

Measuring anonymity in μ CRL

Simona Orzan

with Tom Chothia, Jun Pang and Mohammad Torabi Dashti
(CWI/Birmingham, Oldenburg/Luxemburg, CWI/Zurich)



What is anonymity

- no consensus over the definition
- keeping personal choices secret
- property of group protocols: voting, anonymous broadcast, file sharing

What is anonymity

- no consensus over the definition
- keeping personal choices secret
- property of group protocols: voting, anonymous broadcast, file sharing

More

- meta-trace property, different than reachability
- most natural way to express it: equivalence relation
- identities play an essential part
- quantitative aspect
- passive adversary

Verification of anonymity

- many theoretical studies and formal analysis
- only few automatic approaches, on small examples of specific protocols

Verification of anonymity

- many theoretical studies and formal analysis
- only few automatic approaches, on small examples of specific protocols

Our proposal

- a **general** framework for checking and measuring possibilistic anonymity using an existing powerful toolset
- a **formalization** of anonymity and **two measures** of anonymity w.r.t. an external attacker, possibly supported by coalitions of corrupted protocol participants
- the attacker's power is modeled by a process equivalence, we use **bisimulation**

Verification of anonymity

- many theoretical studies and formal analysis
- only few automatic approaches, on small examples of specific protocols

Our proposal

- a **general** framework for checking and measuring possibilistic anonymity using an existing powerful toolset
- a **formalization** of anonymity and **two measures** of anonymity w.r.t. an external attacker, possibly supported by coalitions of corrupted protocol participants
- the attacker's power is modeled by a process equivalence, we use **bisimulation**

Test results

- Dining Cryptographers: checked anonymity for 17 players (previously in the literature 8)
- first tool-supported analysis of anonymity in FOO

Secret choices

- the **secret choice** of a player i is an action or a piece of data that should not be linked to identity i . It defines the behaviour of player i .
- a **choice vector** defines the behaviour of a system
- a **set of choice vectors** defines all possible behaviour of a system

Secret choices

- the **secret choice** of a player i is an action or a piece of data that should not be linked to identity i . It defines the behaviour of player i .
- a **choice vector** defines the behaviour of a system
- a **set of choice vectors** defines all possible behaviour of a system

A protocol

$$\text{Protocol}(\mathbf{x}) = \tau_I \rho_R \delta_H(\mathbf{P}_1(\mathbf{x}_1) \parallel \mathbf{P}_2(\mathbf{x}_2) \parallel \cdots \parallel \mathbf{P}_n(\mathbf{x}_n) \parallel \mathbf{Q}(\mathbf{n}))$$

- $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2 \cdots \mathbf{x}_n)$ is the vector of secret choices (x_i comes from a known, usually small, domain)
- $P_i(x_i), Q(n)$ are μCRL -processes
- a model of attacker's perspective, therefore some actions are hidden or renamed

Groups of corrupted players

- Observers = the set of malicious participants controlled by the passive intruder
- $\text{Protocol}_{\text{Observers}}(\mathbf{x}) = \tau_{(\mathbf{I}/\text{AO})} \rho_{(\mathbf{R}/\text{AO})} \partial_{\mathbf{H}}(\mathbf{P}_1(\mathbf{x}_1) \parallel \mathbf{P}_2(\mathbf{x}_2) \parallel \dots \parallel \mathbf{P}_n(\mathbf{x}_n) \parallel \mathbf{Q}(\mathbf{n}))$ where AO are the actions observed by the members of Observers.
 $\text{Protocol}_{\emptyset}(\mathbf{x}) = \text{Protocol}(\mathbf{x})$.

Anonymity means different things to different users...



Protocol - the specification of a protocol;

\mathbf{v}_1 and \mathbf{v}_2 - two choice vectors;

Observers - an observer set.

CVS - the set of all possible choice vectors

Choices - the set of all possible choices.

i - a player

c - a secret choice

choice indistinguishability

$$\mathbf{v}_1 \approx_{\text{Observers}} \mathbf{v}_2 \text{ iff } \text{Protocol}_{\text{Observers}}(\mathbf{v}_1) \approx \text{Protocol}_{\text{Observers}}(\mathbf{v}_2).$$

- **choice anonymity degree**

cad of i w.r.t. Observers under \mathbf{x} is:

$$\text{cad}_{\mathbf{x}}(i) = |\{c \in \text{Choices}, \exists \mathbf{v} \in \text{CVS such that } v_i = c \text{ and } \mathbf{v} \approx_{\text{Observers}} \mathbf{x} \text{ and } (\forall j \in \text{Observers. } v_j = x_j)\}|$$

where $\mathbf{v} = \langle v_1, \dots, v_n \rangle$ and $\mathbf{x} = \langle x_1, \dots, x_n \rangle$.

cad of i w.r.t. Observers is $\text{cad}(i) = \min_{\mathbf{x} \in \text{CVS}} \text{cad}_{\mathbf{x}}(i)$.

- **choice anonymity degree**

cad of i w.r.t. Observers under \mathbf{x} is:

$$\text{cad}_{\mathbf{x}}(i) = |\{c \in \text{Choices}, \exists \mathbf{v} \in \text{CVS} \text{ such that} \\ v_i = c \text{ and } \mathbf{v} \approx_{\text{Observers}} \mathbf{x} \text{ and } (\forall \mathbf{j} \in \text{Observers.} \mathbf{v}_{\mathbf{j}} = \mathbf{x}_{\mathbf{j}})\}|$$

where $\mathbf{v} = \langle \mathbf{v}_1, \dots, \mathbf{v}_n \rangle$ and $\mathbf{x} = \langle \mathbf{x}_1, \dots, \mathbf{x}_n \rangle$.

cad of i w.r.t. Observers is $\text{cad}(i) = \min_{\mathbf{x} \in \text{CVS}} \text{cad}_{\mathbf{x}}(i)$.

- **player anonymity degree**

pad of secret choice c , in a protocol with n players, w.r.t. Observers and \mathbf{x} is:

$$\text{pad}_{\mathbf{x}}(c) = |\{i \in \{1, \dots, n\} \setminus \text{Observers}, \exists \mathbf{v} \in \text{CVS} \text{ such that} \\ v_i = c \text{ and } \mathbf{v} \approx_{\text{Observers}} \mathbf{x} \text{ and } (\forall \mathbf{j} \in \text{Observers.} \mathbf{v}_{\mathbf{j}} = \mathbf{x}_{\mathbf{j}})\}|.$$

pad of c w.r.t. Observers is $\text{pad}(c) = \min_{\mathbf{x} \in \text{CVS: pad}_{\mathbf{x}}(c) > 0} \text{pad}_{\mathbf{x}}(c)$.

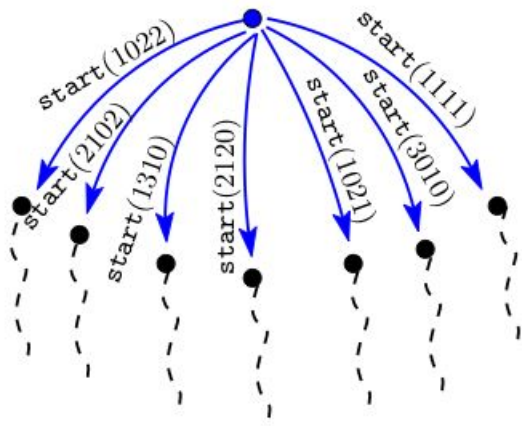
Verification method

- base specification $\text{Protocol}(\mathbf{x})$, given by the user
- *.rn* file with renaming rules, e.g. $\text{assign}(i, \mathbf{x}, \text{true}) \rightarrow \text{assign}(i)$, defining the visibility of the actions of a generic player i (P_i should not contain any further renaming operators.)
- For given \mathbf{x} , Observers and CVS, dedicated tools will generate the μCRL specification corresponding to

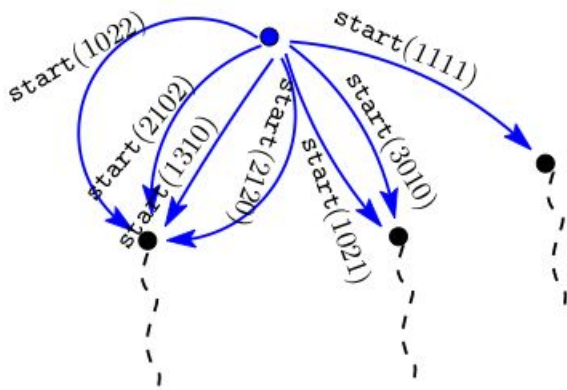
$$\sum_{v \in \text{CVS}: \forall j \in \text{Observers}. v_j = x_j} \text{start}(v). \text{Protocol}_{\text{Observers}}(v),$$

- the state space of this composed specification will be generated by the toolset
- reduction modulo strong bisimilarity will be performed
- on the reduced state space, dedicated tools can extract the **cad** and **pad** information

The tools and examples can be downloaded from
<http://www.win.tue.nl/sorzan/anonymity/>



↓ State space reduction modulo a behavioral equivalence



Equivalence classes

player: 1 2 3 4

1	0	2	2
2	1	0	2
1	3	1	0
2	1	2	0

1	0	2	1
3	0	1	0

1	1	1	1
---	---	---	---

Computing cad for player 2:

$$\text{cad}_{2102}(2) = 3/4$$

$$\text{cad}_{1021}(2) = 1/4$$

$$\text{cad}_{1111}(2) = 1/4$$

$$\text{cad}(2) = 1/4$$

Computing pad for choice 0:

$$\text{pad}_{2102}(0) = 3/4$$

$$\text{pad}_{1021}(0) = 2/4$$

$$\text{pad}_{1111}(0) = 0/4$$

$$\text{pad}(0) = 2/4$$

Dining Cryptographers

- Choices = $\{T, F\}$
- Protocol specification:

$$DC(\mathbf{x}:ChoiceVector) = \partial_{\{\text{tell}, \text{recv}\}}(Crypt_0(x_0) || Crypt_1(x_1) || Crypt_2(x_2))$$

$$Crypt_i(x_i : Bool) = \text{pay}(n, i, x_i) \cdot \sum_{\text{coin_left}:Bool} \text{flip}(i, \text{coin_left}). \\ (\text{tell}(\text{next}(i), \text{coin_left}) || \sum_{\text{coin_right}:Bool} \text{recv}(i, \text{coin_right})). \\ CryptAnnounce(n, 0, id, ch \oplus \text{coin_right} \oplus \text{coin_left})$$

- Renamings: $\{\text{com}(i, \mathbf{X}) \rightarrow \text{com}(i), \text{pay}(n, i, \mathbf{X}) \rightarrow \text{pay}(i)\}$.
- Main question: who pays? (who has the choice T?)
- 3 players, Observers = \emptyset
 The composed process: $\sum_{v \in \Pi(\text{TFF})} \text{start}(v).DC_{\emptyset}(v)$.
 Equivalence classes: $\{\text{FFT}, \text{FTF}, \text{TFF}\}$
 Maximal anonymity: $\text{pad}(\text{T}) = 3:3$, $\text{cad}(0) = 2:2$.
- 5 players, Observers = $\{1, 3\}$
 The composed process: $\sum_{v \in \Pi(\text{TFFFF}):v_1=v_3=F} \text{start}(v).DC_{\{1,3\}}(\text{TFFFF})$
 Equivalence classes: $\{\text{FFFFT}, \text{TFFFF}\} \{\text{FFTFF}\}$
 Anonymity analysis: $\text{pad}(\text{T}) = 1:5$ - there is a scenario when the payer is revealed.
 $\text{pad}_{\text{TFFFF}}(\text{T}) = 2:5$ - anonymity of 0 not broken, but the set of suspects is reduced to 2 players.

	Size	Size after red.	Time	cad(0)	pad(T)
DC3 , Observers = \emptyset	229 469	65 112	1.5s	2:2	3:3
DC3 , Observers = {1}	184 362	71 132	1.3s	2:2	2:3
DC5 , Observers = {1, 3}	5189 14679	1620 4567	4.9s	2:2	2:5
DC7 , Observers = \emptyset	185 769 695 551	27 180 85 763	8m53s	2:2	7:7
DC9 , Observers = \emptyset	5 194 659 22 961 789	1 034 142 4 088 977	(s) - (db) 7h5m	2:2	9:9
DC10 , Observers = \emptyset	27 436 022 130 031 220	5 002 490 21 535 547	(db) 17h20m	2:2	10:10
DC11so , Observers = \emptyset	33 876 41 035	6 156 7 188	(s) 6m (db) 11m	2:2	11:11
DC12so , Observers = {1, 3, 5, 7, 9, 11}	58 749 67 612	17 467 21 219	(s) 7m	1:2	1:12
DC15so , Observers = \emptyset	606 388 721 067	98 320 114 716	(s) 7h2m (db) 44m	2:2	15:15
DC17so , Observers = \emptyset	2 556 144 3 014 887	393 234 458 784	(s) - (db) 5h40m	2:2	17:17

A voting protocol

- choices from a larger domain, complex cryptographic mechanisms
- Main question: for whom has i voted? cad is more informative
- For $\mathbf{x} = \mathbf{1121}$, $\text{Observers} = \{2\}$, $\mathbf{i} = \mathbf{0}$, $\mathbf{v} = \mathbf{1}$, we get $\text{cad}_x(0) = 1$ and $\text{pad}_x(0) > 1$. So cad is more sensitive
- $\text{cad}(i) = 1:n$

	Size	Size after red.	Time	cad	pad
FOO4 , Observers = {2}	58 749 67 612	17 467 21 219	17s	$\text{cad}_{1121}(0) = 1:3$	$\text{pad}_{1121}(1) = 3:4$
FOO6 , Observers = {2}	3 423 841 10 518 810	29 451 92 835	22m36s	$\text{cad}_{010102}(0) = 3:3$	$\text{pad}_{010102}(1) = 5:6$
FOO7 , Observers = {2}	65 282 690 221 299 564	3 676 249 9 628 686	(db) 4h48m	$\text{cad}_{0101022}(0) = 3:3$	$\text{pad}_{0101022}(1) = 6:7$

	Size	Size after red.	Time		Size	Size after red.	Time
MCW 2.7¹	3 297 3 804	1 088 1 398	12s	MC 2.7	1 652 2 285	23 156	11s
MCW 6.3	10 837 081 14 949 720	226 510	1h40m	MC 6.3	4 846 12 795	23 252	1m
MCW 7.1	680 700	17 28	3s	MC 7.1	134 427	10 21	2s
MCW 7.2	28 343 204 41 424 733	695 1 120	(s)- (db) 4h4m	MC 7.2	1205 3598	18 77	3s
MCW 50.1	248 876 250 100	2 652 5 150	(s) 6h10m (db) 40m	MC 50.1	6 326 128 825	53 150	1m10s

Conclusions

Advantages of this method:

- fine-grained picture of the anonymity guaranteed by the protocol
- the state space is generated once for a given Observers
- strong bisimulation can be computed much more efficiently than trace equivalence [SS96]

Conclusions

Advantages of this method:

- fine-grained picture of the anonymity guaranteed by the protocol
- the state space is generated once for a given Observers
- strong bisimulation can be computed much more efficiently than trace equivalence [SS96]

Is strong bisimulation appropriate?

- SB is the same as trace equivalence, for deterministic processes.
- Anyway, using strong bisimulation equivalence is *sound*
- In real systems, it may be possible for the intruder to detect that a process can't perform an action.

Conclusions

Advantages of this method:

- fine-grained picture of the anonymity guaranteed by the protocol
- the state space is generated once for a given Observers
- strong bisimulation can be computed much more efficiently than trace equivalence [SS96]

Is strong bisimulation appropriate?

- SB is the same as trace equivalence, for deterministic processes.
- Anyway, using strong bisimulation equivalence is *sound*
- In real systems, it may be possible for the intruder to detect that a process can't perform an action.

Future work

- probabilistic extensions - not trivial
- theoretical questions: parametrized analysis, decidability