרת הפרוטוקול היא שהמוודא יקבל בהסתברות טובה אם $|S|\geq K$ וידחה בהסתברות טובה אם $|S|\leq\frac{K}{2}$. הסיטואציה הכללית הזו מתאימה לסיטואציה של GNI עם $K=2n!$.

נבחר מספר טבעי k שמקיים $2^{k-2}<K\leq 2^{k-1}$. האינטואיציה היא שלפונקציה שנבחרה באקראי מתוך אוסף של פונקציות תמצות בלתי תלויות בזוגות יש סיכוי סביר "להשאיר בחיים" איבר של S אם $|S|\geq K$ אבל כנראה תחסל את כל אברי S אם $|S|\leq\frac{K}{2}$.

הפרוטוקול הדו שלבי הוא כדלהלן:

1. המוודא מגריל $h\in\mathcal{H}_{n,k}$ ו־ $y\in\{0,1\}^k$ ושולח למוכיח.
2. המוכיח מחפש $x\in S$ כך ש־ $h(x)=y$. אם מצא כזה, הוא שולח את x למוודא בתוספת הוכחה לכך ש־ $x\in S$.
3. המוודא בודק ש־ $x\in S$ (בזמן פולינומי) וש־ $h(x)=y$; אם כן, הוא מקבל, ואחרת דוחה.

נעבור להוכחת הנכונות. בבירור אם קיים $x\in S$ מתאים המוכיח ימצא אותו וישלח למוודא, וכמו כן המוכיח אינו יכול "לעבוד" על המוודא ולגרום לו לקבל מבלי שישלח x מתאים. לכן השאלה היחידה היא מה ההסתברות לכך שב־ S יהיה x המקיים $h(x)=y$.

מתכונות $\mathcal{H}_{n,k}$ מתקיים $\mathbf{Pr}\left[h(x)=y\right]=2^{-k}$ ו־ $\mathbf{Pr}\left[h(x)=y\wedge h(x')=y\right]=2^{-2k}$. לכן מעקרון ההכלה וההפרדה נקבל:

$$\begin{aligned}\mathbf{Pr}\left[\exists x\in S\left(h(x)=y\right)\right] &\geq \sum_{x\in S}\mathbf{Pr}\left[h(x)=y\right]-\sum_{x\neq x'\in S}\mathbf{Pr}\left[h(x)=y\wedge h(x')=y\right]\\ &= |S|\,2^{-k}-\binom{|S|}{2}2^{-2k}\geq|S|\,2^{-k}-\frac{1}{2}\left(|S|\,2^{-k}\right)^2\end{aligned}$$

נסמן $p=|S|\,2^{-k}$, וקיבלנו שההסתברות חסומה מלמטה על ידי $p-\frac{1}{2}p^2=p\left(1-\frac{p}{2}\right)$. עבור $p\leq\frac{1}{2}$ מתקיים $p\left(1-\frac{p}{2}\right)\geq\frac{3}{4}p$ ולכן אם $p\leq\frac{1}{2}$ קיבלנו חסם תחתון של $\frac{3}{4}p$ על ההסתברות.

כמו כן, חסם עליון על ההסתברות הוא טריוויאלי בעזרת Union bound:

$$\mathbf{Pr}\left[\exists x\in S\left(h(x)=y\right)\right]\leq\sum_{x\in S}\mathbf{Pr}\left[h(x)=y\right]=|S|\,2^{-k}=p$$

נסמן כעת $p^*=K2^{-k}$, ונבטא את הסתברות הקבלה של המוודא בשני המקרים האפשריים כפונקציה של p^*. ראשית, אם $|S|=K$ אז $p=|S|\,2^{-k}=K2^{-k}\leq 2^{k-1}\cdot 2^{-k}=\frac{1}{2}$, ולכן הסתברות הקבלה של המוודא חסומה מלמטה על ידי $\frac{3}{4}p=\frac{3}{4}p^*$. אם $|S|>K$ הדבר רק יגדיל את הסתברות הקבלה של המוודא, ולכן גם במקרה זה ההסתברות לקבלה חסומה מלמטה על ידי $\frac{3}{4}p^*$.

לעומת זאת, אם $|S|\leq\frac{K}{2}$ אז $p=|S|\,2^{-k}\leq\frac{K2^{-k}}{2}=\frac{p^*}{2}$, ולכן הסתברות הקבלה של המוודא חסומה מלמעלה על ידי $\frac{p^*}{2}$.

הראינו פער של $\frac{p^*}{4}$ בין הסתברות הקבלה במקרה של קלט בשפה והסתברות הקבלה של קלט שאינו בשפה; זה מספיק כדי שניתן יהיה לבצע הגברה של הפער במספר פולינומי של סיבובים, תוך שימוש בחסם צ'רנוף. הבעיה היא שאנו מוגבלים לשני סיבובים בלבד; הפתרון הוא בכך שכל הסיבובים יתבצעו <u>במקביל</u>. כלומר, המוודא שולח למוכיח שני וקטורים $(h_1,h_2,\ldots,h_t)$ ו־ $(y_1,y_2,\ldots,y_t)$ כאשר t מחושב באמצעות צ'רנוף, והמוכיח צריך להחזיר לו $(x_1,x_2,\ldots,x_t)$ כך ש־ $h_i(x_i)=y_i$ (יחד עם הוכחות שייכות ל־ S כמובן). המוודא מקבל רק אם המוכיח עמד ברוב האתגרים שלו.

הוכחת משפט Goldwasser–Sipser

עד כה ראינו דוגמה בלבד עבור המקרה של GNI, אך הרעיון הכללי של ההוכחה דומה ובפרט פרוטוקול Set Lower Bound הוא אבן היסוד בו ללא שינויים נוספים.

נתחיל מהוכחת המקרה הפרטי $\mathsf{IP}[2] \subseteq \mathsf{AM}[4]$. תהא $L \in \mathsf{IP}[2]$, ויהי פרוטוקול אינטראקטיבי, בעל שני סיבובים, עם מוודא V שרץ בזמן פולינומי, וכך שמתקיים:

$$w \in L \implies \exists P : \mathbf{Pr}[V \leftrightarrow P \text{ accepts } w] \geq \frac{2}{3}$$
$$w \notin L \implies \forall P : \mathbf{Pr}[V \leftrightarrow P \text{ accepts } w] \leq \frac{1}{3}$$

בפרוטוקול כזה על קלט w: המוודא מגריל לעצמו מחרוזת אקראית r; מחשב מתוכה אתגר $q = q(w,r)$; שולח למוכיח את q; מקבל מהמוכיח חזרה תשובה $a = a(w,q)$; ואז מבצע חישוב $V(w,r,a)$ ומקבל או דוחה לפיו.

נגדיר $S_{q,a} := \{r \mid q = q(w,r) \wedge V(w,r,a) = \text{acc}\}$. כלומר, $S_{q,a}$ היא קבוצת כל ה־ r־ים שיגרמו למוודא לשלוח את האתגר q ולקבל אם תשובת המוכיח תהיה a.

נסמן $S_q \triangleq S_{q,a(q)}$ כאשר $a(q) := \arg\max_a |S_{q,a}|$. כלומר, S_q היא הקבוצה הגדולה ביותר של מחרוזות r שיגרמו למוודא לשלוח למוכיח אתגר q ולקבל את תשובת המוכיח, בהנחה שהמוכיח עונה בצורה אופטימלית (הנחה שתמיד אפשר להניח).

ניסיון ראשון

אם $w \in L$ אז $\sum_q |S_q| \geq \frac{2}{3} \cdot 2^{p(n)}$, ואם $w \notin L$ אז $\sum_q |S_q| \leq \frac{1}{3} \cdot 2^{p(n)}$. היינו רוצים לבצע מעין Set Lower Bound שיאפשר לנו להעריך את $\sum_q |S_q|$, אך $\sum_q |S_q|$ זה גודל הקבוצה $\bigcup_q S_{q,a(q)}$ כאשר $a(q)$ מייצג את התשובה אופטימלית עבור האתגר q. לרוע המזל, לא ברור איך אפשר להוכיח שייכות ל־ $\bigcup_q S_{q,a(q)}$, מאחר ולא ברור איך אפשר להוכיח אופטימליות של a ספציפי (ובלעדי זה קיימת סכנת הרמאות הבאה: לכל r מרלין יספק לארתור q ו־ a כזה שגורם למוודא לקבל, גם אם זה לא ה־ a שמוכיח שאינו יודע מהי r היה פולט). אם כן, יש להתחכם שוב.

ניסיון שני (מוצלח)

נוכל, תוך הסתמכות על הגברת הסתברויות (במקביל), להניח ש־ V רץ בזמן פולינומי $p(n)$ וכמו כן מתקיים:

$$w \in L \implies \exists P : \mathbf{Pr}[V \leftrightarrow P \text{ accepts } w] \geq \frac{2}{3}$$
$$w \notin L \implies \forall P : \mathbf{Pr}[V \leftrightarrow P \text{ accepts } w] \leq \frac{1}{6p(n)}$$

כעת, לכל N נגדיר את הקבוצה $T_N = \left\{q \,\middle|\, |S_q| \geq N\right\}$, ונחפש ערך "טיפוסי" של N שמעיד על כך ש־ T_N גדולה.

לצורך זה, נגדיר קבוצות $B_1, \ldots, B_{p(n)}$ באופן הבא: $B_i = \left\{q \,\middle|\, 2^{i-1} \leq |S_q| \leq 2^i\right\}$. שימו לב שלכל q מתקיים $0 \leq |S_q| \leq 2^{p(n)}$ (כי יש $2^{p(n)}$ מחרוזות r אפשריות באופן כללי), כך שכל q נופל אל תוך אחת מהקבוצות B_i. נסמן ב־ t את האינדקס של הקבוצה B_t כך ש־ $\sum_{q \in B_t} |S_q|$ מקסימלי, ונגדיר $N^* \triangleq 2^{t-1}$ (שימו לב ש־ $N^* > 0$).

- נניח ש־ $w \in L$. אזי, משלמות הפרוטוקול מתקיים $\sum_q |S_q| \geq \frac{2}{3} \cdot 2^{p(n)}$, ומעיקרון שובך היונים נובע ש־ $\sum_{q \in B_t} |S_q| \geq \frac{2}{3p(n)} \cdot 2^{p(n)}$ (מחלקים במספר התאים $p(n)$). כעת, מכיוון שלכל $q \in B_t$ מתקיים על פי ההגדרה $2^{t-1} \leq |S_q| \leq 2^t$, מקבלים:

$$\frac{2}{3p(n)} \cdot 2^{p(n)} \leq \sum_{q \in B_t} |S_q| \leq |B_t| \cdot 2^t \implies |B_t| \geq \frac{1}{2^t} \cdot \frac{2}{3p(n)} \cdot 2^{p(n)} = \frac{1}{2^{t-1}} \cdot \frac{1}{3p(n)} \cdot 2^{p(n)} = \frac{1}{3N^* p(n)} \cdot 2^{p(n)}$$

ולכן

$$|T_{N^*}| = \left|\left\{q \,\middle|\, |S_q| \geq N^*\right\}\right| = \left|\left\{q \,\middle|\, |S_q| \geq 2^{t-1}\right\}\right| \geq \left|\left\{q \,\middle|\, 2^{t-1} \leq |S_q| \leq 2^t\right\}\right| = |B_t| \geq \frac{1}{3N^* p(n)} \cdot 2^{p(n)}$$

- נניח ש־ $w \notin L$. אזי, מנאותות הפרוטוקול מתקיים $\sum_q |S_q| \le \frac{1}{6p(n)} \cdot 2^{p(n)}$, ולכן עבור כל N מקבלים:

$$|T_N| \cdot N = \left|\{q \,\big|\, |S_q| \ge N\}\right| \cdot N \le \sum_q |S_q| \le \frac{1}{6p(n)} \cdot 2^{p(n)} \quad \Longrightarrow \quad |T_N| \le \frac{1}{6Np(n)} \cdot 2^{p(n)}$$

בסך הכל קיבלנו שעבור N^* ו־ $K \triangleq \frac{1}{3N^* p(n)} \cdot 2^{p(n)}$ מתקיים:

$$\begin{aligned} w \in L \quad &\Longrightarrow \quad |T_{N^*}| \ge K \\ w \notin L \quad &\Longrightarrow \quad |T_{N^*}| \le \frac{1}{2}K \end{aligned}$$

לכן, מתקיימת ההפרדה הדרושה לפרוטוקול Set Lower Bound. הבעיה היא כיצד נוכל בהינתן q לוודא בזמן פולינומי ש־ $q \in T_{N^*}$, או באופן שקול ש־ $|S_q| \ge N^*$? הפתרון הוא הרצה של פרוטוקול Set Lower Bound נוסף עבור בדיקת שייכות ל־ S_q והפעם עם $K' = N^*$.

נשים לב שבהינתן q ו־ a, ניתן לוודא שייכות ל־ $S_{q,a}$ באופן יעיל: פשוט בודקים אם מתקיים $q = q(w,r)$ ואם החישוב $V(w,r,a)$ מסתיים במצב מקבל. נזכור ש־ $S_q = S_{q,a(q)}$, ולכן על מנת לוודא שייכות ל־ S_q יש לדעת מהו $a(q)$. לשם כך ארתור יצפה שמרלין ישלח לו את $a(q)$. מאחר ו־ $S_{q,a(q)}$ זו הקבוצה הגדולה ביותר מבין כל ה־ $S_{q,a}$־ים האפשריים, ומאחר ומרלין מנסה לשכנע את ארתור ש־ S_q היא קבוצה גדולה, מרלין האופטימלי תמיד ישלח לארתור את $a(q)$ ולא ינסה לרמות.

אם כן, מה שנותר להראות הוא שמתקיימת ההפרדה הדרושה לפרוטוקול ה־ Set Lower Bound הנוסף.

- ברור שאם $q \in T_{N^*}$ אז $|S_q| \ge N^*$ (לפי הגדרה).
- לעומת זאת, אם $q \notin T_{N^*}$ אז לא בהכרח מתקיים $|S_q| \le N^*/2$. למזלנו, מספר ה־ q־ים עבורם זה לא מתקיים זניח, ולכן הנאותות לא תפגע. כמה q־ים קיימים כך ש־ $|S_q| \ge N^*/2$? ראינו קודם שלכל N מתקיים:

$$|T_N| \le \frac{1}{6Np(n)} \cdot 2^{p(n)} \quad \Longrightarrow \quad \left|\{q \,\big|\, |S_q| \ge N/2\}\right| = \left|T_{N/2}\right| \le \frac{1}{3Np(n)} \cdot 2^{p(n)}$$

כלומר, לכל N שבריר ה־ q־ים כך ש־ $|S_q| \ge N/2$ הוא $\frac{1}{3Np(n)}$, וזה הרי זניח.

כעת נוכל לתאר את הפרוטוקול:

1. מרלין שולח לארתור את N^*.
2. ארתור מחשב מתוך N^* את $K = \frac{1}{3N^* p(n)} \cdot 2^{p(n)}$ ומתוך K את k, מגריל פונקציית תמצות $h \in \mathcal{H}_{n.k}$ ו־ $y \in \{0,1\}^k$, ושולח למרלין.
3. מרלין מוצא q כך ש־ $h(q) = y$ ושולח אותו למוכיח בתוספת $a = a(q)$. (מרלין יחשב את a באמצעות המוכיח האופטימלי.)
4. ארתור שולח למרלין אתגר חדש עבור $K' = N^*$: האתגר הוא h', y'.
5. מרלין שולח לארתור r כך ש־ $h'(r) = y'$ וכמו כן $r \in S_{q,a}$. (את זה ארתור יכול לבדוק בעצמו.)
6. ארתור בודק את תשובות מרלין, משתכנע ש־ $q \in T_{N^*}$, ומקבל. (הבדיקה כוללת: $h(q) = y$, $h'(r) = y'$, ו־ $r \in S_{q,a}$.)

נכון לנקודה זו קיבלנו פרוטוקול MA [5]; נרצה להפוך אותו לפרוטוקול AM [4]. הבעיה נעוצה בכך שמרלין שולח ראשון את N^*. עם זאת, כפי שראינו N^* הוא פשוט חזקה של 2 בתחום $0, 1, \ldots, p(n) - 1$, ולכן ניתן לחסוך את שלב 1 על ידי כך שארתור ישלח למרלין אתגר עבור <u>כל</u> אחד מ־ $p(n)$ הערכים האפשריים של N, ומרלין יענה רק לאתגר המתאים ל־ N^*. בכך סיימנו להראות ש־ $\text{IP}[2] \subseteq \text{AM}[4]$.

המקרה הכללי

נעבור למקרה של IP [4]. במקרה זה הפרוטוקול מתנהל כך:

1. בשלב הראשון V מחשב $q_1 = q_1(w,r)$ ושולח למוכיח.
2. המוכיח מחזיר $a_1 = a_1(w, q_1)$.
3. בשלב השני V מחשב $q_2 = q_2(w, r, a_1)$ ושולח למוכיח.

4. המוכיח מחזיר $a_2 = a_2\left(w, q_1, a_1, q_2\right)$.
5. המוודא מחשב $V\left(w, r, a_1, a_2\right)$ ומקבל או דוחה בהתאם.

נשתמש בסימון מקוצר כדי למנוע סרבול בכתיבה: אם אנו "מריצים" את V על קלט חלקי, למשל $V\left(w, r, a_1\right)$, הכוונה היא להתנהגות V בהנחה שהפרוטוקול נמשך מול מוכיח אופטימלי (שימו לב שבהינתן מוכיח אופטימלי ו־ r, המשך הפרוטוקול נקבע באופן דטרמיניסטי).

נגדיר:

- $S_{(q_1,a_1)} := \left\{r \mid q_1 = q_1\left(w, r\right) \wedge V\left(w, r, a_1\right) = \text{acc}\right\}$. במילים אחרות, זוהי קבוצת כל המחרוזות האקראיות שמייצרות פרוטוקול מקבל מול מוכיח אופטימלי שתחילתו ב־ $\left(q_1, a_1\right)$.
- $S_{(q_1,a_1),(q_2,a_2)} := \left\{r \mid q_1 = q_1\left(w, r\right) \wedge q_2 = q_2\left(w, r, a_1\right) \wedge V\left(w, r, a_1, a_2\right) = \text{acc}\right\}$. כלומר, זוהי קבוצת כל מחרוזות האקראיות שמייצרות פרוטוקול מקבל שבו האתגרים הם q_1, q_2 והתשובות הן a_1, a_2.

שימו לב לכך ש־ $S_{(q_1,a_1),(q_2,a_2)} \subseteq S_{(q_1,a_1)}$, בדומה לאופן שבו $S_{q,a}$ הייתה תת קבוצה של אוסף כל המחרוזות האקראיות.

כמקודם, מטרתו של מרלין היא לשכנע את ארתור בכך שיש הרבה q_1־ים כך שהקבוצה $S_{(q_1,a_1)}$ גדולה מערך "טיפוסי" כלשהו עבור a_1 שנבחר בהתאם ל־ q_1.

כדי להוכיח ש־ $S_{(q_1,a_1)}$ גדולה, מרלין ישלח לארתור ערך "טיפוסי" חדש, וינסה להוכיח לארתור שיש הרבה q_2־ים כך שהקבוצה $S_{(q_1,a_1),(q_2,a_2)}$ גדולה מהערך ה"טיפוסי"; ולהוכיח ש־ $S_{(q_1,a_1),(q_2,a_2)}$ גדולה מרלין יוכל בצורה ישירה על ידי הצגת r מתאים.

את הרעיון הזה אפשר להכליל למקרה של $\text{IP}\left[k\right]$ עבור k חשיבה בזמן פולינומי, אך לא נכתוב במפורש את הפרוטוקול למקרה זה.