

Ομάδες, Στατιστική Απόσταση & Pedersen

1. Χρησιμοποιούμε δεσμεύσεις Pedersen για να υλοποιήσουμε τη ρίψη νομισμάτων. Στη συνηθισμένη εφαρμογή των δεσμεύσεων, ο πρώτος παίκτης αντί να φανερώσει το κέρμα του φανερώνει μια δέσμευση, την οποία ανοίγει αφού φανερώσει την τιμή του ο δεύτερος παίκτης.

Κάποιος προτείνει να αλλάξουμε το πρωτόκολλο ως εξής για μεγαλύτερη ασφάλεια: ο δεύτερος παίκτης, αντί να αποκαλύψει άμεσα την τιμή b , θα δώσει μία δέσμευση σε αυτήν. Κατόπιν, ο πρώτος παίκτης ανοίγει τη δεσμευσή του, μετά ο δεύτερος τη δική του και το πρωτόκολλο τερματίζει κατά τα γνωστά. Οι παράμετροι για το σχήμα του Pedersen υποθέτουμε εδώ ότι προήλθαν από έμπιστο τρίτο πρόσωπο.

Γνωρίζοντας από το μάθημα ότι αυτή η παραλλαγή είναι ευπαθής σε replay attacks, **εάν ο δεύτερος παίκτης απλώς επαναλάβει τη δέσμευση του πρώτου, θεωρούμε ότι χάνει χωρίς να υπολογίσουμε το XOR.**

Εξετάστε την περίπτωση της παραπάνω αλλαγής στο πρωτόκολλο (πέρα από την ανάγκη για την ύπαρξη του τρίτου προσώπου).

Λύση Η αλλαγή καταστρέφει την ασφάλεια του πρωτοκόλλου. Ο δεύτερος παίκτης μπορεί να εξαναγκάσει το πρωτόκολλο να βγάλει αποτέλεσμα 0. Ο πιο απλός τρόπος είναι, όπως έχουμε δει, απλά να επαναλάβει τη δέσμευση που έλαβε από τον πρώτο παίκτη, και να επαναλάβει το άνοιγμα που θα κάνει ο πρώτος παίκτης στο επόμενο βήμα. Αφού οι τιμές θα είναι ίδιες, το αποτέλεσμα της αποκλειστικής διάζευξης (XOR) θα είναι 0.

Η παραπάνω «επίθεση» γίνεται όμως εύκολα αντιληπτή, άρα και μπορεί να «απαγορευτεί» όπως περιγράψουμε.

Για το λόγο αυτό, ο δεύτερος παίκτης μπορεί να πολλαπλασιάσει την αρχική δέσμευση με g^z για τυχαίο z , και όταν ανοίγει την (παραλλαγμένη πλέον) δέσμευση να προσθέσει z στο r που έδωσε ο πρώτος παίκτης ως άνοιγμα.

2. Η Καρολίνα και ο Λυκούργος χρησιμοποιούν το σχήμα δέσμευσης του Pedersen. Η Καρολίνα έχει ορίσει τις παραμέτρους και ο Λυκούργος της έχει στείλει τη δέσμευση c . Πριν γίνει οτιδήποτε άλλο, ο Λυκούργος διαπιστώνει ότι από τυπογραφικό λάθος στις σημειώσεις του, υπολόγισε την τιμή $g^m h^r$ αντί $g^r h^m$. Είναι δυνατό με τις ανάλογες διορθώσεις από τη μεριά της Καρολίνας να συνεχίσουν την επικοινωνία τους ή πρέπει ο Λυκούργος να δημιουργήσει νέα δέσμευση;

Λύση Ναι. Ουσιαστικά, πρόκειται για μια παραλλαγή του συστήματος. Μπορούμε να ελέγξουμε ότι οι αποδείξεις και για τις δύο ιδιότητες ασφάλειας (δέσμευση, απόκρυψη) μπορούν να επαναδιατυπωθούν με τυπικές μόνο αλλαγές.

3. Ο Λυκούργος αποφασίζει για καλό και για κακό να δημιουργήσει νέα δέσμευση c' έχοντας διορθώσει το λάθος στον υπολογισμό, αλλά για οικονομία χρόνου επαναχρησιμοποιεί την ίδια τιμή r . Δεδομένου ότι δεσμεύεται στο ίδιο μήνυμα με πριν (άρα και η τιμή m είναι ίδια), είναι σωστή η αποφασή του;

Λύση Όχι. Θα έχει αποστείλει δύο δεσμεύσεις $c = g^m h^r$ και $c' = g^r h^m$ οι οποίες έχουν συγκεκριμένη σχέση μεταξύ τους. Αυτό κατά κανόνα πρέπει να αποφεύγεται. Συγκεκριμένα, στις δεσμεύσεις Pedersen, ο ορισμός του commit προβλέπει ότι το r θα είναι μια «φρέσκια» τιμή από το \mathbb{Z}_q και όχι κάποια τιμή που δεν έχει χρησιμοποιηθεί. Αυτό έχει σαν αποτέλεσμα η απόδειξη της ασφάλειας να μη μας καλύπτει.

Πέρα από τη θεωρητική διερεύνηση, θα δώσουμε και ένα συγκεκριμένο υπολογισμό που μπορεί να εκτελέσει η Αλίκη: $q = c/(c'^t)$ όπου $h = g^t$. Το t της είναι γνωστό αφού αυτή έφτιαξε τις παραμέτρους. Άρα έχουμε:

$$\begin{aligned} q &= g^m h^r / (h^m g^r)^t \\ q &= g^m h^r h^{-tm} g^{-tr} \\ q &= g^m g^{tr} g^{-t^2 m} g^{-tr} \\ q &= g^{m+tr-t^2 m-tr} \\ q &= g^{m-t^2 m} \\ q &= g^{m(1-t^2)} \end{aligned}$$

Με βάση το q , και αφού το $1 - t^2$ είναι γνωστό, η Αλίκη είναι σε θέση να ελέγξει αν το c περιέχει κάποια τιμή m^* ή όχι, παραβαιάζοντας έτσι την ιδιότητα της απόκρυψης.

4. Έστω f, g δύο αμελητέες συναρτήσεις (βλ. ορισμό 2.6.3). Δείξτε ότι οι συναρτήσεις $h_1 := f + g$ και $h_2 := f \cdot g$ είναι επίσης αμελητέες. Επίσης, εάν το $q(x)$ είναι πολυώνυμο, δείξτε ότι η συνάρτηση $q \cdot f$ είναι και αυτή αμελητέα.

Λύση

Από τον ορισμό, μια συνάρτηση f είναι αμελητέα εάν για κάθε c υπάρχει n_0 ώστε για κάθε $n > n_0$:

$$f(n) \leq \frac{1}{n^c}$$

- Για την $f \cdot g$. Έστω κάποιο c_* , θα βρούμε κατάλληλο n_* ώστε για κάθε $n > n_*$:

$$f(n) \leq \frac{1}{n^{c_*}}$$

Για το συγκεκριμένο c_* , αφού η f είναι αμελητέα υπάρχει n_f τέτοιο ώστε για κάθε $n > n_*$:

$$f(n) \leq \frac{1}{n^c}$$

Επίσης, η g είναι αμελητέα, οπότε για $c = 0$ υπάρχει n_g ώστε για κάθε $n > n_g$:

$$g(n) \leq \frac{1}{n^0}$$

$$g(n) \leq 1$$

Από τα προηγούμενα έχουμε ότι για $n_* = \max\{n_g, n_f\}$ ισχύει ότι για κάθε $n > n_*$:

$$f(n) \cdot g(n) \leq \frac{1}{n^c} \cdot 1$$

- Για την $f + g$:

Έστω κάποιο c_p , θα βρούμε κατάλληλο n_p .

Θεωρούμε το $c_p + 1$, και προσδιορίζουμε κατάλληλα n_f, n_g ώστε για $n > \max\{n_f, n_g\}$:

$$f(n) \leq \frac{1}{n^{c_p+1}}$$

$$g(n) \leq \frac{1}{n^{c_p+1}}$$

Θέτουμε $n_{max} = \max(n_f, n_g, 2)$. Τότε έχουμε για όλα τα $n > n_{max}$:

$$f(n) \leq \frac{1}{2 \cdot n^{c_p}}$$

$$g(n) \leq \frac{1}{2 \cdot n^{c_p}}$$

και άρα:

$$f(n) + g(n) \leq \frac{1}{n^{c_p}}$$

- Για την $f \cdot q$:

Έστω κάποιο c'' , θα βρούμε κατάλληλο n'' .

Το q θα έχει κάποιο βαθμό d , άρα για μεγάλα n (μεγαλύτερα από κάποιο n_s) θα ισχύει:

$$\forall n > n_q : q(n) < n^{d+1}$$

Αφού f αμελητέα, για το $n^{c''+d+1}$ θα υπάρχει n_s τέτοιο ώστε:

$$\forall n > n_s : f(n) < \frac{1}{n^{c''+d+1}}$$

Οπότε, για $n'' = \max\{n_q, n_s\}$ έχουμε:

$$\forall n > n'' : f(n) \cdot q(n) < \frac{n^{d+1}}{n^{c''+d+1}}$$

5. Για τους παρακάτω αλγορίθμους, να υπολογιστεί η στατιστική απόσταση της εξόδου τους από την ομοιόμορφη κατανομή U στο $S = \{0, 1, 2, \dots, A-1\}$.

Sampler 1. $n := \lceil \log_2 A \rceil$

$x_0, x_1, \dots, x_{n-1} \leftarrow \{0, 1\}$

$y := \sum_{i=0}^{n-1} 2^i \cdot x_i$

return y

Sampler 2. $x_0, x_1, \dots, x_{A-1} \leftarrow \{0, 1\}$

$y := \sum_{i=0}^{A-1} x_i$

return y

Υπόδειξη: Γνωρίζουμε ότι για τη διωνυμική ο μέσος όρος $E(X)$ είναι $n \cdot p = \frac{A}{2}$, και η διακύμανση $Var[X]$ είναι $n \cdot p \cdot q = \frac{A}{4}$. Επιπλέον, από τις σημειώσεις, ανατρέξτε στην ανισότητα του Chebychev (2.5.2):

$$\Pr[|X - E(X)| \geq t] \leq \frac{Var[X]}{t^2}$$

Sampler 3. $n := \lceil \log_2 A \rceil$

repeat:

$x_0, x_1, \dots, x_{n-1} \leftarrow \{0, 1\}$

$y := \sum_{i=0}^{n-1} 2^i \cdot x_i$

if $y < A$: return y

Υπόδειξη: Χρησιμοποιήστε την έννοια της δεσμευμένης πιθανότητας.

Sampler 1. Ορίζουμε $B = 2^n$, και ελέγχουμε ότι η έξοδος του sampler είναι μια τυχαία μεταβλητή Y που ακολουθεί την ομοιόμορφη κατανομή στο $\{0, 1, 2, \dots, B-1\}$. Παρατηρούμε επίσης ότι $B \geq A$, άρα το σύνολο τιμών της Y υπερκαλύπτει αυτό της U . Από τον ορισμό της στατιστικής απόστασης έχουμε:

$$\begin{aligned}
 \Delta(U, Y) &= \frac{1}{2} \sum_{i=0}^B |\Pr(U = 1) - \Pr(Y = i)| \\
 &= \frac{1}{2} \sum_{i=0}^A |\Pr(U = 1) - \Pr(Y = i)| + \frac{1}{2} \sum_{i=A}^B |\Pr(U = 1) - \Pr(Y = i)| \\
 &= \frac{1}{2} \sum_{i=0}^A \left| \frac{1}{A} - \frac{1}{B} \right| + \frac{1}{2} \sum_{i=A}^B \left| 0 - \frac{1}{B} \right| \\
 &= \frac{1}{2} A \left| \frac{1}{A} - \frac{1}{B} \right| + \frac{1}{2} (B - A) \left| 0 - \frac{1}{B} \right| \\
 &= \frac{1}{2} A \left(\frac{1}{A} - \frac{1}{B} \right) + \frac{1}{2} (B - A) \frac{1}{B} \\
 &= \frac{1}{2} \left(\frac{A}{A} - \frac{A}{B} \right) + \frac{1}{2} \frac{B - A}{B} \\
 &= \frac{1}{2} \left(1 - \frac{A}{B} + 1 - \frac{A}{B} \right) \\
 &= 1 - \frac{A}{B}
 \end{aligned}$$

Όπως αναμένουμε, όταν $A = B$ η στατιστική απόσταση είναι μηδενική, ενώ όταν $A = 2^k - 1$ η διαφορά είναι σχεδόν $\frac{1}{2}$. Όταν το A είναι κοντά στο B , παρατηρούμε επίσης ότι η απόσταση είναι αμελητέα ως προς το n .

Sampler 2. Αναγνωρίζουμε ότι η κατανομή Y του sampler είναι η διωνυμική με παραμέτρους $p = q = \frac{1}{2}$ και $n = A$ το πλήθος δοκιμές. Ξέρουμε ότι η κατανομή αυτή έχει σχήμα καμπύλης, άρα πιστεύουμε ότι μακριά από το μέσο όρο, θα εμφανίζει χαμηλή πιθανότητα. Αφού ο μέσος όρος είναι $p \cdot A = \frac{A}{2}$, επιλέγουμε να εξετάσουμε τις τιμές από το $\frac{3A}{4}$ ως το A . Προφανώς για τη στατιστική απόσταση θα ισχύει ότι:

$$\begin{aligned}
 \Delta(U, Y) &= \frac{1}{2} \sum_{i=0}^A |\Pr(U = 1) - \Pr(Y = i)| \\
 &\geq \frac{1}{2} \sum_{i=3A/4}^A |\Pr(U = 1) - \Pr(Y = i)| \quad (\text{Παραλείπουμε θετικούς όρους, το άθροισμα δεν αυξάνεται}) \\
 &\geq \frac{1}{2} \sum_{i=3A/4}^A (\Pr(U = 1) - \Pr(Y = i)) \quad (\text{Αφαιρούμε απόλυτα, άρα το άθροισμα δεν αυξάνεται}) \\
 &= \frac{1}{2} \sum_{i=3A/4}^A \Pr(U = 1) - \frac{1}{2} \sum_{i=3A/4}^A \Pr(Y = i) \\
 &= \frac{1}{2} \cdot \frac{1}{4} - \frac{1}{2} \sum_{i=3A/4}^A \Pr(Y = i) \\
 &= \frac{1}{8} - \frac{1}{2} \sum_{i=3A/4}^A \Pr(Y = i)
 \end{aligned}$$

Για να έχουμε λοιπόν μια εκτίμηση για την απόσταση, μένει να υπολογίσουμε (προσεγγιστικά) το άθροισμα πιθανοτήτων της Y για τιμές από $3A/4$ ως A , ή ισοδύναμα την πιθανότητα $P[Y >= \frac{3A}{4}]$. Από τις σημειώσεις γνωρίζουμε την ανισότητα του Chebyshev:

$$\Pr[|Y - E(Y)| \geq t] \leq \frac{Var[Y]}{t^2}$$

Γνωρίζουμε ότι για τη διωνυμική ο μέσος όρος $E(Y)$ είναι $n \cdot p = \frac{A}{2}$, η διακύμανση $Var[Y]$ είναι $n \cdot p \cdot q = \frac{A}{4}$. Αντικαθιστούμε για $t = \frac{A}{4}$ και έχουμε:

$$\Pr[Y \geq \frac{3A}{4}] = \Pr[Y - \frac{A}{2} \geq \frac{A}{4}] = \Pr[Y - E(Y) \geq t] \leq \Pr[|Y - E(Y)| \geq t] \leq \frac{\frac{A}{4}}{\frac{A}{4} \cdot \frac{A}{4}} = \frac{1}{A} = \frac{1}{A}$$

Επιστρέφοντας στο φράγμα για τη στατιστική απόσταση έχουμε ότι:

$$\begin{aligned}
 \Delta(U, Y) &= \frac{1}{2} \sum_{i=0}^A |\Pr(U = 1) - \Pr(Y = i)| \\
 &\geq \frac{1}{8} - \frac{1}{2} \sum_{i=3A/4}^A \Pr(Y = i) \\
 &\geq \frac{1}{8} - \frac{1}{2} \frac{1}{A} \\
 &= \frac{1}{8} - \frac{1}{2A}
 \end{aligned}$$

Άρα για μεγάλες τιμές του A , η στατιστική απόσταση είναι σημαντική, και αυξάνεται όσο μεγαλώνει το A .

Παρατήρηση: Σε τρία σημεία της λύσης μπορούμε επικαλούμενοι τη συμμετρία να πάρουμε σημαντικά καλύτερες τιμές για τα φράγματα (περίπου 3 διπλασιασμούς). Πρώτα: μπορούμε να εξετάσουμε και το διάστημα από 0 έως $\frac{A}{4}$. Μετά: μπορούμε να ισχυριστούμε ότι αφού οι πιθανότητες και των δύο κατανομών αθροίζονται στο 1, η διαφορά που έχουν στο διάστημα $[0, A]$ θα είναι συνολικά 0, άρα η διαφορά στο κεντρικό διάστημα (χωρίς απόλυτη τιμή) θα είναι η αντίθετη της διαφοράς στα άκρα (οπότε εφαρμόζοντας την απόλυτη τιμή διπλασιάζουμε τη διαφορά που είχαμε). Τέλος, στην ανισότητα Chebychev φράξαμε την πιθανότητα να ξεπεράσουμε τα $3/4$ με την πιθανότητα να απομακρυνθούμε από το $1/2$ κατά $1/4$ ή παραπάνω, η οποία λόγω συμμετρίας θα είναι περίπου διπλάσια.

Sampler 3. Θα κάνουμε χρήση της δεσμευμένης πιθανότητας. Οι τιμές του y επιλέγονται ομοιόμορφα από το $\{0, 1, 2, \dots, B - 1\}$ αλλά επιστρέφονται μόνο εάν ισχύει $y < A$. Ονομάζουμε Z την τυχαία μεταβλητή της εξόδου του sampler, και θεωρούμε την Y όπως στον πρώτο sampler. Από την κατασκευή του προγράμματος έχουμε ότι: $\Pr(Z = x) = \Pr(Y = x | Y < A)$. Από τον ορισμό της δεσμευμένης πιθανότητας:

$$\Pr(Y = x | Y < A) = \Pr(Y = x | x < A) = \frac{\Pr((Y = x) \cap (x < A))}{\Pr(Y < A)}$$

Όταν $x \geq A$ η παραπάνω πιθανότητα είναι μηδενική. Στη μη τετριμμένη περίπτωση όπου $x < A$, έχουμε:

$$\frac{\Pr(Y = x)}{\Pr(Y < A)} = \frac{\frac{1}{B}}{\frac{A}{B}} = \frac{1}{B} \cdot \frac{B}{A} = \frac{1}{A}$$

Άρα, για $x \geq A$, $\Pr(Z = x) = 0$ και για $x < A$, $\Pr(Z = x) = \frac{1}{A}$, και η στατιστική απόσταση είναι προφανώς 0. Παρατηρούμε όμως ότι ο χρόνος εκτέλεσης δεν είναι σταθερός.