Minimum Disclosure as Boolean Optimization: New Results

Alessandro Armando[1,2], Angela Contento[1], Daniele Costa[1], Marco Maratea[1]

[1] DIBRIS, University of Genova
[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

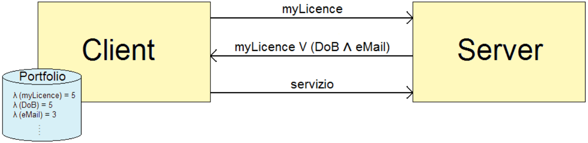
There is thus the need for

- 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)

Our contribution

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.

Client Portfolio example (C.A Ardagna at al., 2010)

Properties			
Id	Type	Value	
Name	*Name*	Bob	
DoB	*DoB*	1975/10/23	
Address	*Address*	3155, 5th Ave, New York, USA	
Country	*Country*	USA	
VISANum	*CCNum*	4353. . . 21	
MCNum	*CCNum*	5643. . . 18	
Phone	*Phone*	789-231-044	
eMail	*eMail*	bob@abc.com	
NickName	*NickName*	bob75	
Credentials			
Id	**Type**	**Certified properties**	**Atomic**
myId	*id-card*	Name,DoB,Address	$\surd$
myLicense	*driver-lic*	Name,DoB,Country	
myVISA	*credit-card*	Name,VISANum	$\surd$
myMC	*credit-card*	Name,MCNum	$\surd$

There possibly can be a hierarchy of credential types.

Atomic credentials can only be released as a whole.

Modeling the Client Portfolio (1)

There are some constraints on the client portfolio that must be satisfied

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

  $$p \rightarrow \bigvee_{c \in C, p \in properties(c)} c$$

  $$Name \rightarrow (myId \vee myLicense \vee myVISA \vee myMC)$$

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

  $$c \rightarrow \bigwedge_{p \in properties(c)} p$$

  $$myId \rightarrow (Name \wedge DoB \wedge Address)$$

Modeling the Client Portfolio (2)

- 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. $\{Address, Phone, eMail\}_2$)

    $\bigvee_{s \in S, |s| \leq n} (\bigwedge_{x \in s} x \wedge \bigwedge_{x \notin s} \neg x)$

    $(Address \wedge Phone \wedge \neg eMail) \vee (Address \wedge \neg Phone \wedge eMail) \vee (\neg Address \wedge Phone \wedge eMail) \vee (Address \wedge \neg Phone \wedge \neg eMail) \vee (\neg Address \wedge Phone \wedge \neg eMail) \vee (\neg Address \wedge \neg Phone \wedge eMail) \vee (\neg Address \wedge \neg Phone \wedge \neg eMail)$
  - Our formulation

    $\bigwedge_{I' \subseteq I, |I'| = n+1} \bigvee_{x \in I'} \neg x$

    $(\neg Address \vee \neg Phone \vee \neg eMail)$
- Forbidden views: Given a view v (e.g. $\{Name, NickName\}$),

  $\bigvee_{x \in v} \neg x$

  $\neg(Name \wedge NickName)$

Modeling the Server Request

- Terms and term satisfaction
  - Each term r: $type.\{pt_1, \ldots, pt_m\}$. A credential c satisfies r iff
    - $type(c) \preceq_{isa} type(r) = type$;
    - $\forall pt \in properties(r), \exists p \in properties(c) : type(p) = pt$
  - $TermSAT(r) = \bigvee_{c \in C, type(c) \preceq\ type(r)} c \wedge (\bigwedge_{p \in properties(c), pt \in properties(r), type(p) = pt}\ p)$
- Server Request and SRs satisfaction:
  - $\mathcal{R} = R_1 \vee R_2 \cdots \vee R_n$ (SR)
  - $R = r_1 \wedge r_2 \cdots \wedge r_m$ (simple request)
  - Example: $\mathcal{R} = r_1 \wedge r_2 =$ $id.\{Name, Address\} \ \wedge \ cc.\{Name, CCNum\}$
  - $\bigvee_{R \in \mathcal{R}} \bigwedge_{r \in R}\ TermSAT(r)$

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

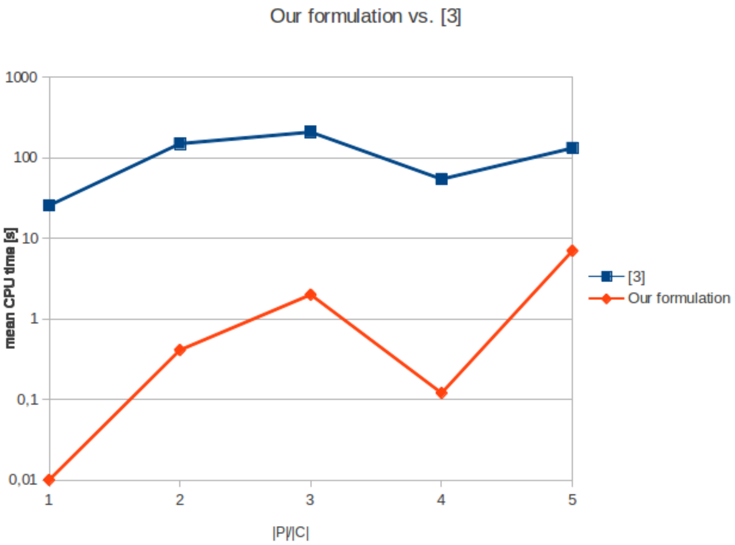
Benchmarks

- 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

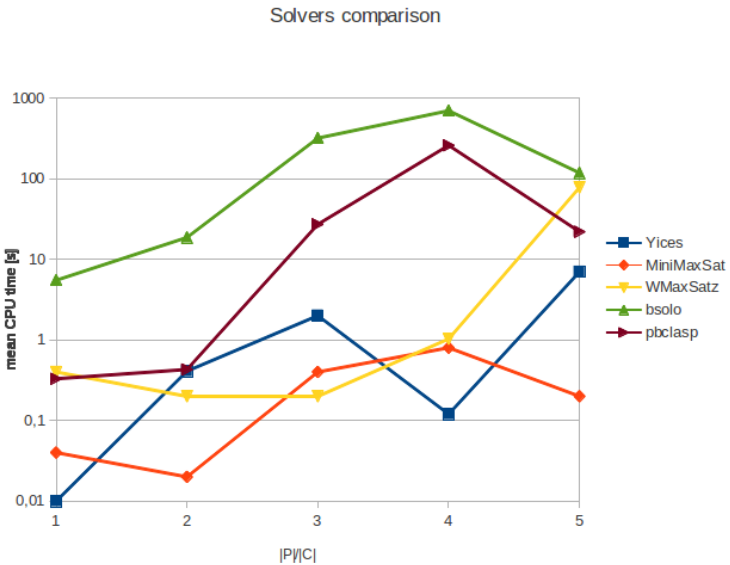
Encodings and solvers

- Max-SAT, PB, SMT
- MiniMaxSAT, WMaxSatz, Bsolo, Pbclasp, Yices

Our formulation vs SOTA (C.A Ardagna at al., 2010)



Solvers comparison



Conclusions and future work

In this work we have

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

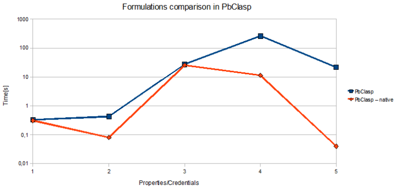
Results obtained

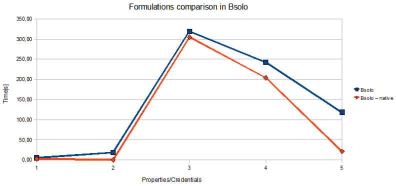
- more than one order of magnitude improvement in the modeling
- up to one order of magnitude improvement in the solving

Current work

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

Further results





Some references

(Bonatti and Samarati, 2002) P.A. Bonatti, P. Samarati: A Uniform Framework for Regulating Service Access and Information Release on the Web. Journal of Computer Security 10(3): 241-272

(Ardagna et al., 2010) C.A Ardagna, S. De Capitani di Vimercati, S. Foresti, S. Paraboschi, P. Samarati: Supporting privacy preferences in credential-based interactions. Proc. of the ACM Workshop on Privacy in the Electronic Society, ACM, 83-92

(Sinz, 2005) C. Sinz: Towards an Optimal CNF Encoding of Boolean Cardinality Constraints. Proc. of CP 2005, 827-831

Max-SAT Example

p wcnf 14 46 72			Map:
72 -1 10 11 12 13 14 0	72 12 13 0	1 -1 0	c 1 Name
72 -2 10 11 14 0	72 12 1 0	5 -2 0	c 2 DoB
72 -3 10 14 0	72 12 6 0	5 -3 0	c 3 Address
72 -4 11 14 0	72 1 13 0	2 -4 0	c 4 Country
72 -5 12 14 0	72 1 6 0	10 -5 0	c 5 VISANum
72 -6 13 14 0	72 5 13 0	15 -6 0	c 6 MCNum
72 -7 14 0	72 5 1 0	9 -7 0	c 7 Phone
72 -8 14 0	72 5 6 0	3 -8 0	c 8 eMail
72 -9 14 0	72 10 0	1 -9 0	c 9 NickName
72 -10 1 0	72 1 0	1 -10 0	c 10 myID
72 -10 2 0	72 3 0	5 -11 0	c 11 myLicense
72 -10 3 0		3 -12 0	c 12 myVISA
72 -12 1 0		8 -13 0	c 13 myMC
72 -12 5 0		5 -1 -3 -12 0	c 14 decl
72 -13 1 0		-2 -3 -4 0	
72 -13 6 0			
72 -1 -9 0			
72 -3 -7 0			
72 -3 -8 0			
72 -7 -8 0			72=1+5+5+. . . +8+5-2

Dependencies expressed in Max-SAT

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

Instead of expressing it as a constraint

- $\neg(Address \wedge Country)$ with weight $-\lambda_C$

this is rewritten with the following set of constraints

- $\neg(Address \wedge \neg Country)$ with weight λ_A
- $\neg(\neg Address \wedge Country)$ with weight λ_C
- $\neg(Address \wedge Country)$ with weight λ_A

Max-SAT Example

p wcnf 14 45 84			Map:
84 -1 10 11 12 13 14 0	84 12 13 0	1 -1 0	c 1 Name
84 -2 10 11 14 0	84 12 1 0	5 -2 0	c 2 DoB
84 -3 10 14 0	84 12 6 0	5 -3 0	c 3 Address
84 -4 11 14 0	84 1 13 0	2 -4 0	c 4 Country
84 -5 12 14 0	84 1 6 0	10 -5 0	c 5 VISANum
84 -6 13 14 0	84 5 13 0	15 -6 0	c 6 MCNum
84 -7 14 0	84 5 1 0	9 -7 0	c 7 Phone
84 -8 14 0	84 5 6 0	3 -8 0	c 8 eMail
84 -9 14 0	84 10 0	1 -9 0	c 9 NickName
84 -10 1 0	84 1 0	1 -10 0	c 10 myID
84 -10 2 0	84 3 0	5 -11 0	c 11 myLicense
84 -10 3 0		3 -12 0	c 12 myVISA
84 -12 1 0		8 -13 0	c 13 myMC
84 -12 5 0		5 -1 -3 -12 0	c 14 decl
84 -13 1 0		5 -3 4 0	
84 -13 6 0		2 3 -4 0	
84 -1 -9 0		5 -3 -4 0	
84 -3 -7 0			
84 -3 -8 0			
84 -7 -8 0			84=1+5+5+. . . +2+5

Size of the formulas