

Beágyazott intelligens rendszerek

Aktivítások felismerése

Tervfelismerés

- **Fizikai mozgás**
 - Mozgás szenzor jelei
- **Viselkedések**
 - Futás, megfogás, emelés, ...
- **Tervek**
 - Egy pohár vizet szerezni
- **Célok**
 - Szomjúság csillapítása

AAL – AAC

- emberi viselkedés megértése
szenzor információ alapján
- emberi modellek tanulása
 - ötletadás, segítségnyújtás
- emberi gondozók riasztása,
teendőkre utalás, ...

Tervfelismerési probléma dimenziói

Leselkedő $\Leftrightarrow$ interaktív

Leselkedő (keyhole)

Meghatározni, hogy az ágens cselekvései hogyan járulnak hozzá a feltételezett céljaihoz

Modell: világ

agens hiedelmei

megfigyelő modellje – nincs

Interaktív (kooperatív)

Ágens szándékosan jelzi másoknak a hiedelmeit és a kívánalmait
(ACL, beszéd aktusok - inform, request, ...)

Párbeszéd konvenciók

Együttműködés alakulása

Tervfelismerés

Tervfelismerési probléma dimenziói

Ideális $\Leftrightarrow$ hibázó ágensek

- elhibázott hiedelmek
- kognitív hibák (kognitív képességek nem megfelelő használata)
- irracionális viselkedések

Megbízható $\Leftrightarrow$ nem megbízható megfigyelések

Zárt $\Leftrightarrow$ nyitott világok

Rögzített tervtár?

Rögzített célhalmaz?

Cselekvések „természetes” $\Leftrightarrow$ „nem természetes” időzítése

Egyedi $\Leftrightarrow$ többszörös egyidejű tervek

Kívánt eredmény:

- konzisztens tervek, vagy célok halmaza?
- a leginkább valószínű terv, vagy cél?
- a legkritikusabb terv, vagy cél?
- beavatkozások a megfigyelt ágens (tervének, céljának) megsegítésére, vagy ellehetetlenítésére?

Megközelítések:

- konzisztencia alapú
- valószínűségi alapú

Tervfelismerés

= **abduktív magyarázatadás**
minimális (a legvalószínűbb)
magyarázatok szerkesztése

Problémák szándékfelismeréssel HC Aml-ben

Tervek (célok) elhagyása: minden emberre jellemző

– **szándékos**, tudatos: megváltozott körülmények,

- **nem szándékos**: figyelemelterelés, felejtés, ...

(pl. telefon gyógyszerzedés közepette, figyelmeztetéshez fel kell először ismerni, hogy a terv/cél elmaradt)

Ellenséges ágensek: szokásos megoldások = kooperatív ágensek, **a tervük felismerését nem veszik zokon**, sőt segítenek ebben.

Paciens = „ellenséges” ágens, függetlensége érdekében a megfigyelésnek ellenáll.

A cselekvések teljes megfigyelhetősége = nem tartható.

Nem megfigyelhető cselekvések kikövetkeztetése megfigyelt cselekvések és változások alapján.

Sikertelen cselekvések megfigyelése: sikertelen cselekvés egyformán „beszédes”, ami a rejtett célokat illeti.

Részben rendezett tervek: a szokásos emberi tervek (ADL) általában rugalmasak az egyes (elemi) cselekvések tekintetében.

Problémák szándékfelismeréssel HC Aml-ben

Többszörös konkurens célok: emberi terekben megszokott (TV + étkezés, ...)

Nem elég a megfigyelt cselekvéseket legjobban magyarázó, de egyetlenegy célt tudni kiszámítani.

Több hatás érdekében végzett cselekvések: emberi terekben megszokott (pl. elmenni otthonról, benyúlni a spájzba, ...).

Megfigyelés kudarca: cselekvések közé „logikusan” beékelődő cselekvések hiánya.

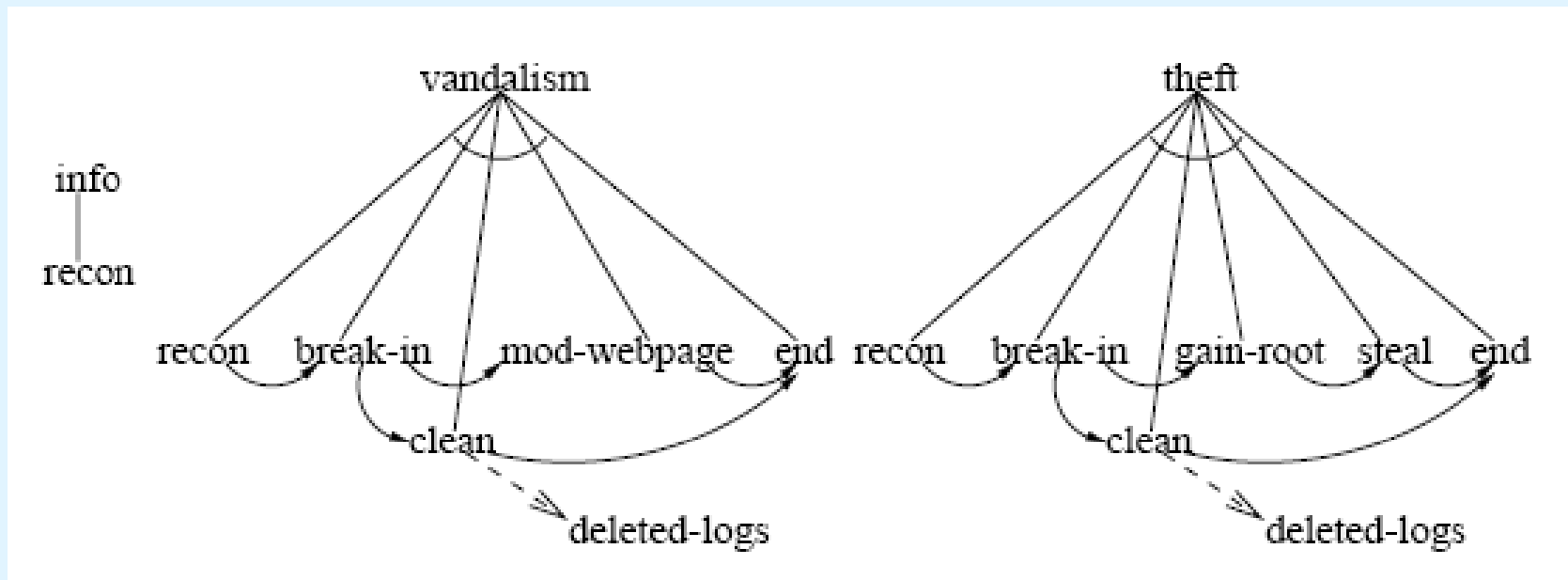
Következtetés egy cselekvés nem tapasztalásáról igen fontos.

Konkrét pozitív bizonyítékok konkrét célokról viszonylag ritkák.

Világ(környezet)állapot hatása a tervek végrehajtására: bizonyos tevékenységek, célok erős korrelációja időszakokkal, lokációkkal, emocionális-, kognitív állapotokkal, ...

Lehetséges többszörös hipotézis rangsorolása: a megfigyelt cselekvéshez nem egyetlenegy magyarázat, hanem a lehetséges magyarázatok rangsorolása.

Pending sets - PHATT Probabilistic Hostile Agent Task Tracker



recon

scan the system to determine vulnerabilities

break-in

exploit the systems weaknesses to gain entry

clean

hide traces of presence on computer

gain-root

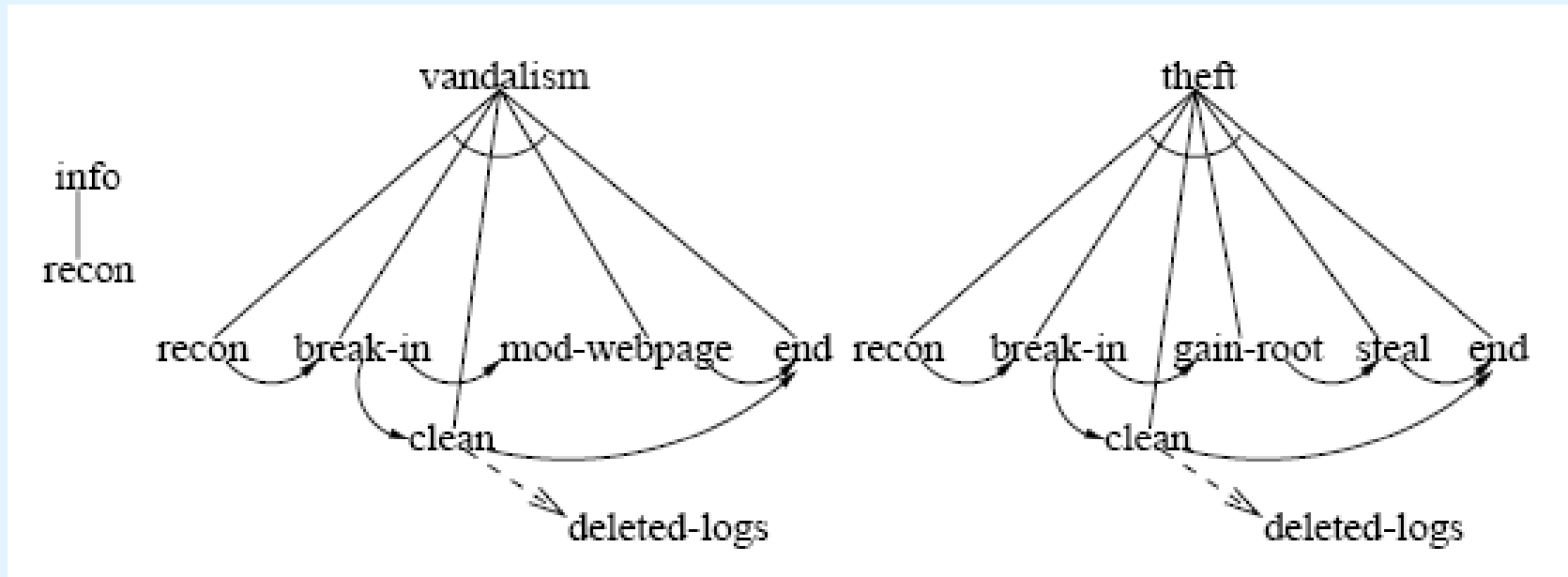
escalate privileges

steal

export desired data

terv: taszkok, rendezés, hatások

Pending sets - PHATT Probabilistic Hostile Agent Task Tracker



Pl. a **recon** megfigyelését meg lehetne magyarázni:

(theft) V (vandalism) V (info)

de lehetne tovább is ...

(theft) V (vandalism) V (info) V ((theft) ∧ (vandalism)) V ((theft) ∧ (info)) V ((info) ∧ (vandalism)) V ((info) ∧ (theft) ∧ (vandalism))

hány célja van a rendszernek?

Pending sets - PHATT Probabilistic Hostile Agent Task Tracker

Tervfelismerés:

ágens több tervet hajt végre dinamikusan
minden pillanatban ágens minden olyan cselekvést hajthat végre,
amit a korábbi cselekvései lehetővé tesznek

Pillanatnyi helyzet: (Pending Set)

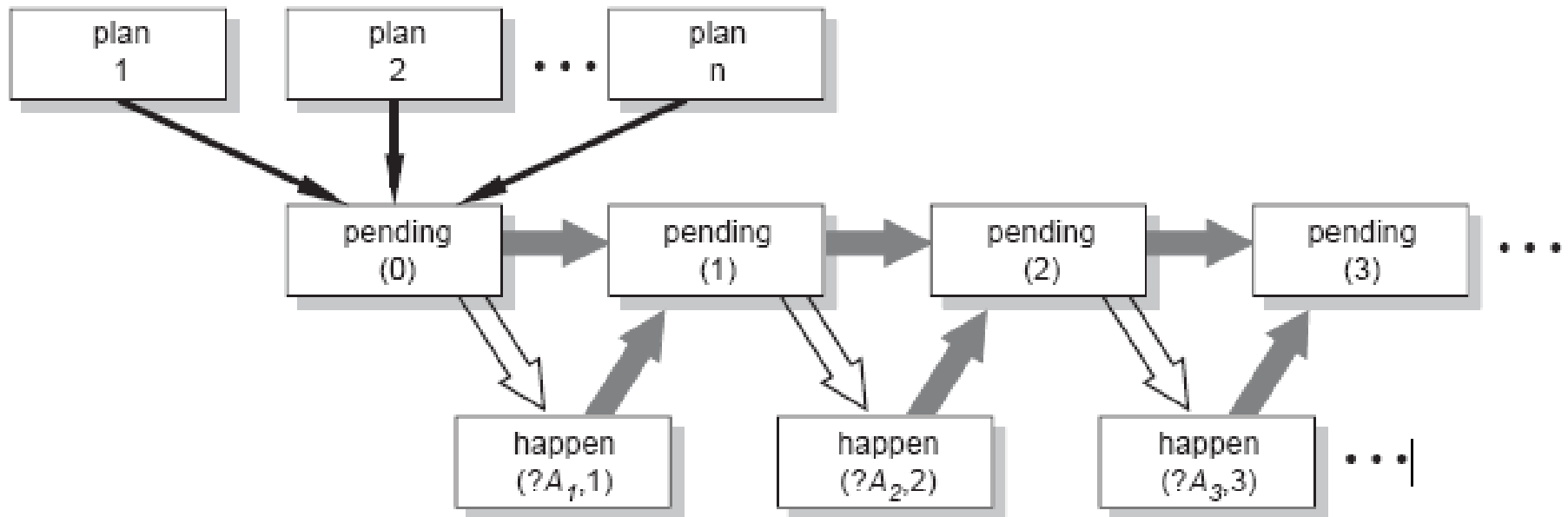
az előbbi cselekvések nyomán lehetségessé vált cselekvések halmaza

Kiindulás:

kiinduló célhalmaz, azokhoz tartozó tervek, azokhoz tartozó végrehajtható elemi cselekvések

Ciklus: valamelyik végrehajtható cselekvés végrehajtása,
állapotváltozások,
végrehajtott cselekvés törlése,
új végrehajtható cselekvések hozzáadása

Pending sets - PHATT Probabilistic Hostile Agent Task Tracker



Feltételezések a megfigyelésekről:

- (1) A **nem megfigyelt** cselekvések számának **felső korlátja ismert**,
(korlát hatása: több alternatíva, jobb probabilisztikus becslés)
- (2) A megfigyelések **helyesek** (mi történt és mikor).

A. Nem megfigyelt cselekvések kikövetkeztetése megfigyelt cselekvések alapján

Legyen a megfigyelés: (**gain-root**, **mod-webpage**)

Hacker célja mind a lopás, mind a rongálás, mert a két cselekvés két diszjunkt terv része (egyetlenegy cél mindkettőt megmagyarázni nem tudja).

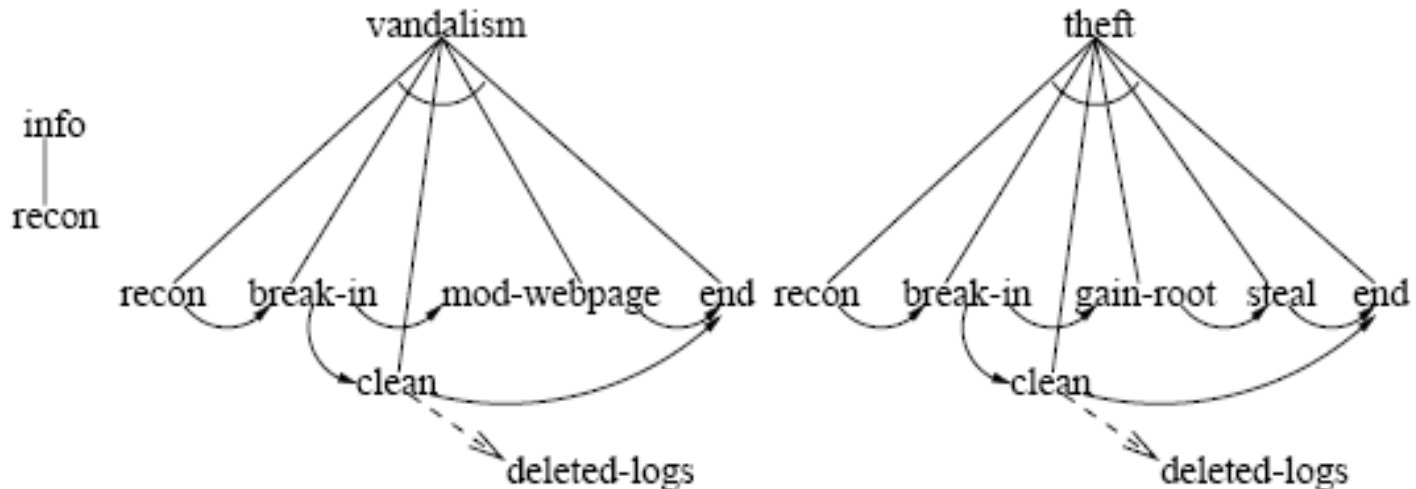
Rejtett információ: mindkét cselekvés kiváltója a nem megfigyelt cselekvések!

Nem engedélyezett cselekvés: megfigyelt, mielőtt a tervbeli előzményét megfigyeltük volna.

Tervkönyvtár: **recon** és **break-in** megelőzi a **gain-root**, vagy a **mod-webpage** cselekvéseket.

Evidencia, hogy: (**recon**, **break-in**, **mod-webpage**) $\wedge$ (**recon**, **break-in**, **gain-root**) tervekkel számolni kell, más szekvenciákat meg lehet nyesni.

egyes cselekvéseket könnyű, más nehezebb elrejteni – szenzorok, fúzió, kontextus, ...



B. Nem megfigyelt cselekvések kikövetkeztetése megfigyelt állapotváltozások alapján

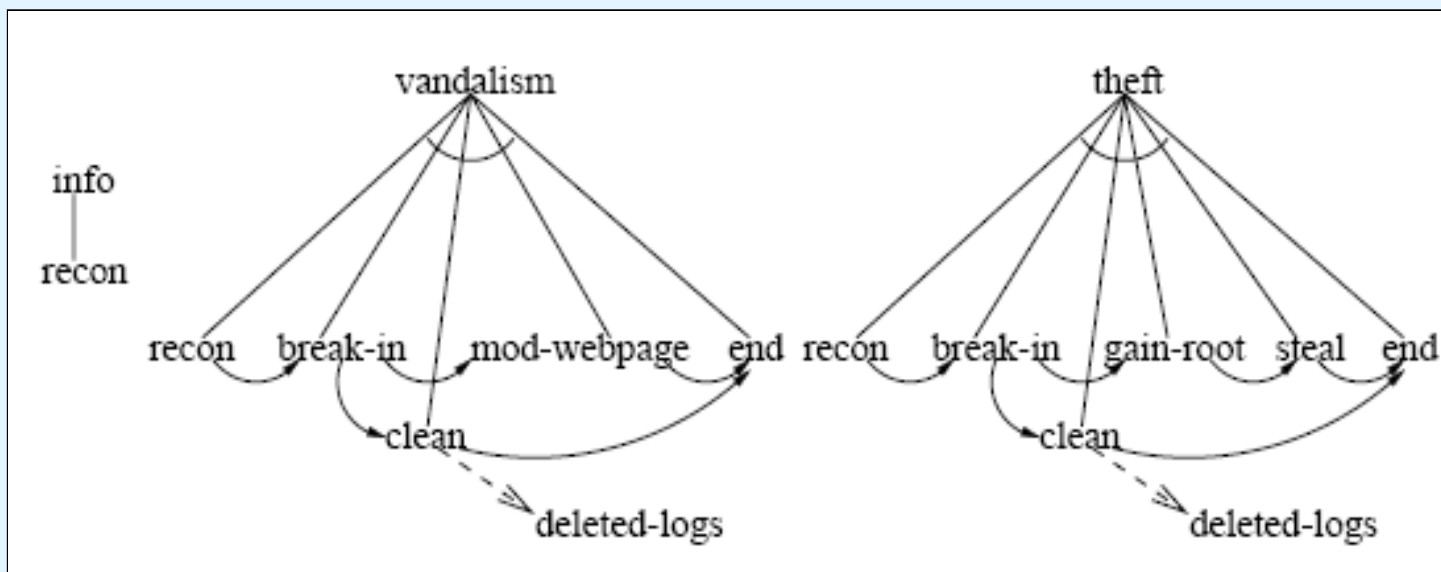
Ráció: ha egy cselekvés elrejtése még megy is, elrejtteni a következményeit sokkal **nehezebb**. Legyen a megfigyelés: (**recon**, **break-in**, **deleted-logs**)

A törölt log fájlok implikálják a nem megfigyelt **clean** cselekvés megtörténését.
Tervkönyvtár alapján: időben a **break-in** végrehajtása és a **deleted-logs** bejelentése között.

Megoldás

Valószínűségi eloszlás definiálása a Pending Set-ek egész halmaza felett.

Megfigyelt és kikövetkezhető cselekvések – végrehajtási nyomok – azok alapján a célok (hipotézisek) – a terv együttesek valószínűségei.



Példa: Legyen a megfigyelt cselekvés: (**break-in**₁) happen(**break-in**,1)

A nem megfigyelt cselekvések számának felső korlátja legyen: 3.

A nem engedélyezett cselekvések előzményeinek hozzáadása.

Konzisztens (megfigyelés alapján) végrehajtható végrehajtási nyomok szerkesztése:

(recon , break-in),	konzisztens: (theft) V (vandalism).
(recon , break-in, clean),	konzisztens: (theft) V (vandalism).
(recon , break-in, clean , deleted-logs),	konzisztens: (theft) V (vandalism).
(recon , break-in, clean , mod-webpage),	konzisztens: (vandalism).
(recon , break-in, clean , gain-root),	konzisztens: (theft).
(recon , break-in, recon),	konkurrens akármelyik 2 cél a 3-ból.
(recon , recon , break-in),	konkurrens akármelyik 2 cél a 3-ból.
(recon , break-in, recon , clean),	konzisztens: ((theft) V (vandalism)) $\wedge$ (info).
(recon , recon , break-in, clean),	konzisztens: ((theft) V (vandalism)) $\wedge$ (info).
... ..	

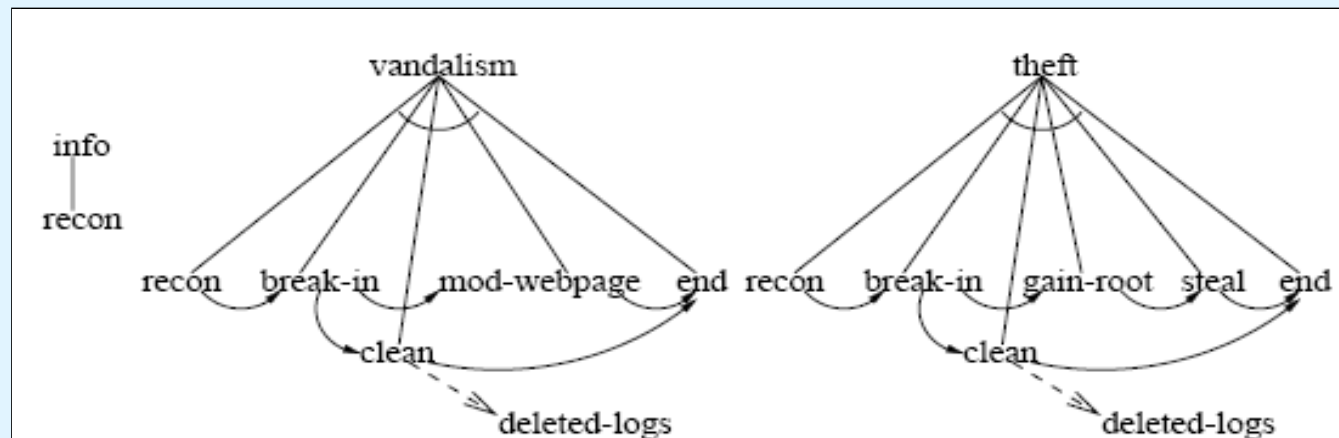
Pending Set1:

(**clean**)

$\Rightarrow$ (**deleted-logs**)

(**mod-webpage**)

(**gain-root**)



Legyen a megfigyelt cselekvés: (**break-in**₁, **deleted-logs**₂)
happen(**deleted-logs**,2)

Lehetséges végrehajtási nyomok:

(**recon**, **break-in**, **clean**, **deleted-logs**), konzisztens: (**theft**) V (**vandalism**).

(**recon**, **recon**, **break-in**, **clean**, **deleted-logs**),
konzisztens: konkurens **akármelyik 2 cél a 3-ból**.

(**recon**, **break-in**, **recon**, **clean**, **deleted-logs**), konzisztens: $\neq$.

(**recon**, **break-in**, **clean**, **recon**, **deleted-logs**), konzisztens: $\neq$.

(**recon**, **break-in**, **clean**, **deleted-logs**, **recon**), konzisztens: $\neq$.

(**recon**, **break-in**, **gain-root**, **clean**, **deleted-logs**), konzisztens: (**theft**).

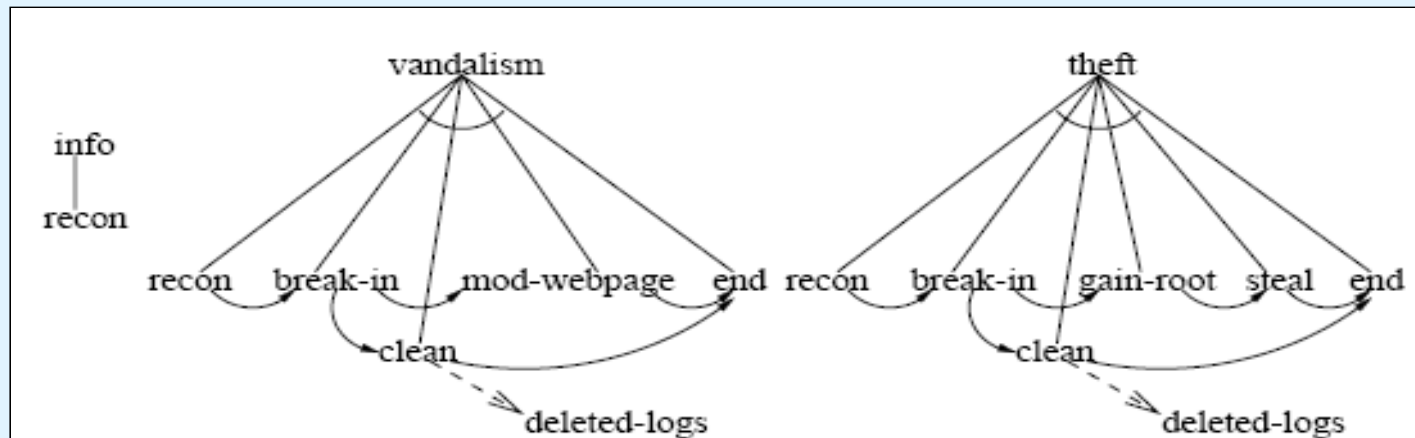
(**recon**, **break-in**, **clean**, **deleted-logs**, **gain-root**), konzisztens: (**theft**).

(**recon**, **break-in**, **mod-webpage**, **clean**, **deleted-logs**), konzisztens: (**vandalism**)

(**recon**, **break-in**, **clean**, **deleted-logs**, **mod-webpage**), konzisztens: (**vandalism**)

... ..

Pending Set2:
(**mod-webpage**)
(**gain-root**)



Legyen a megfigyelt cselekvés: (**break-in**₁, **deleted-logs**₂, **gain-root**₃)
happen(**gain-root**,3)

Lehetséges végrehajtási nyomok:

(**recon**, **break-in**, **clean**, **deleted-logs**, **gain-root**), konzisztens: (**theft**).

(**recon**, **recon**, **break-in**, **clean**, **deleted-logs**, **gain-root**),
konzisztens: (**theft**) $\wedge$ (**info**).

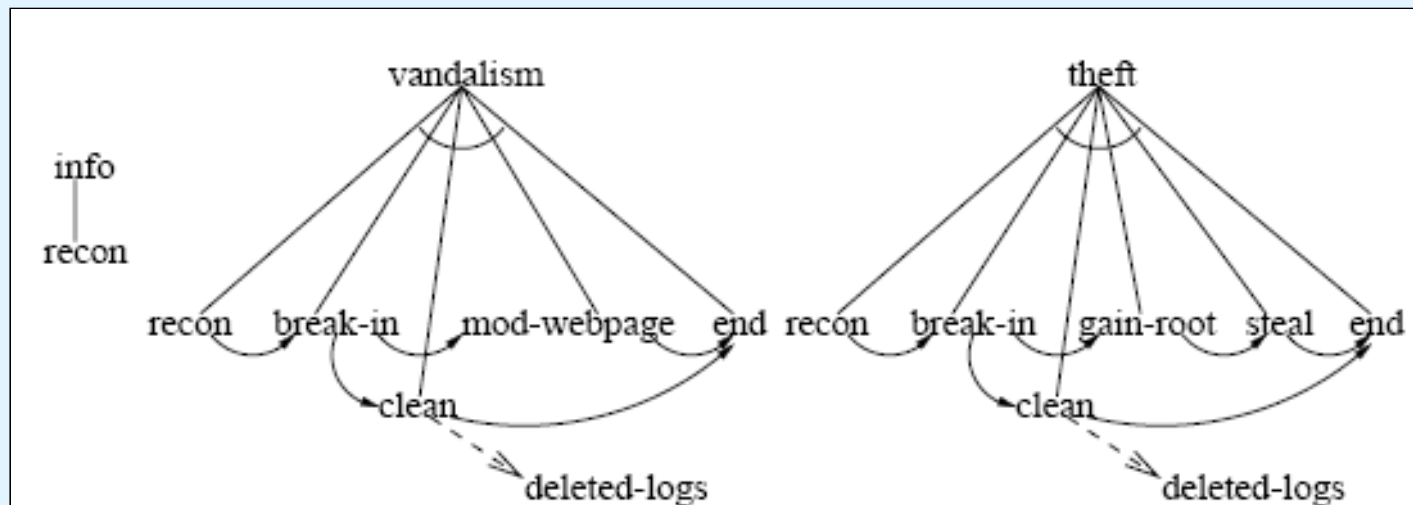
(**recon**, **break-in**, **recon**, **clean**, **deleted-logs**, **gain-root**), konzisztens: $\neq$.

(**recon**, **break-in**, **clean**, **recon**, **deleted-logs**, **gain-root**), konzisztens: $\neq$.

(**recon**, **break-in**, **clean**, **deleted-logs**, **recon**, **gain-root**), konzisztens: $\neq$.

... ..

Pending Set3:
(**steal**)



Valószínűségek bizonytalanságok kifejezésére

Keresett:

$$P(C \mid o) = ? = \sum_{e_i} P(e_i \mid o) \quad \text{ahol a } C \text{ cél benne van az } e_i \text{ magyarázatban}$$

Ennek alapján rangsorolni tudjuk az egyes célokat.

A számítás elemei:

$$P(e \mid o) = \alpha P(e) P(o \mid e) \quad \text{Bayes tétel}$$

$$P(e) = P(\text{célok}) P(\text{tervek} \mid \text{célok}) \quad \begin{array}{l} \text{magyarázatok a célokhoz tartozó tervek} \\ \text{(tervrészletek)} \end{array}$$

$$P(\text{célok}) = \prod_i P(C_i) \quad \text{a célok halmaza független}$$

$$P(o \mid e) = P(o_1 \mid e) P(o_2 \mid e, o_1) \dots P(o_n \mid e, o_1, \dots, o_{n-1})$$

megfigyelés egy n db. megfigyelés sorozat

$$P(o_i \mid e, o_1, \dots, o_{i-1}) = 1 / |\text{PS}(D_{i-1})|$$

D a hipotetizált tervek részfája, PS a soron lévő
Pendig Set, ill. abban feltüntetett cselekvések
számossága.