On Protocol Divertibility

Gerrit Bleumer

AT&T Labs-Research, Florham Park NJ 07932-0971, USA

Abstract. In this paper, we establish the notion of divertibility as a protocol property as opposed to the existing notion as a language property (see Okamoto, Ohta [OO90]). We give a definition of protocol divertibility that applies to arbitrary 2-party protocols and is compatible with Okamoto and Ohta's definition in the case of interactive zero-knowledge proofs. Other important examples falling under the new definition are blind signature protocols. A sufficient criterion for divertibility is presented and found to be satisfied by many examples of protocols in the literature. The generality of the definition is further demonstrated by examples from protocol classes that have not been considered for divertibility before. We show diverted El-Gamal encryption and diverted Diffie-Hellman key exchange.

Keywords: interactive protocol, divertibility, zero-knowledge proof, Fiat-Shamir identification, blind signature, Diffie-Hellman key-exchange, El-Gamal encryption.

1 Introduction

The idea of divertibility entered the cryptographic literature during the mid 80's by examples of identification protocols. The basic observation was that some 2-party identification protocols could be extended by placing an intermediate party—called Warden for historical reasons [S84]—between the prover and the verifier so that they cannot distinguish talking directly to each other from talking indirectly through the Warden. Since identification protocols were developed in close relation to interactive zero-knowledge proofs (ZKP), Okamoto and Ohta [OO90] (and later Desmedt and Burmester [BD91] and Ihto et al [ISS91]) established the notion of divertibility as a *language property*, i.e., a language is called divertible if it can be recognized by a diverted interactive zero-knowledge proof system.

In this paper, we establish divertibility as a *2-party protocol property*, which is orthogonal to zero knowledge or any other particular protocol property (Section 2). Informally, we call a protocol between Alice and Bob perfectly (computationally) divertible if there exists a Warden W so that Alice and Bob cannot distinguish talking to each other from talking to W when they have unlimited (polynomial) computing power. We suggest a definition of perfect divertibility that is slightly stronger than the earlier one's. The difference is illustrated by one of our new examples in Section 5 and is discussed in the Appendix. Our main result is a sufficient criterion for perfect divertibility of 2-party protocols (Section 3). We have found this criterion satisfied by many diverted zero knowledge proofs and other protocols in the literature. Our criterion is constructive in the sense that if it is shown for a given 2-party protocol, then it also gives a diverted protocol, rather than only stating that one exists. So the criterion is helpful both (i) for proving given 3-party protocols to be perfectly diverted and (ii) for designing new perfectly diverted protocols for given 2-party protocols. In Section 4, we demonstrate (i) by applying the criterion to a diverted ZKP protocol that Okamoto and Ohta used for their important result [OO90] and to an interesting blind modified El-Gamal signature by Horster et al [HMP95]. In the literature, little proof of divertedness is given for each. In Section 5, we demonstrate (ii) by showing a perfectly diverted public key encryption protocol. In addition, we show a computationally diverted key exchange protocol. Both illustrate our claim that divertibility is an independent concept rather than a special topic of zero knowledge proof theory.

2 Definitions

In order to deal with protocols of more than two parties, we generalize the notion of *interactive Turing machine* (ITM) by Goldwasser et al [GMR89]. Then we define connections of ITMs and finally give the definition of protocol divertibility.

Definition 1 ((m, n)-Interactive Turing Machine).
An (m, n)-*Interactive Turing Machine* ((m, n)-*ITM*) is a Turing machine with $m \in \mathbb{N}$ read-only *input tapes*, m write-only *output tapes*, m read-only *random tapes*, a *work tape*, a read-only *auxiliary tape*, and $n \in \mathbb{N}_0$ pairs of *communication tapes*. Each pair consists of one read-only and one write-only tape that serves for reading in-messages from or writing out-messages to another ITM. (The purpose of allowing $n = 0$ will become clear below.) The random tapes each contain an infinite stream of bits chosen uniformly at random. Read-only tapes are readable only from left to right. If the string to the right of a read-only head is empty, then we say the tape is *empty*.

Associated to an ITM is a *security parameter* $k \in \mathbb{N}$, a family $D = \{D_\pi\}_\pi$ of tuples of domains, a probabilistic *picking algorithm* $pick(k)$ and an encoding scheme S. Each member

$$D_\pi = (In_\pi^{(1)}, \ldots, In_\pi^{(m)}, Out_\pi^{(1)}, \ldots, Out_\pi^{(m)}, Rnd_\pi^{(1)}, \ldots, Rnd_\pi^{(m)},$$
$$(IM_\pi^{(1)}, OM_\pi^{(1)}), \ldots, (IM_\pi^{(n)}, OM_\pi^{(n)}))$$

of D contains one input (output, choice, in-message, out-message) domain for each of the m input (output, random) tapes and n (read-only, write-only) communication tapes. The algorithm $pick(k)$ on input some security parameter k outputs a family index π. Finally, there is a polynomial $P(k)$ so that for each π chosen by $pick(k)$, S encodes all elements of all domains in D_π as bitstrings of length $P(k)$.

ITMs proceed in rounds. During each round, an ITM first reads all its in-messages from its read-only communication tapes, then performs some computations and finally writes a message to each of its write-only communication tapes. It may write an empty string—denoted ε. If, at the beginning of a round, an ITM finds all its input tapes and all its read-only communication tapes empty, then it performs a last computation, writes empty strings to all its write-only communication tapes, writes results to all its output tapes, and then stops. The overall number of reading, writing and computation steps during an execution of an ITM is bound by a polynomial in the security parameter k.

An (m, n)-ITM is called *m-party protocol* if $n = 0$, and linear if $n \leq 2$. The *native functions* of an ITM A are defined as the family

$$nativ_\pi : \prod_{i=1}^{m} Rnd_{\pi,i} \times \prod_{i=1}^{m} In_{\pi,i} \times \prod_{j=1}^{n} IM_{\pi,j} \to \prod_{j=1}^{n} OM_{\pi,j}$$

of functions that, on input (rnd, in, im), return the respective out-messages that A would write to its write-only communication tapes would it read this data from its random, input and read-only communication tapes.

An (m_A, n)-ITM A and an (m_B, n)-ITM B ($m_A \leq m_B$) with equal input and in-message domains, respectively, are said to be *interaction equivalent*, denoted $A \equiv_i B$, iff there is a family of bundling functions[1]

$$\begin{aligned} f_\pi \ : \ & \prod_{i=1}^{m_B} Rnd_{B,\pi,i} \to \prod_{j=1}^{m_A} Rnd_{A,\pi,j} \quad \text{such that} \\ & nativ_{B,\pi}(\beta_1, \ldots, \beta_{m_B}, in_1, \ldots, in_{m_A}, i_1, \ldots, i_n) \\ = \ & nativ_{A,\pi}(f_\pi(\beta_1, \ldots, \beta_{m_B}), in_1, \ldots, in_{m_A}, i_1, \ldots, i_n). \ \diamond \end{aligned}$$

In order to enhance readability, we denote in-messages and out-messages of an ITM A by $(r-1)$-dimensional column vectors, where r is the number of rounds that A takes. (The dimension of the vectors is one less than the number of rounds because there is no message received in round 1 and no message sent in round r.) Two out-messages are written as an $(n-1, 2)$-matrix and so on. For m-party protocols P, we adopt the following interface notation:

$$(out_1, \ldots, out_m) \leftarrow P(in_1, \ldots, in_m),$$

where the left arrow indicates a probabilistic assignment. If the inputs or outputs consist of several components, we delimit them by square brackets.

Definition 2 (Connections of ITMs).
Let A be an (m_A, n_A)-ITM and B be an (m_B, n_B)-ITM with equal picking algorithm *pick*. Then a connection $C = \langle A, B \rangle$ is any ITM consisting of A and B sharing $c \leq \min\{n_A, n_B\}$ pairs of their communication tapes. The picking algorithm of C is *pick*, and the domains of C are defined as the cartesian products of the respective domains of A and B. $\diamond$

Obviously, the linear connection operator $\langle \bullet, \bullet \rangle$ is associative and we can therefore omit brackets in the usual way:

$$\langle A, B, C \rangle \stackrel{\text{def}}{=} \langle \langle A, B \rangle, C \rangle = \langle A, \langle B, C \rangle \rangle.$$

[1] i.e., each image of f_π has equally many preimages.

All connections we consider in the following are linear and have a small constant number of rounds.

Definition 3 (Divertibility of Protocols).
Let $P = \langle A, B\rangle$ be a two-party protocol with interface $P([y, x_A]^A, [y, x_B]^B)$ and input domains $In_\pi = (Y_\pi \times X_{A,\pi}) \times (Y_\pi \times X_{B,\pi})$. Common inputs y are taken from Y_π, whereas private inputs x_A, x_B are taken from $X_{A,\pi}$ and $X_{B,\pi}$, respectively. The product domain of private inputs is denoted $X_\pi = X_{A,\pi} \times X_{B,\pi}$. Furthermore, let $R = \{R_\pi\}_\pi$ be a family of relations $R_\pi \subseteq Y_\pi \times X_\pi$.

The protocol P is called *perfectly (computationally) divertible* iff a (1,2)-ITM W exists such that the following properties are met:

Invariance: Connecting A (B) to W is interaction equivalent to A (B):

$$\langle A, W\rangle \equiv_i A \quad \text{and} \quad \langle W, B\rangle \equiv_i B.$$

Perfect (Computational) Indistinguishability: For all polynomial-time actively adversary ITMs $\tilde{A}$, $\tilde{B}$, for all indices π, all common and private inputs $(y, (x_A, x_B)) \in R_\pi$ and all polynomial size strings q representing shared a-priori knowledge of $\tilde{A}$ and $\tilde{B}$, the ensembles of simultaneous views of $\tilde{A}$ and $\tilde{B}$ upon W and of their views upon honest B and A, i.e.,

$$view_W^{(\tilde{A},\tilde{B})}\langle \tilde{A}, W, \tilde{B}\rangle([y, x_A, q]^{\tilde{A}}, [y]^W, [y, x_B, q]^{\tilde{B}}) \quad \text{and}$$
$$(view_B^{(\tilde{A})}\langle \tilde{A}, B\rangle([y, x_A, q]^{\tilde{A}}, [y, x_B]^B), view_A^{(\tilde{B})}\langle A, \tilde{B}\rangle([y, x_A]^A, [y, x_B, q]^{\tilde{B}}))$$

are equal (polynomially indistinguishable).[2]

An ITM W that satisfies invariance and perfect (computational) indistinguishability is said to *perfectly (computationally) divert* protocol P. $\diamond$

Divertibility as defined by Okamoto, Ohta [OO90] and almost equivalently by Itoh et al [ISS91] has been introduced as a *language property.* A language L is called divertible, if there exists a diverted zero knowledge proof system for proving membership in L. In contrast, we define divertibility as a *2-party protocol property.* The main difference between the two definitions is that we ask for a concrete protocol P to be divertible, whereas they ask for existence of a divertible protocol meeting a certain specification S (namely to be a zero-knowledge proof). Consequently, Definition 3 (Invariance) relates the two interfaces of the diverted protocol P' to the interface of the given protocol P, where their definition relates them to S. Another difference is, that we suggest a stronger definition than Okamoto and Ohtas. We require Indistinguishability even for two attackers $\tilde{A}$ and $\tilde{B}$ who *know of each other* and who therefore know which of their views result from the same protocol instance. We illustrate this by an example in Section 5 and discuss it further in Appendix 7.1).

[2] By $view_B^{(A)}P$, we denote the *view* of A on B in a protocol P. This notion as well as that of *polynomial indistinguishability* of families of random variables is defined, e.g., by Goldwasser, Micali and Rackoff [GMR89].

An immediate consequence of the definition is that if a protocol P is divertible, then we can insert second and third wardens and we, again, obtain a diverted protocol.

3 Main Result

Theorem 4 (Criterion for Perfect Divertibility).
Let $P = \langle A, B\rangle$ be a two-party protocol with interface $P([y, x_A]^A, [y, x_B]^B)$. Let the input domains be $(Y_\pi \times X_{A,\pi}) \times (Y_\pi \times X_{B,\pi})$, the random domains be $Rnd_{A,\pi} \times Rnd_{B,\pi}$, the out-message domains be $OM_{A,\pi} \times OM_{B,\pi}$, and let the native functions of A and B be

$$\begin{aligned} nativ_{A,\pi} &: Rnd_{A,\pi} \times Y_\pi \times OM_{B,\pi} \times X_{A_\pi} \to OM_{A,\pi}, \\ nativ_{B,\pi} &: Rnd_{B,\pi} \times Y_\pi \times OM_{A,\pi} \times X_{B,\pi} \to OM_{B,\pi}. \end{aligned}$$

Furthermore, let $R = \{R_\pi\}_\pi$ be a family of relations $R_\pi \subseteq (Y_\pi \times X_{A,\pi} \times X_{B,\pi})$, which capture a possible correspondence between common and private inputs.

Then P is perfectly divertible if only there exist:

(i) a family $(Rnd_\pi, \odot, 1)$ of (not necessarily commutative) groups, and
(ii) three families of functions

$$\begin{aligned} base_\pi &: Y_\pi \times X_{A,\pi} \times X_{B,\pi} \to OM_{A,\pi} \times OM_{B,\pi}, \\ join_\pi &: Rnd_{A,\pi} \times Rnd_{B,\pi} \times Y_\pi \times X_{A,\pi} \times X_{B,\pi} \to Rnd_\pi \\ divrt_\pi &: Rnd_\pi \times Y_\pi \times OM_{A,\pi} \times OM_{B,\pi} \to OM_{A,\pi} \times OM_{B,\pi}, \end{aligned}$$

where for each y, x_A, x_B, $join_\pi(\alpha, \beta, y, x_A, x_B)$ is injective as a function of (α, β). So we may write: $join_\pi^{-1}(join_\pi(\alpha, \beta, y, x_A, x_B), y, x_A, x_B) = (\alpha, \beta)$,

that satisfy the following three conditions:

For every π, for all random choices $\alpha \in Rnd_{A,\pi}, \beta \in Rnd_{B,\pi}$, all common and corresponding private inputs $(y, (x_A, x_B)) \in R_\pi$, and all out-messages $o_A \in OM_{A,\pi}$, $o_B \in OM_{B,\pi}$ (the index π is omitted in the following):

DECOMPOSITION:

$$\begin{aligned} &(nativ_A(\alpha, y, o_B, x_A), nativ_B(\beta, y, o_A, x_B)) \\ &= divrt(join(\alpha, \beta, y, x_A, x_B), y, base(y, x_A, x_B)), \end{aligned}$$

GROUND:

$$\omega = 1 \quad \Leftrightarrow \quad divrt(\omega, y, (o_A, o_B)) = (o_A, o_B),$$

MIXED ASSOCIATIVITY:

$$divrt(\omega', y, divrt(\omega, y, (o_A, o_B))) = divrt(\omega \odot \omega', y, (o_A, o_B)).$$

If the above conditions are met, then a perfectly diverting ITM W can be constructed as follows: Choose $\omega \in_R Rnd_\pi$, and then, given the out-messages o_A, o_B written by A and B to W, determine the outmessages o'_A, o'_B written by W to A and B by applying $(o'_A, o_B) = divrt(\omega, y, (o_A, o'_B))$. ◇

Proof. First observe that if *divrt* satisfies the premises Ground and Mixed Associativity, then it is injective as a function of ω: Assume two choices $\omega_1, \omega_2 \in Rnd_\pi$ with equal images, i.e.,

$$divrt(\omega_1, y, (o_A, o_B)) = divrt(\omega_2, y, (o_A, o_B)). \tag{1}$$

Now let $\omega \stackrel{\text{def}}{=} \omega_1^{-1} \odot \omega_2$ so that ω_2 splits into $\omega_1 \odot \omega$. Then we rewrite the right hand side of (1) using Mixed Associativity:

$$divrt(\omega_1, y, (o_A, o_B)) = divrt(\omega, y, divrt(\omega_1, y, (o_A, o_B)).$$

From the Ground premise, we conclude that $\omega = 1$, and so $\omega_1 = \omega_2$.

Now, consider a protocol $P = \langle A, W, B\rangle$ that satisfies Theorem 4. Then we can derive Invariance (Definiton 3) of P as follows:

$$\begin{aligned}
&(nativ_{\langle A,W\rangle}(\omega, \alpha, y, o_B, x_A), nativ_B(\beta, y, o_A, x_B))\\
&= divrt(\omega, y, (nativ_A(\alpha, y, o_B, x_A), nativ_B(\beta, y, o_A, x_B)))\\
&= divrt(\omega, y, divrt(\omega', y, base(y, x_A, x_B))), \quad \text{where } \omega' \stackrel{\text{def}}{=} join(\alpha, \beta, y, x_A, x_B)\\
&= divrt(\omega' \odot \omega, y, base(y, x_A, x_B))\\
&= divrt(\omega^{-1}, y, divrt(\omega' \odot \omega \odot \omega^{-1}, y, base(y, x_A, x_B)))\\
&= divrt(\omega^{-1}, y, divrt(\omega', y, base(y, x_A, x_B)))\\
&= divrt^{-1}(\omega, y, (nativ_A(\alpha, y, o_B, x_A), nativ_B(\beta, y, o_A, x_B)))\\
&= (nativ_A(\alpha, y, o_B, x_A), nativ_{\langle W,B\rangle}(\omega, \beta, y, o_A, x_B)).
\end{aligned}$$

Note that the maps $(\omega, \alpha) \mapsto \alpha$ and $(\omega, \beta) \mapsto \beta$ are each bundling by definition.

In order to show perfect divertibility, we can show that for every π, every common input and corresponding private input $(y, (x_A, x_B)) \in R_\pi$, and every two out-messages $o_A, o'_A \in OM_A$ and $o_B, o'_B \in OM_B$, there is at most one choice ω such that

$$divrt(\omega, y, (o_A, o_B)) = (o'_A, o'_B).$$

This is immediate from the fact that *divrt* as a function of ω is injective. □

4 Known Examples of Diverted Protocols

The most prominent examples of diverted protocols in the literature are diverted interactive proofs and blind signatures. Since divertibility has been introduced only in the former context, blind signatures are a good example to illustrate the more general concept of divertibility of protocols as proposed in Definition 3. In this Section, we investigate two examples in more detail: (i) the diverted ZKP that Okamoto and Ohta used to prove their main theorem [OO90] and (ii) a blind modified El-Gamal Signature, which was presented by Horster, Michels and Petersen [HMP95] who built on ideas of Camenisch, Piveteau and Stadler [CPS95]. More examples, all satisfying the divertibility criterion, can be found in Appendix 7.2.

Since all the following protocols are based on the intractability of computing discrete logarithms, the following definitions are useful. Let p be a k-bit prime ($k \in \mathbb{N}$), q be a large prime diviser of $p-1$ and G_q be the unique (multiplicative) subgroup of order q in $\mathbb{Z}_p^*$. Furthermore, $g \neq 1$ denotes a randomly chosen element of G_q. (The restriction to $g \neq 1$ asserts that g generates G_q). Arithmetic operations are either in G_q, i.e., multiplication mod p or in $\mathbb{Z}_q$, i.e., addition and multiplication mod q. We dare to omit the "(mod p)" and "(mod q)" whenever they are clear from the context.

4.1 Okamoto-Ohta ZKP

In their seminal paper [OO90] (Theorem 1, p138), Okamoto and Ohta used a diverted zero knowledge proof protocol in order to prove that any commutative random self-reducible language has a diverted perfect zero knowledge proof. We reconsider their diverted protocol and show that it satisfies our divertibility criterion. A side-effect of this analysis is a tigthtening of their result: In fact, every commutative random self-reducible relation has a *perfectly* diverted perfect zero-knowledge proof.

The diverted proof protocol is restated in Figure 1. Only the prover has a private input, namely her secret x, whereas the common input is y. Note, that in their protocol, Okamoto and Ohta named the secret y and the common input x! In order to keep a somewhat harmonised presentation throughout this paper, we instantiate their meta-protocol for discrete logarithms. This is what they singled out as their example E2. With this choice, their bullet and power operation, i.e. "$\bullet$" and "a^b" in $\mathbb{Z}_q$, instantiate to addition and multiplication mod q. Although we show our result for a particular instantiation, it generalizes to all commutative random self-reducible relations.

Proposition 5. *The Warden in protocol DOO perfectly diverts the 2-party protocol between Alice and Bob.* ◇

Proof. The native functions of Alice, Bob and Warden follow immediately from protocol *DOO* in Figure 1. From the Warden's native function, we deduce the functions *divrt* by expressing the out-messages of Warden to Bob and back as a function of the out-messages of Alice to Warden and back (and, of course, the common input and choices of Warden). Furthermore, we choose the injective functions *join* and the functions *base* as follows:

$$nativ_A(r, y, o_B, x) \stackrel{\text{def}}{=} \begin{pmatrix} g^r y \\ \varepsilon \\ r + x^{o_{B2}} \end{pmatrix} \quad , \quad nativ_B(\beta, o_A) \stackrel{\text{def}}{=} \begin{pmatrix} \varepsilon \\ \beta \\ \varepsilon \end{pmatrix} ,$$

$$([accept]^B) \leftarrow DOO([y,x]^A, [y]^W, [y]^B)$$

	Alice		**Warden**		**Bob**
(1)	$r \in_R \mathbb{Z}_q$		$e \in_R \{0,1\}, u \in_R \mathbb{Z}_q$		
(2)	$x' \leftarrow g^r y$				
		$\xrightarrow{x'}$			
(3)			$x'' \leftarrow g^u y^e x'^{1-2e}$		
				$\xrightarrow{x''}$	
(4)					$\beta \in_R \{0,1\}$
				$\xleftarrow{\beta}$	
(5)			$\beta' \leftarrow \beta \oplus e$		
		$\xleftarrow{\beta'}$			
(6)	$z \leftarrow r + x^{\beta'}$				
		$\xrightarrow{z}$			
(7)			$z' \leftarrow u + z(1-2e)$		
				$\xrightarrow{z'}$	
(8)					$accept \leftarrow$ $x'' \stackrel{?}{=} g^{z'} y^{1-\beta}$

Fig. 1. Diverted Okamoto-Ohta ZKP

$$base(y,x) \stackrel{\text{def}}{=} \begin{pmatrix} y^{\frac{1}{2}} & \varepsilon \\ \varepsilon & 0 \\ -\frac{x}{2} & \varepsilon \end{pmatrix} \quad , \quad join(\beta, r, x) \stackrel{\text{def}}{=} (\beta, r + \tfrac{x}{2}),$$

$$divrt((e,u), y, (o_A, o_B)) \stackrel{\text{def}}{=} \begin{pmatrix} g^u y^e o_{A1}^{1-2e} & o_{B1} \\ o_{A2} & o_{B2} \oplus e \\ u + o_{A3}(1-2e) & o_{B3} \end{pmatrix}.$$

Furthermore, we define the non-Abelian group $(\{0,1\} \times \mathbb{Z}_q, \odot, (0,0))$, where

$$(e_1, u_1) \odot (e_2, u_2) \stackrel{\text{def}}{=} (e_1 \oplus e_2, u_1(1 - 2e_2) + u_2).$$

Associativity of the operation $\odot$ is not trivial yet immediate from the definition. The inverse of (e,u) is $(e, u(2e-1))$. (Here, we make use of the fact that for all $q > 1$ the following equation holds: $e_1 + e_2 - 2e_1e_2 \bmod q = e_1 \oplus e_2$. Now, the Ground premise of Theorem 4 is immediately satisfied, and the other two premises are checked as follows:

Decomposition:

$$(nativ_A(r, y, o_B, x), nativ_B(\beta, o_A)) = \begin{pmatrix} g^r y & \varepsilon \\ \varepsilon & \beta \\ r + x^\beta & \varepsilon \end{pmatrix}$$
$$= divrt(join(\beta, r, x), base(y, x)).$$

Mixed Associativity:

$$\begin{aligned}
&\mathit{divrt}((e_2,u_2),y,\mathit{divrt}((e_1,u_1),y,(o_A,o_B)))\\
&=\mathit{divrt}((e_2,u_2),\begin{pmatrix} g^{u_1}y^{e_1}o_{A1}^{1-2e_1} & o_{B1}\\ o_{A2} & o_{B2}\oplus e_1\\ u_1+o_{A3}(1-2e_1) & o_{B3}\end{pmatrix})\\
&=\begin{pmatrix} g^{u_2}y^{e_2}(g^{u_1}y^{e_1}o_{A1}^{1-2e_1})^{1-2e_2} & o_{B1}\\ o_{A2} & o_{B2}\oplus e_1\oplus e_2\\ u_2+(u_1+o_{A3}(1-2e_1))(1-2e_2) & o_{B3}\end{pmatrix}\\
&=\begin{pmatrix} g^{u_1(1-2e_2)+u_2}y^{e_1\oplus e_2}o_{A1}^{1-2(e_1\oplus e_2)} & o_{B1}\\ o_{A2} & o_{B2}\oplus e_1\oplus e_2\\ u_1(1-2e_2)+u_2+o_{A3}(1-2(e_1\oplus e_2)) & o_{B3}\end{pmatrix}\\
&=\mathit{divrt}((e_1\oplus e_2,u_1(1-2e_2)+u_2),y,(o_A,o_B))\\
&=\mathit{divrt}((e_1,u_1)\odot(e_2,u_2),(o_A,o_B)).
\end{aligned}$$

Hence, the claim follows from Theorem 4. □

4.2 Modified El-Gamal Signature

The blind version of a modified El-Gamal signature protocol according to Horster, Michels and Petersen [HMP95] is restated in Figure 2. Here, both the signer and the recipient have a private input, namely her private signing key x and his message m to be signed, respectively. The common input is the public verification key y corresponding to x.

This protocol—as it is—cannot be divertible: Alice could distinguish talking directly to Bob from talking to a Warden just by checking if the challenge c she receives before step 6 is the hash value $h(m,z,a,b)$ of the message she sends after step 3. Therefore, we consider the idealized protocol *DMEG∗*, where Bob really chooses his challenge at random.[3]

Proposition 6. *The Warden of protocol DMEG∗ perfectly diverts the underlying 2-party protocol of Alice and Bob.* ◇

Proof. The native functions of Alice, Bob and Warden follow immediately from protocol *DMEG∗*. The functions *divrt* are deduced analogously as in Section4.1. Here, we choose the injective functions *join* and the functions *base* as follows:

$$\mathit{nativ}_A(\alpha,o_B,x)\stackrel{\text{def}}{=}\begin{pmatrix} g^\alpha\\ \varepsilon\\ g^\alpha x+\alpha o_{B2}\end{pmatrix}\quad,\quad \mathit{nativ}_B(\beta)\stackrel{\text{def}}{=}\begin{pmatrix}\varepsilon\\ \beta\\ \varepsilon\end{pmatrix},$$

[3] We note, that *DMEG∗* is not a signature protocol because Bob could not convince a third party that he instead of Alice herself has chosen the challenge. Also note that the message m no longer occurs as input to *DMEG∗* because m occurs in *DMEG* only as an input to the hash function h in step (4).

$$([a', b']^B) \leftarrow DMEG([y, x]^A, []^W, [y, m]^B)$$

	Alice		**Warden**		**Bob**
(1)	$\alpha \in_R \mathbb{Z}_q$		$\omega \in_R \mathbb{Z}_q^*, \phi \in_R \mathbb{Z}_q$		
(2)	$a \leftarrow g^\alpha$				
		$\xrightarrow{a}$			
(3)			$a' \leftarrow a^\omega g^\phi$		
				$\xrightarrow{a'}$	
(4)					$c' \leftarrow h(m, a')$
				$\xleftarrow{c'}$	
(5)			$c \leftarrow \frac{\omega a}{a'} c' \bmod q$		
		$\xleftarrow{c}$			
(6)	$b \leftarrow ax + \alpha c$				
		$\xrightarrow{b}$			
(7)			$b' \leftarrow \frac{a'}{a} b + c'\phi$		
				$\xrightarrow{b'}$	
(8)					accept iff $g^{b'} \stackrel{?}{=} y^{a'} a'^{c'}$

Fig. 2. Diverted Modified El-Gamal Signature

$$base(x) \stackrel{\text{def}}{=} \begin{pmatrix} 1 & \varepsilon \\ \varepsilon & 1 \\ x & \varepsilon \end{pmatrix} \quad , \quad join(\alpha, \beta) \stackrel{\text{def}}{=} (\beta^{-1} g^\alpha, \alpha),$$

$$divrt((\omega, \phi), (o_A, o_B)) \stackrel{\text{def}}{=} \begin{pmatrix} o_{A1}^\omega g^\phi & o_{B1} \\ o_{A2} & \frac{1}{\omega} o_{B2} o_{A1}^{\omega-1} g^\phi \\ (o_{A3} + \frac{\phi}{\omega} o_{B2}) o_{A1}^{\omega-1} g^\phi & o_{B3} \end{pmatrix}.$$

Furthermore, we define the non-Abelian group $(\mathbb{Z}_q^* \times \mathbb{Z}_q, \odot, (1, 0))$, where

$$(\omega_1, \phi_1) \odot (\omega_2, \phi_2) \stackrel{\text{def}}{=} (\omega_1 \omega_2, \phi_1 \omega_2 + \phi_2).$$

Associativity of operation $\odot$ is again immediate from the definition, and the inverses are as follows: $(\omega, \phi)^{-1} = (\omega^{-1}, -\phi\omega^{-1})$. Now, the Ground premise of Theorem 4 is immediately satisfied, and the other two premises are checked as follows:

Decomposition:

$$(nativ_A(\alpha, o_B, x), nativ_B(\beta)) = \begin{pmatrix} g^\alpha & \varepsilon \\ \varepsilon & \beta \\ g^\alpha x + \alpha\beta & \varepsilon \end{pmatrix} = divrt(join(\alpha, \beta), base(x))$$

Mixed Associativity:

$$divrt((\omega_2, \phi_2), divrt((\omega_1, \phi_1), (o_A, o_B)))$$

$$
\begin{aligned}
&= divrt((\omega_2,\phi_2), \begin{pmatrix} o_{A1}^{\omega_1} g^{\phi_1} & o_{B1} \\ o_{A2} & \frac{1}{\omega_1} o_{B2} o_{A1}^{\omega_1 - 1} g^{\phi_1} \\ (o_{A3} + \frac{\phi_1}{\omega_1} o_{B2}) o_{A1}^{\omega_1 - 1} g^{\phi_1} & o_{B3} \end{pmatrix})\\
&= \begin{pmatrix} (o_{A1}^{\omega_1} g^{\phi_1})^{\omega_2} g^{\phi_2} & o_{B1} \\ o_{A2} & \begin{array}{c} \frac{1}{\omega_2}(\frac{1}{\omega_1} o_{B2} o_{A1}^{\omega_1-1} g^{\phi_1}) \\ \cdot (o_{A1}^{\omega_1} g^{\phi_1})^{\omega_2 - 1} g^{\phi_2} \end{array} \\ \begin{pmatrix} (o_{A3} + \frac{\phi_1}{\omega_1} o_{B2}) o_{A1}^{\omega_1 - 1} g^{\phi_1} \\ + \frac{\phi_2}{\omega_2}(\frac{1}{\omega_1} o_{B2} o_{A1}^{\omega_1-1} g^{\phi_1}) \end{pmatrix} (o_{A1}^{\omega_1} g^{\phi_1})^{\omega_2 - 1} g^{\phi_2} & o_{B3} \end{pmatrix}\\
&= \begin{pmatrix} o_{A1}^{\omega_1\omega_2} g^{\phi_1\omega_2 + \phi_2} & o_{B1} \\ o_{A2} & \frac{1}{\omega_1\omega_2} o_{B2} o_{A1}^{\omega_1\omega_2 - 1} g^{\phi_1\omega_2+\phi_2} \\ (o_{A3} + (\frac{\phi_1}{\omega_1} + \frac{\phi_2}{\omega_1\omega_2} o_{B2}) o_{A1}^{\omega_1\omega_2 - 1} g^{\phi_1\omega_2+\phi_2} & o_{B3} \end{pmatrix}\\
&= divrt((\omega_1\omega_2, \phi_1\omega_2 + \phi_2), (o_A, o_B))\\
&= divrt((\omega_1, \phi_1) \odot (\omega_2, \phi_2), (o_A, o_B)).
\end{aligned}
$$

Hence, the claim follows from Theorem 4. □

5 New Examples

5.1 El-Gamal Encryption

A diverted version of the El-Gamal encryption protocol [EG85] is suggested in Figure 3. Here, Alice's private input is the message m to be encrypted, Bob's private input x is the decryption key, and the common input y is the public encryption key.

$([m']^B) \leftarrow DEE([y, m]^A, []^W, [x]^B)$

	Alice		Warden		Bob
(1)	$\alpha \in_R \mathbb{Z}_q$		$\omega, \phi \in_R \mathbb{Z}_q$		
(2)	$a \leftarrow (g^\alpha, my^\alpha)$	$\xrightarrow{a}$			
(3)			$a' \leftarrow (a_1^\omega, \phi a_2^\omega)$	$\xrightarrow{a'}$	
(4)					$m' \leftarrow \frac{a'_2}{a_1'^x}$

Fig. 3. Diverted El-Gamal Encryption

Proposition 7. *The Warden of protocol DEE perfectly diverts the 2-party protocol between Alice and Bob.* ◇

A proof by applying our divertibility criterion is given in Appendix 7.3.

5.2 Diffie-Hellman Key Exchange

A diverted version of the Diffie-Hellman key-exchange protocol [DH76] is suggested in Figure 4. Neither Alice nor Bob have a private input, nor is there any common input.

$$([a'']^A, [b'']^B) \leftarrow DDH([]^A, []^W, []^B)$$

$$\begin{array}{lcccccc}
 & \textbf{Alice} & & \textbf{Warden} & & \textbf{Bob} \\
 & \alpha \in_R \mathbb{Z}_q & & \omega \in_R \mathbb{Z}_q & & \beta \in_R \mathbb{Z}_q \\
(1) & a \leftarrow g^\alpha & & & & b \leftarrow g^\beta \\
 & & \xrightarrow{a} & & \xleftarrow{b} & \\
(2) & & & (a', b') \leftarrow (a^\omega, b^\omega) & & \\
 & & \xleftarrow{b'} & & \xrightarrow{a'} & \\
(3) & a'' \leftarrow b'^\alpha & & & & b'' \leftarrow a'^\beta
\end{array}$$

Fig. 4. Diverted Diffie-Hellman Key Exchange Protocol

Proposition 8. *The warden of protocol DDH computationally diverts the Diffie-Hellman protocol between Alice and Bob.* $\diamond$

Proof (Sketch). If for given (a, b), an attacker could distinguish valid from invalid diverted out-messages (a', b') with non-negligible probability, i.e., probability $\geq \frac{1}{P(k)}$ for some polynomial P, then he had broken the simultaneous discrete log assumption [CEG88]. $\square$

6 Conclusions and Open Questions

We have introduced the notion of perfect and computational protocol divertibility, and have given a sufficient criterion for the former. All diverted protocols we have found in the literature (see Appendix 7.2) turned out to satisfy this criterion. Examples of a public key encryption and a key distribution protocol have been shown to be divertible under the new definition. The latter is the first computationally divertible protocol we know of. Interesting open questions remain: (i) Is the divertibility criterion necessary? (ii) Is there an analogous criterion for computational divertibility? (iii) Are there (at least computationally) diverted protocols whose diverting function is significantly less complex than one of the native functions? (iv) What applications are there for diverted key-exchange and encryption protocols?

References

[B93] Stefan Brands: An Efficient Off-line Electronic Cash System Based On The Representation Problem; Centrum voor Wiskunde en Informatica, Computer Science/Departement of Algorithmics and Architecture, Report CS-R9323, March 1993.

[B94] Stefan Brands: Untraceable Off-line Cash in Wallet with Observers; Crypto '93, LNCS 773, Springer-Verlag, Berlin 1994 302-318.

[BD91] Mike V. D. Burmester, Yvo Desmedt: All languages in NP have divertible zero-knowledge proofs and argurments under cryptographic assumptions; Eurocrypt '90, LNCS 473, Springer-Verlag, Berlin 1991, 1-10.

[C85] David Chaum: Security without Identification: Transaction Systems to make Big Brother Obsolete; Communications of the ACM 28/10 (1985) 1030-1044.

[C94] Lidong Chen: Witness Hiding Proofs and Applications; DAIMI PB-477, Computer Science Department Aarhus University, August 1994.

[CEG88] David Chaum, Jan.-Hendrik Evertse, Jeroen van de Graaf: An improved protocol for demonstrating possession of discrete logarithms and some generalizations; Eurocrypt '87, LNCS 304, Springer-Verlag, Berlin 1988, 127-141.

[CP92] David Chaum, Torben Pryds Pedersen: Wallet Databases with Observers. Crypto '92, LNCS 740, Springer Verlag, Berlin 1993, 89-105.

[CPS95] Jan L. Camenisch, Jean-Marc Piveteau, Markus A. Stadler: Blind Signatures Based on the Discrete Logarithm Problem; Eurocrypt '94, LNCS 950, Springer-Verlag, Berlin 1995, 428-432.

[DGB88] Yvo Desmedt, Claude Goutier, Samy Bengio: Special uses and abuses of the Fiat-Shamir passport protocol; Crypto '87, LNCS 293, Springer-Verlag, Berlin 1988, 21-39.

[DH76] Whitfield Diffie, Martin E. Hellman: New Directions in Cryptography; IEEE Transactions on Information Theory 22/6 (1976) 644-654.

[EG85] Taher ElGamal: A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms; IEEE Transactions on Information Theory 31/4 (1985) 469-472.

[GMR89] Shafi Goldwasser, Silvio Micali, Charles Rackoff: The Knowledge Complexity of Interactive Proof Systems; SIAM Journal on Computing 18/1 (1989) 186-207.

[HMP95] Patrick Horster, Markus Michels, Holger Petersen: Meta-Message Recovery and Meta-Blind Signature Schemes Based on the Discrete Logarithm Problem and Their Applications; Asiacrypt '94, LNCS 917, Springer-Verlag, Berlin 1995, 224-237.

[ISS91] Toshija Itoh, Kouichi Sakurai, Hiroki Shizuya: Any Language in IP has a Divertible ZKIP, AsiaCrypt '91, Springer-Verlag, Berlin 1993.

[OO90] Tatsuaki Okamoto, Kazuo Ohta: Divertible zero-knowledge interactive proofs and commutative random self-reducibility. Eurocrypt '89, LNCS 434, Springer-Verlag, Berlin 1990, 134-149.

[S84] Gustavus J. Simmons: The Prisoners' Problem and the Subliminal Channel. Crypto '83, Plenum Press, New York 1984, 51-67.

[S91] Claus P. Schnorr: Efficient Signature Generation by Smart Cards; Journal of Cryptology 4/3 (1991) 161-174.

7 Appendix

7.1 Why the Previous Definition of Divertibility is a Little too Weak

The previous definition of divertibility by Okamoto and Ohta [OO90], and by Itoh et al [ISS91] as well, requires that two attackers $\tilde{A}, \tilde{B}$ who on the one hand form a linear 3-party protocol P' with an intermediate Warden and on the other hand form 2-party protocols $P_{\tilde{A}}$ with an honest B and $P_{\tilde{B}}$ with an honest A cannot distinguish their views in $\langle \tilde{A}, B\rangle$ and $\langle A, \tilde{B}\rangle$ from those in *separate* instances of $\langle \tilde{A}, W, \tilde{B}\rangle$. More formally, they require indistinguishability of the two ensembles (protocol inputs exactly as in Definiton 3) before:

$$(\mathit{view}_W^{(\tilde{A})}\langle \tilde{A}, W, \tilde{B}\rangle, \mathit{view}_W^{(\tilde{B})}\langle \tilde{A}, W, \tilde{B}\rangle) \tag{2}$$

$$\text{and} \quad (\mathit{view}_B^{(\tilde{A})}\langle \tilde{A}, B\rangle, \mathit{view}_A^{(\tilde{B})}\langle A, \tilde{B}\rangle). \tag{3}$$

However, the attacker model that seems to underly the literature on divertibility is stronger than expressed by the above requirement. The attackers A and B are considered to know when they engage in a protocol with the Warden and so they know which of their views result from the same protocol instances.

A good example to illustrate this difference is protocol *DDH* in Section 5.2. The two ensembles according to (2) and (3) above are equal and thus protocol DDH would have to be regarded as perfectly diverted. This is counterintuitive because the warden in DDH uses less random coins than Alice and Bob together. On the other hand, according to Definition 3, DDH is only computionally diverted, which is the most we would expect.

7.2 More Known Examples

In the following, we list more known examples of perfectly diverted protocols from the cryptographic literature. All of them satisfy the divertibility criterion in Theorem 4, but due to the page limit we do not unfold them completely. We only show how the group must be chosen in order to apply the divertibility criterion. We use precisely the variable names from the original presentations. Where present, $k \in \mathbb{N}$ denotes some global system parameter—not the security parameter.

A simple early example of a perfectly diverted signature protocol was presented by Chaum [C85]. The group can be chosen to be $(\mathbb{Z}_N^*, 1)$, where $N = pq$ is some RSA modulus.

Also well before the formal definition of divertibility [OO90] appeared, Desmedt, Goutier and Bengio [DGB88] suggested a perfectly diverted interactive proof of knowledge of square roots modulo a composite $N = pq$, where the prime factors p, q are known only to the prover (Fiat-Shamir identification protocol). They provided no proof of divertibility, but the criterion applies if the group is chosen as follows: $(\{0,1\}^k, \mathbb{Z}_n, \odot, (0,1))$, where $y \in \mathbb{Z}_n$ and:

$$(\boldsymbol{f}_1, r_1) \odot (\boldsymbol{f}_2, r_2) = (\boldsymbol{f}_1 \oplus \boldsymbol{f}_2, r_1 r_2 \textstyle\prod_{i=1}^k y_i^{-f_{1i} f_{2i}}),$$

In her thesis [C94], Sect. 3.4.1, Chen showed an alternative and slightly more flexible way of diverting the protocol of Okamoto and Ohta (Section 4.1). The divertibility criterion applies to her protocol as well.

In [B93], Sect.16.1, Brands gave a diverted interactive proof of knowledge of discrete representations, which built on previous work of Chaum, Evertse, van de Graaf [CEG88]. For a given tuple of generators $(g_1, \ldots, g_k)$ and a residue $y \in \mathbb{Z}_p$, the prover demonstrates knowledge of a discrete representation $(x_1, \ldots, x_k)$ of $y = \prod_{i=1}^{k} g_i^{x_i}$. We can choose the following non-commutative group: $(G_q^k \times G_q^k \times \mathbb{Z}_q, \odot, (0, 0, 0))$, where the operation is:

$$(\boldsymbol{x}'_1, \boldsymbol{w}_1, d_1) \odot (\boldsymbol{x}'_2, \boldsymbol{w}_2, d_2) = (\boldsymbol{x}'_1 + \boldsymbol{x}'_2, \boldsymbol{w}_1 + \boldsymbol{w}_2 + d_2 \boldsymbol{x}'_1, d_1 + d_2).$$

In [B93,B94], Brands has also proposed a (restrictive) blind signature protocol for an untraceable electronic coin system. The only difference between the signature protocols is that in [B93], Sect. 11.2, the message is taken to be $m = Id$, whereas in [B94] it is $m = Ig_2$, where I is the account holder's identity and d, g_2 are global constants. Here, we can choose the non-commutative group $(\mathbb{Z}_q^{*2} \times \mathbb{Z}_q, \odot, (0, 1, 0))$ with

$$(s_1, u_1, v_1) \odot (s_2, u_2, v_2) = (s_1 s_2, u_1 u_2, v_1 u_2 + v_2).$$

7.3 Proof for Diverted El-Gamal Encryption

Proof. The native functions of Alice, Bob and Warden follow immediately from the protocol DEE in Figure 3:

$$\begin{aligned} \mathit{nativ}_A(\alpha, y, m) &\stackrel{\text{def}}{=} (g^\alpha, my^\alpha)\ , \quad \mathit{nativ}_B() \stackrel{\text{def}}{=} \varepsilon, \\ \mathit{base}(y) &\stackrel{\text{def}}{=} (g, y) \qquad\quad , \mathit{join}(\alpha, m) \stackrel{\text{def}}{=} (\alpha, m), \\ \mathit{divrt}((\alpha, \beta), (o_A, o_B)) &\stackrel{\text{def}}{=} (o_{A1}^\alpha, \beta o_{A2}^\alpha). \end{aligned}$$

Here, we take the group $(\mathbb{Z}_q^* \times G_q, \odot, (1, 1))$, with:

$$(\alpha_1, \beta_1) \odot (\alpha_2, \beta_2) \stackrel{\text{def}}{=} (\alpha_1 \alpha_2, \beta_1^{\alpha_2} \beta_2).$$

Associativity is again not trivial but straightforward and the inverse of (α, β) is $(\alpha^{-1}, \beta^{-\alpha^{-1}})$.

Again, the Ground premise is straightforward to see, so we only check the two premises:

Decomposition:

$$(\mathit{nativ}_A(\alpha, y, m), \mathit{nativ}_B()) = (g^\alpha, my^\alpha) = \mathit{divrt}(\mathit{join}(\alpha, m), \mathit{base}(y)).$$

Mixed Associativity

$$\begin{aligned} & \mathit{divrt}((\alpha_2, \beta_2), \mathit{divrt}((\alpha_1, \beta_1), (o_A, o_B))) \\ = {} & \mathit{divrt}((\alpha_2, \beta_2), (o_{A1}^{\alpha_1}, \beta_1 o_{A2}^{\alpha_1}) \\ = {} & ((o_{A1}^{\alpha_1})^{\alpha_2}, \beta_2 (\beta_1 o_{A2}^{\alpha_1})^{\alpha_2}) \\ = {} & \mathit{divrt}((\alpha_1, \beta_1) \odot (\alpha_2, \beta_2), (o_A, o_B)). \end{aligned}$$

□