Proving outcome of private statistical signal testing

Vadym Fedyukovych
Intropro, Kiev, Ukraine

Abstract—**We consider a decision problem recognising whether signal correlation with a reference pattern is above a threshold on condition that no information about reference is released to verifying party other than validity of the statement tested. It is desirable sometimes to keep the reference private avoiding releasing it even to verifying party while testing validity of a statement. Furthermore, exact value of signal-to-reference correlation can be a sensitive data. Releasing the threshold might be unwanted as well. At last, weighted matching might be an intellectual property and should be protected.**

A protocol is introduced to prove knowledge of features of reference, weights and threshold committed such that a weighted Euclidean-like distance between features of signal and that of reference is below the threshold.

I. Introduction

Deciding whether signal correlation with a reference pattern is above a threshold is the core of statistical signal testing. With private testing, verifying party is receiving limited information about the data involved in testing.

Interactive proof systems is the well-established method of probabilistic testing of validity of statements without releasing additional information about statement tested.

Lagrange theorem on four squares was suggested to handle 'less than' relation over integers. Namely, one can always find four integers replacing 'less than' with equality relation considered 'a solution' to original relation. Given commitment to such a solution, one can run a protocol as a verifier. Such a protocol can be elaborated to also prove knowledge of the threshold, in case it should be also kept private.

We approach correlation between the signal and reference by comparing features derived from signal and reference respectively. We model custom matching with a kind of Euclidean distance in feature space. We refine matching with weighted metric to specify which features are more 'significant' for matching.

A. Contribution

We introduce a protocol to let verifying party test validity of 'less or equal statement' for a set of integer-valued features derived from signal and another set of reference features in weighted Euclidean metric, with the other party proving his knowledge of reference features, weights and threshold. Signal features and commitments to reference features, weights and threshold are common input of the parties. Commitments are produced according to Damgard-Fujisaki integer commitment scheme. Protocol has negligible soundness error in a single run, is statistically witness indistinguishable and statistical special honest verifier zero knowledge.

B. Previous results

Interactive proof systems [1] and proofs of knowledge [2] were initially introduced in the contaxt of computational complexity theory. Proofs with 'algebraic' responses [3] were introduced for proving knowledge of discrete logarithm. Proofs with challenges chosen from a large set [4] have negligible soundness error in a single run. Any homomorphic commitment scheme admits [5] Schnorr-like proofs, including Pedersen commitment scheme [6] (for an element of a prime field) and Damgard-Fujisaki integer commitment scheme [7].

Range proofs [8] was the first successful attempt to introduce proofs for relations of 'more than' type, followed by Lipmaa proof [9] derived from Lagrange theorem.

A challenge-response system for proving statement about polynomials was derived from Schwartz-Zippel lemma [10], ans was used to design a number of proofs, including proof of knowledge protocol for 'almost the same' sets and an electronic signature scheme tolerating errors with private signing keys [11], for a codeword of Goppa code [12], [13], for deciding Hamiltonicity [14], [15] and for multiple substring matching [16], [17].

An alternative protocol proving knowledge of 4-tuple of integers following Lipmaa setup and with challenge-response system for polynomials was presented at Information Security conference, Kiev, 2008. An enhancement protocol hiding threshold and weights is presented in this paper.

II. Definition

Interactive proof system [1] for deciding a language is a pair of probabilistic interactive Turing machines such that Verifier machine always outputs yes/no decision in polynomial time and completeness and soundness properties holds. We only consider polynomial time Prover machines with private input tape with *witness* on it. We say a party following the protocol is *honest* and we say *any party* otherwise. *Completeness error* is probability for a honest Verifier to accept for a honest Prover and input from language. We only consider perfect completeness such that honest Verifier always accepts for a honest Prover. *Soundness error* is probability for a honest Verifier to accept for any Prover and input not from language taken over random choices of Verifier; we require soundness error to be negligible (for example, exponentially small) in security parameter. Protocol is of *proof* type if probability estimates are unconditional, and is an *argument* if it depends on additional assumptions (like hardness of factorisation). Protocol is a *proof of knowledge* if *extractor* algorithm exists [2] producing with some propability witness of Prover given acceptable responses of Prover controlled by a machine with rewind capability. We only consider protocols always producing either witness of Prover or solution to an instance of a hard problem from any pair of acceptable responses. Protocol is *zero knowledge* if an expected polynomial time *simulator* algorithm exists [18] producing *simulated transcript* indistinguishable from all protocol transcripts with a Prover. Protocol with 3 messages is *special honest verifier zero knowledge* [19] if simulated transcript is indistinguishable from all protocol transcripts having the same challenge. Protocol is *witness hiding* if no polynomial time algorithm exists [20] producing witness of Prover from protocl transcript. Protocol is *witness indistinguishable* if there is only a negligible advantage for any polynomial time algorithm [20] deciding which witness was used by Prover while running the protocol.

Factorisation problem is an assumption that no efficient algorithm exists producing the set of primes such that their product is the number on input tape. *Strong RSA problem* is an assumption that no efficient algorithm exists producing e, A for given random N, B such that $B = A^e \pmod N$.

III. Protocol

Let $\vec{a} = \{a_i, i = 1 \ldots n\}$ and $\vec{b} = \{b_i, i = 1 \ldots n\}$ be feature vectors of the signal and the reference, $\{m_i\}$ be non-negative weights of an Euclidean-like feature vector norm $d(\vec{a}) = \sum_{i=1}^{n} m_i a_i^2$ and t be a threshold such that $d(\vec{a}-\vec{b}) \leq t$. We consider feature vector components, weights and threshold to be integers.

Consider a game for two parties that are willing to test validity of a statement of good enough matching of the signal and the reference. Verifying party is only interested in testing the statement, and proving party is interesting to keep reference private avoiding sending it to verifying party for testing. Even more, signal and reference should be compared by evaluating feature difference in Euclidean-like metric with weights, and proper weights could been established with a case study and are considered an intellectual property. Parties agree to follow a reasonable testing technique to meet seemingly contradicting requirements, in case it exists.

Let $N = p_1 p_2$ be a product of two strong primes $p_1 = 2p_1' + 1$, $p_2 = 2p_2' + 1$, and let $\mathcal{G}$ be a multiplicative group of square residues modulo N. Let g, h be two elements of $\mathcal{G}$. We assume Prover does not know factorisation of N and Strong RSA and factorisation problems are hard for Prover. In thr following, we omit $\pmod N$ in case group operation is assumed.

A. Protocol description

Let n be number of components of feature vectors; $\{B_i\}$, $\{M_i\}$, T be Damgard-Fujisaki commitments to reference feature vector components $\{b_i\}$, weights $\{m_i\}$ and threshold t: $B_i = g^{b_i} h^{\zeta_{Bi}}$, $M_i = g^{m_i} h^{\zeta_{Mi}}$, $T = g^t h^{\zeta_T}$.

Let γ_c be a security parameter such that 2^{γ_c} is less that group order $p_1' p_2'$. This parameter defines an interval to choose Verifier challenge from. Soundness error depends, as we will see, on this parameter.

Common input of Prover and Verifier is (N, g, h), $\{a_i\}$, $\{B_i\}$, $\{M_i\}$, T. Private input of Prover is $\{b_i\}$, $\{\zeta_{Bi}\}$, $\{m_i\}$, $\{\zeta_{Mi}\}$, t, ζ_T. Protocol goes as follows:

1) Prover produces $\{r_j, j = 1 \ldots 4\}$ such that

$$\sum_{j=1}^{4} r_j^2 = t - \sum_{i=1}^{n} m_i (a_i - b_i)^2 \qquad (1)$$

Prover chooses some $\{\zeta_{Rj}\}$ at random and produces $R_j = g^{r_j} h^{\zeta_{Rj}}$. Prover chooses $\{\phi_i, i = 1 \ldots n\}$, $\{\psi_i\}$, $\{\theta_i\}$, ω, $\{\xi_j, j = 1 \ldots 4\}$, produces $\{C_{Bi}\}$,$\{C_{Mi}\}$,C_T,$\{C_{Rj}\}$: $C_{Bi} = g^{\psi_i} h^{\delta_{Bi}}$, $C_{Mi} = g^{\theta_i} h^{\delta_{Mi}}$, $C_T = g^{\omega} h^{\delta_T}$, $C_{Rj} = g^{\xi_j} h^{\delta_{Rj}}$. Prover also produces a polynomial $F(z)$ of degree 2. Let $\{u_k, k = 0 \ldots 2\}$ be coefficients of $F(z)$:

$$F(z) = z^2(zt+\omega) - \sum_{i=1}^{n} (zm_i+\theta_i)(za_i-(zb_i+\psi_i))^2 - z\sum_{j=1}^{4} (zr_j + \xi_j)^2 = u_2 z^2 + u_1 z + u_0 \qquad (2)$$

Prover chooses ζ_{Uk} at random and produces $U_k = g^{u_k} h^{\zeta_{Uk}}$. Prover sends $\{R_j\}$, $\{C_{Bi}\}$, $\{C_{Mi}\}$, C_T, $\{C_{Rj}\}$, $\{U_k\}$ to Verifier.

2) Verifier chooses a challenge c from interval $[1 \ldots 2^{\gamma_c}]$ and sends it to Prover.

3) Prover produces and sends responses:

$$\Psi_i = cb_i + \psi_i, \quad \Delta_{Bi} = c\zeta_{Bi} + \delta_{Bi} \qquad (3)$$

$$\Theta_i = cm_i + \theta_i, \quad \Delta_{Mi} = c\zeta_{Mi} + \delta_{Mi} \qquad (4)$$

$$\Omega = ct + \omega, \quad \Delta_T = c\zeta_T + \delta_T \qquad (5)$$

$$\Phi_j = cr_j + \phi_j, \quad \Delta_{Rj} = c\zeta_{Rj} + \delta_{Rj} \qquad (6)$$

$$\Delta_c = c^2 \zeta_{U2} + c\zeta_{U1} + \zeta_{U0} \qquad (7)$$

4) Verifier produces

$$F_c = c^2 \Omega - \sum_{i=1}^{n} \Theta_i (ca_i - \Psi_i)^2 - c\sum_{j=1}^{4} \Phi_j^2 \qquad (8)$$

Verifier accepts if

$$g^{\Psi_i} h^{\Delta_{Bi}} B_i^{-c} = C_{Bi} \qquad (9)$$

$$g^{\Omega} h^{\Delta_T} T^{-c} = C_T \qquad (10)$$

$$g^{\Theta_i} h^{\Delta_{Mi}} M_i^{-c} = C_{Mi} \qquad (11)$$

$$g^{\Phi_j} h^{\Delta_{Rj}} R_j^{-c} = C_{Rj} \qquad (12)$$

$$g^{F_c} h^{\Delta_c} U_2^{-c^2} U_1^{-c} U_0^{-1} = 1 \qquad (13)$$

B. Protocol properties

It is clear that honest Verifier always accepts for a honest Prover so protocl has zero completeness error.

Theorem 1 *Protocol described in section III-A has negligible soundness error on assumption of hardness of factorisation and Strong RSA problems.*

Proof: Consider the case of reference and signal features that differ more than the threshold in the weighted metric. It is clear that no 4-tuple of integers exists such that sum of their squares is a negative number. It follows that for any 4-tuple (r_1, r_2, r_3, r_4), any 3-tuple (u_2, u_1, u_0) and any ω, $\{\psi_i\}$, $\{\theta_i\}$, $\{\phi_j\}$ polynomial $F_u(z)$ is of degree exactly 3:

$$F_u(z) = z^2(zt + \omega) - \sum_{i=1}^{n}(zm_i + \theta_i)(za_i - (zb_i + \psi_i))^2 - z\sum_{j=1}^{4}(zr_j + \phi_j)^2 - z^2u_2 - zu_1 - u_0 \quad (14)$$

There is at most $\frac{1}{2^{\gamma_c}}$ probability for any Prover to provide acceptable responses passing (9), (10), (11), (12) without knowledge of reference feature vector, threshold, weights and 4-tuple from Lagrange theorem committed at $\{B_i\}$, T, $\{M_i\}$, $\{R_j\}$. As will be shown later, it is not the best strategy for any Prover to cheat while passing (9), (10), (11), (12), so we consider responses $\{\Psi_i\}$, Ω, $\{\Theta_i\}$, $\{\Phi_j\}$ to be linear polynomials evaluated at some point $z = c$ chosen as a challenge by Verifier.

From Schwartz-Zippel lemma [10] it follows there is at most $\frac{3}{2^{\gamma_c}}$ probability to choose a root of $F_u(z)$ by choosing some $z = c$ at random with flat distribution from interval $[1 \ldots 2^{\gamma_c}]$. Let $F_u(c) \neq 0$, $\Delta_u = \Delta_c - c^2\zeta_2 - c\zeta_1 - \zeta_0$, $d = \gcd(F_u(c), \Delta_u)$. Then if $d \neq 1$ it should be a multiple of group order: $(g^{\frac{F_u(c)}{d}} h^{\frac{\Delta_u}{d}})^d = 1 \pmod N$, which is equivalent to solving factorisation of N. Let $d = 1$. Then some (α, β) can always be found with extended Euclidean algorithm such that $\alpha F_u(c) + \beta\Delta_u = 1$, which is a solution of an instance of Strong RSA problem: $h = (h^{\alpha} g^{-\beta})^{F_u(c)}$.

It follows protocol soundness error is $\frac{3}{2^{\gamma_c}}$ and protocol is an argument. ■

We only outline our proof for statistical witness indistinguishablility. We can always set proper large intervals for choosing random values for $\{\psi_i\}$, ω, $\{\theta_i\}$, $\{\phi_j\}$ such that initial random values would be expected much larger than the witness $\{b_i\}$, t, $\{m_i\}$, $\{r_j\}$ making it hard to distinguish it from responses. The same reasoning with larger intervals applies for choosing random values while producing commitments.

Theorem 2 *Protocol described in section III-A is special honest Verifier zero knowledge.*

Proof: Protocol has a simulator shown below. Challenge c is an input to this algorithm.

1) Verifier chooses random group elements $\{R_j\}$,$(U_2, U_1,$
2) Verifier chooses $\{\Psi_i\}$, $\{\Delta_{Bi}\}$, Ω, Δ_T, $\{\Theta_i\}$, $\{\Delta_{Mi}\}$, $\{\Phi_j\}$, $\{\Delta_{Rj}\}$, Δ_c at random.
3) Verifier produces

$$F_c = c^2\Omega - \sum_{i=1}^{n}\Theta_i(ca_i - \Psi_i)^2 - c\sum_{j=1}^{4}\Phi_j^2 \quad (15)$$

$$C_{Bi} = g^{\Psi_i}h^{\Delta_{Bi}}B_i^{-c} \quad (16)$$

$$C_T = g^{\Omega}h^{\Delta_T}T^{-c} \quad (17)$$

$$C_{Mi} = g^{\Theta_i}h^{\Delta_{Mi}}M_i^{-c} \quad (18)$$

$$C_{Rj} = g^{\Phi_j}h^{\Delta_{Rj}}R_j^{-c} \quad (19)$$

$$U_0 = g^{F_c}h^{\Delta_c}U_2^{-c^2}U_1^{-c} \quad (20)$$

Simulated transcript is $\{R_j\}$,$\{C_{Bi}\}$,$\{C_{Mi}\}$,C_T,$\{C_{Rj}\}$,$\{U_k\}$, c, $\{\Psi_i\}$, $\{\Delta_{Bi}\}$, Ω, Δ_T, $\{\Theta_i\}$, $\{\Delta_{Mi}\}$, $\{\Phi_j\}$, $\{\Delta_{Rj}\}$, Δ_c. ■

IV. Conclusion

A constant-round protocol is presented to prove a statement facilitating statistical signal analysis. In particular, verifying party can test that feature difference is below a threshold in weighted Euclidean metric without learning some other information about threshold, weights and reference features. Protocol is of argument type, on condition of hardness of factorisation and Strong RSA problems.

References

[1] S. Goldwasser, S. Micali, and C. Rackoff, "The knowledge complexity of interactive proof systems," *SIAM J. Comput.*, vol. 18, no. 1, pp. 186–208, 1989.

[2] M. Bellare and O. Goldreich, "On defining proofs of knowledge," in *CRYPTO*, 1992, pp. 390–420.

[3] D. Chaum, J.-H. Evertse, and J. van de Graaf, "An improved protocol for demonstrating possession of discrete logarithms and some generalizations." in *EUROCRYPT*, 1987, pp. 127–141.

[4] C.-P. Schnorr, "Efficient identification and signatures for smart cards." in *CRYPTO*, 1989, pp. 239–252.

[5] V. Fedyukovych, "On taking decisions with finite automata (in Russian)," in *Ukrainian Mathematical Congress*, 2009, http://www.imath.kiev.ua/~congress2009/en/.

[6] T. P. Pedersen, "Non-interactive and information-theoretic secure verifiable secret sharing." in *CRYPTO*, 1991, pp. 129–140.

[7] I. Damgård and E. Fujisaki, "A statistically-hiding integer commitment scheme based on groups with hidden order," in *ASIACRYPT*, 2002, pp. 125–142.

[8] F. Boudot, "Efficient proofs that a committed number lies in an interval." in *EUROCRYPT*, 2000, pp. 431–444.

[9] H. Lipmaa, "On diophantine complexity and statistical zero-knowledge arguments," in *ASIACRYPT*, 2003, pp. 398–415.

[10] J. T. Schwartz, "Fast probabilistic algorithms for verification of polynomial identities," *J. ACM*, vol. 27, no. 4, pp. 701–717, 1980.

[11] V. Fedyukovych, "A signature scheme with approximate key matching (in Russian)," in *Information Technology and Systems Conference*, 2009, pp. 396–400, http://www.iitp.ru/ru/conferences/539.htm.

[12] ——, "Argument of knowledge of a bounded error," Cryptology ePrint Archive, Report 2008/359, 2008.

[13] ——, "Argument of knowledge of a codeword of Goppa code and a bounded error (in Russian)," *Applied Discrete Mathematics (PDM)*, no. 4, pp. 64–71, 2009, http://mi.mathnet.ru/pdm151.

[14] ——, "An argument for Hamiltonicity," Cryptology ePrint Archive, Report 2008/363, 2008.

[15] ——, "Protocols for graph isomorphism and hamiltonicity," in *Central European Conference on Cryptography*, 2009.

[16] V. Fedyukovych and V. Sharapov, "A protocol for K-multiple substring matching," Cryptology ePrint Archive, Report 2008/357, 2008.

[17] ——, "A protocol for K-multiple substring matching (in Russian)," in *Information Technology and Systems Conference (ITaS)*, 2008, pp. 459–466, http://www.iitp.ru/ru/conferences/340.htm.

[18] O. Goldreich, S. Micali, and A. Wigderson, “Proofs that yield nothing but their validity for all languages in NP have zero-knowledge proof systems,” *J. ACM*, vol. 38, no. 3, pp. 691–729, 1991.

[19] R. Cramer, I. Damgård, and B. Schoenmakers, “Proofs of partial knowledge and simplified design of witness hiding protocols,” in *CRYPTO*, 1994, pp. 174–187.

[20] U. Feige and A. Shamir, “Witness indistinguishable and witness hiding protocols,” in *STOC*, 1990, pp. 416–426.