dAEL
Distributed Autoepistemic Logic

IJCAI 2016 – NEW YORK

PIETER VAN HERTUM, MARCOS CRAMER, BART BOGAERTS, MARC DENECKER
Distributed Autoepistemic Logic and its application to **Access Control**

- **Access Control Policy**
  - A set of norms defining which principal is to be granted access to which resource under which circumstances

- **Access Control logic**
  - Represent policies
  - Represent requests
  - Reason about requests
  - *Access is granted if it is entailed by the policy*
Motivation: Example 1

- Agents
  - $A$: Professor
  - $B$: Student of $A$
  - $C$: Postdoc of $A$, supervising $B$
- $A$ owns resource $r$, $s$
- $A$ gives $B$ access to $s$
- $A$ delegates to the decision whether $B$ has access to $r$
Motivation: Example 1

- An agent grants access if the request is a **logical consequence** of his theory.
- Agent knows if other agents grants access.
  = **POSITIVE MUTUAL INTROSPECTION**

\[
\text{Access}(B, s) \\
\text{Access}(B, r) \leftarrow C \text{ says } \text{Access}(B, r)
\]

\[
\text{Access}(B, r)
\]
Motivation: Example 2

- **Agents**
  - \( A \): Professor
  - \( B \): Student of \( A \)
  - \( C \): Postdoc of \( A \), supervising \( B \)
- \( A \) owns resource \( r \)
- \( A \) gives \( B \) access to \( r \)
- \( A \) gives \( C \) permission to revoke \( B \)’s access to \( r \)

**Access**

\[ Access(B, r) \leftarrow \neg (C \text{ says } \neg Access(B, r)) \]
Motivation: Example 2

- An agent’s statements are a complete characterization of what he supports
- To give revocation rights, agent needs to know what an agent doesn’t support!
  \[\text{NEGATIVE MUTUAL INTROSPECTION}\]
Distributed Autoepistemic Logic and its application to Access Control

- Needed for our logic:
  - An agent grants access if the request is a logical consequence of his theory.
  - An agent’s statements are a complete characterization of what he supports.
  - Positive and negative mutual introspection needed.

- Autoepistemic logic (AEL)
  - Logic to model knowledge (single agent).
  - Reason about knowledge and knowledge derived of (lack of) knowledge.
  - A theory is a complete characterization of what is known.
  - K operator: I know → I support.
Autoepistemic logic: $\mathcal{L}_k$

- Syntax of $\mathcal{L}_k$ over $\Sigma$
  - First order logic
  - $K(\psi) \in \mathcal{L}_k$ if $\psi \in \mathcal{L}_k$

- Structure $I$
  - As defined in FO
  - Potential state of affairs

- Possible world structure $Q$
  - Set of structures
  - All structures that are deemed possible

- Semantics: $\varphi^{Q,I} =$
  - Rules for FO
    - $(K\psi)^{Q,I} = t$ if $\psi^{Q,J} = t$ for each $J \in Q$
Autoepistemic logic
Semantics

- A possible world structure $Q$ is consistent with a theory $T$ iff
  
  $T^{Q,I} = t \text{ for each } I \in Q$

- Define revision operator $D$:
  
  $D_T(Q) = \{I|T^{Q,I} = t\}$

  What do I derive from $T$ if I assume $Q$ represents my current belief?

- $T$-Consistent possible world structures = fixpoints for $D_T$
Distributed Autoepistemic Logic

- $\mathcal{L}_d$ over $\Sigma$ and $\mathcal{A}$
- First order logic
- $K_A(\psi) \in \mathcal{L}_k$ if $\psi \in \mathcal{L}_k, A \in \mathcal{A}$

- Distributed possible world structure

$$Q = \langle Q_A \rangle_{A \in \mathcal{A}}$$
One pws per agent

- Valuation as AEL, but:

$$K_A(\psi)^{Q,I} = t \quad \text{if} \quad \psi^{Q,I} = t$$
for each $J \in Q_A$
Distributed Autoepistemic logic Semantics

- A distributed possible world structure $Q$ is **consistent with a theory $T$** iff
  \[ T^Q_I = t \text{ for each } I \in Q_A \]

- Define revision operator $\mathcal{D}$:
  \[ \mathcal{D}_T(Q) = \langle \{ I | T^Q_A, I = t \} \rangle_{A \in \mathcal{A}} \]

What do I derive from $T$ if I assume $Q$ represents my current belief?

- $T$ – Consistent **distributed** possible world structures = fixpoints for $\mathcal{D}_T$
dAEL example: *Child wants candy*

- Assume 2 agents: {Mom, Dad} and voc ={c}
- Child wants candy
  - Mom: You can have candy if it’s ok for your father
  - Dad: You can have candy if it’s ok for mom
- \( T_M = \{K_D(c) \Rightarrow c\} \) and \( T_D = \{K_M(c) \Rightarrow c\} \)
- Child knows dAEL and knows the 4 possible situations:
  - The empty possible world (inconsistent belief)
  - The belief of c
  - The disbelief of c
  - The lack of knowledge
dAEL
example: Child wants candy

- $T_M = \{K_D(c) \Rightarrow c\}$
- $T_D = \{K_M(c) \Rightarrow c\}$
- 4 possible situations
  - The empty possible world (inconsistent belief)
  - The belief of c
  - The disbelief of c
  - The lack of knowledge

- $T$ – Consistent possible world structures:
  - One where nothing is known
  - One where they both know c
  - Or they both agree to candy, or none of them does

(= What Moore called autoepistemic expansions)

Not all fixpoints are interesting. Is consistent a good notion?
Our paper

- We study which fixpoints are interesting in the context of dAEL
- We find them using an *approximator* of revision operator
  - This is certainly known (by A)
  - This is certainly not known (by A)
- Approximation Fixpoint Theory

- Inductive definitions in dAEL
  - Allow us define access control policies
Different Semantics for dAEL example: Child wants candy

- $T_M = \{K_D(c) \Rightarrow c\}$
- $T_D = \{K_M(c) \Rightarrow c\}$
- 4 possible situations
  - The empty possible world (inconsistent belief)
  - The belief of $c$
  - The disbelief of $c$
  - The lack of knowledge

- **Kripke-Kleene model**
  - $K_D(c) = u$
  - $K_D(\neg c) = f$
  - $K_M(c) = u$
  - $K_M(\neg c) = f$
  - They don’t know whether to give candy, but know that they will never derive to not give candy.
Different Semantics for dAEL example: Child wants candy

- \( T_M = \{ K_D(c) \Rightarrow c \} \)
- \( T_D = \{ K_M(c) \Rightarrow c \} \)

4 possible situations
- The empty possible world (inconsistent belief)
- The belief of \( c \)
- The disbelief of \( c \)
- The lack of knowledge

- **Stable model**
  - Only 1: nothing is known
  - They know that they will never derive that they will give candy

- **Well-founded model**
  - Exact: nothing is known (=stable model)
  - They know that they will never derive that they will give candy

Stable and well-founded semantics are grounded
⇒ knowledge only derived if non-self supporting
Conclusion

- We propose a new logic: $dAEL$
  - Full mutual introspection
  - Good for delegation and revocation of access rights
  - $AEL$ in a multi-agent case
  - Inductive definitions for $dAEL$: $dAEL(ID)$

- Future work: Decision procedure for $dAEL$