Minimum Disclosure as Boolean Optimization: New Results

Alessandro Armando\textsuperscript{1,2}, Angela Contento\textsuperscript{1}, Daniele Costa\textsuperscript{1}, Marco Maratea\textsuperscript{1}

\textsuperscript{1} DIBRIS, University of Genova
\textsuperscript{2} Security & Trust Unit, FBK-irst, Trento

RCRA2012: Rome, Italy, June 14-15th 2012
Context and Motivation

The treatment of personal information is becoming more and more critical in several contexts e.g.

- Home-banking
- e-Commerce

There is thus the need for

- Solutions to allow transmitting, receiving and process information
- in a fast and efficient ways, given also the increasing amount of data available

Two main actors interact in this context

1. servers that offer services
2. clients/users that request services
Client/server interaction based on credentials and preferences

- Portfolio: information items (and related constraints) that a client can expose in order to access the service
- Requests: of a service (from the client); of information (from the server) to grant the service
- Preferences: how much the client values her information
- Disclosure: set of information items that satisfy client constraints and the server request

Problem faced

To find a “minimum” disclosure, i.e. a disclosure that exposes the “minimum” information.
State of the art

First work in this context, among the ones based on logic-based languages, dates back to 10 years ago (Bonatti and Samarati, 2002).

Recent effective approaches:

- Heuristic graph-based approaches

- Exact approaches that represent the problem as a Max-SAT problem (Ardagna et al., 2010)
State of the art

First work in this context (based on logic) dates back to 10 years ago (Bonatti and Samarati, 2002).

Recent effective approaches:

- Heuristic approaches based on graph

- Exact approaches that represent the problem as a Max-SAT problem (Ardagna et al., 2010)
Starting from (Ardagna et al., 2010)

1. Simplification and optimization of this proposal
   - more intuitive modeling
   - reduced size formulas

2. Use of different encodings and Boolean Optimization solvers that can solve the problem

Efficiency is critical in this context given this is a run-time task.
There possibly can be a hierarchy of credential types. Atomic credentials can only be released as a whole.
There are some constraints on the client portfolio that must be satisfied

- **Certifiability:** Each disclosed property must be certified by (at least) a credential

  \[ p \rightarrow \lor_{c \in C, p \in \text{properties}(c)} c \]

  \[ \text{Name} \rightarrow (\text{myId} \lor \text{myLicense} \lor \text{myVISA} \lor \text{myMC}) \]

- **Atomicity:** If an *atomic* credential is disclosed, all of its properties are disclosed

  \[ c \rightarrow \land_{p \in \text{properties}(c)} p \]

  \[ \text{myId} \rightarrow (\text{Name} \land \text{DoB} \land \text{Address}) \]
Disclosure limitations: Given a DL $I$ of which at most $n$ information items can be disclosed

Formulation in (C.A Ardagna at al., 2010): Given $S$ to be the power set of $I$ (e.g. $\{\text{Address, Phone, eMail}\}_2$)

\[ \forall s \in S, |s| \leq n \ (\bigwedge x \in s \ x \wedge \bigwedge x \notin s \ \neg x) \]

\[ (\text{Address} \wedge \text{Phone} \wedge \neg \text{eMail}) \lor (\text{Address} \wedge \neg \text{Phone} \wedge \text{eMail}) \lor \]
\[ (\neg \text{Address} \wedge \text{Phone} \wedge \text{eMail}) \lor (\text{Address} \wedge \neg \text{Phone} \wedge \neg \text{eMail}) \lor (\neg \text{Address} \wedge \neg \text{Phone} \wedge \text{eMail}) \lor (\neg \text{Address} \wedge \neg \text{Phone} \wedge \neg \text{eMail}) \]

Our formulation

\[ \wedge I' \subseteq I, |I'| = n + 1 \ \forall x \in I' \ \neg x \]

\[ (\neg \text{Address} \lor \neg \text{Phone} \lor \neg \text{eMail}) \]

Forbidden views: Given a view $v$ (e.g. $\{\text{Name, NickName}\}$),

\[ \forall x \in v \ \neg x \]
\[ \neg (\text{Name} \wedge \text{NickName}) \]
Modeling the Server Request

- Terms and term satisfaction
  - Each term \( r: type.\{pt_1, \ldots, pt_m\} \). A credential \( c \) satisfies \( r \) iff
    - \( type(c) \preceq_{isa} type(r) = type \)
    - \( \forall pt \in properties(r), \exists p \in properties(c): type(p) = pt \)
  - \( TermSAT(r) = \bigvee_{c \in C, type(c) \preceq type(r)} c \land \left( \land_{p \in properties(c), pt \in properties(r)} type(p) = pt \right) \)

- Server Request and SRs satisfaction:
  - \( R = R_1 \lor R_2 \cdots \lor R_n \) (SR)
  - \( R = r_1 \land r_2 \cdots \land r_m \) (simple request)
  - Example: \( R = r_1 \land r_2 = id.\{Name, Address\} \land cc.\{Name, CCNum\} \)
  - \( \bigvee_{R \in R} \land_{r \in R} TermSAT(r) \)
User preferences and disclosure

How much the user values her information items

- Costs of properties and credentials (if exposed)
- Sensitivity view: set of information items that brings a sensitivity which is higher than the sum of the cost of its element
- Dependence: set of information items that brings a sensitivity which is lower than the sum of the cost of its element
  - \{Address, Country\}

Given a disclosure, its cost is obtained by summing

- the costs of properties and credentials in the disclosure
- the costs of the exposed sensitivity views
- the (negative) costs of dependencies exposed

Disclosure

- A set of credentials and properties than satisfies the SR and the client portfolio constraints.
- Our goal is to find a minimum disclosure, i.e. such that each other disclosure does not have lower cost.
### Experimental analysis

#### Benchmarks
- Randomly generated instances following (Ardagna et al., 2010)
- Setting considered in our analysis
  - 20 credentials
  - \{20, 40, 60, 80, 100\} properties
  - 10 instances per point

#### Encodings and solvers
- Max-SAT, PB, SMT
- MiniMaxSAT, WMaxSatz, B solo, Pbclasp, Yices
Our formulation vs SOTA (C.A Ardagna et al., 2010)
Conclusions and future work

In this work we have

- simplified and optimized the SOTA modeling
- evaluated a number of Boolean optimization solvers

Results obtained

- more than one order of magnitude improvement in the modeling
- up to one order of magnitude improvement in the solving

Current work

- model disclosure limitations with polynomial encoding e.g. (Sinz, 2005)
- evaluate other Boolean optimization solvers (e.g. WPM, SAT4J, CPLEX)
Further results

Formulations comparison in PbClasp

Formulations comparison in Bsolo
Some references


### Max-SAT Example

```
p wcnf 14 46 72
72 -1 10 11 12 13 14 0
72 -2 10 11 14 0
72 -3 10 14 0
72 -4 11 14 0
72 -5 12 14 0
72 -6 13 14 0
72 -7 14 0
72 -8 14 0
72 -9 14 0
72 -10 1 0
72 -10 2 0
72 -10 3 0
72 -12 1 0
72 -12 5 0
72 -13 1 0
72 -13 6 0
72 -1 -9 0
72 -3 -7 0
72 -3 -8 0
72 -7 -8 0
72 12 13 0
72 12 1 0
72 12 6 0
72 1 13 0
72 1 6 0
72 5 13 0
72 5 1 0
72 5 6 0
72 10 0
72 1 0
72 3 0
1 -1 0
5 -1 0
5 -2 0
5 -3 0
2 -4 0
10 -5 0
15 -6 0
9 -7 0
3 -8 0
1 -9 0
1 -10 0
5 -11 0
3 -12 0
8 -13 0
5 -1 -3 -12 0
-2 -3 -4 0
5 -2
```

**Map:**
- c 1 Name
- c 2 DoB
- c 3 Address
- c 4 Country
- c 5 VISANum
- c 6 MCNum
- c 7 Phone
- c 8 eMail
- c 9 NickName
- c 10 myID
- c 11 myLicense
- c 12 myVISA
- c 13 myMC
- c 14 decl

\[72 = 1 + 5 + 5 + \ldots + 8 + 5 - 2\]
Dependencies expressed in Max-SAT

Our portfolio contains Address and Country with weights $\lambda_A$ and $\lambda_C$.
Assuming to have a dependency $\{\text{Address, Country}\}$ whose weight is $-\lambda_C$.

Instead of expressing it as a constraint

- $\neg(\text{Address} \land \text{Country})$ with weight $-\lambda_C$

this is rewritten with the following set of constraints

- $\neg(\text{Address} \land \neg \text{Country})$ with weight $\lambda_A$
- $\neg(\neg \text{Address} \land \text{Country})$ with weight $\lambda_C$
- $\neg(\text{Address} \land \text{Country})$ with weight $\lambda_A$
## Max-SAT Example

<table>
<thead>
<tr>
<th>p wcnf 14 45 84</th>
<th>84 12 13 0</th>
<th>1 -1 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>84 -1 10 11 12 13 14 0</td>
<td>84 12 1 0</td>
<td>5 -2 0</td>
</tr>
<tr>
<td>84 -2 10 11 14 0</td>
<td>84 12 6 0</td>
<td>5 -3 0</td>
</tr>
<tr>
<td>84 -3 10 14 0</td>
<td>84 1 13 0</td>
<td>2 -4 0</td>
</tr>
<tr>
<td>84 -4 11 14 0</td>
<td>84 1 6 0</td>
<td>10 -5 0</td>
</tr>
<tr>
<td>84 -5 12 14 0</td>
<td>84 5 13 0</td>
<td>15 -6 0</td>
</tr>
<tr>
<td>84 -6 13 14 0</td>
<td>84 5 1 0</td>
<td>9 -7 0</td>
</tr>
<tr>
<td>84 -7 14 0</td>
<td>84 5 6 0</td>
<td>3 -8 0</td>
</tr>
<tr>
<td>84 -8 14 0</td>
<td>84 10 0</td>
<td>1 -9 0</td>
</tr>
<tr>
<td>84 -9 14 0</td>
<td>84 1 0</td>
<td>1 -10 0</td>
</tr>
<tr>
<td>84 -10 1 0</td>
<td>84 3 0</td>
<td>5 -11 0</td>
</tr>
<tr>
<td>84 -10 2 0</td>
<td>84 10 0</td>
<td>3 -12 0</td>
</tr>
<tr>
<td>84 -10 3 0</td>
<td>84 1 0</td>
<td>8 -13 0</td>
</tr>
<tr>
<td>84 -12 1 0</td>
<td>84 5 -1 -3 -12 0</td>
<td></td>
</tr>
<tr>
<td>84 -12 5 0</td>
<td>84 -3 4 0</td>
<td></td>
</tr>
<tr>
<td>84 -13 1 0</td>
<td>2 3 -4 0</td>
<td></td>
</tr>
<tr>
<td>84 -13 6 0</td>
<td>5 -3 -4 0</td>
<td></td>
</tr>
<tr>
<td>84 -1 -9 0</td>
<td>84 -3 -8 0</td>
<td></td>
</tr>
<tr>
<td>84 -3 -7 0</td>
<td>84 -7 -8 0</td>
<td></td>
</tr>
<tr>
<td>84 -3 -8 0</td>
<td>84 =1+5+5+. . .+2+5</td>
<td></td>
</tr>
</tbody>
</table>

### Map:
- c 1 Name
- c 2 DoB
- c 3 Address
- c 4 Country
- c 5 VISANum
- c 6 MCNum
- c 7 Phone
- c 8 eMail
- c 9 NickName
- c 10 myID
- c 11 myLicense
- c 12 myVISA
- c 13 myMC
- c 14 decl

---

M. Maratea  
Minimum Disclosure as Boolean Optimization: New Results
Size of the formulas

Confronto delle dimensioni Nostra formulazione vs. SOTA

![Graph comparing the size of formulas with SOTA and Nostra Formulazione](image-url)