

## 10 zkSNARKs from Extended Algebraic Programs

In this section we provide a new type of NIZK arguments which allows to prove knowledge of a secret key, given a public key, with only one 1 gate instead of  $>864$  as in current state of the art implementations (disregarding the constraints they require in addition to assure that the input is valid), for the price of a drastic reduction of the statistical knowledge soundness. It is an extension of Groths[23] QAP based zkSNARK construction, where we roughly speaking add the word for 'point multiplication on an elliptic curve', to the words '+' and '×'. Therefore the language of the arithmetic circuit, R1CS and QAP had to be extended consequently to what we name R1CS\* and extended algebraic program (EAP) instead of the commonly used QAPs.

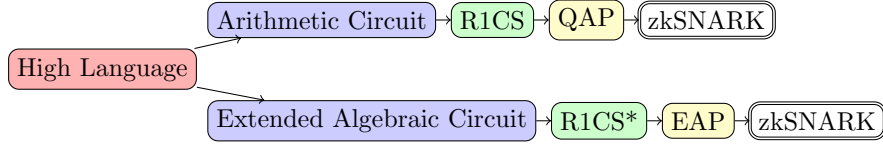


Figure 6: Shows the necessary steps to make a program written in a high level language (could be a subset of common C), provable with a zkSNARK. The upper path is the standard procedure. The path below our contribution.

The verifier work increases from 3 to 4 pairing function calls, 1 exponentiation in  $\mathbb{G}_T$ , 1 in  $\mathbb{G}_K$  and 2 in  $\mathbb{G}_K$ . Before this extension, proving knowledge of the discrete logarithm (e.g. I prove I know  $x$  s.t.  $pk = g^x$  without revealing  $x$ ) had to be done by expressing the point multiplication as arithmetic circuit. This required at least 864 gates for each point multiplication within a circuit,  $N \cdot 864$  for  $N$  point multiplications, what increases CRS by at least  $N \cdot 3 \cdot 864$  elements. Here its  $N$  gates and therefore  $3N$  CRS elements. The prover time improves drastically as he does not need to compute the coefficients of the polynomials that were needed for proving knowledge of a valid arithmetic circuit assignment. Consequently the prover also does not need to perform the corresponding blind evaluations, which requires a point multiplications for each coefficient and therefore the most largest source of overhead. The polynomial division effort decrease as the degree is reduced by  $N \cdot 864$ . Before this scheme, it was almost infeasible to pack multiple point multiplications into one SNARK. Performing a shielded transaction in zCash could be done in a few milliseconds instead of seconds now. To make the proof elements uniformly random and therefore zero knowledge, only one additional gate is required.

Our NIZK arguments for EAPs considers a circuit consisting of addition, multiplication and elliptic curve point multiplication gates. The computations are performed over a finite field  $\mathbb{F}$ . Figure 10 shows a how a EAP enforced R1CS\* constraint for a specific statement looks like. In the following lines we write  $\langle \mathbf{A}, \mathbf{B} \rangle_{[N_i, N_j]} := \sum_{i=N_i}^{N_j} A_i B_i$ , and  $\langle \mathbf{A}, \mathbf{B} \rangle_{[N_j]} := \langle \mathbf{A}, \mathbf{B} \rangle_{[0, N_j]}$  for shortness.

Also we use multiplicative notation for the elliptic curves group operation.

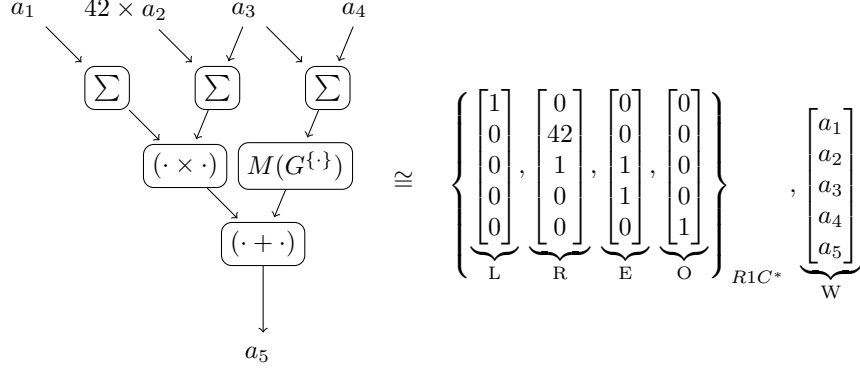


Figure 7: Shows how the example statement  $a_1 \times (42 \times a_2 + a_3) + M(G^{a_3+a_4}) = a_5$  is realizable with one EAP gate and hence can be expressed as a single extended Rank 1 Constraint (R1C\*). The prover will convince the verifier that he has knowledge of  $W$  s.t. it satisfies:  $\mathcal{G}_M^A(\langle E, W \rangle) + \langle L, W \rangle \cdot \langle R, W \rangle = \langle O, W \rangle$ .

The description of a EAP is

$$EAP := (\mathbb{F}, e(\cdot, \cdot), \mathcal{G}_M^A(\cdot), l, \{L_i(X), R_i(X), E_i(X), O_i(X)\}_{i=0}^m, D(X)),$$

where

- $\mathbb{F}$  is a finite field.
- $e$  is a bilinear mapping  $e : \mathbb{G}_1 \times \mathbb{G}_2 \rightarrow \mathbb{G}_T$  as described in section 8.  $G$  as  $\mathbb{G}_1$  and  $H$  as  $\mathbb{G}_2$  generators.
- $n$  is the number of gates in a circuit
- $m$  is the number of wires of the circuit
- $\mathcal{A} \subset \mathbb{N}_{\geq 0}$  defines the set of gate indexes. We enumerate the gates from 0 to  $n-1$  e.g.  $\mathcal{A} = \{0, \dots, n-1\}$
- $L_i(X), R_i(X), E_i(X), O_i(X)$  are polynomials of degree  $n-1$  that express the scale factor the  $i$ -th wire at gate  $X$ . For example:  $L_3(42) = 1$  states that the 3rd wire is the 'left' multiplication input of gate with index 42.  $R_7(42) = 2$  implies, that the 7th wire is 'right' multiplication input of gate with index 42 and that this wires assigned signal will be scaled by the linear factor of 2.  $E_7(42) = 1$  implies, that the 7th wire is exponentiation input of gate with index 42. We write  $\mathbf{L}$  for the vector of polynomials  $(L_0(X), L_1(X), \dots, L_m(X))$  (same for  $\mathbf{R}, \mathbf{E}, \mathbf{O}$ ).
- Each EAP-gate has a uniquely assigned index  $r_i \in \mathcal{A}$ , which compose the domain polynomial  $D(X) = \prod_{r_g \in \mathcal{A}} (X - r_g)$ . The degree  $\deg(D(X)) = n = |\mathcal{A}|$  consequently.

- A crucial role plays the projection function  $M(\cdot) : \mathbb{G}_1 \mapsto \mathbb{F}$ . Its best possible design currently exceed our understanding. Its most trivial realization would be taking  $x$  coordinate of the curve point. However: This would reduce security by factor 2 since for each curve point  $(X, Y)$  (except the roots) exists  $(X, -Y)$ . It reduces security even further by a smaller factor since we have to compensate the outputs of of the point multiplication results with the assignments which end up in a linear combination with  $O_i$ . The assignments are computed on the arithmetic circuit field  $\mathbb{F}$  which is preferably of the same order then the curve is defined on. In cryptographic constructions the curve order is co-prime to  $|\mathbb{F}|$ . Therefor we lose information as point multiplication input is modulo the curve order eg.  $G^x = G^{x \bmod |\langle G \rangle|}$ . It has turned out to be very beneficial if the field order  $|\mathbb{F}|$  where the arithmetic operations are performed on is equivalent to the order of the field on which the curve is defined. This brings the advantage that curve operations can be performed within the circuit using far fewer gates.
- $\mathcal{G}_M^{\mathcal{A}}(P(X)) \rightarrow \tilde{P}(X)$ , where  $\mathcal{A}$  is the set of gate indexes and  $|\mathcal{A}| = k + 1$  takes as input a polynomial of degree  $k$  and returns a polynomial of degree  $k$  s.t.  $\forall x \in \mathcal{A} : \tilde{P}(x) = M(G^{P(x)})$ .

A EAP is defined by the binary relation

$$R = \left\{ (\phi, w) \left| \begin{array}{l} \phi = (a_1, \dots, a_l) \in \mathbb{F}^l \\ w = (a_{l+1}, \dots, a_m) \in \mathbb{F}^{m-1} \\ \mathcal{G}_M^{\mathcal{A}}(\langle \mathbf{a}, \mathbf{E} \rangle) + \langle \mathbf{a}, \mathbf{L} \rangle_{[m]} \cdot \langle \mathbf{a}, \mathbf{R} \rangle_{[m]} - \langle \mathbf{a}, \mathbf{O} \rangle_{[m]} = Q(X)D(X) \end{array} \right. \right. \quad (19)$$

for some degree  $n - 2$  quotient polynomial  $Q(X)$ . This gives us the NIZK argument:

$(\sigma, \tau) \leftarrow \text{Setup}(R, 1^\lambda)$ : Pick  $\alpha, \beta, \gamma, \delta \xleftarrow{\$} \mathbb{Z}_p^*, x \xleftarrow{\$} \mathcal{A}$ . Define  $\tau = (\alpha, \beta, \gamma, \delta, x)$ . The common reference string is:

$$\sigma = \left( \left\{ G^{\frac{\alpha R_i(x) + \beta(L_i(x) + E_i(x)) + O_i(x)}{\gamma}} \right\}_{i=0}^l, \left\{ G^{\frac{\alpha R_i(x) + \beta(L_i(x) + E_i(x)) + O_i(x)}{\delta}} \right\}_{i=l+1}^m, G^\alpha, G^\beta, G^\gamma, H^\beta, H^\gamma, H^\delta, \left\{ G^{\frac{x^i D(x)}{\delta}} \right\}_{i=0}^{n-1}, \left\{ G^{x^i} \right\}_{i=0}^{n-1}, \left\{ H^{x^i} \right\}_{i=0}^{n-1} \right)$$

$\pi = (A, B, C, D) \leftarrow \text{Prove}(EAP, \sigma, \phi, w)$ : Compute

$$\begin{aligned} A &= G^{\alpha + \langle \mathbf{a}, \mathbf{L}(x) \rangle_{[m]} + \langle \mathbf{a}, \mathbf{E}(x) \rangle_{[m]}} \\ B &= H^{\langle \mathbf{a}, \mathbf{R}(x) \rangle_{[m]}} \\ C &= G^{\frac{\langle \mathbf{a}, \beta \mathbf{L}(x) + \alpha \mathbf{R}(x) + \beta \mathbf{E}(x) + \mathbf{O}(x) \rangle_{[l+1, m]} + Q(x)D(x)}{\delta}} \\ F &= G^{\langle \mathbf{a}, \mathbf{E}(x) \rangle_{[m]}} \end{aligned}$$

$0/1 \leftarrow \text{Vfy}(EAP, \sigma, \phi, \pi)$  accept iff:

$$e(A, B \cdot H^\beta) \cdot e(G, H)^{M(F)} = \\ \left( e(G^\alpha, H^\beta) \cdot e\left(G^{\frac{\langle \mathbf{a}, \alpha \mathbf{R}(x) + \beta(\mathbf{L}(x) + \mathbf{E}(x)) + \mathbf{O}(x) \rangle_{[l]}}{\gamma}}, H^\gamma \right) \cdot e(C, H^\delta) \cdot e(F, B) \right)$$

Regarding efficiency we observe that the one-time setup  $\sigma$  runs in time linear to the circuits size  $\mathcal{O}(|C|)$ . The prover must perform  $\mathcal{O}(|C|)$  cryptographic work and  $\mathcal{O}(|C| \log^2(|C|))$  to compute  $Q(x)$ . The polynomial representation  $L_i(X), R_i(X)$  etc. is not needed for proving and the prover rather works with the vectors  $l_i = (L_i(1), L_i(2), \dots, L_i(n)), r_i = (R_i(1), R_i(2), \dots, R_i(n))$  where most elements are 0. This sparsity of the evaluation vectors is then exploited. The  $\mathcal{O}(|C| \log^2(|C|))$  runtime is achieved by using FFT techniques and binary tree based polynomial interpolation algorithms as it is done in the Pinocchio-Protocol[27] instead of naive Lagrange interpolation and polynomial division which would take  $\mathcal{O}(n^2)$ .

Note that  $e(G, H)$  and  $e(G^\alpha, H^\beta)$  can be precomputed. Furthermore verification requires  $l$  exponentiations e.g. proofing time is quasi linear in the statement size. Constructing the pairing-friendly elliptic curves such that  $\mathbb{G}_1$  representations of group elements are smaller,  $C$  is assigned to the first group for efficiency. For the same reason the verifier includes the statement  $\phi$  into the verification process over the group  $\mathbb{G}_1$ .

We proof perfect completeness by direct verification:

$$\begin{aligned} & e(A, B \cdot H^\beta) \cdot e(G, H)^{M(F)} = \\ & e(A, B \cdot H^\beta) \cdot e(G, H)^{\mathcal{G}_M^A(\langle \mathbf{a}, \mathbf{E} \rangle)} = \\ & e(G, H)^{(\alpha + \langle \mathbf{a}, \mathbf{L} \rangle_{[m]} + \langle \mathbf{a}, \mathbf{E} \rangle_{[m]})(\langle \mathbf{a}, \mathbf{R} \rangle_{[m]} + \beta) + M(F)} = \\ & e(G, H)^{\alpha \langle \mathbf{a}, \mathbf{R} \rangle_{[m]} + \langle \mathbf{a}, \mathbf{L} \rangle_{[m]} \langle \mathbf{a}, \mathbf{R} \rangle_{[m]} + M(F) + \langle \mathbf{a}, \mathbf{E} \rangle_{[m]} \langle \mathbf{a}, \mathbf{R} \rangle_{[m]} + \alpha \beta + \beta \langle \mathbf{a}, \mathbf{E} + \mathbf{L} \rangle_{[m]}} = \\ & e(G, H)^{\alpha \langle \mathbf{a}, \mathbf{R} \rangle_{[m]} + \langle \mathbf{a}, \mathbf{O} \rangle_{[m]} + QD + \langle \mathbf{a}, \mathbf{E} \rangle_{[m]} \langle \mathbf{a}, \mathbf{R} \rangle_{[m]} + \alpha \beta + \beta \langle \mathbf{a}, \mathbf{E} + \mathbf{L} \rangle_{[m]}} = \\ & e(G^\alpha, H^\beta) \cdot e(G, H)^{\langle \mathbf{a}, \alpha \mathbf{R} + \beta \mathbf{E} + \beta \mathbf{L} + \mathbf{O} \rangle_{[m]} + QD + \langle \mathbf{a}, \mathbf{E} \rangle_{[m]} \langle \mathbf{a}, \mathbf{R} \rangle_{[m]}} = \\ & e(G^\alpha, H^\beta) \cdot e(G, H)^{\langle \mathbf{a}, \alpha \mathbf{R} + \beta \mathbf{E} + \beta \mathbf{L} + \mathbf{O} \rangle_{[m]} + QD} \cdot e(F, B) = \\ & e(G^\alpha, H^\beta) \cdot e\left(G^{\frac{\langle \mathbf{a}, \alpha \mathbf{R} + \beta \mathbf{E} + \beta \mathbf{L} + \mathbf{O} \rangle_{[l]}}{\gamma}}, H^\gamma\right) \cdot e(C, H^\delta) \cdot e(F, B) \quad \square \end{aligned}$$

Note that the first transformation  $M(F(x)) = M(G^{\langle \mathbf{a}, \mathbf{E}(x) \rangle_{[m]}}) = \mathcal{G}_M^A(\langle \mathbf{a}, \mathbf{E} \rangle)(x)$  can only be applied because  $x \in \mathcal{A}$ . For that reason this scheme is not secure. If the  $E_i$  polynomials are all 0 however, e.g. we don't use the point multiplication in the entire circuit, then we can safely pick  $x \xleftarrow{\$} \mathbb{Z}_p^*$  from the entire group and end up with a scheme similar to the original[23] which is statistically knowledge sound again.

Since we do not know how to proof soundness, we try to argue why it might be sound (disregarding the issue with the small space from  $x$  can be chosen). Lets call the exponents of the proof elements  $a, b, c, f$  where  $A = G^a, B =$

$H^b, C = G^c, F = G^f$ . We also define  $In = \frac{\langle \mathbf{a}, \alpha \mathbf{R}(x) + \beta \mathbf{E}(x) + \gamma \mathbf{L}(x) + \mathbf{O}(x) \rangle_{[l]}}{\gamma}$  as the input that will end up in the exponent a verifier will produce given the statement  $\phi = (a_1, \dots, a_l)$ . The equation for which an adversary effectively has to find a satisfying assignment is

$$a(b + \beta) + M(F) = \alpha \cdot \beta + In \cdot \gamma + c \cdot \delta + f \cdot b. \quad (20)$$

We now make the following observation:

- $a, b, c, f$  are under 100% control of the adversary. He can forge them at will.
- $In$  is depended on  $\phi = (a_1, \dots, a_l)$  and the identity  $a_0 = 1$ . An PPT adversary cannot forge  $In$  at will. He can however set it to 0 if  $a_1 = a_2 = \dots = a_l = 0 \wedge L_0 = R_0 = E_0 = O_0 = 0 \implies In = 0$ . If the identity  $a_0$  is used in the circuit e.g. at least one of the wire polynomials  $L_0, R_0, E_0, O_0$  is unequal to 0 then  $In \neq 0$ . Notice that the unlikely possibility exists, where there is  $k \in \{1, \dots, l\}$  s.t.  $L_k = L_0, R_k = R_0, E_k = E_0, O_k = O_0$ . Then setting  $a_k = |\mathbb{F}|$  pushes  $In = 0$  again.
- $M(F)$  is defined as  $M(G^f)$ . DLP on  $\mathbb{G}_1$  is assumed to be hard, therefore a PPT adversary cannot forge  $M(F)$  at will.
- $\alpha, \beta, \gamma, \delta$  are unknown to the attacker, however  $G^\alpha$  etc. are known and part of the CRS. Therefor he can set  $a, c, f \in \{\alpha, \beta, \delta\}$  and  $b \in \{\beta, \gamma, \delta\}$ .
- If there was an efficiently computable isomorphism  $\psi : \mathbb{G}_2 \rightarrow \mathbb{G}_1$ ,

We now try to reduce equation 20 to the seemingly easiest satisfiable form by setting  $A = G^\alpha, b = c = 0$ . Then it remains for an attacker to find an assignment for:

$$M(F) = In \cdot \gamma. \quad (21)$$

In case  $M$  is the trivial function that takes a curve point and returns its X-coordinate and  $G^0 = (0, 1)$ , an adversary can create a satisfying assignment by setting  $\phi = (0, \dots, 0)$ . Consequently a verifier needs to check the proof elements or the statement, and reject if among them is one of the trivial assignments. If the trivial cases are rejected, finding a valid assignment therefore always ends up facing the DLP, which is assumed to be hard.

We want to point out, that Groth's protocol [23] upon we build this work, has the same trivial satisfying assignment. Using the same tricks and argumentation leads to:

$$0 = In \cdot \gamma, \quad (22)$$

which is satisfiable if  $\phi = (0, \dots, 0), A = G^\alpha, B = H^\beta, C = G^0$ . In both schemes the trivial assignment is not an immanent threat however since  $a_0 = 1$  and its very unlikely, that the identity element is never used in any real world application circuit.

### 10.0.1 Perfekt zero-knowledge

Since the prover does not include any randomness at this point, the proof elements are not zero-knowledge. Proofing the same statement twice leads to equivalent proof elements and an adversary could learn more then the just the correctness of the statement. To make the proof elements uniformly random, we introduce a randomization gate. This gate has no connection to the rest of the circuit and is suggested to be the last gate regarding its index. The witness vector is then extended by two more elements  $a_m$  and  $a_{m-1}$  to satisfy this gate. Its structure and constraint is shown in figure 8. The prover picks  $r \xleftarrow{\$} \mathbb{Z}_p^*$  and set  $a_m = r * r + M(G^r)$  and  $a_{m-1} = r$ . The proof elements are now uniformly random.

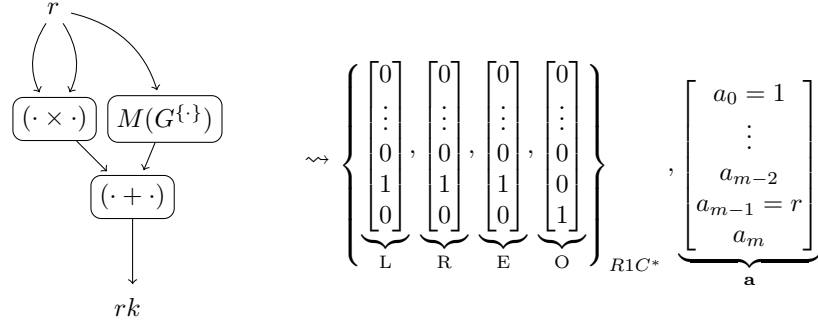
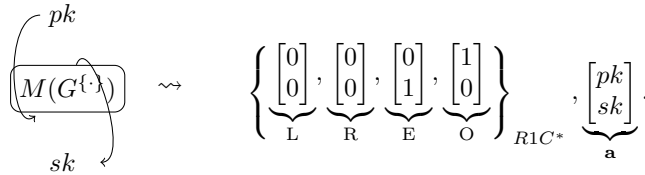


Figure 8: Shows the randomization gate and constraint. It enforces  $a_m \equiv a_{m-1} \cdot a_{m-1} + M(G^{a_{m-1}}) \bmod p$ . Since the  $r^2$  might reveal some information (gut feeling) the randomization circuit could use three different random inputs for the price of increasing  $|\mathbf{a}|$  by 4 instead of 2 elements.

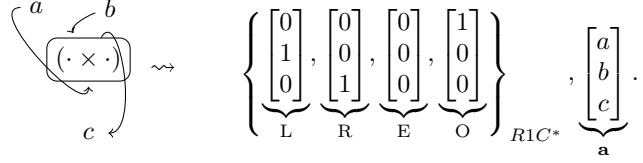
## 10.1 Example

### 10.1.1 Inverse Gates

In case a public key is passed into the circuit as an argument and we want to proof knowledge of the secret key, we immediately observe that one gate to perform this inversion is sufficient since gates are only a guarantee that the input-output relation is satisfied and therefor they can be used in either direction.

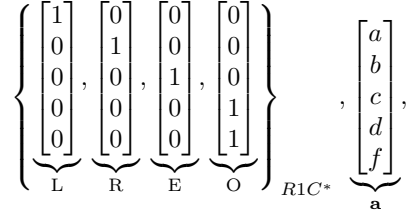


Division can be understood likewise. Consider the example where we need to compute  $c = a/b$  given  $a, b$ . The corresponding circuit and R1C\* is



### 10.1.2 Combined Gate

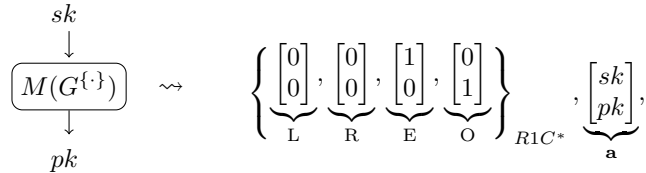
Consider an example where we want to apply multiplication and curve group multiplication at once in one gate. It can be done but it is **very important** to notice possible side effects! Let's say we have  $a, b, c$  and want to compute  $d = a * b$  and  $f = M(G^c)$ :



would be **insufficient** since all we could derive from this assignment is that  $d + f = a * b + M(G^c)$ . If we continue to use  $d$  and  $f$  we lose their determinism, however they remain entangled. If another constraint assigns a value to one of them, the other one will be fixed too. This phenomenon we call 'un-spooky action in a circuit'.

### 10.1.3 Single point multiplication circuit

Let's consider the case, where only one point multiplication defines the circuit e.g. the prover wants to convince the verifier, that he knows  $sk$  s.t.  $pk = M(G^{sk})$  without revealing  $sk$ .  $M$  maps a curve point onto its  $x$  coordinate for example. We therefore know that there must exist a second distinct  $sk$  that map to  $pk$ , however this is still sufficiently secure for this example. The circuit, its corresponding constraint and the assignment vector become



therefor the polynomials  $L_1(X) = L_2(X) = R_1(X) = R_2(X) = E_1(X) = O_2(X) = 0$ ,  $E_2(X) = O_1(X) = 1$  are simple constant lines. The domain polynomial times the extracted polynomial therefor also  $Q(X)D(X)=0$ . The statement  $\phi = (pk)$ , the witness  $w = (sk)$ .

The proof elements become:

$$\begin{aligned} A &= G^{\alpha+sk} \\ B &= H^0 \\ C &= G^{\frac{\beta sk}{\delta}} \\ F &= G^{sk} \end{aligned}$$

The verifier checks if

$$\begin{aligned} &e(A, B \cdot H^\beta) \cdot e(G, H)^{M(F)} = \\ &e(G^\alpha, H^\beta) \cdot e\left(G^{\frac{pk}{\gamma}}, H^\gamma\right) \cdot e(C, H^\delta) \cdot e(F, B) \end{aligned}$$

what is indeed true since

$$\begin{aligned} &e(A, B \cdot H^\beta) \cdot e(G, H)^{M(F)} = \\ &e(G, H)^{(\alpha+sk)\beta} \cdot e(G, H)^{pk} = \\ &e(G, H)^{\alpha\beta} \cdot e(G, H)^{\frac{pk}{\gamma} \cdot \gamma} \cdot e(G, H)^{\frac{\beta sk}{\delta} \cdot \delta} \cdot e(G, H)^{sk \cdot 0} = \\ &e(G^\alpha, H^\beta) \cdot e\left(G^{\frac{pk}{\gamma}}, H^\gamma\right) \cdot e(C, H^\delta) \cdot e(F, B) \quad \square \end{aligned}$$

From this we derive a simple cryptographic protocol: Proving knowledge of  $sk$ , s.t.  $pk = G^{sk}$ : Verifier: pick random  $\alpha \xleftarrow{\$} \mathbb{Z}_p^*$  send it to the prover. Prover: set the proof  $\pi = (A) = (G^{sk+\alpha})$  Verifier take  $\pi = (A), \alpha, pk$  and accept iff:

$$e(pk, H) \cdot e(G, H)^\alpha = e(A, H). \quad (23)$$

#### 10.1.4 Point multiplication and randomization circuit

