



# A formal analysis of the Neuchâtel e-voting protocol

Véronique Cortier, David Galindo, Mathieu Turuani

► **To cite this version:**

Véronique Cortier, David Galindo, Mathieu Turuani. A formal analysis of the Neuchâtel e-voting protocol. [Research Report] Inria Nancy - Grand Est. 2017. hal-01616425

**HAL Id: hal-01616425**

**<https://hal.inria.fr/hal-01616425>**

Submitted on 13 Oct 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A formal analysis of the Neuchâtel e-voting protocol

Véronique Cortier<sup>1</sup>, David Galindo<sup>2</sup>, and Mathieu Turuani<sup>3</sup>

<sup>1</sup>CNRS, Loria, France

<sup>2</sup>University of Birmingham, UK

<sup>3</sup>INRIA, Pesto team, France

October 13, 2017

## Abstract

Remote electronic voting is used in several countries for legally binding elections. Unlike academic voting protocols, these systems are not always documented and their security is rarely analysed rigorously.

In this paper, we study a voting system that has been used for electing political representatives and in citizen-driven referenda in the Swiss canton of Neuchâtel. We design a detailed model of the protocol in ProVerif for both privacy and verifiability properties. Our analysis mostly confirms the security of the underlying protocol: we show that the Neuchâtel protocol guarantees ballot privacy, even against a corrupted server; it also ensures cast-as-intended and recorded-as-cast verifiability, even if the voter's device is compromised. To our knowledge, this is the first time a full-fledged automatic symbolic analysis of an e-voting system used for politically-binding elections has been realized.

## 1 Introduction

Remote electronic voting (or Internet voting) allows voters to vote from their home or when they are travelling. It is also seen as a means to get feedback from citizens on a more regular basis. Therefore, Internet voting has been used in legally binding elections in several countries, including e.g. Estonia [20], Australia [9], France [29] or Switzerland [18]. Of course, designing and implementing a secure electronic voting system is a difficult task and many attacks or weaknesses have been discovered on deployed systems (see e.g. [33, 32]). Voting protocols should offer some basic security guarantees, such as ballot privacy (no one knows how I voted) as well as verifiability (voters can check the voting process) with as few trust assumptions as possible.

On an academic level, several Internet voting protocols have been developed and some of them offer a prototype implementation or even a voting platform. This is for

example the case of Helios [1], Civitas [11], Belenios [12], or Select [26]. These protocols are well documented and typically come with a proof of their security, at least w.r.t. privacy, in a symbolic or a cryptographic model (see e.g. [4, 22, 26, 12]). On the other hand, industrial scale protocols are being deployed and now also aim at offering some verifiability properties. For instance some systems offer voters to check their vote. Some examples are: the Estonian [21] protocol, where voters can check their vote during a short period of time; the so-called Norwegian protocol [19], that does so using return codes; or the New South Wales iVote protocol [9]. The systems deployed in Norway and New South Wales have been (co)-developed by Scytl, a company specialized in e-voting solutions. Each of those systems has been adapted to suit the requirements and needs of each jurisdiction.

**Contributions.** In this paper, we analyze the next generation of the Norwegian protocol [19, 18], that has been deployed and is being used in the Swiss cantons of Neuchâtel and Fribourg [30]. Our main contribution is a thorough analysis of the Neuchâtel protocol (as specified in [18, 31]) in ProVerif. The tool ProVerif [5, 6] is a state-of-the-art tool for the formal analysis of security protocols. We are able to prove *ballot privacy* (modelled as an equivalence property) as well as *cast-as-intended* and *recorded-as-cast verifiability* (modelled as a reachability property).

It should be noted that the Neuchâtel protocol, as well as most industrial-scale protocols, is not fully verifiable according to the academic tradition, since the content of the ballot box is not publicly disclosed. Instead, the protocol aims at providing *cast-as-intended* verifiability: if the voting server registers a ballot in the name of a voter then the ballot contains the vote *intended* by the voter, even if the voting device is corrupted and tries to cast a vote for another voting option. Cast-as-intended is achieved through return codes: after casting a vote, a voter receives a code and checks (on her voting card) that it matches the code displayed next to her intended choice. The Neuchâtel protocol additionally guarantees *recorded-as-cast* verifiability: if a voter completes the voting process then she is guaranteed that her ballot, as built by her voting device, has reached the voting server. These two verifiability properties hold under the assumption that the voting server is not compromised. Note that cast-as-intended verifiability is not offered by academic systems such as Civitas [11] and Belenios [12]. Often when it is offered quite a burden is placed on the voter: this is the case for Helios (cast-or-audit mechanism [1, 24]), where the voter needs to use two voting devices that are not simultaneously compromised. Namely, the device used for auditing the vote needs to be different from the device used to cast the vote.

In order to prove our security claims, we present a detailed model of the protocol, that includes for example the authentication phase after which the voter retrieves her voting credentials (i.e. by opening a password-protected keystore). Whereas such an initialization phase must be present in any real-world e-voting system, it is omitted in virtually every security analysis. Even worse, such authentication mechanism is typically not specified by the protocols in the academic literature. This is not necessarily surprising, and it might stem from the fact academic research artefacts often do not get used in practice and thus do not need to be described as a fully detailed system. Additionally, we capture in our model elections where voters can select  $k$  options among  $n$  voting options (while systems in the literature are often analysed in the case of elections

where voters selects 1 option among  $n$  or even 1 among 2).

We chose to perform an automated security analysis using ProVerif (instead of a manual proof) precisely to be able to model as many details as possible. We also believe that our model could serve as a basis for further studies of other voting protocols, as it is easier to adapt a symbolic model than a manual proof.

One difficulty we had to face resides in the fact that ProVerif does not handle well protocols with global states. This is of particularly critical importance in the case of the Neuchâtel protocol, since it becomes insecure (w.r.t. cast-as-intended) as soon as revoting is allowed. It is therefore crucial to model the fact that each voter votes “at most once”, and to do it in such a way that ProVerif can still handle the resulting model and provide a proof. Again, we believe that the techniques developed here to circumvent this issue are likely to be found useful elsewhere.

Our analysis mostly confirms the security of the protocol: we prove in ProVerif ballot privacy against a dishonest voting server as well as cast-as-intended and recorded-as-cast against a dishonest voting device, for an *unbounded number* of voting options and of voters. However, while modelling the protocol, we also discovered small variations thereof, which could realistically come up when implementing the protocol in a real scenario, that would render the protocol insecure in practice.

**Related Work.** The study most closely related to this paper is the analysis of the Norwegian protocol [16], that solely studies ballot privacy. The Norwegian protocol is an ancestor of the Neuchâtel protocol. The goal of the analysis in [16] was to provide a modelling as precise as possible of the protocol’s underlying primitives (with associativity and commutativity properties) and prove ballot privacy by hand, which was accomplished by developing some general lemmas regarding equivalence. A brief analysis was also performed in ProVerif but in a quite abstract model. Reusing the previous model was deemed to be not possible, as the Neuchâtel protocol has evolved quite significantly and the resulting equational theories for the atomic primitives are different.

Earlier research proposed the first symbolic models of electronic protocols. This includes a model of JCJ [3] and Helios [17, 15]. These models solely study privacy properties and consider a simpler scenario where voters select one candidate among a finite number of options. These protocols also allow revoting, which significantly simplifies the analysis in ProVerif (cf. the discussion above on dealing with a global state).

Computational proofs of privacy and/or some verifiability properties have been provided for Helios [4], Civitas [22], Select [26], Belenios [12] and [23] for example. Cryptographic models are more accurate w.r.t. the underlying primitives and consider a more powerful attacker, which may for example exploit algebraic properties of the primitives. Most of these proofs are done by hand, with the exception of [14], that provides a mechanized proof of ballot privacy for Helios-like protocols. Given their complexity, these proofs focus on the core of the protocol, abstracting away many details, including detailed analysis of the high-level interactions between the different parties of the protocol (e.g. computational proofs would typically assume a secure channel between the voter and the ballot box, without being explicit nor studying how this is done). The intrinsic complexity of computational proofs of elaborated protocols,

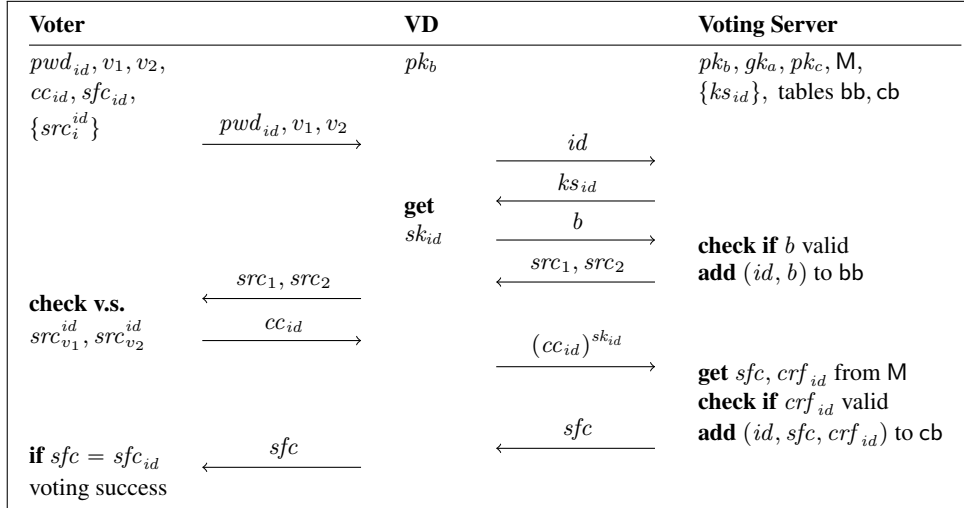


Figure 1: Overview of Neuchâtel Voting Protocol

as those used in e-voting systems, make those proofs even more error-prone and thus harder to verify.

## 2 Overview of the Neuchatel's system

The Neuchâtel voting protocol involves four main participants: the Voter (V), that casts a vote with the help of a Voting Device (VD); the Voting Server (S), that interacts with the voter's device to store the voter's ballot in a database, and next it computes return codes that need to be approved by the voter V; finally the Tallying Authority (T), that computes the result of the election by tallying the ballots database built by the server.

For the sake of clarity, we summarize the Neuchâtel protocol in the case of a selection of  $k = 2$  choices among  $n$  voting options  $v_1, \dots, v_n$ . Our analysis accounts however for several values of  $k$  other than  $k = 2$ . A synthetic view of the protocol is provided in Figure 1.

**Tallying Authority (T).** Creates an ElGamal asymmetric encryption key pair  $(pk_b, sk_b)$ . The public encryption key  $pk_b$  is communicated to the Registrar. Once the election is closed, the authority T computes and outputs the result of the election as follows. From each ballot to be counted, it extracts the ElGamal encryption of the vote. Next, T applies a mixnet to the resulting ciphertext list, decrypts every entry with the corresponding election decryption key  $sk_b$  and computes a zero-knowledge proof of correct decryption. The result is simply the multiset of the decrypted votes.

**Voter (V).** The voter V associated to  $id$  enters into her voting device VD her password  $pwd_{id}$  and her preferred voting choices  $v_1, v_2$ . At some point, VD displays short return codes  $src_1, src_2$  to the voter. The voter retrieves the codes  $src_{v_1}^{id}$  and  $src_{v_2}^{id}$  linked to  $v_1, v_2$  from her voting card. If the displayed and the retrieved code sets coincide, the

voter confirms her votes by entering the confirmation code  $cc_{id}$ . Finally, VD displays a short finalization code  $sfc$ , which should be equal to the code  $sfc_{id}$  on the voter's card. In this case, V is ensured that her ballot contains her intended voting options and that it has been accepted by the voting server S.

**Voting Device (VD).** The device VD uses the voter's password  $pwd_{id}$  to obtain a keystore  $ks_{id}$ , from which it retrieves the voters' key pair  $(pk_{id}, sk_{id})$ . Next, the device computes a ballot  $b = (\text{ctxt}, (v_1)^{sk_{id}}, (v_2)^{sk_{id}}, ec_1, ec_2, pk_{id}, P)$  that is received by the Voting Server. The ballot consists of several parts:

- $\text{ctxt} = aenc(pk_b, \phi(v_1, v_2), r)$  is an encryption of voting choices  $v_1, v_2$  using an ElGamal encryption algorithm  $aenc$ , and a bijective compacting function  $\phi$  that maps any list of integers to a single integer (this relies on the uniqueness of prime factors decomposition).
- so-called partial return codes  $(v_1)^{sk_{id}}, (v_2)^{sk_{id}}$  allow the Voting Server to compute the short return codes  $src_{v_1}^{id}, src_{v_2}^{id}$  that appear in the voter's voting card;
- the remaining components serve to guarantee consistency between ciphertext  $\text{ctxt}$  and the partial return codes  $(v_1)^{sk_{id}}, (v_2)^{sk_{id}}$ , by using a zero-knowledge proof P.

If the ballot is accepted by the server S, the voting device VD receives short return codes  $src_1, src_2$  from the server and displays them to the voter V. Next, on input of the finalization code  $cc_{id}$  entered by the voter, it computes and sends  $(cc_{id})^{sk_{id}}$  to the server. Finally, the server sends a short finalization code  $sfc$ , to be displayed to the voter.

**Voting Server (S).** It interacts with a voter V with identifier  $id$  through her voting device VD as follows. Firstly, the server receives the voter's identifier  $id$  from VD and replies with the corresponding keystore  $ks_{id}$ . Next it receives from VD a ballot  $b = (\text{ctxt}, (v_1)^{sk_{id}}, (v_2)^{sk_{id}}, ec_1, ec_2, pk_{id}, P)$ , and checks that it is a valid ballot. In particular, it verifies the zero-knowledge proof P and checks that voter V did not vote already. If valid, the ballot is stored in a database  $bb$ . From the partial return codes  $(v_1)^{sk_{id}}$  and  $(v_2)^{sk_{id}}$ , the server can compute the values  $rc_1$  and  $rc_2$  through a keyed pseudo-random function, and retrieve their corresponding short return codes  $src_1, src_2$  by looking into a table M. These short codes are sent to VD. Next, if the voter is satisfied with the return codes, S receives the value  $(cc_{id})^{sk_{id}}$ , from which it can compute  $fc$  (same way as  $rc$ ) and retrieve its short code  $sfc$  by looking again into M. The server S also retrieves a validity proof  $crf_{id}$  from M, that tells whether the retrieved code  $sfc$  is valid or not. If all the tests pass, S adds the confirmation values  $sfc$  and  $crf_{id}$  to the database (cb here), to keep track that the ballot  $b$  was successfully confirmed, and sends  $sfc$  to VD.

### 3 Framework

In the coming sections we present a ProVerif model of the Neuchatel protocol. A detailed presentation of the syntax and semantics of ProVerif can be found in [6]. For

$M, N, U ::=$	terms
$x \mid n \mid f(M_1, \dots, M_k)$	where $x \in \mathcal{V}, n \in \mathcal{N}$ , and $f \in \mathcal{C}$
$D ::=$	expressions
$M \mid h(D_1, \dots, D_k)$	where $h \in \mathcal{C} \cup \mathcal{D}$
$\phi ::=$	formula
$M = N \mid \phi_1 \wedge \phi_2 \mid \phi_1 \vee \phi_2 \mid \neg \phi$	
$P, Q ::=$	processes
$0$	nil
$\text{out}(N, M); P$	output
$\text{in}(N, x : T); P$	input
$P \mid Q$	parallel composition
$!P$	replication
$\text{new } a : T; P$	restriction
$\text{let } x : T = D \text{ in } P$	assignment
$\text{if } \phi \text{ then } P$	conditional
$\text{event}(M); P$	event

Figure 2: Syntax of the core language of ProVerif.

the sake of readability, we give next an overview of the protocol, focusing on the parts that are more relevant to our model. Notations and definitions are mainly borrowed from [6].

### 3.1 Syntax

We assume a set  $\mathcal{V}$  of variables, a set  $\mathcal{N}$  of names, a set  $\mathcal{T}$  of types. By default in ProVerif, types include *channel* for channel's names, and *bitstrings* for bitstrings (also written *any*). The syntax for *terms*, *expressions*, and *processes* is displayed in Figure 2.

*Terms and expressions.* Symbols for functions are split into two sets of constructors  $\mathcal{C}$  and destructors  $\mathcal{D}$  respectively. Terms are built over names, variables and constructors and represent actual messages sent over the network, while expressions may also contain destructors and represent cryptographic computations. Function symbols are given with their types:  $g(T_1, \dots, T_n) : T$  means that the function  $g$  takes  $n$  arguments as input of types respectively  $T_1, \dots, T_n$  and returns a result of type  $T$ . A substitution is a mapping from variables to terms, denoted  $\{U_1/x_1, \dots, U_n/x_n\}$ . The application of a substitution  $\sigma$  to a term  $U$ , denoted  $U\sigma$ , is obtained by replacing variables by the corresponding terms and is defined as usual. We only consider well typed substitutions.

The evaluation of an expression is defined through rewrite rules. Specifically, each destructor  $d$  is associated with a rewrite rule of the form  $d(U_1, \dots, U_n) \rightarrow U$ , over

terms. Then the evaluation of an expression is recursively defined as follows:

- $g(D_1, \dots, D_n)$  evaluates to  $U$ , denoted  $g(D_1, \dots, D_n) \Downarrow U$ , if  $\forall i, D_i \Downarrow U_i$ , and  $g$  is a constructor ( $g \in \mathcal{C}$ ) and  $U = g(U_1, \dots, U_n)$ ; or  $g$  is a destructor ( $g \in \mathcal{D}$ ) and there exists a substitution  $\sigma$  such that  $U_i = U'_i\sigma$ ,  $U = U'\sigma$ , where  $d(U'_1, \dots, U'_n) \rightarrow U'$  is the rewrite rule associated to  $g$ .
- $g(D_1, \dots, D_n)$  evaluates to fail, denoted  $g(D_1, \dots, D_n) \Downarrow \text{fail}$ , otherwise.

The evaluation  $\llbracket \phi \rrbracket$  of a formula is defined by  $\llbracket M = M \rrbracket = \top$ ,  $\llbracket M = N \rrbracket = \perp$  otherwise, and is then extended to  $\wedge, \vee, \neg$  as expected.

**Example 1.** To model the simple theory of encryption and concatenation, we consider a type *symkey* for symmetric keys and the sets of constructors and destructors with their associated rewrite rules as follows:

$$\begin{aligned} \mathcal{C}_{\text{basic}} &= \left\{ \begin{array}{l} \text{pair}(\text{any}, \text{any}) : \text{any}, \\ \text{enc}(\text{symkey}, \text{any}) : \text{any} \end{array} \right\} \\ \mathcal{D}_{\text{basic}} &= \left\{ \begin{array}{l} \text{proj1}(\text{any}) : \text{any}, \text{proj2}(\text{any}) : \text{any}, \\ \text{dec}(\text{symkey}, \text{any}) : \text{any} \end{array} \right\} \\ \mathcal{R}_{\text{basic}} &= \left\{ \begin{array}{l} \text{proj1}(\text{pair}(x, y)) \rightarrow x, \\ \text{proj2}(\text{pair}(x, y)) \rightarrow y, \\ \text{dec}(x, \text{enc}(y, x)) \rightarrow x \end{array} \right\} \end{aligned}$$

We often write  $(m_1, m_2)$  instead of  $\text{pair}(m_1, m_2)$  and  $(m_1, m_2, \dots, m_k)$  stands for  $(m_1, (m_2, (\dots, m_k)))$ .

*Processes.* Figure 2 provides a convenient abstract language for describing protocols (formally modeled as processes). The output of a message  $M$  on channel  $N$  is represented by  $\text{out}(N, M); P$  while  $\text{in}(N, x : T); P$  represents an input on channel  $N$ , stored in variable  $x$ . Process  $P \mid Q$  models the parallel composition of  $P$  and  $Q$ , while  $!P$  represents  $P$  replicated an arbitrary number of time.  $\text{new } a : T; P$  generates a fresh name of type  $T$  and behaves like  $P$ .  $\text{let } x : T = D \text{ in } P \text{ else } Q$  evaluates  $D$  and behaves like  $P$  unless the evaluation fails, in which case it behaves like  $Q$ .  $\text{event}(M); P$  is used to specify security property: the process emits an *event* (not observable by an attacker) to reflect that fact that it reaches some specific state, with some values, stored in  $M$ .

The set of free names of a process  $P$  is denoted  $\text{fn}(P)$ , and the set of its free variables by  $\text{fv}(P)$ . A *closed* process is a process with no free variables. Following ProVerif's handy notations, we may write  $\text{in}(c, = x).P$  instead of  $\text{in}(c, y : T).\text{if } x = y \text{ then } P$ , where  $T$  is the type of  $x$ . Similarly, we may write  $\text{in}(c, (x : T, y : T')).P$  instead of  $\text{in}(c, z : \text{any}).\text{let } x : T = \text{proj1}(z) \text{ in let } y : T' = \text{proj2}(z) \text{ in } P$ .

**Example 2.** In the Neuchâtel voting protocol, the Voter interacts with a Voting Device (e.g. her computer or cell phone) to cast her vote. Initially, the voter receives a voting card with her personal data for the election, including a password *pwd* (used to derive the voter's key and id), one short return code *src* for each candidate in the



```

Voter( $c, pwd, j_1, j_2, src_1, src_2, cc_{id}, sfc_{id}$ ) :=
  out( $c, (pwd, j_1, j_2)$ );           (* Sends password & choices. *)
  in( $c, (src'_1 : any, src'_2 : any)$ ); (* Gets the return codes. *)
  if ( $src'_1 = src_1 \wedge src'_2 = src_2$ )  $\vee$  ( $src'_1 = src_2 \wedge src'_2 = src_1$ ) then
    event( $confirmed(pwd, j_1, j_2)$ ); (* Checks ok; Reaches 'confirmed'. *)
    out( $c, cc_{id}$ );                 (* Confirms the vote. *)
    in( $c, =sfc_{id}$ );                (* Final confirmation. *)
    event( $happy(pwd, j_1, j_2)$ ).   (* All ok; Reaches 'happy'. *)

```

Figure 3: The Voter Process

list, a confirmation code  $cc_{id}$  (sent if the received return codes are valid), and a short finalization code  $sfc_{id}$  (that should correspond to the server's last acknowledgement message). For simplicity, we model an election where voters have to select two options. We model a voter that votes for two options  $j_1, j_2$ , with corresponding return codes  $src_1, src_2$  (read from the voting card). The corresponding process is defined in Figure 3. It communicates on channel  $c$  with the voting device. It includes two events that witness some important states of the voter. They will be used later to formally state security properties (see Section 5), and are defined by:

$$C_{\text{voter}} = \{ \text{confirmed}(\text{password}, \text{int}, \text{int}) : \text{any}, \\ \text{happy}(\text{password}, \text{int}, \text{int}) : \text{any} \}$$

### 3.2 Semantics

A configuration  $E, \mathcal{P}$  is given by a multiset  $\mathcal{P}$  of processes, representing the current state of the processes, and a set  $E = (\mathcal{N}_{\text{pub}}, \mathcal{N}_{\text{priv}})$  representing respectively the public and private names used so far. The semantics of processes is defined through a reduction relation  $\rightarrow$  between configuration, defined in Figure 4. A trace is a sequence of reductions between configurations  $E_0, \mathcal{P}_0 \rightarrow \dots \rightarrow E_n, \mathcal{P}_n$ . We say that a trace  $E_0, \mathcal{P}_0 \rightarrow^* E', \mathcal{P}'$  executes an event  $M$  if it contains a reduction  $E, \mathcal{P} \cup \{\text{event}(M); P\} \rightarrow E, \mathcal{P} \cup \{P\}$  for some  $E, \mathcal{P}, P$ .

### 3.3 Properties

As usual, we assume that protocols are executed in an untrusted network, meaning that communications over a public network are fully controlled by an attacker who may eavesdrop, intercept, or send messages. This is easily modeled by executing a protocol  $\mathcal{P}_0$  in parallel with an arbitrary process  $Q$ . Formally, we assume given a set of *public constructors*, subset of the constructors. An *adversarial process* w.r.t. to a set of names  $\mathcal{N}_{\text{pub}}$  is a process  $Q$  such that  $fn(Q) \subset \mathcal{N}_{\text{pub}}$  and  $Q$  uses only public constructors (and destructors). In what follows, all constructors are public, unless otherwise specified.

$$\begin{aligned}
E, \mathcal{P} \cup \{0\} &\rightarrow E, \mathcal{P} \\
E, \mathcal{P} \cup \{P \parallel Q\} &\rightarrow E, \mathcal{P} \cup \{P, Q\} \\
E, \mathcal{P} \cup \{!P\} &\rightarrow E, \mathcal{P} \cup \{P, !P\} \\
(\mathcal{N}_{pub}, \mathcal{N}_{priv}), \mathcal{P} \cup \{\text{new } a : T; P\} &\rightarrow (\mathcal{N}_{pub}, \mathcal{N}_{priv} \cup \{a'\}), \mathcal{P} \cup \{P[a'/a]\} \quad \text{where } a' \notin \mathcal{N}_{pub} \cup \mathcal{N}_{priv} \\
E, \mathcal{P} \cup \{\text{out}(N, M); Q, \text{in}(N, x); P\} &\rightarrow \\
&E, \mathcal{P} \cup \{Q, P[M/x]\} \\
E, \mathcal{P} \cup \{\text{let } x = D \text{ in } P\} &\rightarrow E, \mathcal{P} \cup \{P[M/x]\} \\
&\quad \text{if } D \Downarrow M \text{ and } M \neq \text{fail} \\
E, \mathcal{P} \cup \{\text{if } \phi \text{ then } P\} &\rightarrow E, \mathcal{P} \cup \{P\} \quad \text{if } \llbracket \phi \rrbracket = \top \\
E, \mathcal{P} \cup \{\text{event}(M); P\} &\rightarrow E, \mathcal{P} \cup \{P\}
\end{aligned}$$

Figure 4: Transitions between configurations.

### 3.3.1 Correspondence

Many security properties can be stated as “if Alice reaches some state (e.g. finishes her session) then Bob must have engage a conversation with her”. This is for example the case of many variants of agreement properties [27]. ProVerif allows to specify *correspondence* properties between events.

**Definition 1.** *A closed process  $P_0$  satisfies the correspondence*

$$P_0 \rightsquigarrow \bigwedge_{i=1}^m \bigvee_{j=1}^{l_i} \text{event}(M_{ij})$$

where the  $M_{i,j}$  do not contain names, if for any (adversarial) closed process  $Q$  such that  $\text{fn}(Q) \subset \text{fn}(P_0)$ , for any trace  $tr$ , for any substitution  $\sigma$ , if  $tr$  executes event  $M\sigma$ , then there exists  $\sigma'$  such that  $M\sigma = M\sigma'$  and for any  $i$ , there exists  $j$  such that  $tr$  executes event  $M_{ij}\sigma'$ .

Examples can be found in Section 5.

### 3.3.2 Equivalence

Observational equivalence of two processes  $P$  and  $Q$  models the fact that an adversary cannot distinguish between the two processes. Slightly more precisely, whenever  $P$  may emit on some channel  $c$  (interacting with an adversarial process  $R$ ), then  $Q$  can emit on  $c$  as well. For readability, we summarise here the definition of exquivalence from [6]. We write  $\mathcal{C} \downarrow_N$  when a configuration  $\mathcal{C} = E, \mathcal{P}$  with  $E = (\mathcal{N}_{pub}, \mathcal{N}_{priv})$  can output on some channel  $N$ , i.e. if there exists  $\text{out}(N, M); P \in \mathcal{P}$  such that  $\text{fn}(N) \in \mathcal{N}_{pub}$ . Also, an adversarial context  $C[-]$  is a process of the form  $\text{new } n : \text{bitstrings}; \_ \mid Q$  where  $\text{fv}(Q) = \emptyset$  and all functional symbols in  $Q$  are public, with  $\_$  being a ‘hole’ expected to be filled by a configuration  $\mathcal{C} = (\mathcal{N}_{pub}, \mathcal{N}_{priv}), \mathcal{P}$ . Therefore, and assuming that  $\mathcal{N}_{priv} \cap \text{fn}(Q) = \emptyset$ , the application of one to the other is

defined by :

$$\begin{aligned}
C[\mathcal{C}] &= (\mathcal{N}'_{pub}, \mathcal{N}'_{priv}), \mathcal{P} \cup \{Q\} \\
\text{with } \mathcal{N}'_{pub} &= (\mathcal{N}_{pub} \cup fn(Q)) \setminus \{n\} \\
\text{and } \mathcal{N}'_{priv} &= \mathcal{N}_{priv} \cup \{n\}
\end{aligned}$$

From this, the definition of observational equivalence follows :

**Definition 2.** *The Observational equivalence between configurations, denoted by  $\approx$ , is the largest symmetric relation such that  $\mathcal{C} \approx \mathcal{C}'$  implies :*

- if  $\mathcal{C} \downarrow_N$  then  $\exists \mathcal{C}'_1$  s.t.  $\mathcal{C}' \rightarrow^* \mathcal{C}'_1$  and  $\mathcal{C}'_1 \downarrow_N$ ;
- if  $\mathcal{C} \rightarrow \mathcal{C}_1$ , then  $\exists \mathcal{C}'_1$  s.t.  $\mathcal{C}' \rightarrow^* \mathcal{C}'_1$  and  $\mathcal{C}_1 \approx \mathcal{C}'_1$ ;
- $C[\mathcal{C}] \approx C[\mathcal{C}']$ , for any adversarial context  $C[-]$ .

## 4 Formal model of the Neuchâtel's voting system

We present here the main parts of our formal model. For simplicity, we present a model for elections with  $k = 2$  choices to be made amongst  $n$  voting options. In our automated analysis, we generate the model corresponding to any given particular value  $k$  automatically, and then run ProVerif on several values of  $k$  (up to  $k = 4$  due to ProVerif time out).

### 4.1 Standard primitives

The Neuchâtel protocol makes use of the standard primitives: symmetric and asymmetric encryption, signatures, hashes and concatenation. We consider the corresponding types *agentId*, *int*, *ekey*, *epkey*, *skey*, *spkey*, *symkey*. The constructors and the associated rewrite rules are defined as follows.

$$\begin{aligned}
\mathcal{C}_{\text{stand}} &= \{ \\
&\quad sk_e(\text{agentId}) : ekey, \\
&\quad pub_e(ekey) : epkey, \\
&\quad aenc(epkey, any, int) : any, \\
&\quad enc(symkey, any) : any, \\
&\quad pair(any, any) : any, \\
&\quad sk_s(\text{agentId}) : skey, \\
&\quad pub_s(skey) : spkey, \\
&\quad sign(skey, any) : any \quad \} \\
\mathcal{R}_{\text{stand}} &= \{ \\
&\quad adec(k, aenc(pub_e(k), m, r)) \rightarrow m \\
&\quad dec(k, enc(k, m)) \rightarrow m \\
&\quad proj1(pair(a, b)) \rightarrow a \\
&\quad proj2(pair(a, b)) \rightarrow b \\
&\quad verify(pub_s(k), m, sign(k, m)) \rightarrow \text{ok} \quad \}
\end{aligned}$$

where all function symbols are public, except  $sk_e$  and  $sk_s$  that are private. The set of destructors  $\mathcal{D}_{\text{stand}}$  can be inferred easily. The term  $sk_\alpha(id)$  represents the private key of user  $id$  w.r.t. scheme  $\alpha$ , where  $\alpha = e$  stands for asymmetric encryption while  $\alpha = s$  stands for signature. The rewrite rules are the standard ones for these primitives. For example  $adec(k, aenc(pub_e(k), m, r)) \rightarrow m$  models the fact that the plaintext of an (asymmetric) encryption can be retrieved by decrypting with the corresponding private key.

## 4.2 Voting device

A voter  $id$  is provided with a password  $pwd$  and codes (return codes as well as a confirmation code). When she connects to the voting server through her voting device, she first needs to retrieve her personal private key  $sk_e(\delta_{Id}(pwd))$ . The identifier  $id$  of the voter is actually derived from her password, that is  $id = \delta_{Id}(pwd)$ . This key is stored in a keystore (on the server's side), encrypted with a key that can be derived from the password:  $\delta_{Key}(pwd)$ . We therefore introduce the following theory to model the key store.

$$\begin{aligned} \mathcal{C}_{ks} &= \{ \delta_{Id}(password) : agentId, \\ &\quad \delta_{Key}(password) : symkey, \\ &\quad c_{ekey}(ekey) : any \} \\ \mathcal{R}_{ks} &= \{ c_{any}(c_{ekey}(k)) \rightarrow k \} \end{aligned}$$

The corresponding set of destructors  $\mathcal{D}_{ks}$  can be inferred easily. The functions  $c_{any}$  and  $c_{ekey}$  are auxiliary functions that convert private keys to bitstring and conversely.

We denote

$$ks(pwd) := enc(\delta_{Key}(pwd), c_{ekey}(sk_e(\delta_{Id}(pwd))))$$

the encrypted value stored in the key store for voter  $\delta_{Id}(pwd)$ .

As explained in Section 2, the voting device builds a ballot as follows. It encrypts the choices  $j_1, j_2$  of the voters; builds pre-return codes for each choice  $prc(sk_e, j_1)$ ,  $prc(sk_e, j_2)$  using the private key of the voter; and proves that the return codes correspond to the encrypted votes, through a zero-knowledge proof. This is modeled as follows.

$$\mathcal{C}_{zk} = \{ zkp(epkey, epkey, any, int, int, int, ekey) : any, \\ prc(ekey, int) : int \}$$

$$\begin{aligned} \mathcal{R}_{zk} &= \{ verif_{zkp}(pk_b, pk, e, p_1, p_2, \\ &\quad zkp(pk_b, pk, e, p_1, p_2, r, sk_{id}) \rightarrow ok \} \\ &\quad \text{with } pk = pub_e(sk_{id}), e = aenc(pk_b, v, r), \\ &\quad p_1 = prc(sk_{id}, v_1), p_2 = prc(sk_{id}, v_2) \end{aligned}$$

with  $v = (v_1, v_2)$ . Note that here, our model abstracts some properties of the primitives. Indeed, as explained in Section 2, the two choices  $j_1, j_2$  of the voters are not

encrypted as a list but are “compacted” in a single integer  $\Phi(j_1, j_2)$ , which is actually commutative:  $\Phi(j_1, j_2) = \Phi(j_2, j_1)$ . Similarly, the zero-knowledge proof compacts the pre-return codes. Since ProVerif cannot handle associative and commutative properties, we abstract away these properties, assuming a slightly stronger proof system.

We can now provide the Device process, which is shown in Figure 5.

### 4.3 Voting server

When the voting server is contacted by some voter  $id$  (through her voting device), the server first needs to retrieve the personal keystore associated to  $id$ . Only valid  $ids$ , of the form  $id = \delta_{Id}(pwd)$ , are registered. The server also recovers (for later use) a signature  $sign(sk_c, sfc(gk_a, id))$  of the short finalization code  $sfc(gk_a, id)$  that should be rebuilt and sent at the end by the server (defined in  $\mathcal{C}_{rc}$  below). The signing key  $sk_c$  is a fixed long term key of the setup authorities and  $gk_a$  is the global audit key of the server. We model this data retrieval by considering the following rewrite rule:

$$\mathcal{R}_{\text{retrieve}} = \{ \text{Get}(\delta_{Id}(pwd), gk_a) \rightarrow (ks(pwd), sign(sk_c, sfc(gk_a, id))) \}$$

The server needs to compute a long return code  $f(gk_a, prc_1)$  from the partial return code  $prc_1$  sent by the voter. This long return code is too long to be human readable and is therefore associated (in a table) to a short return code. This table also provides a correspondence for long and short finalization codes. This is modeled through the following theory:

$$\mathcal{C}_{rc} = \{ \begin{array}{ll} src(symkey, agentId, int) & : any \\ sfc(symkey, agentId) & : any \\ f(symkey, int) & : symkey \\ cc(agentId) & : int \end{array} \}$$

$$\mathcal{R}_{rc} = \{ \begin{array}{l} read_{RC}(f(gk_a, prc(sk_e(id), j))) \\ \rightarrow src(gk_a, id, j) \\ read_{FC}(f(gk_a, prc(sk_e(id), cc(id)))) \\ \rightarrow sfc(gk_a, id) \end{array} \}$$

with  $cc$ ,  $src$  and  $sfc$  private function symbols. Actually, short return codes are stored encrypted using the long return code as a symmetric key. For the sake of clarity, we omit this part here but it is reflected in our ProVerif’s model.

*No revote.* One of the main challenging tasks when modeling the voting server is the fact that it accepts at most one request from each voter. Note that the protocol is insecure otherwise. Indeed, in case revoting was allowed, a malicious voting device could first vote as queried by the voter and display the (correct) return codes and then revotes for the candidate of its choice (discarding the corresponding return codes). Therefore, it is crucial to model accurately that no voter can vote twice.

A first approach (as described in [18]) is to use a table that stores whether a voter already voted or not. Then, intuitively, the code of the server is (informally) as follows:

```

if  $id \notin \text{Table}$     then proceed and add  $id$  to Table
                        else stop

```

However, since the voting device may process several requests at the same time, checking whether  $id \notin \text{Table}$  is actually insufficient. Indeed, if two requests from the same voter reach the server at the same time, they would both pass the test  $id \notin \text{Table}$  and both ballots would be accepted. We therefore need a clean lock mechanism, for which there is a lot of implementation support.

However, when it comes to modeling this lock mechanism in ProVerif, we have two options. The first option is to encode the lock mechanism directly in ProVerif, for example using private channels, that can be used as “tokens”. This approach presents two drawbacks. Firstly, this encoding would necessarily be ad-hoc. And in principle there would be no assurance that the protocol is secure if the lock mechanism is implemented in another way (and of course, real lock mechanisms will never use private channels). Secondly, it is also known that ProVerif, due to its internal behavior, cannot properly handle private channels when used as tokens or, more generally, cannot handle properly events that happen “at most once”.

Instead, we take a different point of view. We model a voting server that does *not* prevent revoting and adds blindly new ballots. Then, instead of asking ProVerif whether some property  $\phi$  holds, we query a property of the form

$\phi \vee$  two ballots have been accepted for the same voter

If ProVerif proves this query, it guarantees that for any execution trace where no 2 ballots are accepted by the server, then  $\phi$  holds. It is then up to the implementation to ensure that any execution is such that no 2 ballots for the same voter are accepted by the server.

This yields a more flexible result: for any realization of the protocol that further ensures that “no two ballots from the same voters are accepted”, then the realization satisfies  $\phi$ , no matter how the “no revoke policy” is actually implemented. We believe that this approach is of independent interest and could re-used in other contexts when modeling e.g. lock mechanisms in ProVerif.

Formally, the server simply issues events built from  $\mathcal{C}_{\text{BB}} = \{\text{InsertBB}(\text{agentId}, \text{any}) : \text{any}\}$  to record the fact that a ballot has been added to the box. The process corresponding to the voting server is then defined in Figure 6. These events are used later to specify security properties as just discussed.

#### 4.4 Tally process

Once the voting process is over, the tally phase can start. The tally authorities check the validity of each ballot (same checks as the server); mix the ciphertexts containing the votes; decrypt the mixed ciphertexts and publish the election result. In our analysis, we consider that either the tally authorities are corrupted (for cast-as-intended and

```

Device( $c_1, c_2, pk_b$ ) := (* channels  $c_1$  with Voter and  $c_2$  with Server. *)
  in( $c_1, (pwd : password, j_1 : int, j_2 : int)$ ); (* Get password & choices. *)
  out( $c_2, \delta_{Id}(pwd)$ ); (* Ask for the keystore, *)
  in( $c_2, ks : any$ ); (* ... receive keystore, *)
  let  $sk_{id} : ekey = c_{any}(dec(\delta_{Key}(pwd), ks))$  in (* ... retrieve the key. *)
  new  $r : int$ ; let  $e = aenc(pk_b, (j_1, j_2), r)$  in (* Encrypt voter's choices. *)
  let  $p = zkp(pk_b, pub_e(sk_{id}), e, prc(sk_{id}, j_1), prc(sk_{id}, j_2), r, sk_{id})$  in
  out( $c_2, (e, prc(sk_{id}, j_1), prc(sk_{id}, j_2), pub_e(sk_{id}), p)$ ); (* Sends the ballot. *)
  in( $c_2, (src_1 : any, src_2 : any)$ ); (* Get the short return codes *)
  out( $c_1, (src_1, src_2)$ ); (* ... transmit codes. *)
  in( $c_1, cc_{id} : int$ ); (* Get confirmation code *)
  out( $c_2, prc(sk_{id}, cc_{id})$ ); (* ... transmit it. *)
  in( $c_2, sfc_{id} : any$ ); (* Get short finalization code *)
  out( $c_1, sfc_{id}$ ). (* ... transmit it. *)

```

Figure 5: The Device Process

```

Server( $c : channel, pk_b : epkey, gk_a : symkey, pk_c : spkey, c_t : channel$ ) :=
  in( $c, id : agentId$ ); (* New voting requests. *)
  let ( $ks : any, crf : any$ ) = Get( $id, gk_a$ ) in (* Recovers the keystore, *)
  out( $c, ks$ ); (* ... and transmits it. *)
  in( $c, b : any$ ); (* Waits for a ballot. *)
  let ( $e : any, prc_1 : int, prc_2 : int, =pk_e(id), p : any$ ) =  $b$  in (* Parse it. *)
  if  $verify_{zkp}(pk_b, pk_e(id), e, prc_1, prc_2, p)$  then (* Checks the proof. *)
    event(InsertBB( $id, e$ )); (* Table addition. *)
    let  $src_1 = read_{RC}(f(gk_a, prc_1))$  in (* Gets the short return codes. *)
    let  $src_2 = read_{RC}(f(gk_a, prc_2))$  in
    out( $c, (src_1, src_2)$ ); (* Sends them to the Voter. *)
    ! in( $c, cm : int$ ); (* Waits for confirmation. *)
    let  $sfc_{id} : any = read_{FC}(f(gk_a, cm))$  in (* Gets the finalization code. *)
    if  $verify(pk_c, sfc_{id}, crf)$  in (* Checks the signature. *)
    event(HasVoted( $id, e$ )); (* Vote approval. *)
    out( $c, sfc_{id}$ ); out( $c_t, (id, b, sfc_{id}, crf)$ ) (* Confirms; Feeds the Tally. *)

```

Figure 6: The Server Process

recorded-as-cast properties) or that the overall tally process is honest (for ballot privacy). As we shall see in Section 5 (security properties), we only need to consider two honest voters, as well as arbitrary many dishonest voters.

Since privacy is ensured as soon as the two honest ballots are mixed, we model a tally process that mixes the two honest ballots only. The mixnet is modeled in a standard way, by sending the ballots over a private channel called `mix` concurrently, thus without fixing the order, and reading them back from that same channel. Formally, we define two processes: *TallyH*, as shown in Figure 7 for the honest voters (whose ballots are mixed); and *TallyD* as shown in Figure 8 for any dishonest voter (that can be executed arbitrarily). The process *TallyD* simply decrypts any (valid) ballot provided the corresponding *id* is not an honest voter, that is, is neither  $id_a$  nor  $id_b$ .

## 5 Security properties

The Neuchâtel protocol is designed to achieve cast-as-intended verifiability: even if the voter’s device is corrupted, the ballot registered in the name of a voter corresponds to the vote *intended* by *that* voter. This offers a strong protection against attackers-controlled personal computers and smartphones (e.g. through malware). The voting server shall be trusted for this step. Conversely, the Neuchâtel’s protocol also guarantees vote privacy against a dishonest voting server. Note however that the protocol is not publicly verifiable: the content of the ballot box is not public and therefore voters cannot check that the result corresponds to the received ballots. We define next the security properties proved in ProVerif.

### 5.1 Verifiability properties

The protocol ensures cast-as-intended and recorded-as-cast: if a voter successfully completes the voting procedure, she is guaranteed that her ballot has been property recorded by the voting server.

*Cast-as-Intended.* The Neuchâtel’s protocol provides *Cast-as-Intended* verifiability: if the server registers a ballot for some voter *id* then this ballot contains the votes *intended* by the voter. This can be formalized by the following correspondence property. Remember that the identity *id* of a voter is derived from her password ( $id = \delta_{Id}(pwd)$ ).

$$\begin{aligned}
& \text{event}(\text{HasVoted}(pk_b, \delta_{Id}(pwd_a), e)) \Rightarrow \\
& \exists v_1, v_2, j_1, j_2, r, \\
& \quad \text{event}(\text{confirmed}(pwd_a, v_1, v_2)) \\
& \quad \wedge e = \text{aenc}(pk_b, (j_1, j_2), r) \\
& \quad \wedge ((j_1 = v_1 \wedge j_2 = v_2) \vee (j_1 = v_2 \wedge j_2 = v_1))
\end{aligned} \tag{1}$$

Intuitively, the above reads as follows: if the server issues an event  $\text{HasVoted}(pk_b, \delta_{Id}(pwd_a), e)$ , meaning that he accepted a ballot containing an encryption  $e$ , then the voter with password  $pwd_a$  must have had cast a vote  $(v_1, v_2)$  that corresponds to  $e$ . Note that we cannot exclude the case where a malicious device swaps



the vote (that is, casts  $(v_2, v_1)$  instead of  $(v_1, v_2)$ ) and then swaps the received return codes. This is captured in the property above by allowing the two cases (the option order has no impact on the way votes are counted).

Unfortunately, ProVerif fails to prove this property. Indeed, cast-as-intended cannot be guaranteed as soon as the server may answer two requests from the same voter, as the attacker would then get two sets of return codes and could show the wrong one. This is explicitly forbidden by the Neuchâtel’s protocol: the server does not answer to re-vote queries. However, ProVerif over-approximates the behaviors and takes into account the case where the server would answer twice (yielding “cannot be proved”).

Instead, we consider the following correspondence property.

$$\begin{aligned}
& \text{event}(\text{Has Voted}(pk_b, \delta_{Id}(pwd_a), e)) \Rightarrow \\
& \exists v_1, v_2, j_1, j_2, j_3, j_4, j_5, j_6, r, r_1, r_2, \\
& \quad \text{event}(\text{confirmed}(pwd_a, v_1, v_2)) \\
& \quad \wedge e = \text{aenc}(pk_b, (j_1, j_2), r) \\
& \quad \wedge \text{event}(\text{InsertBB}(\delta_{Id}(pwd_a), e)) \\
& \quad \wedge \text{event}(\text{InsertBB}(\delta_{Id}(pwd_a), \text{aenc}(pk_b, (j_3, j_4), r_1))) \\
& \quad \wedge \text{event}(\text{InsertBB}(\delta_{Id}(pwd_a), \text{aenc}(pk_b, (j_5, j_6), r_2))) \\
& \quad \wedge (j_3 = v_1 \vee j_4 = v_1) \wedge (j_5 = v_2 \vee j_6 = v_2)
\end{aligned} \tag{2}$$

This property states that if the server accepts a ballot containing an encryption  $e$  then the voter corresponding to  $pwd_a$  must have cast a vote  $(v_1, v_2)$  such that the encryption  $e$  has been inserted in the ballot box, on behalf of  $\delta_{Id}(pwd_a)$ . Moreover, there must have been two other insertions in the ballot box: one with  $(j_3, j_4)$  and one with  $(j_5, j_6)$ , such that  $v_1$  is equal to either  $j_3$  or  $j_4$  and  $v_2$  is equal to either  $j_5$  or  $j_6$ .

Why is this useful? Because we know that the protocol’s implementation further guarantees that there is at most one insertion for each voter. Combined with Property 2, this implies  $j_1 = j_3 = j_5 \wedge j_2 = j_4 = j_6$ , hence the desired Property 1 (since  $v_1 \neq v_2$ ).

More formally, for any trace  $\text{tr}$  of a process, if  $\text{tr}$  satisfies (2) and is such that there no distinct insertion for the same voter, that is:

$$\begin{aligned}
& \text{event}(\text{InsertBB}(\delta_{Id}(pwd_a), \text{aenc}(pk_b, (j_3, j_4), r_1))) \in \text{tr} \\
& \wedge \text{event}(\text{InsertBB}(\delta_{Id}(pwd_a), \text{aenc}(pk_b, (j_5, j_6), r_2))) \in \text{tr} \\
& \Rightarrow (j_3 = j_5) \wedge (j_4 = j_6) \wedge (r_1 = r_2)
\end{aligned}$$

then  $\text{tr}$  satisfies (1). The proof is immediate.

The first interesting feature of this encoding is of course that it circumvents the issue that ProVerif over-approximates the no-revote policy. The second feature is that Property 1 is proved *independently* of the exact implementation of the no-revote policy. Assuming that the implementation guarantees that there is at most one insertion for each voter, then the protocol achieves cast-as-intended, no matter how this is implemented in practice.

*Recorded-as-cast.* The protocol further guarantees that if a voter completes the voting process then she is ensured that her vote has been recorded by the server. This

property can be formally stated as follows.

$$\begin{aligned}
& \text{event}(\text{happy}(pwd_a, v_1, v_2)) \Rightarrow \\
& \exists j_1, j_2, r, \\
& \quad \text{event}(\text{HasVoted}(pk_b, \delta_{Id}(pwd_a), e)) \\
& \quad \wedge e = \text{aenc}(pk_b, (j_1, j_2), r) \\
& \quad \wedge (j_1 = v_1 \wedge j_2 = v_2) \vee (j_1 = v_2 \wedge j_2 = v_1)
\end{aligned} \tag{3}$$

In our ProVerif model, we further show that the ballot registered by the server is well-formed and will therefore be accepted at the tally phase. Similarly to cast-as-intended, this property cannot be proved in ProVerif. So instead, we prove an amended property which implies the desired property as soon as the implementation guarantees that there is at most one ballot insertion per voter.

$$\begin{aligned}
& \text{event}(\text{happy}(pwd_a, v_1, v_2)) \Rightarrow \\
& \exists j_1, j_2, j_3, j_4, j_5, j_6, r, r_1, r_2, \\
& \quad \text{event}(\text{HasVoted}(pk_b, \delta_{Id}(pwd_a), e)) \\
& \quad \wedge e = \text{aenc}(pk_b, (j_1, j_2), r) \\
& \quad \wedge \text{event}(\text{InsertBB}(\delta_{Id}(pwd_a), e)) \\
& \quad \wedge \text{event}(\text{InsertBB}(\delta_{Id}(pwd_a), \text{aenc}(pk_b, (j_3, j_4), r_1))) \\
& \quad \wedge \text{event}(\text{InsertBB}(\delta_{Id}(pwd_a), \text{aenc}(pk_b, (j_5, j_6), r_2))) \\
& \quad \wedge (j_3 = v_1 \vee j_4 = v_1) \wedge (j_5 = v_2 \vee j_6 = v_2)
\end{aligned} \tag{4}$$

*Trust Assumptions.* We prove cast-as-intended (Property 2) and recorded-as-cast (Property 4) even if the voting device and the tally process are corrupted. We assume however the voting server to be honest. Formally, we consider the following process:

$$\begin{aligned}
& \text{Init}; !\text{Server}(c, pk_b, gk_a, sk_c, c) \mid !\text{Corr} \\
& \quad \mid \text{Voter}(c, pwd_a, v_1, v_2, \text{src}(gk_a, id_a, v_1), \\
& \quad \quad \text{src}(gk_a, id_a, v_2), \text{cc}(id_a), \text{sfc}(gk_a, id_a))
\end{aligned}$$

where *Init* is an initialization process: it broadcasts  $id_a = \delta_{Id}(pwd_a)$  and  $pk_e(id_a)$  on some public channel  $c$ , generates the elections keys  $sk_b$ ,  $gk_a$  and  $sk_c$ , and publishes  $pk_b = \text{pub}_e(sk_b)$ ,  $\text{pub}_s(sk_c)$ , and  $sk_b$ . The process *!Corr* models an arbitrary number of dishonest voters: each generates a password  $pwd_i$  for some voter  $i$  and broadcasts all the corresponding (public and private) data.

Note that we do not need to include the tally process since it is assumed to be dishonest. Instead, we simply provide the election key ( $sk_b$ ) to the adversary: cast-as-intended and recorded-as-cast properties are guaranteed even if the decryption key is lost.

## 5.2 Privacy

Intuitively, a voting protocol guarantees *ballot privacy* if an attacker cannot learn any information about how a voter voted. In symbolic models, this is typically formalized as follows [25, 17]:

$$V_A(0) \mid V_B(1) \approx V_A(1) \mid V_B(0)$$

An attacker should not be able to distinguish the case where Alice votes 0 and Bob votes 1 from the case where the votes are swapped.

We show that the Neuchâtel’s protocol ensures ballot privacy, even if the voting server and all but two voters (and their voting devices) are corrupted. Formally, we consider a set up process  $I$  similar to  $Init$ , but with two honest voters  $A, B$  with passwords  $pwd_A, pwd_B$ , and a corrupter  $Server$  process that leaks its private data  $gk_a, sk_c$ . The election key  $sk_b$  (generated during the setup and distributed to election authorities) is assumed to be secret. We call  $V_A(a, b)$  the process for voter  $A$  that votes for  $a$  and  $b$ .

$$\begin{aligned} V_A(a, b) &:= Device(c_A, c, pk_b) \\ &| Voter(c_A, pwd_A, a, b, src(gk_a, id_A, a), \\ &\quad src(gk_a, id_A, b), cc(id_A), sfc(gk_a, id_A)) \end{aligned}$$

Then we prove (in ProVerif) that:

$$\begin{aligned} &I | V_A(a, b) | V_B(c, d) | !Corr | T \\ \approx &I | V_A(c, d) | V_B(a, b) | !Corr | T \end{aligned} \quad (5)$$

$$\begin{aligned} \text{with } T &:= TallyH(c, sk_b, pk_c, id_A, id_B) \\ &| !TallyD(c, sk_b, pk_c, id_A, id_B) \end{aligned}$$

ProVerif cannot prove arbitrary equivalences. Instead, it proves diff-equivalence (a stronger notion of equivalence [7]) of pairs of processes that only differ in treatment of terms. More formally, ProVerif considers bi-processes  $P$  that may contain bi-terms  $choice(t_1, t_2)$  instead of pure terms. Then ProVerif proves equivalence of  $proj_1(P)$  and  $proj_2(P)$  where  $proj_i(P)$  is obtained from  $P$  by replacing any occurrence of a bi-term  $choice(t_1, t_2)$  by  $t_i$ . We refer the reader to [7] for a detailed and formal definition of bi-processes and diff-equivalence.

So proving the equivalence  $V_A(0) | V_B(1) \approx V_A(1) | V_B(0)$  amounts into considering the process

$$V_A(choice(0, 1)) | V_B(choice(1, 0))$$

However, applying directly this transformation to the equivalence (5) yields a process that ProVerif cannot prove. Instead, we need to further transform it by also swapping the output of the tally. This is a usual technique, as devised e.g. in [8]. Formally, we replace the  $out(mix, e_a) | out(mix, e_b)$  part in the  $TallyH$  process by:

$$out(mix, choice(e_a, e_b)) | out(mix, choice(e_b, e_a))$$

Note that since the resulting processes are equivalent (since  $P | Q \approx Q | P$ ), we still prove the desired equivalence (Property 5).

## 6 Results and lessons learned

Previous symbol models of electronic voting protocols [3, 17, 15, 16] consider a simple scenario where the voter selects one candidate among finitely many options. In

	Voting Device	Server	Tally
Cast-as-Intended	D	H	D
Recorded-as-Cast	D	H	D
Ballot Privacy	H	D	H

Table 1: Properties and trust assumptions. D stands for dishonest while H stands for honest.

Number of option’s choices	1	2	3	4	5
Cast-as-Intended	< 1s	< 1s	2s	8m	time out > 48h
Recorded-as-Cast	< 1s	< 1s	3s	20m	time out > 48h
Ballot Privacy	14s	49m	time out > 48h	time out > 48h	time out > 48h

Table 2: Security analysis in ProVerif.

this study and for the sake of clarity, we have presented the Neuchâtel’s protocol for the particular case of an election where  $k = 2$  options among  $n$  options need to be selected. In our ProVerif model we have considered an arbitrary number of options  $n$ , an arbitrary number of voters  $m$ , and several values for the number of selections  $k$ . To be able to cope with an arbitrary number of selections, we would need to handle lists of arbitrary size (representing the selection of a voter). While there are some preliminary results for protocols with lists [10, 2, 28], none of them can be applied to our symbolic model for the Neufchâtel protocol. This is why we consider several fixed values for  $k$ .

The security properties together with corresponding the trust assumptions are summarized in Table 1 while the experiments are presented in Table 2.

## 6.1 Results

We run ProVerif version 1.94 on a Xeon E5-2687W v3 @ 3.10GHz. We were able to analyze cast-as-intended and individual verifiability up to  $k = 4$  and ballot privacy up to  $k = 2$ . The detailed analysis times are reported in Table 2. The models in ProVerif can be found in [13]. As explained in Section 5.2, ballot privacy is expressed as an equivalence property of the form:

$$V_A(a, b) \mid V_B(c, d) \approx V_A(c, d) \mid V_B(a, b)$$

for  $k = 2$ , where  $a, b, c, d$  are constants. This implicitly means that  $A$  and  $B$  vote for distinct options. So in the case  $k = 2$  we further prove privacy when the two

honest voters were respectively voting  $(a, b)$ ,  $(b, c)$ , or  $(a, b)$ ,  $(c, b)$ , or  $(a, b)$ ,  $(b, a)$ , to check that no attack appears when  $A$  and  $B$  share one or two options. ProVerif proves these cases in exactly the same time than the case where the four options are pairwise distinct.

## 6.2 Lessons learned

Our analysis mostly confirms the security of the Neuchâtel protocol w.r.t. both privacy and verifiability properties. However, while modeling the protocol we discovered that flaws may occur in case of small but realistic deviations of the protocol. We reported these subtleties to the the company that designed the system who confirmed to be aware of them and that they have been taken care of for the actual implementation.

*Exactly  $k$  choices.* In case of an election where voters select  $k$  options out of  $n$ , with  $k \geq 2$ , voters (and authorities) should be aware (and properly instructed) that voters should select *exactly*  $k$  options (and not less). Otherwise, a dishonest voting device may use the remaining “unused” choices for other unexpected voting options (and discard their return codes).

*No duplicate.* Still in case of an election where voters select  $k$  options out of  $n$ , voters should not be offered the possibility to vote twice for the same option (for example, the election rules may allow voters to give two choices to the same candidate). Indeed, the protocol would then be vulnerable to an attack where the intruder uses the duplicated choice (say Alice votes twice for  $a$ , that is, she votes  $a, a$ ) to make her vote for  $a, b$  (and manually duplicate the return code corresponding to  $a$ , to make Alice happy).

*Blank vote.* It becomes particularly tricky for elections that allow voters to abstain (that is, vote “blank”). In that case,  $k$  different blank voting options must be provided to voters. Those blank voting options shall have different individual return codes, and voters shall be advised that they need to check a return code for each blank option. In other words, if Alice wishes to abstain in an election where voters can select  $k$  options out of  $n$ , then she must receive (and check)  $k$  distinct return codes, corresponding to  $k$  blank voting options. Of course, in case the election includes several questions (e.g., several sub-elections) then these blank options have to be specific to each question. This may be difficult to understand for voters.

*Synchronization.* As pointed before, the protocol is no longer secure w.r.t. the cast-as-intended property as soon as the voting server answers two different requests from the same voters. (Note that this is explicitly forbidden by the Neuchâtel protocol since revoting is not allowed). Therefore, the voting server must implement some form of thread synchronization to guarantee that two different ballots will never be accepted for the same *id*, even if none of them has yet been confirmed by the voter. This should be enforced even when voting servers are duplicated for efficiency reasons. In particular, the use of tables as described in [18] is insufficient and further requires a proper lock mechanism.

## 7 Conclusion

We provide an automated proof of an e-voting protocol in use for politically-binding elections in the Swiss cantons of Neuchâtel and Fribourg. Our analysis confirms the security of the protocol: it ensures cast-as-intended and recorded-as-cast against a dishonest voting device (assuming an honest voting server) and it guarantees privacy against a dishonest voting server (assuming an honest voting device). Previous analysis of other e-voting protocols (eg [3, 17, 15, 16] ) left several parts of the protocol undefined because they study protocols that have not been deployed for actual government elections. Our ProVerif model covers the authentication phase, voters' cryptographic keys derivation from passwords (as voters cannot be asked to copy long strings), no revote, as well as elections where voters may select several options. One particular challenging aspect of the Neuchâtel protocol is the fact that it forbids re-voting since it would be insecure otherwise. Due to ProVerif's over-approximations, we had to propose several ideas in order to still obtain automated proofs of all the desired properties.

As future work, we plan to explore how we could extend ProVerif in order to cope with protocols where some events happen only once. This is the case as soon as a protocol embeds some lock mechanisms or uses of counters.

## Acknowledgments

This work has been partially supported by the European Research Council (ERC) under the European Union's Horizon 2020 research (grant agreements No 645865-SPOOC).

## References

- [1] B. Adida. Helios: Web-based Open-Audit Voting. In P. C. van Oorschot, editor, *Proceedings of the 17th USENIX Security Symposium*, pages 335–348. USENIX Association, 2008.
- [2] M. Arnaud, V. Cortier, and S. Delaune. Deciding security for protocols with recursive tests. In N. Bjørner and V. Sofronie-Stokkermans, editors, *Proceedings of the 23rd International Conference on Automated Deduction (CADE'11)*, Lecture Notes in Artificial Intelligence, pages 49–63, Wrocław, Poland, July 2011. Springer.
- [3] M. Backes, C. Hritcu, and M. Maffe. Automated verification of remote electronic voting protocols in the applied pi-calculus. In *CSF 2008*, pages 195–209, 2008.
- [4] D. Bernhard, O. Pereira, and B. Warinschi. How Not to Prove Yourself: Pitfalls of the Fiat-Shamir Heuristic and Applications to Helios. In X. Wang and K. Sako, editors, *Advances in Cryptology - ASIACRYPT 2012*, volume 7658 of *Lecture Notes in Computer Science*, pages 626–643. Springer, 2012.
- [5] B. Blanchet. Automatic verification of security protocols in the symbolic model: The verifier proverif. In *Foundations of Security Analysis and Design VII* -

- FOSAD 2012/2013 Tutorial Lectures*, volume 8604 of *Lecture Notes in Computer Science*, pages 54–87. Springer, 2014. [prosecco.gforge.inria.fr/personal/bblanche/proverif](http://prosecco.gforge.inria.fr/personal/bblanche/proverif).
- [6] B. Blanchet. Modeling and verifying security protocols with the applied pi calculus and ProVerif. *Foundations and Trends in Privacy and Security*, 1(1–2):1–135, Oct. 2016.
- [7] B. Blanchet, M. Abadi, and C. Fournet. Automated verification of selected equivalences for security protocols. In *20th IEEE Symposium on Logic in Computer Science (LICS 2005)*, pages 331–340. IEEE Computer Society, June 2005.
- [8] B. Blanchet and B. Smyth. Automated reasoning for equivalences in the applied pi calculus with barriers. In *CSF 2016*, 2016.
- [9] I. Brightwell, J. Cucurull, D. Galindo, and S. Guasch. An overview of the ivote 2015 voting system. Available through <https://www.elections.nsw.gov.au>, 2015.
- [10] N. Chridi, M. Turuani, and M. Rusinowitch. Decidable analysis for a class of cryptographic group protocols with unbounded lists. In *CSF 2009*, pages 277–289, 2009.
- [11] M. R. Clarkson, S. Chong, and A. C. Myers. Civitas: Toward a Secure Voting System. In *2008 IEEE Symposium on Security and Privacy (S&P 2008)*, pages 354–368. IEEE Computer Society, 2008.
- [12] V. Cortier, D. Galindo, S. Glondu, and M. Izabachene. Election Verifiability for Helios under Weaker Trust Assumptions. In *Proceedings of the 19th European Symposium on Research in Computer Security (ESORICS’14)*, LNCS. Springer, 2014.
- [13] V. Cortier, D. Galindo, and M. Turuani. Proverif model for the neuchâtel e-voting protocol. <https://drive.google.com/open?id=0B88EHBgxIrkJMG93UmlTOW9SQmc>. Experiments.
- [14] V. Cortier, B. Schmidt, C. C. Dragan, P.-Y. Strub, F. Dupressoir, and B. Warinschi. Machine-checked proofs of privacy for electronic voting protocols. In *Proceedings of the 37th IEEE Symposium on Security and Privacy (S&P’17)*. IEEE Computer Society Press, 2017. To appear.
- [15] V. Cortier and B. Smyth. Attacking and fixing Helios: An analysis of ballot secrecy. *Journal of Computer Security*, 21(1):89–148, 2013.
- [16] V. Cortier and C. Wiedling. A formal analysis of the norwegian e-voting protocol. *Journal of Computer Security*, 25(15777):21–57, 2017.
- [17] S. Delaune, S. Kremer, and M. Ryan. Verifying privacy-type properties of electronic voting protocols. *Journal of Computer Security*, 17(4):435–487, 2009.

- [18] D. Galindo, S. Guasch, and J. Puiggali. 2015 Neuchâtel’s Cast-as-Intended Verification Mechanism. In *5th International Conference, (VoteID 2015)*, pages 3–18, 2015.
- [19] K. Gjøsteen. The norwegian internet voting protocol. Cryptology ePrint Archive, Report 2013/473, 2013. <http://eprint.iacr.org/2013/473>.
- [20] S. Heiberg, P. Laud, and J. Willemson. The application of i-voting for estonian parliamentary elections of 2011. In *VoteID 2011*, volume 7187 of *Lecture Notes in Computer Science*, pages 208–223. Springer, 2012.
- [21] S. Heiberg, T. Martens, P. Vinkel, and J. Willemson. Improving the verifiability of the estonian internet voting scheme. In *E-Vote-ID 2016*, volume 10141 of *Lecture Notes in Computer Science*, pages 92–107. Springer, 2017.
- [22] A. Juels, D. Catalano, and M. Jakobsson. Coercion-Resistant Electronic Elections. In *Proceedings of Workshop on Privacy in the Eletronic Society (WPES 2005)*, pages 61–70. ACM Press, 2005.
- [23] A. Kiayias, T. Zacharias, and B. Zhang. DEMOS-2: scalable E2E verifiable elections without random oracles. In *ACM SIGSAC Conference on Computer and Communications Security*, pages 352–363. ACM, 2015.
- [24] A. Kiayias, T. Zacharias, and B. Zhang. Ceremonies for end-to-end verifiable elections. In *Public-Key Cryptography - PKC 2017*, volume 10175, pages 305–334. Springer, 2017.
- [25] S. Kremer and M. Ryan. Analysis of an electronic voting protocol in the applied pi-calculus. In M. Sagiv, editor, *Programming Languages and Systems – Proceedings of the 14th European Symposium on Programming (ESOP’05)*, volume 3444 of *Lecture Notes in Computer Science*, pages 186–200, Edinburgh, U.K., April 2005. Springer.
- [26] R. Küsters, J. Müller, E. Scapin, and T. Truderung. sElect: A Lightweight Verifiable Remote Voting System. In *IEEE 29th Computer Security Foundations Symposium (CSF 2016)*, pages 341–354. IEEE Computer Society, 2016.
- [27] G. Lowe. A hierarchy of authentication specifications. In *Proc. of the 10th Computer Security Foundations Workshop (CSFW’97)*. IEEE Computer Society Press, 1997.
- [28] M. Paiola and B. Blanchet. Verification of security protocols with lists: From length one to unbounded length. *Journal of Computer Security (JCS)*, pages 781–816, 2013.
- [29] T. Pinault and P. Courtade. E-voting at expatriates’ mps elections in france. In *5th International Conference on Electronic Voting 2012, (EVOTE 2012)*, volume 205 of *LNI*, pages 189–195. GI, 2012.



- [30] Swiss. Post. Swiss post's e-voting solution. <https://www.post.ch/en/business/a-z-of-subjects/industry-solutions/swiss-post-e-voting>, 2017.
- [31] Scytl. Swiss On-line Voting Protocol. 2016. Manuscript.
- [32] D. Springall, T. Finkenauer, Z. Durumeric, J. Kitcat, H. Hursti, M. MacAlpine, and J. A. Halderman. Security Analysis of the Estonian Internet Voting System. In G. Ahn, M. Yung, and N. Li, editors, *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 703–715. ACM, 2014.
- [33] S. Wolchok, E. Wustrow, D. Isabel, and J. A. Halderman. Attacking the Washington, D.C. Internet Voting System. In *16th Intl. Conference on Financial Cryptography and Data Security (FC'12)*, 2012.

```

TallyH( $c_t : channel, sk_b : ekey, pk_c : spkey, id_a : agentId, id_b : agentId$ ) :=
  in( $c_t, (= id_a, b_a : any, sfc_a : any, crf_a : any)$ );
  in( $c_t, (= id_b, b_b : any, sfc_b : any, crf_b : any)$ );
  let ( $e_a : any, prc_a^1 : int, prc_a^2 : int, = pk_e(id_a), p_a : any$ ) =  $b_a$  in
  let ( $e_b : any, prc_b^1 : int, prc_b^2 : int, = pk_e(id_b), p_b : any$ ) =  $b_b$  in
  if  $verif_{zkp}(pk_b, pk_e(id_a), e_a, prc_a^1, prc_a^2, p_a) \wedge verify(pk_c, sfc_a, crf_a)$ 
   $\wedge verif_{zkp}(pk_b, pk_e(id_b), e_b, prc_b^1, prc_b^2, p_b) \wedge verify(pk_c, sfc_b, crf_b)$  then
    out(mix,  $e_a$ ) | out(mix,  $e_b$ ) | in(mix,  $e'_a : any$ ); in(mix,  $e'_b : any$ );
    out( $c, (adec(sk_b, e'_a), adec(sk_b, e'_b))$ );

```

Figure 7: The Tally Process – Honest version

```

TallyD( $c_t : channel, sk_b : ekey, pk_c : spkey, id_a : agentId, id_b : agentId$ ) :=
  in( $c_t, (id, b : any, sfc : any, crf : any)$ );
  let ( $e : any, prc^1 : int, prc^2 : int, = pk_e(id), p : any$ ) =  $b$  in
  if  $verif_{zkp}(pk_b, pk_e(id), e, prc^1, prc^2, p) \wedge verify(pk_c, sfc, crf)$ 
   $\wedge id \neq id_a \wedge id \neq id_b$  then
    out( $c, adec(sk_b, e)$ );

```

Figure 8: The Tally Process – Dishonest version