

## Differential Privacy and Security

**Damas P. Gruska**<sup>\*†</sup>

*Institute of Informatics, Comenius University*  
*Mlynska dolina, 842 48 Bratislava, Slovakia*  
*gruska@fmph.uniba.sk*

---

**Abstract.** A quantification of process’s security by differential privacy is defined and studied in the framework of probabilistic process algebras. The resulting (quantitative) security properties are investigated and compared with other (qualitative and quantitative) security notions.

**Keywords:** differential privacy, probabilistic process algebra, information flow, security, opacity, min-entropy

### 1. Introduction

Several formulations of system security can be found in the literature. Many of them are based on a non-interference (see [8]) which assumes an absence of any information flow between private and public systems activities. More precisely, systems are considered to be secure if from observations of their public activities no information about private activities can be deduced. This approach has found many reformulations for different formalisms, computational models and nature or “quality” of observations. For many applications such properties could be criticized for being either too restrictive or too benevolent. They are too restrictive in the case that there exists some information flow between public and private activities (or data) but this flow is reasonable small. For example, usually access control processes exhibit some information flow (mostly) showing which password is not correct but they are still considered to be secure under reasonable password policy, namely it is not meaningful to consider such systems insecure in the case that a number of possible passwords is sufficiently large. On the other side, qualitative security properties could be too benevolent. For example, if an intruder cannot

---

<sup>\*</sup>Work supported by the grant VEGA 1/1333/12.

<sup>†</sup>Address for correspondence: Institute of Informatics, Comenius University, Mlynska dolina, 842 48 Bratislava, Slovakia

learn the whole secret (password, private key, etc) they could consider a system to be safe despite the fact, that the intruder could still learn almost all the secret (for example, significant number of bits of a private key). Hence there is a need to quantify an amount of information flow which can be gained from observations of public system activities.

An amount of possibly leaked information could be expressed by means of Shannon's information theory as it was done, for example, in [3, 4] for simple imperative languages and in [11] for process algebras. Another possibility is to exploit probabilistic theory as it was used for process algebras in [10]. Resulting techniques could express how many bits of private information can leak or how probable is that an intruder can learn some secret property of processes. In [14] an information flow is studied in a framework of process algebra and is investigated how much information, i.e. a number of bits, can be transmitted by observing some system's timed activities. In [9] it is investigated which private actions can be gained or excluded by observations of public actions.

The aim of this paper is to quantify an amount of information flow by differential privacy (see [5]) in a framework of probabilistic process algebra. The concept of differential privacy was originally developed to "provide means to maximize the accuracy of queries from statistical databases while minimizing the chances of identifying its records". Later on it was used also for other applications. In [19] differential privacy is studied for probabilistic automata and in [18] it is applied in a framework of probabilistic process algebra where probabilities of a given output produced by inputs which differ in one position are compared. Here we extend and further develop this approach and we propose several other security properties based on  $\epsilon$ -differential privacy for a (different) probabilistic process algebra. We generalize differential privacy also for arbitrary inputs incorporating a distance between them, as well as we weaken differential privacy in such a way, that we require that only a subset of inputs satisfies  $\epsilon$ -differential privacy constraint. We prove some compositionality properties as well as undecidability and decidability results for the resulting security properties. Moreover, we show a relationship among the resulting properties as well as how they are related to some traditional qualitative security properties (namely, Non-Deducibility on Composition [6] and opacity [1, 2]) and quantitative one, based on min-entropy in the style similar to that from [17].

The paper is organized as follows. In Section 2 we describe our working formalism - probabilistic process algebra. In Sections 3 we recall some (qualitative) security properties based on an absence of information flow which will serve as a motivation for our subsequent work. Section 4 is devoted to differential privacy. Here we define and investigate various security properties based on  $\epsilon$ -differential privacy.

## 2. Probabilistic Process Algebra

In this section we define the Probabilistic Process Algebra, pCCS for short, which is based on Milner's CCS (see [15]). First we assume a set of atomic action symbols  $A$  not containing symbol  $\tau$  and such that for every  $a \in A$  there exists  $\bar{a} \in A$  and  $\bar{\bar{a}} = a$ . We define  $Act = A \cup \{\tau\}$ . We assume that  $a, b, \dots$  range over  $A$  and  $u, v, \dots$  range over  $Act$ .

To add probabilities to CCS calculus we will follow alternating model (the approach presented in [12]) which is neither reactive nor generative nor stratified (see [13]). Probabilistic transitions are not associated with actions but they are labeled with probabilities. In so called probabilistic states a next transition is chosen according to probabilistic distribution. For example, process  $a.(0.3.b.Nil \oplus 0.7.(a.Nil +$

$b.Nil$ )) can perform action  $a$  and after that it reaches the probabilistic state and from this state it can reach with probability 0.3 the state where only action  $b$  can be performed or with probability 0.7 it can reach the state where it can perform either  $a$  or  $b$ .

Formally, we introduce a new operator  $\bigoplus_{i \in I} q_i.P_i$ ,  $q_i$  being real numbers in  $(0, 1]$  such that  $\sum_{i \in I} q_i = 1$ . Processes which can perform as the first action probabilistic transition will be called probabilistic processes or states (to stress that  $P$  is non-probabilistic process we will sometimes write  $P_N$  if necessary). Hence we assume the signature  $\Sigma = \bigcup_{n \in \mathbb{N}} \Sigma_n$ , where

$$\begin{aligned} \Sigma_0 &= \{Nil\} \\ \Sigma_1 &= \{x. \mid x \in Act\} \cup \{[S] \mid S \text{ is a relabeling function}\} \\ &\quad \cup \{\backslash M \mid M \subseteq A\} \\ \Sigma_2 &= \{\mid\} \\ \Sigma_n &= \{\bigoplus_{i \in I} q_i, |I| = n\} \end{aligned}$$

with the agreement to write unary action operators in prefix form, the unary operators  $[S]$ ,  $\backslash M$  in postfix form, and the rest of operators in infix form. Relabeling functions,  $S : Act \rightarrow Act$  are such that  $S(\bar{a}) = S(\bar{a})$  for  $a \in A$  and  $S(\tau) = \tau$ . Actions  $a$  and  $\bar{a}$  represent input and output communications, respectively, by means of channel  $a$ . Sometimes when we say channel we mean the corresponding input or output action. The set of pCCS terms over the signature  $\Sigma$  is defined by the following BNF notation:

$$P ::= X \mid op(P_1, P_2, \dots, P_n) \mid \mu X P$$

where  $X \in Var$ ,  $Var$  is a set of process variables,  $P, P_1, \dots, P_n$  are pCCS terms,  $\mu X$  – is the binding construct,  $op \in \Sigma$ . We require that all  $P_i$  processes in  $\bigoplus_{i \in I} q_i.P_i$  are non-probabilistic ones. By pCCS we will denote the set of all probabilistic and non-probabilistic processes and all definitions and notations for CCS processes (see [15]) are extended for pCCS ones. Structural operational semantics is given by labeled transition systems. The transition relation  $\rightarrow$  is a subset of  $pCCS \times Act \cup (0, 1] \times pCCS$ . We just mention the new transition rules for probabilistic.

$$\begin{array}{c} \frac{}{P_N \xrightarrow{1} P_N} \quad A1 \qquad \frac{}{\bigoplus_{i \in I} q_i.P_i \xrightarrow{q_i} P_i} \quad A2 \\[10pt] \frac{P \xrightarrow{q} P', Q \xrightarrow{r} Q'}{P \mid Q \xrightarrow{q \cdot r} P' \mid Q'} \quad Pa \end{array}$$

For probabilistic choice we have the rule  $A2$  and for a probabilistic transition of two processes running in parallel we have the rule  $Pa$ . The technical rule  $A1$  enables parallel run of probabilistic and non-probabilistic processes by allowing to non-probabilistic processes to perform  $\xrightarrow{1}$  transition and hence the rule  $Pa$  could be applied.

We will use an usual definition of opened and closed terms where  $\mu X$  is the only binding operator. Closed terms which are guarded (each occurrence of  $X$  is within some subexpression  $u.A$ ) are called pCCS processes. Note that  $Nil$  will be often omitted from processes descriptions and hence, for example, instead of  $a.b.Nil$  we will write just  $a.b$ . We write  $P \xrightarrow{x} P'$  instead of  $(P, x, P') \in \rightarrow$  and  $P \not\xrightarrow{x}$  if there is no  $P'$  such that  $P \xrightarrow{x} P'$ . The meaning of the expression  $P \xrightarrow{x} P'$  is that the term  $P$  can evolve to  $P'$  by performing action  $x$ , by  $P \xrightarrow{x}$  we will denote that there exists a term  $P'$  such that  $P \xrightarrow{x} P'$ .

To express what an observer can see from system behaviour we will define modified transitions  $\xRightarrow{x}$  which hide the action  $\tau$  and probabilities. Formally, we will write  $P \xRightarrow{x} P'$  iff  $P \xrightarrow{s_1} \xrightarrow{x} \xrightarrow{s_2} P'$  for  $s_1, s_2 \in (\{\tau\} \cup (0, 1])^*$  and  $P \xRightarrow{s}$  instead of  $P \xrightarrow{x_1} \xrightarrow{x_2} \dots \xrightarrow{x_n}$ . We will write  $P \xRightarrow{x}$  if there exists  $P'$  such that  $P \xRightarrow{x} P'$ . By  $\epsilon$  we will denote the empty sequence of actions and by  $s \sqsubseteq s'$ ,  $s, s' \in (Act \cup (0, 1])^*$  we will denote that  $s$  is a prefix of  $s'$ . By  $Sort(P)$  we will denote the set of actions from  $A$  which can be performed by  $P$  i.e.  $Sort(P) = \{x | P \xrightarrow{s.x} \text{ for some } s \in (Act \cup (0, 1])^* \text{ and } x \in A\}$ .

Let  $s \in (Act \cup (0, 1])^*$ . By  $s|_B$  we will denote the sequence obtained from  $s$  by removing all actions not belonging to  $B$  and we will write  $x \in s$  if the sequence  $s$  contains  $x$  as its element. We will write  $\bar{s}$  for sequence obtained from  $s$  in such a way that all actions are replaced by its complementary action if it exists and left as they are otherwise. For example, if  $s = a.\bar{b}.\tau.c$  we have  $\bar{s} = \bar{a}.b.\tau.\bar{c}$ .

As regards behavioral semantics, we will work with the weak trace equivalence.

**Definition 2.1.** The set of weak traces of process  $P$  is defined as  $Tr_w(P) = \{s \in A^* | \exists P'. P \xRightarrow{s} P'\}$ . Two processes  $P$  and  $Q$  are weakly trace ( $P \approx_w Q$ ) iff  $Tr_w(P) = Tr_w(Q)$ .

We conclude this section with a definition of probabilities of traces for a given process. Let  $P$  be a pCCS process and let  $P \xrightarrow{x_1} P_1 \xrightarrow{x_2} P_2 \xrightarrow{x_3} \dots \xrightarrow{x_n} P_n$ , where  $x_i \in Act \cup (0, 1]$  for every  $i, 1 \leq i \leq n$ . The sequence  $P.x_1.P_1.x_2 \dots x_n.P_n$  will be called a finite computational path of  $P$  (path, for short), its label is a subsequence of  $x_1 \dots x_n$  consisting of those elements which belong to  $Act$  i.e.  $label(P.x_1.P_1.x_2 \dots x_n.P_n) = x_1 \dots x_n|_{Act}$  and its probability is defined as a multiplication of all probabilities contained in it, i.e.  $Prob(P.x_1.P_1.x_2 \dots x_n.P_n) = 1 \times q_1 \times \dots \times q_k$  where  $x_1 \dots x_n|_{(0, 1]} = q_1 \dots q_k$ . The multiset of finite paths of  $P$  will be denoted by  $Path(P)$ . For example, the path  $(0.5.a.Nil \oplus 0.5.a.Nil).0.5.(a.Nil).a.(Nil)$  is contained in  $Path(0.5.a.Nil \oplus 0.5.a.Nil)$  two times. There exist a few techniques how to define this multiset. For example, in [16] a technique of schedulers are used to resolve the nondeterminism and in [7] all transitions are indexed and hence paths can be distinguished by different indexes. In the former case, every scheduler defines (schedules) a particular computation path and hence two different schedulers determine different paths, in the later case, the index records which transition was chosen in the case of several possibilities. The set of indexes for process  $P$  consists of sequences  $i_1 \dots i_k$  where  $i_j \in \{0, \dots, n\} \cup \{0, \dots, n\} \times \{0, \dots, n\}$  where  $n$  is the maximal cardinality of  $I$  for subterms of  $P$  of the form  $\bigoplus_{i \in I} q_i.P_i$ . An index records how a computation path of  $P$  could be derived, i.e. it records which process was chosen in case of several nondeterministic possibilities. If there is only one possible successor transitions are indexed by 1 (i.e. corresponding  $i_l = 1$ ). If transition  $P_i \xrightarrow{x} P'$  is indexed by  $k$  then transition  $\bigoplus_{i \in I} q_i.P_i \xrightarrow{x} P'$  is indexed by  $k.i$ , and if transitions  $P \xrightarrow{x} P'$  and  $Q \xrightarrow{x} Q'$  are indexed by  $k$  and  $l$ , respectively, then transitions of  $P|Q$  have indexes from  $\{(k, 0), (0, l), (k, l)\}$  depending on which transition rule for parallel composition was applied. Every index defines at most one path and the set of all indexes defines the multisets of paths  $Path(P)$ . Let  $C, C' \subseteq Path(P)$  be a finite multiset. We define  $Pr(C) = \sum_{c \in C} Prob(c)$  if  $C \neq \emptyset$  and  $Pr(\emptyset) = 0$ . For  $s \in Tr_w(P)$  we will denote by  $Pr(s)$  the probability of performing  $s$  (i.e. it is the sum of probabilities of all paths  $c \in Path(P)$  such that  $label(c) = s$ ).

### 3. Information Flow

In this section we recall two (qualitative) security properties for CCS (i.e. non-probabilistic process algebra). The first inspiration for our work is the security property Non-Deducibility on Composition

(NDC for short, see in [6]). Suppose that all actions are divided into two groups, namely public (low level) actions  $L$  and private (high level) actions  $H$  i.e.  $A = L \cup H, L \cap H = \emptyset$ . Process  $P$  has property NDC if for every high level user  $A$ , the low level view of the behaviour of  $P$  is not modified (in terms of weak trace equivalence) by the presence of  $A$ . The idea of NDC can be formulated as follows.

**Definition 3.1. (NDC)**  $P \in NDC$  iff for every  $A, Sort(A) \subseteq H \cup \{\tau\}$

$$(P|A) \setminus H \approx_w P \setminus H.$$

Now we introduce another information flow notion, which is based on a more general concept of observation and opacity. This concept was exploited in [1] and [2] for Petri Nets and transition systems, respectively. First we assume an observation function  $\mathcal{O} : Act^* \rightarrow Act^*$ . Now suppose that we have some security property of process's traces. This might be an execution of one or more classified actions, an execution of actions in a particular classified order which should be kept hidden, etc. Suppose that this property is expressed by a predicate  $\phi$ . Contrary to the original definition we do not require that the predicate is total. We would like to know whether an observer can deduce the validity of the property  $\phi$  just by observing sequences of actions from  $Act^*$  performed by a given process. The observer cannot deduce the validity of  $\phi$  if there are two traces  $w, w' \in Act^*$  such that  $\phi(w), \neg\phi(w')$  hold and the traces cannot be distinguished by the observer i.e.  $\mathcal{O}(w) = \mathcal{O}(w')$ . We formalize this concept by opacity.

**Definition 3.2. (Opacity)**

Given process  $P$ , a predicate  $\phi$  over  $Act^*$  is opaque w.r.t. the observation function  $\mathcal{O}$  if for every sequence  $w, w' \in Tr_w(P)$  such that  $\phi(w)$  holds and  $\mathcal{O}(w) \neq \epsilon$ , there exists a sequence  $w', w' \in Tr_w(P)$  such that  $\neg\phi(w')$  holds and  $\mathcal{O}(w) = \mathcal{O}(w')$ . The set of processes for which the predicate  $\phi$  is opaque with respect to  $\mathcal{O}$  will be denoted by  $Op_{\mathcal{O}}^{\phi}$ .

Now we are prepared to define several quantitative security properties based on differential privacy. Actually, as we will see later, two of them are really quantitative variants of the above mentioned qualitative properties.

## 4. Differential Privacy

Differential privacy was originally developed for privacy protection of statistical databases (see [5]). In the original definition, a query mechanism  $A$  is  $\epsilon$ -differentially private if for any two databases  $D_1$  and  $D_2$  which differ only for one individual (one row, for example, data of one person), and any property  $S$ , the probability distributions of  $A(D_1), A(D_2)$  differ on  $S$  at most by  $e^\epsilon$ , namely,

$$\Pr(A(D_1) \in S) \leq e^\epsilon \times \Pr(A(D_2) \in S).$$

### 4.1. Differential Privacy for Processes

Now we will reformulate  $\epsilon$ -differential privacy for our process algebra framework. Sequences of high level actions  $s$  (i.e.  $s \in H^*$ ) will represent secrete inputs. Public outputs will be given by sequences  $o$  of low level actions (i.e.  $o \in L^*$ ). First we start with formulation of  $\epsilon$ -differential privacy for a given secrete input and public output. Note that this definition is similar to the one which appeared in [18].

We will write for a given process  $P$  conditional probability  $Pr(o|s)$  as probability  $Pr(o)$  for process  $(P|s.Nil) \setminus H$  provided that the input  $s$  produces the output  $o$ . To be sure that the secrete input  $s$  is "consumed" to produce public output  $o$  we could employ an auxiliary action  $\sqrt{\phantom{x}}$  not belonging to  $A$  and check  $o.\sqrt{\phantom{x}} \in Tr_w((P|s.\sqrt{\phantom{x}}.Nil) \setminus H)$  but to simplify notation we will omit this in the subsequent text and we will write just  $o \in Tr_w((P|s.Nil) \setminus H)$ . Moreover, we assume the set  $I_o, I_o \subseteq H^*$  of all inputs producing the given output  $o$ , i.e.  $s \in I_o$  iff  $o \in Tr_w((P|s.Nil) \setminus H)$ . Usually we will omit to write  $I_o$  if it is clear from the context as well as a reference of  $I_o$  to  $P$ .

**Definition 4.1.**  $P \in DF_\epsilon(o, s)$  iff  $o \in Tr_w((P|s.Nil) \setminus H)$  and

$$Pr(o|s) \leq e^\epsilon \times Pr(o|s')$$

for every  $s' \in I_o$  which differs from  $s$  in one position.

Note that in the previous definition we assume that if  $s = x_1 \dots x_n$   $s' = x'_1 \dots x'_n$  then there exists  $j$  such that  $x_j \neq x'_j$  and  $x_i = x'_i$  for  $i \neq j$ . The property  $DF_\epsilon(o, s)$  requires that by observing the public output  $o$  an intruder cannot be sure (expressed by  $\epsilon$ ) whether the secrete input was  $s$  or any  $s'$  which differs from  $s$  in one position i.e. cannot be sure about any  $x_i, i = 1, \dots, n$ . Note that for  $\epsilon = 0$  the inputs  $s$  and  $s'$  do not lead to different probabilities for the corresponding output. For convenience we put automatically  $P \in DF_\epsilon(o, s)$  if  $o \notin Tr_w((P|s.Nil) \setminus H)$ . Now we will formulate several properties of  $\epsilon$ -differential privacy. First, differential privacy is not sensitive to length of observations (public outputs) i.e. a longer observation can leak less, as well as more, on private inputs as it is stated by the following proposition.

**Proposition 4.2.** For every  $\epsilon$  there exist processes  $P, P'$ , public outputs  $o_1, o_2, o_3, o_4 \in L^*$  such that  $o_1 \sqsubseteq o_2$  and  $o_3 \sqsubseteq o_4$  and secrete inputs  $s, s', s \in I_{o_1} \cap I_{o_2}, s' \in I_{o_3} \cap I_{o_4}$ , such that  $P \in DF_\epsilon(o_1, s), P \notin DF_\epsilon(o_2, s)$  and  $P' \in DF_\epsilon(o_4, s'), P' \notin DF_\epsilon(o_3, s')$ .

**Proof:**

Let  $P = (1/2).(h_2.l_1.(p.l_2.Nil \oplus (1-p).l_3.Nil)) \oplus (1/2).h_1.l_1.l_2.Nil$ ,  $s = h_1$  and  $o_1 = l_1, o_2 = l_1.l_2$ . Then for  $p < 1/e^\epsilon$  we get  $P \in DF_\epsilon(o_1, s), P \notin DF_\epsilon(o_2, s)$ . As regards the second case, let  $P' = ((1-p)/2).h_1.l_1.l_2.Nil \oplus ((1-p)/2).h_2.l_1.l_2.Nil \oplus p.h_2.l_1.Nil$ ,  $s' = h_1$  and  $o_3 = l_1, o_4 = l_1.l_2$ . Then for  $p < (1 - e^\epsilon)/(1 + e^\epsilon)$  we have  $P' \in DF_\epsilon(o_4, s'), P' \notin DF_\epsilon(o_3, s')$  for  $e^\epsilon < 1$ . For  $e^\epsilon \geq 1$  we exchange  $h_1$  and  $h_2$  in  $P'$ .  $\square$

Differential privacy is neither sensitive to length of secrete inputs as it is stated by the following proposition.

**Proposition 4.3.** For every  $\epsilon$  there exist processes  $P, P', o \in L^*$  and  $s_1, s_2, s_3, s_4 \in I_o$  such that  $s_1 \sqsubseteq s_2$  and  $s_3 \sqsubseteq s_4$  and such that  $P \in DF_\epsilon(o, s_1), P \notin DF_\epsilon(o, s_2)$  and  $P' \in DF_\epsilon(o, s_4), P' \notin DF_\epsilon(o, s_3)$ .

**Proof:**

Let  $P = ((1-p)/3).h_1.l.Nil \oplus ((1-p)/3).h_1.l.Nil \oplus ((1-p)/3).h_1.h_2.l.Nil \oplus p.h_1.h_1.h_1.l.Nil$ ,  $o = l$  and  $s_1 = h_1, s_2 = h_1.h_1$ . Then for  $p > e^\epsilon/(3 + e^\epsilon)$  we get  $P \in DF_\epsilon(o, s_1), P \notin DF_\epsilon(o, s_2)$ . As regards the second case, let  $P' = ((1-p)/3).h_1.h_1.l.Nil \oplus ((1-p)/3).h_1.h_2.l.Nil \oplus ((1-p)/3).h_2.l.Nil \oplus p.h_1.h_1.l.Nil$ ,  $o = l$  and  $s_3 = h_1, s_4 = h_1.h_1$ . Then for  $p > e^\epsilon/(3 + e^\epsilon)$  we have  $P' \in DF_\epsilon(o, s_4)$  and  $P' \notin DF_\epsilon(o, s_3)$ .  $\square$

The property  $DF_\epsilon(o, s)$  is rather strong and in general it is undecidable as it is stated by the following proposition.

**Proposition 4.4.** Property  $DF_\epsilon(o, s)$  is undecidable i.e. the question whether  $P \in DF_\epsilon(o, s)$  for arbitrary  $o$  and  $s, s \in I_o$  is undecidable.

**Proof:**

We exploit Turing power of pCCS and hence we reduce the property to the halting problem. Let  $R$  be an arbitrary process and let  $T = \mu X. \sum_{y \in Act} y.X$ . By deciding  $(P | ((R|T) \setminus Act)) \in DF_\epsilon(o, s)$  we could decide halting problem for  $R$  since to compute the probabilities we should know when the process  $(R|T) \setminus Act$  halts.  $\square$

We could put some restrictions on processes in such a way that the property  $DF_\epsilon(o, s)$  becomes decidable for restricted processes. First we need an auxiliary definition.

**Definition 4.5.** Process variable  $X$  is sequential in  $P$  if every subterm of  $P$  containing  $X$  (except  $X$  itself) is of a form  $y.P'$  or  $\sum P_i$ . Otherwise  $X$  is called parallel. Let  $M \subseteq Act$ . Process variable  $X$  is  $M$ -guarded in  $P$  if for every  $u, u \in M$  every occurrence of  $X$  is contained in a subterm of  $P$  of the form  $u.P'$ .

**Proposition 4.6.** Property  $DF_\epsilon(o, s)$  is decidable for finite processes and for such processes  $P$  that if  $X$  is parallel in  $P$  then  $X$  is guarded by some  $l \in Sort(o.Nil)$  and moreover, no other parallel occurrence of  $X$  is guarded by  $M, M = \{\bar{l}\}$ .

**Proof:**

Only the case of infinite processes is interesting but restrictions put on such processes is such that they cannot produce sequences with arbitrary number of  $\tau$  action between two observable actions from  $o$ . Hence, there are only finitely many cases to be checked to compute corresponding probabilities.  $\square$

Now we will formulate and prove some compositional properties of  $DF_\epsilon(o, s)$  property.

**Proposition 4.7.** Let  $P \in DF_\epsilon(o, s)$ . Then  $l.P \in DF_\epsilon(l.o, s)$  and  $h.P \in DF_\epsilon(o, h.s)$ .

**Proof:**

Clearly, every observation of the process  $l.P$  has to start with  $l$  and probabilities of all traces with the proper prefix  $l$  do not change. Similarly for the process  $h.P$ .  $\square$

**Proposition 4.8.** Let us assume processes  $P_i$  and let  $p = \min_{s', i} \{(q_i.Pr(o|s')_i) \mid \text{for } s' \text{ which differs from } s \text{ in one position and } i = 1, \dots, n\}$ ,  $p' = \max_{s', i} \{(q_i.Pr(o|s')_i) \mid \text{for } s' \text{ which differs from } s \text{ in one position and } i = 1, \dots, n\}$ . Let  $P = \bigoplus_{i \in \{1, \dots, n\}} q_i.P_i$  then  $P \in DF_\epsilon(o, s)$  for  $\epsilon \geq \ln(p/p')$ .

**Proof:**

From the choice of  $p, p'$  we have  $Pr(o|s)/Pr(o|s') \leq p'/p$  for process  $P$  and any  $s'$  which differs from  $s$  in one position. Hence  $P \in DF_\epsilon(o, s)$  for  $\epsilon \geq \ln(p/p')$ .  $\square$

**Proposition 4.9.** Let  $P \in DF_\epsilon(o, s)$  and  $S$  be a bijection on  $L$  and on  $H$ . Then  $P[S] \in DF_\epsilon(S(o), S(s))$ . Let  $o \in Tr_w((P \setminus M|s.Nil) \setminus H)$ . Then  $P \setminus M \in DF_\epsilon(o, s)$ .

**Proof:**

The first part of the proof follows directly from the definition of relabeling. The second part follows from the fact that the restriction has no influence on performing  $o$  and hence the corresponding probabilities are not changed by the restriction.  $\square$

**Proposition 4.10.** Let  $P \in DF_\epsilon(o, s)$ ,  $P$  is sequential and process variable  $X$  is  $M$ -guarded in  $P$  for some nonempty  $M$  such that  $Sort(o.Nil) \cap M = \emptyset$ . Then  $\mu X.P \in DF_\epsilon(o, s)$ .

**Proof:**

We show that whenever input  $s$  or  $s'$  which differs from  $s$  in one position, produces output  $o$ , the recursion is not involved and hence probabilities are not changed by recursion. From the assumption of the proposition we know that every occurrence of  $X$  is guarded (also) by some public action which appears in  $o$ . Moreover, this action cannot be involved in an internal communication since  $X$  is sequential in  $P$ .  $\square$

Now we can define the property which expresses security of a given input  $s$  with respect to  $\epsilon$ -differential privacy. Processes have this property if there is no observation (output) which could distinguish between the input  $s$  and input  $s'$  (which differs from  $s$  in one element). The formal definition of  $DF_\epsilon(s)$  property is the following.

**Definition 4.11.**  $P \in DF_\epsilon(s)$  iff for every  $o \in L^*$  it holds  $P \in DF_\epsilon(o, s)$ .

Note that property  $DF_\epsilon(s)$  is undecidable but can be decided under similar restrictions as property  $DF_\epsilon(o, s)$  (see Propositions 4.4 and 4.6). Property  $DF_\epsilon(s)$  strongly depends on choice of  $s$  as it is stated by the following corollary.

**Corollary.** Let  $s_1 \sqsubseteq s_2$  and  $s_1 \neq s_2$ . Then neither  $DF_\epsilon(s_1) \subseteq DF_\epsilon(s_2)$  nor  $DF_\epsilon(s_2) \subseteq DF_\epsilon(s_1)$  holds. Moreover, let  $P \in DF_\epsilon(s_1) \cap DF_\epsilon(s_2)$  then, in general, it does not hold that  $P \in DF_\epsilon(s)$  for  $s_1 \sqsubseteq s \sqsubseteq s_2$ .

**Proof:**

The first part of the proof follows from Proposition 4.3. Now let  $P = p.h.l.Nil \oplus p.h'.l.Nil \oplus p.h.h.l.Nil \oplus p.h.h.h'.l.Nil \oplus p.h.h.l.Nil \oplus (1 - 5p).h.h'.l.Nil$ ,  $s_1 = h$ ,  $s_2 = h.h.h$  and  $s = h.h$ . Then we have  $P \in DF_\epsilon(s_1) \cap DF_\epsilon(s_2)$  but  $P \notin DF_\epsilon(s)$  for  $\epsilon < \ln(p/(1 - 5p))$ .  $\square$

Suppose that  $P \notin DF_\epsilon(s)$ . This means that there exists at least one observation  $o$  which could leak secrete  $s$  with respect to  $\epsilon$ -differential privacy. We can define the non-empty set of such observations for every  $P$  which does not belong to  $DF_\epsilon(s)$ .

**Definition 4.12.**  $DF_\epsilon(P, s) = \{o \mid \text{there exists } s' \text{ which differs from } s \text{ in one position and such that } Pr(o|s) > e^\epsilon \times Pr(o|s') \text{ and } o \in Tr_w((P|s.Nil) \setminus H)\}.$



Clearly,  $P \in DF_\epsilon(s)$  iff  $DF_\epsilon(P, s) = \emptyset$ . On the other side, if  $DF_\epsilon(P, s) \neq \emptyset$  we can ask what is the minimal length of  $o, o \in DF_\epsilon(P, s)$ . Usually, longer  $o$  (a higher value of  $|o|$ ) means that the secrete  $s$  could be considered safer. Formally, let  $n = \min\{|o| \mid \text{for } o \in DF_\epsilon(P, s)\}$ . Then a greater  $n$  means safer  $P$  with respect to  $s$ , hence we denote value of  $n$  by  $DPS_\epsilon(P, s)$  i.e.  $DPS_\epsilon(P, s) = n$  is the security level of  $P$  with respect to the secrete input  $s$  (we define  $DPS_\epsilon(P, s) = \infty$  if  $P \in DF_\epsilon(s)$ ). Unfortunately, the property is undecidable in general as it is stated by the following proposition, but becomes decidable under the similar conditions as they are contained in Proposition 4.6.

**Proposition 4.13.**  $DPS_\epsilon(P, s)$  is undecidable even if we know some  $o, o \in DF_\epsilon(P, s)$ .

**Proof:**

Undecidable follows from undecidability of  $DF_\epsilon(P, s)$  despite the fact that only finitely many  $o', |o'| < |o|$  has to be checked.  $\square$

Now we formulate some simple compositional properties of  $DPS_\epsilon(P, s)$ . Note that other ones could be stated as well in the style of Propositions 4.7 - 4.10.

**Proposition 4.14.** Let  $DPS_\epsilon(P, s) = n$ . Then  $DPS_\epsilon(l.P, s) = n + 1$ ,  $DPS_\epsilon(h.P, h.s) = n$ . Moreover, let  $DPS_\epsilon(Q, s) = m$  then  $DPS_\epsilon(p.P \oplus (1-p).Q, s) \geq \min(n, m)$ .

**Proof:**

Clearly, prefixing only increase public outputs or secrete inputs but does not change probabilities of the sequences longer than 1. Now suppose that  $DPS_\epsilon(p.P \oplus (1-p).Q, s) = k, k < \min(n, m)$ . Hence, there exists observation  $o, |o| = k$  and  $o \in DF_\epsilon(P, s)$ . Let for  $P$  holds  $Pr(o|s) = p_1, Pr(o|s') = p_2$  and for  $Q$  it holds  $Pr(o|s) = q_1, Pr(o|s') = q_2$ . Hence for  $p.P \oplus (1-p).Q$  we have  $Pr(o|s) = p.p_1 + (1-p).q_1$  and  $Pr(o|s') = p.p_2 + (1-p).q_2$ . Since  $o \notin DF_\epsilon(P, s) \cup DF_\epsilon(Q, s)$  we have  $p.p_1 + (1-p).q_1 \leq e^\epsilon.p.p_2 + (1-p).e^\epsilon.q_2 = e^\epsilon(p.p_2 + (1-p).q_2)$  what is in the contradiction to our choice of  $o$ .  $\square$

Similarly to the Definition 4.11, we can define a set of processes (denoted by  $DF_\epsilon(o)$ ) for which by observing the public output  $o$  no secrete can leak (in the sense of  $\epsilon$ -differential privacy).

**Definition 4.15.**  $P \in DF_\epsilon(o)$  iff for every  $s \in I_o$  it holds  $P \in DF_\epsilon(o, s)$ .

For  $DF_\epsilon(o)$  we could formulate similar properties as we did for  $DF_\epsilon(s)$ . Similarly to the Definition 4.12, we can specify which secretes could leak (with respect to  $\epsilon$ -differential privacy) by the given observation  $o$ .

**Definition 4.16.**  $DF_\epsilon(P, o) = \{s \mid s \in I_o \text{ for which there exists } s' \text{ which differs from } s \text{ in one position, } o \in Tr_w((P|s.Nil) \setminus H) \text{ and } Pr(o|s) > e^\epsilon \times Pr(o|s')\}$ .

There is a simple relation between sets from the Definition 4.12 and 4.16, namely,  $o \in DF_\epsilon(P, s)$  iff  $s \in DF_\epsilon(P, o)$ . Now we define another generalization of above mentioned concepts as overall security of processes with respect to  $\epsilon$ -differential privacy which requires that processes are secure with respect to every secrete input and public output (denoted by  $DF_\epsilon$ ). The formal definition follows.

**Definition 4.17.**  $DF_\epsilon = \{P | P \in DF_\epsilon(o, s) \text{ for every } o \in L^*, s \in I_o\}$ .

Note that for  $P \in DF_\epsilon$  it holds also  $P \in DF_\epsilon(o), P \in DF_\epsilon(s)$  for every  $s$  and  $o$  and vice versa.

Till now we have investigated an impact of a difference in probabilities of public outputs for two secret inputs which differ only in one position. This approach could be too restrictive in many cases, so now we will extend it. We assume a distance between two secret inputs  $s, s' \in I_o$ , denoted by  $\rho(s, s')$ . At the moment, we do not put any restrictions on  $\rho$ . Hence we can relate probabilities of output  $o$  produced by arbitrary secrets  $s, s'$  not only those ones which differ only in one position. We assume that with the bigger distance between two secret inputs also the difference between corresponding probabilities could be bigger, hence we multiply  $\epsilon$  by  $\rho(s, s')$ . Note that the resulting security property (denoted by  $DF_{\epsilon, \rho}(o, s)$ ) has to be parameterized by  $\rho$  as well as by  $\epsilon$ .

**Definition 4.18.**  $P \in DF_{\epsilon, \rho}(o, s)$  iff  $o \in Tr_w((P|s.Nil) \setminus H)$  and

$$Pr(o|s) \leq e^{\epsilon \times \rho(s, s')} \times Pr(o|s')$$

for every  $s' \in I_o$ .

Similarly to the Definition 4.17 we can define sets of security properties with respect to  $\rho$  and  $\epsilon$ -differential privacy as well as the set of secrets which are secure under the observation  $o$ .

**Definition 4.19.**  $DF_{\epsilon, \rho} = \{P | P \in DF_{\epsilon, \rho}(o, s) \text{ for every } o \in L^*, s \in I\}$ .  $DF_{\epsilon, \rho}(o) = \{P | P \in DF_{\epsilon, \rho}(o, s) \text{ for every } s \in I\}$ .

Now we can relate qualitative security property NDC to the quantitative one, namely to  $\epsilon$ -differential privacy.

**Proposition 4.20.** Let  $P$  be a process and  $\rho$  be a distance on sequences of  $H$  actions such that  $\rho(x, y) \neq 0$  whenever  $x \neq y$ . Then if  $P \in NDC$  then for every  $o \in L^*$ , such that  $I_o$  is finite and every  $s \in I_o$ , there exists  $\epsilon$  such that  $P \in DF_{\epsilon, \rho}(o, s)$ . Moreover, if for every  $o \in L^*, s \in I_o$  there exists  $\epsilon$  such that  $P \in DF_{\epsilon, \rho}(o, s)$  then  $P \in NDC$ .

**Proof:**

Let  $P \in NDC$ , i.e.  $(P|A) \setminus H \approx_w P \setminus H$  for every  $A$  such that  $Sort(A) \subseteq H \cup \{\tau\}$ . This means that also  $(P|s.Nil) \setminus H \approx_w (P|s'.Nil) \setminus H$  and hence  $Pr(o|s) = 0$  iff  $Pr(o|s') = 0$  for every  $o$  i.e. it cannot happen that one of these probabilities is non-zero and another one is equal to zero, hence there exists  $\epsilon$  (since the number of possible inputs leading to  $o$  is finite) such that  $P \in DF_{\epsilon, \rho}(o, s)$ .

Now suppose that for every  $o \in L^*, s \in I_o$  there exists  $\epsilon$  such that  $P \in DF_{\epsilon, \rho}(o, s)$ . This means that for any two secrets if one could output  $o$  then also another one can do the same and hence  $P \in NDC$ .  $\square$

As regards the distance  $\rho$ , there are several meaningful choices how to measure the distance between two secrets. First we consider a variant of Hamming distance.

**Definition 4.21.** Let  $s, s' \in Act^*$  and  $s = x_1.x_2 \dots x_n, s' = x'_1.x'_2 \dots x'_m$ . We define metrics  $\rho_0$  as a number of positions where  $s$  and  $s'$  differ, i.e.  $\rho_0(s, s') = |m - n| + \sum_{i=1, x_i \neq x'_i}^{\min(n, m)} 1$ .

For the metric  $\rho_0$  we have the following result which relates  $DF_\epsilon(o, s)$  and  $DF_{\epsilon, \rho_0}(o, s)$  properties.

**Proposition 4.22.**  $P \in DF_{\epsilon, \rho_0}(o, s)$  iff  $P \in DF_\epsilon(o, s')$  for every  $s' \in I_o$ .

**Proof:**

Let  $s$  and  $s'$  differ in one position, i.e.  $\rho_0(s, s') = 1$  and let  $P \in DF_{\epsilon, \rho_0}(o, s)$ . This means that  $Pr(o|s) \leq e^{\epsilon \times 1} \times Pr(o|s')$  i.e.  $P \in DF_\epsilon(o, s)$ . Suppose that  $\rho_0(s, s') = n$ , then there exist  $s_1, \dots, s_{n-1}$  such that  $s_i, s_{i+1}$  differ by one element and the same holds for the pairs  $s, s_1$  and  $s_{n-1}, s'$ . Since we have  $P \in DF_\epsilon(o, s)$ ,  $P \in DF_\epsilon(o, s_i)$  for all  $i, 1 \leq n-1$  then we have  $Pr(o|s) \leq e^{\epsilon \times n} \times Pr(o|s')$  i.e.  $P \in DF_{\epsilon, \rho_0}(o, s)$ .  $\square$

The metric  $\rho_0$  does not take into account the length of inputs. If we have two completely different inputs of length 2 and inputs which differ in two positions but both of length 128, in the both cases the metric is 2 what does not express an amount of secrecy which could leak or which is protected. In the first case the whole secrete is protected and in the second case only a fraction of secrecy could be protected if  $P \in DF_{\epsilon, \rho_0}(o, s)$ . Hence we could consider more elaborated metrics, for example  $\rho_{min}(s, s') = \rho_o / \min(|s|, |s'|)$ ,  $\rho_+(s, s') = \rho_o / (|s| + |s'|)$  for cases that longer secretes are more valuable. On the other side, it also makes a difference if a secrete could leak by short observation or it could leak only by very long observations. This could lead us to further generalization of  $\epsilon$ -differential privacy. We could consider function  $f(s, s', o)$  which would take into account a distance between secrete inputs, their length, as well as length of outputs. Moreover, it can incorporate also a cost of observations (it could be different from it length) and other relations and properties.

## 4.2. Weak $\epsilon$ -differential Privacy

Now let us weaken the definition of  $\epsilon$ -differential privacy. The original concept assumes that probabilities of the public output  $o$  does not differ very much (in  $\epsilon$  sense) for "every" secrete input. Now we assume that it is enough that the probability does not differ only for inputs from a given set  $S, S \subseteq I_o$ . This means that an intruder (observing the output) still cannot be sure about the secret input  $s$  if  $S$  contains an element different from  $s$ . The formal definition follows.

**Definition 4.23.**  $P \in DF_\epsilon(o, s, S)$  iff  $o \in Tr_w((P|s.Nil) \setminus H)$  and

$$Pr(o|s) \leq e^\epsilon \times Pr(o|s')$$

for every  $s' \in S$ .

If  $S = \emptyset$  or  $S = \{s\}$  then for every  $P, P \in DF_\epsilon(o, s, S)$  i.e. the property says nothing. Suppose that  $|S| \geq 2$  and  $s \in S$  or  $S = \{s'\}, s' \neq s$ . Then if  $P \in DF_\epsilon(o, s, S)$  that means that there exists  $s'$  such that probabilities of output  $o$  with inputs  $s$  and  $s'$ , respectively are very close (expressed by  $\epsilon$ ). Hence by observing  $o$  we cannot be sure whether input was  $s$  or  $s'$ .

Note that  $P \in DF_\epsilon(o, s)$  iff  $P \in DF_\epsilon(o, s, S)$  where  $S = \{s' | \text{where } s' \text{ differs from } s \text{ in one position}\}$ . If  $S = I_o$  then  $DF_{\epsilon, \rho_f}(o, s) = DF_\epsilon(o, s, S)$  where  $\rho_f$  is the flat distance defined as  $\rho_f(s, s') = 0$  if  $s = s'$  and 1 otherwise.

**Example 4.24.** Let  $P \in DF_\epsilon(o, s, S)$  for every  $s$  and  $S$  such that  $S = \{s' | \rho_0(s, s') = n\}$ . Then by observing  $o$  an intruder cannot learn at least  $n$  components, i.e. elements, of  $s$ .

In general, we have a hierarchy of security properties given by set inclusion of sets  $S$ .

**Proposition 4.25.** Let  $S_1 \subseteq S_2$ . Then  $DF_\epsilon(o, s, S_2) \subseteq DF_\epsilon(o, s, S_1)$ .

**Proof:**

Let  $P \in DF_\epsilon(o, s, S_2)$  then  $Pr(o|s) \leq e^\epsilon \times Pr(o|s')$  for every  $s' \in S_2$  and since  $S_1 \subseteq S_2$  we have  $P \in DF_\epsilon(o, s, S_1)$ .  $\square$

Clearly, for  $S = \emptyset$  there are no restrictions put on  $P$ . If  $S$  contains all private inputs the the property coincide with the strong variant of  $\epsilon$ -differential privacy. Hence as the direct consequence of the previous proposition we obtain the following corollary.

**Corollary.** For every  $S, S \subseteq I_o$  it holds

$$DF_\epsilon(o, s) = DF_\epsilon(o, s, I_o) \subseteq DF_\epsilon(o, s, S) \subseteq DF_\epsilon(o, s, \emptyset) = \{P \mid \text{for which } o \in Tr_w(P)\}.$$

### 4.3. Quantification of Opacity

With the help of  $DF_\epsilon(o, s, S)$  property we can quantify opacity (see Definition 3.2). Opacity is based on predicates over process traces. Since we are interested in secrete inputs we will consider only predicates which depend only on these inputs. The formal definition follows.

**Definition 4.26.** Let  $\phi$  is a predicate over sequences from  $Act^*$ . We say that  $\phi$  is a security predicate if its value depends only on high level actions, i.e.  $\phi(t) \Leftrightarrow \phi(t')$  iff  $t|_H = t'|_H$ .

**Proposition 4.27.** Suppose that for every  $o \in L^*, s \in I_o$  such that  $\phi(s)$  holds, there exists  $\epsilon, S, S \subseteq I_o$  and  $s' \in S$  for which  $\neg\phi(s')$  holds such that  $P \in DF_\epsilon(o, s, S)$ . Then  $P \in Op_\mathcal{O}^\phi$  for security predicate  $\phi$  and for  $\mathcal{O}$  which maps high level actions, probabilities as well as  $\tau$  action to empty sequence. And vice versa.

**Proof:**

Let us assume that  $t$  is a trace of  $P$  and  $\phi(t)$  holds. Let  $t|_L = o$  and  $t|_H = s$ . Hence also  $\phi(s)$  holds since  $\phi$  is the security predicate. Since  $P \in DF_\epsilon(o, s, S)$  and there exists  $s' \in S$  for which  $\neg\phi(s')$  holds then also  $t'$  is a trace of  $P$ ,  $t'|_L = o$  and  $t'|_H = s'$  and  $\neg\phi(t')$  holds. Since  $\mathcal{O}$  maps high level actions, probabilities as well as  $\tau$  action to empty sequence we have  $\mathcal{O}(t) = \mathcal{O}(t')$ . Hence  $P \in Op_\mathcal{O}^\phi$ .

Now let us suppose that  $P \in Op_\mathcal{O}^\phi$ . This means, that for every trace  $t$  of  $P$  for which  $\phi(t)$  holds there exists trace  $t'$  for which  $\phi(t')$  does not hold and  $\mathcal{O}(t) = \mathcal{O}(t')$ . By choice of  $\mathcal{O}$  we have  $t|_L = t'|_L = o$ . Let  $t|_H = s$  and  $t'|_H = s'$ . Since  $\phi$  is a security predicate,  $\phi(s)$  holds and  $\phi(s')$  does not hold. Hence  $P \in DF_\epsilon(o, s, S)$  for some  $\epsilon$  and  $S = \{s, s'\}$ .  $\square$

Quantification from the previous proposition assumes an existence of one  $\epsilon$  for every given  $o$  and  $s$ . Now we present stronger quantification of opacity called  $\epsilon$ -Opacity.

**Definition 4.28. ( $\epsilon$ -Opacity)**

Given process  $P$ , a predicate  $\phi$  over  $Act^*$  is  $\epsilon$ -opaque w.r.t. the observation function  $\mathcal{O}$  if  $P \in Op_{\mathcal{O}}^{\phi} \cap DF_{\epsilon}(\rho)$ . The set of processes for which the predicate  $\phi$  is  $\epsilon$ -opaque with respect to  $\mathcal{O}$  will be denoted by  $\epsilon\rho Op_{\mathcal{O}}^{\phi}$ .

The combination of opacity and  $\epsilon$ -differential privacy gives us a qualitative new security property which cannot be obtained separately by these two properties.

**4.4. Min-entropy and  $\epsilon$ -differential Privacy**

Information flow can be quantified also by means of min-entropy leakage which is based on the vulnerability of the secrete to be guessed by one try (see [17]). Let us assume random variable  $X$ . Min-entropy of  $X$  is defined as

$$H_{\infty}(X) = \log_2 \max_{x \in X} (1/Pr(x)).$$

Let us assume random variable with uniform distribution  $S$  with values  $s, s \in I_o, |I_o| = n$  for a given output  $o$ . Clearly  $H_{\infty}(S) = \log_2(n)$ . Suppose that we have process  $P$  for which an intruder can see the output  $o$ . Now we define min-entropy for  $Pr(s|o)$  for this output  $o$ . Note that here we differ from [17] where it is defined for all possible outputs.

**Definition 4.29.** Let us assume process  $P$  and its public output  $o$ . Then we define min-entropy of  $S$  when  $o$  is observed as

$$H_{\infty}^P(S|o) = \log_2 \max_{s \in S} (1/Pr(s|o)).$$

Here we cannot directly define information flow as  $H_{\infty}(S) - H_{\infty}^P(S|o)$  (as it is done in [17]) since it can be negative, but the difference in absolute value expresses quantification of security of  $P$  when  $o$  is observed. In the next proposition we relate this quantification with  $\epsilon$ -differential privacy.

**Proposition 4.30.** Let  $P \in DP_{\epsilon, \rho_f}(o)$ . Then

$$H_{\infty}(S) - \log_2(e^{\epsilon}) \leq H_{\infty}^P(S|o) \leq H_{\infty}(S) + \log_2(e^{\epsilon}).$$

**Proof:**

We have  $Pr(s|o) \cdot Pr(o) = Pr(s, o) = Pr(o|s) \cdot Pr(s)$ . Moreover,  $Pr(o) = \sum_{s' \in I_P} Pr(o|s') \cdot Pr(s')$ . Hence  $Pr(s|o) = (Pr(o|s) \cdot Pr(s)) / \sum_{s' \in I_P} Pr(o|s') \cdot Pr(s')$ . Since  $P \in DP_{\epsilon, \rho_f}(o)$  we have  $Pr(o|s)/e^{\epsilon} \leq Pr(o|s')$  and  $Pr(o|s') \leq e^{\epsilon} \cdot Pr(o|s)$  for every  $s'$ . Then we have  $1/(n \cdot e^{\epsilon}) \leq Pr(s|o) \leq e^{\epsilon}/n$  for every  $s$ .  $\square$

Note the difference:  $\epsilon$ -differential privacy has been defined by means of probabilities  $Pr(o|s)$  but min-entropy by probabilities  $Pr(s|o)$ .

**5. Conclusions**

We have presented several (quantitative) security concepts based on  $\epsilon$ -differential privacy. They could be seen as quantifications of some qualitative properties, namely non-deducibility on composition [6] and opacity [1, 2]). They express how secure is a secrete input  $s$  with respect to the corresponding public output  $o$ , which secrete could leak by observing the public output  $o$ , by which output the secrete  $s$  could

leak or which processes are completely safe i.e. there is no secret and output which could leak it. Even very basic of these properties are undecidable in general but we have shown under which conditions they become decidable. But since also in this case complexity remains very high we propose some compositional properties to manage it at least somehow. It is shown that there is no relation between length of secrets or public outputs with respect to security of processes, i.e., for example, a longer observation could leak less than a shorter one. We propose also some metrics on inputs which could be exploited to obtain more realistic security properties. As it was mentioned, one should consider also length of inputs and relate it to the length of public outputs. Without this we could obtain too restrictive security notions. The price of leakage - as a relation between amount of leaked secrecy with respect to the length of observation is the crucial security characterization. Otherwise, for example, no access control process based on passwords would be considered safe (if a number of attempts to guess the password is not limited). We propose and study also another quantitative parameter - a length of the shortest observation which could leak a given secret input. Later, we have weakened differential privacy in such a way, that we require that only a subset of inputs satisfies  $\epsilon$ -differential privacy constraints. This still ensures that observing some public output, an intruder cannot be sure about a secret input which has produced it. At the end we have compared proposed  $\epsilon$ -differential privacy with another quantification of security based on min-entropy. Our approach is different from the one presented in [17] since we concentrate on change of expected probability of secret inputs if we already know public outputs.

This work represents theoretical research in the field of information flow based security but the proposed security properties offer also some application scenarios. Critical (with respect to security) parts of systems could be described by the probabilistic process algebra. Then we can check basic property  $DF_\epsilon(o, s)$  for finite or also infinite processes under some restrictions. To avoid state explosions and an enormous number of traces to be checked, we recommend to simplify the process algebra description as much as possible. In the case that no leakage of information is detected, one could try to use more detailed description and proceed. In general, there is no guaranty that there still could not be a leakage of information on very detailed (or complete) level of description but that is not a typical situation. Another possibility is to check only parts of the system description and to try to exploit some compositional properties of proposed security properties. In any case, there is no one universal manual how to practically exploit the proposed theory but one should try various approaches depending on the particular application. However, critical points usually represent those ones where probabilities of the "next action" could strongly depend on secret ones.

## References

- [1] Bryans J., M. Koutny and P. Ryan: Modelling non-deducibility using Petri Nets. Proc. of the 2nd International Workshop on Security Issues with Petri Nets and other Computational Models, 2004.
- [2] Bryans J., M. Koutny, L. Mazare and P. Ryan: Opacity Generalised to Transition Systems. In Proceedings of the Formal Aspects in Security and Trust, LNCS 3866, Springer, Berlin, 2006.
- [3] Clark D., S. Hunt and P. Malacaria: A Static Analysis for Quantifying the Information Flow in a Simple Imperative Programming Language. The Journal of Computer Security, 15(3), 2007.
- [4] Clarkson M.R., A.C. Myers, F.B. Schneider: Quantifying Information Flow with Beliefs. Journal of Computer Security, to appear, 17 (5), 2009.

- [5] Dwork C.: Differential Privacy: A Survey of Results. Proc. Theory and Applications of Models of Computation, LNCS 4978, 2008.
- [6] Focardi R., R. Gorrieri and F. Martinelli: Real-Time information flow analysis. IEEE Journal on Selected Areas in Communications 21 (2003).
- [7] Glabbeek R. J. van, S. A. Smolka and B. Steffen: Reactive, Generative and Stratified Models of Probabilistic Processes Inf. Comput. 121(1): 59-80, 1995.
- [8] Goguen J.A. and J. Meseguer: Security Policies and Security Models. Proc. of IEEE Symposium on Security and Privacy, 1982.
- [9] Gruska D.P.: Gained and Excluded Private Actions by Process Observations. To appear in Fundamenta Informaticae, 2011.
- [10] Gruska D.P.: Quantifying Security for Timed Process Algebras, Fundamenta Informaticae, vol. 93, Numbers 1-3, 2009.
- [11] Gruska D.P.: Probabilistic Information Flow Security. Fundamenta Informaticae, vol. 85, Numbers 1-4, 2008.
- [12] Hansson H. a B. Jonsson: A Calculus for Communicating Systems with Time and Probabilities. In Proceedings of 11th IEEE Real - Time Systems Symposium, Orlando, 1990.
- [13] López N. and Núñez: An Overview of Probabilistic Process Algebras and their Equivalences. In Validation of Stochastic Systems, LNCS 2925, Springer-Verlag, Berlin, 2004.
- [14] Lowe G.: Quantifying information flow”. In Proc. IEEE Computer Security Foundations Workshop, 2002.
- [15] Milner R.: *Communication and concurrency*. Prentice-Hall International, New York, 1989.
- [16] Segala R. and N. Lynch: Probabilistic Simulations for Probabilistic Processes. Nord. J. Comput. 2(2): 250-273, 1995.
- [17] Smith G.: Quantifying Information Flow Using Min-Entropy. QEST, 2011.
- [18] Xu L.: Modular reasoning about differential privacy in a probabilistic process calculus. In TGC, pages 198212, 2012.
- [19] Xu L., K. Chatzikokolakis, H. Lin and Catuscia Palamidessi: Metrics for Differential Privacy in Concurrent Systems, In Proceedings of HotSpot, 2014.