DYNAMIC ENFORCEMENT OF KNOWLEDGE-BASED SECURITY POLICIES

Piotr (Peter) Mardziel

Joint work with Stephen Magill, Michael Hicks (UMD), and Mudhakar Srivatsa (IBM TJ Watson)
Where is your information?
Bad + Good

Getaways in DC 83% Off
homerun.com
The best Getaways in Washington DC. Huge discounts for our members. Click now to join free!

M.S. – Computer Forensics
online.uab.edu
Become a leader in the rapidly growing field of computer forensics with a Masters degree from UAB!

DC Bucket List
partners.livingsocial.com
365 Things to do in Washington, DC before you die.

Cointreau Noir
Remy Martin Cognac blended with orange liqueur: Cointreau Noir. Do you LIKE it?
👍 Like

Need a Local Plumber?
redbeacon.com
Redbeacon gives up to 44% off on local plumbers. Sign up today!

Do you have Anxiety?
healingllc.com
Is your mind spinning with worries and you just want a break? Counseling can help! Email us today at AnotherLook @ HealingLLC.com.
Take back control
out = 24 \leq \text{Age} \leq 30 \\
\text{Æ Female?} \\
\text{Æ Engaged?}
out = (ssn, ccn, fav-color)
The big problem: how to decide to run a query or not?

Q1
out = \(24 \leq \text{age} \leq 30\)
Æ female?
Æ engaged?

Q2
out = \text{age}

Q3
out = \(\text{ssn, ccn, fav-color}\)
Outline

- Clarkson’s work on quantified information flow
  - (adversary) belief/knowledge and belief revision (Bayesian revision)
  - probability of adversary guessing secret
- Suitable knowledge-based policy formulation
  - maintain bound while revealing query outputs or rejecting queries
- Clarkson’s probabilistic semantics
- Computational feasibility
  - approximation of knowledge
  - abstract interpretation (of probabilistic semantics)
  - soundness in terms of policy
- Experimental results
  - compare to prob-scheme, enumeration-based probabilistic evaluation
Clarkson’s Quantitative Information Flow

= belief (probability distribution) over secret data

- Given belief, query, determine revised belief given the output of the query (Bayesian revision)
  - Pr[gender = male | output = o]
- Pr[Bad] < t =
  - adversary holding belief has bounded probability of guessing my secret(s) in one try
    - sufficient: Pr[gender = male | output = o] < t

- Assumption
  - initial belief is correct
    - our approach might help here (later)
  - adversary guesses by sampling

Ex:
Pr[gender = male] = 0.5
Pr[gender = female] = 0.5
• More about Clarkson’s semantics later
  • Let us consider an example
  • Flawed initial policy
    • Pr[my secret] < t
    • demonstrate problem with rejection
    • revise policy to address problem
Meet Bob

Bob (born September 27, 1980)
bday = 270
byear = 1980

Secret

0 \cdot \text{bday} \cdot 364
1956 \cdot \text{byear} \cdot 1992

Policy
Pr[\text{bday}] < 0.2
Pr[\text{bday,byear}] < 0.05

Currently
Pr[\text{bday}] = 1/365
Pr[\text{bday,byear}] = 1/(365*37)
bday-query1

```plaintext
today := 260;
if bday ≥ today \Æ bday < (today + 7)
then out := true
else out := false
```

Bob
bdays = 270
byear = 1980
**Potentially**
Pr[bday] = 1/358 < 0.2
Pr[bday,byear] = 1/(358*37) < 0.05

**Bday-query1**
today := 260;
if bday ≤ today ∧ bday < (today + 7)
then out := true
else out := false

Bob
bday = 270
byear = 1980

= (out = false)
Next day …

\[
\begin{array}{c}
\text{bday-query2} \\
\text{today := 261;} \\
\text{if bday \geq today \land bday < (today + 7)} \\
\text{then out := true} \\
\text{else out := false}
\end{array}
\]

\[\text{Bob} \]
\[
\begin{array}{c}
bday = 270 \\
\text{byear = 1980}
\end{array}
\]
Potentially
Pr[bday] = 1/357 < 0.2
Pr[bday,byear] = 1/(357*37) < 0.05

```plaintext
bday-query2
today := 261;
if bday ≥ today ∧ bday < (today + 7)
then out := true
else out := false
```

Bob
bday = 270
byear = 1980
Meet Bob’

Bob’
bday = 267
byear = 1980

**bdy-query2**
today := 261;
if bday \leq today \land bday < (today + 7)
then out := true
else out := false

So reject?
Querier’s perspective

Assume querier knows policy

if \( \text{bday} \neq 267 \)

will get answer

if \( \text{bday} = 267 \)

will get reject
Querier’s perspective

- **Solution?**
  - Decide policy independently of secret
  - Revised **policy**
    - For every possible output o, for every possible bday, Pr[bday | out = o] < t
    - So the real bday in particular
    - Therefore Pr[bad] < t
**bdy-query2**
today := 261;
if bday ≥ today ∧ bday < (today + 7) 
    then out := true
    else out := false

**reject**
(regardless of what bday actually is)
Clarkson’s Probabilistic Interpretation

• Given $\pm: \text{States} \rightarrow \mathbb{R}$
  • probability distribution on program states (including secrets)
  • $\pm(\frac{3}{4}) = \text{probability of state } \frac{3}{4}$
• Given program $S$
• Compute
  • $\langle S \neg \pm$
    • probability distribution on resulting program states
  • $\pm \mathcal{A} \mathcal{B}$
    • (sub)probability distribution of $\pm$ only on states consistent with $\mathcal{B}$
  • $\pm | \text{out} = \text{true}$
    • Bayesian revision, post belief
Probabilistic Interpretation

- **Semantics**
  - «skip» = ±
  - «S₁; S₂» = «S₂ » «S₁ » ±
  - «if B then S₁ else S₂ » = «S₁ » (± AE B) + «S₂ » (± AE : B)
  - «pif p then S₁ else S₂ » = «S₁ » (p*±) + «S₂ » ((1-p)*±)
  - «x := E » = ±[x ! E]
  - «while B do S » = lfp(±. F («S » (± AE B)) + (± AE : B))

- **Operations**
  - p*± – scale probabilities by p
  - ± AE B – remove mass inconsistent with B
  - ±₁ + ±₂ – combine mass from both
  - ±[x ! E] – transform mass
Subdistribution operations

\[ \pm \, \& \, B \text{ – remove mass inconsistent with } B \]
\[ \pm \, \& \, B = \frac{3}{4}. \text{ if } \neg \frac{3}{4} \text{ is true then } \pm(\frac{3}{4}) \text{ else } 0 \]

\[ B = x , y \]

\[ \pm \, \& \, B \]

\[ \pm_1 + \pm_2 \text{ – combine mass from both } \]
\[ \pm_1 + \pm_2 = \frac{3}{4}. \pm_1(\frac{3}{4}) + \pm_2(\frac{3}{4}) \]

\[ \text{if } x \cdot 5 \text{ then } y := y + 3 \text{ else } y := y - 3 \]
\[ \pm_1 \]

\[ \pm_2 \]

\[ \pm_1 + \pm_2 \]

\[ S \neg \pm \]
Infeasibility

- Computational trouble
  - $\pm[x ! E] = \frac{3}{4} \cdot \sum_{\hat{\theta}} | \hat{\theta}[x ! [E]\hat{\theta}] = \frac{3}{4} \pm(\hat{\theta})$
  - $\max_{\frac{3}{4}} \pm(\frac{3}{4}) = ?$ (for policy check)
    - enumeration?

- Sampling (prob-scheme, IBAL, …)
  - evaluate statement for some set of input states
  - poor probability bounds if evaluated on small subset of possible states (later)
  - prohibitive (time, memory) for large state space

- Let’s try an approximation
Abstraction

• Approximate representation $P$
  • $P$ abstracts a set of distributions, $\circ(P)$
  • sound probability bound
    • if $\pm 2 \circ(P)$ then $\max_{\frac{3}{4}} \pm (\frac{3}{4}) \cdot \maxprob(P)$

• $((S)) P$ – abstract interpretation
  • if $\pm 2 \circ(P)$ then $\llbracket[S]\rrbracket \pm 2 \circ((S)P)$

• $P | (\text{out} = X)$ – abstract conditioning
  • if $\pm 2 \circ(P)$ then $(\pm | (\text{out} = X)) 2 \circ(P | (\text{out} = X))$

• (more) computationally feasible

• let us revisit Bob to motivate a suitable abstraction
Representation

$P_1$: 0 · bday · 364, 1956 · byear · 1992
  $p = 0.000074$
  $s = 13505$ (# of points)
  $m = 1$ (total mass)

$P_2$: 267 · bday · 364, 1956 · byear · 1992
  $p = 0.000074$
  $s = 3626$
  $m = 0.268$

Can determine ($\exists\ e$ | out = false)
spec-byear-query
age := 2011 – byear;
if age = 20 Ç … Ç age = 60
then out := true
else out := false;
pif 0.1 then out := true

pif p then $S_1$
• evaluate $S_1$ with probability $p$
Approximation

\( \mathcal{E} \text{ out} = \text{true} \)

\( P_1: 0 \cdot \text{bday} \cdot 259, 1992 \cdot \text{byear} \cdot 1992 \)
\( p = 0.0000074 \)
\( s = 260 \)
\( m = 0.0019 \)

\( P_2: 0 \cdot \text{bday} \cdot 259, 1991 \cdot \text{byear} \cdot 1991 \)
\( p = 0.0000074 \)
\( s = 260 \)
\( m = 0.019 \)

soundness

\( P_1: 0 \cdot \text{bday} \cdot 259, 1956 \cdot \text{byear} \cdot 1992 \)
\( p = 0.0000074 \)
\( s = 260 \)
\( m = 0.019 \)

\( P_2: 267 \cdot \text{bday} \cdot 364, 1956 \cdot \text{byear} \cdot 1992 \)
\( p = 0.0000074 \)
\( s = 3626 \)
\( m = 0.053 \)
Approximation

\[ \text{AE out} = \text{false} \]

\[ P_1: 0 \cdot \text{bday} \cdot 259, 1992 \cdot \text{byear} \cdot 1992 \]
\[ p = 0.000067 \]
\[ s = 260 \]
\[ m = 0.019 \]

\[ P_2: 0 \cdot \text{bday} \cdot 259, 1982 \cdot \text{byear} \cdot 1990 \]
\[ p = 0.000067 \]
\[ s = 2340 \]
\[ m = 0.173 \]

\[ P_1 \]
\[ P_2 \]

\[ p \text{ and } s \text{ only refer to possible (non-zero probability) points in region} \]
Approximation

For each $P_i$, store
- region (polyhedron)
- upper bound on probability of each possible point
- upper bound on the number of points
- upper bound on the total probability mass (useful)

Also store
- lower bounds on the above

$$\Pr[A | B] = \frac{\Pr[A \land B]}{\Pr[B]}$$
Approximation

± 2 ° (P) iff

\[ \text{support}(±) = \{ \frac{3}{4} | ±(\frac{3}{4}) > 0 \} \]

\[ \text{support}(±) \mu^°(C) \]
\[ p_{min} \cdot ±(\frac{3}{4}) \cdot p_{max} \] for every \( \frac{3}{4} \) 2 support(±)
\[ s_{min} \cdot | \text{support}(±) | \cdot s_{max} \]
\[ m_{min} \cdot \text{mass}(±) \cdot m_{max} \]

\[ \max_{\frac{3}{4}} ±(\frac{3}{4}) \cdot p_{max} \]
Abstract Interpretation

• **Need**
  - \(((S))\ P\)
    - define identically to \([[S]]\ P\) but using abstract operations
    - if \(\pm 2 \circ(P)\) then \([[S]]\pm 2 \circ((S))P\)

• **Need Abstract operations**
  - \(P_1 + P_2\)
    - if \(\pm_i 2 \circ(P_i)\) for \(i = 1,2\) then \(\pm_1 + \pm_2 2 \circ(P_1 + P_2)\)
  - \(P \leftarrow B\)
    - if \(\pm 2 \circ(P)\) then \(\leftarrow 2 \circ(P \leftarrow B)\)
  - \(p^*P\)
    - if \(\pm 2 \circ(P)\) then \(p^* \pm 2 \circ(p^*P)\)
  - ...

Abstract operation example

$\pm_1 + \pm_2$ – combine mass from both

$\pm_1 + \pm_2 = \frac{3}{4} \cdot \pm_1(\frac{3}{4}) + \pm_2(\frac{3}{4})$

What is the maximum number of possible points in the sum?

- determine minimum overlap (10)

$| C_1 \cap C_2 | = 20$

$P_3 = P_1 + P_2$

$s_{3max} \cdot s_{1max} + s_{2max} = 130$

$s_{3max} = 120$
Operations

- $P_3 = P_1 + P_2$
  - $C_3$ – convex hull of $C_1$, $C_2$
  - $s_3^{\text{max}}$ – what is the smallest overlap?
  - $s_3^{\text{min}}$ – what is the largest overlap?
  - $p_3^{\text{max}}$ – is overlap possible?
  - $p_3^{\text{min}}$ – is overlap impossible?
  - $m_3^{\text{max}}$ – simple sum $m_1^{\text{max}} + m_2^{\text{max}}$
  - $m_3^{\text{min}}$ – simple sum $m_1^{\text{min}} + m_2^{\text{min}}$

- Other operations, similar, complicated formulas abound

- Need to
  - count number of integer points in a convex polyhedra
    - Latte
  - maximize a linear function over integer points in a polyhedron
    - Latte
  - convex hull, intersection, affine transform
    - Parma
Precision

- Extend abstraction to a set of Probabilistic Polyhedrons
  - $\pm 2^\circ(P_1, P_2)$ iff $\pm = \pm_1 + \pm_2$ with $\pm_1 2^\circ(P_1)$ and $\pm_2 2^\circ(P_2)$
  - similarly for more than two
  - $\max_{\frac{3}{4}} \pm(\frac{3}{4}) \cdot \sum_i p_i^{\max}$
    - can do better with a bit more work

- performance / precision tradeoff
Implementation and Results

\[ 0 \cdot \text{bday} \cdot 364 = 1956 \cdot \text{byear} \cdot 1992 \]
Policy

Pr[bday, byear] \cdot 0.05

\textbf{bdy-query1}

today := 260;
if bday \geq \today \land bday < (\today + 7)
then out := true
else out := false

1. prob (our implementation)
2. prob-scheme (sampling/enumeration)
   - provides sound estimation after partial enumeration
   - measure time and bound on \( \max_{\frac{3}{4}} \pm \left(\frac{3}{4}\right) \) produced
Implementation and Results

\[ 0 \cdot \text{bday} \cdot 364 \]
\[ 1956 \cdot \text{byear} \cdot 1992 \]

\[ 0 \cdot \text{bday} \cdot 364 \]
\[ 1910 \cdot \text{byear} \cdot 2010 \]
Precision?

\[ 0 \cdot \text{bday} \cdot 364 \]
\[ 1956 \cdot \text{byear} \cdot 1992 \]

\[ \text{byear-query1} \]
then
\[ \text{spec-byear-query} \]
age := 2011 – byear;
if age = 20 Ç … Ç age = 60
then out := true
else out := false;
\[ \text{pif 0.1 then out := true} \]

- Limiting number of prob. polyhedra requires merging two into one at various points
- Deciding which ones to merge is troublesome
  - likely reason for the strangeness above
Conclusions

• Knowledge-based policies
  • quantitative information flow, probabilistic semantics
  • bound on probability of specific bad events (guess of secret)

• Dynamic enforcement, via abstract interpretation
  • policy formulation safe for rejection
  • resistant to state space explosion
  • can be sound in respect to many initial distributions
    • alleviates the problem of determining what is the correct initial distribution

• Drawback
  • restricted language, integer linear expressions

• Potential (future work)
  • simpler domains (Octagons) can replace Polyhedra
    • increased performance?
Thank you!

Go back.