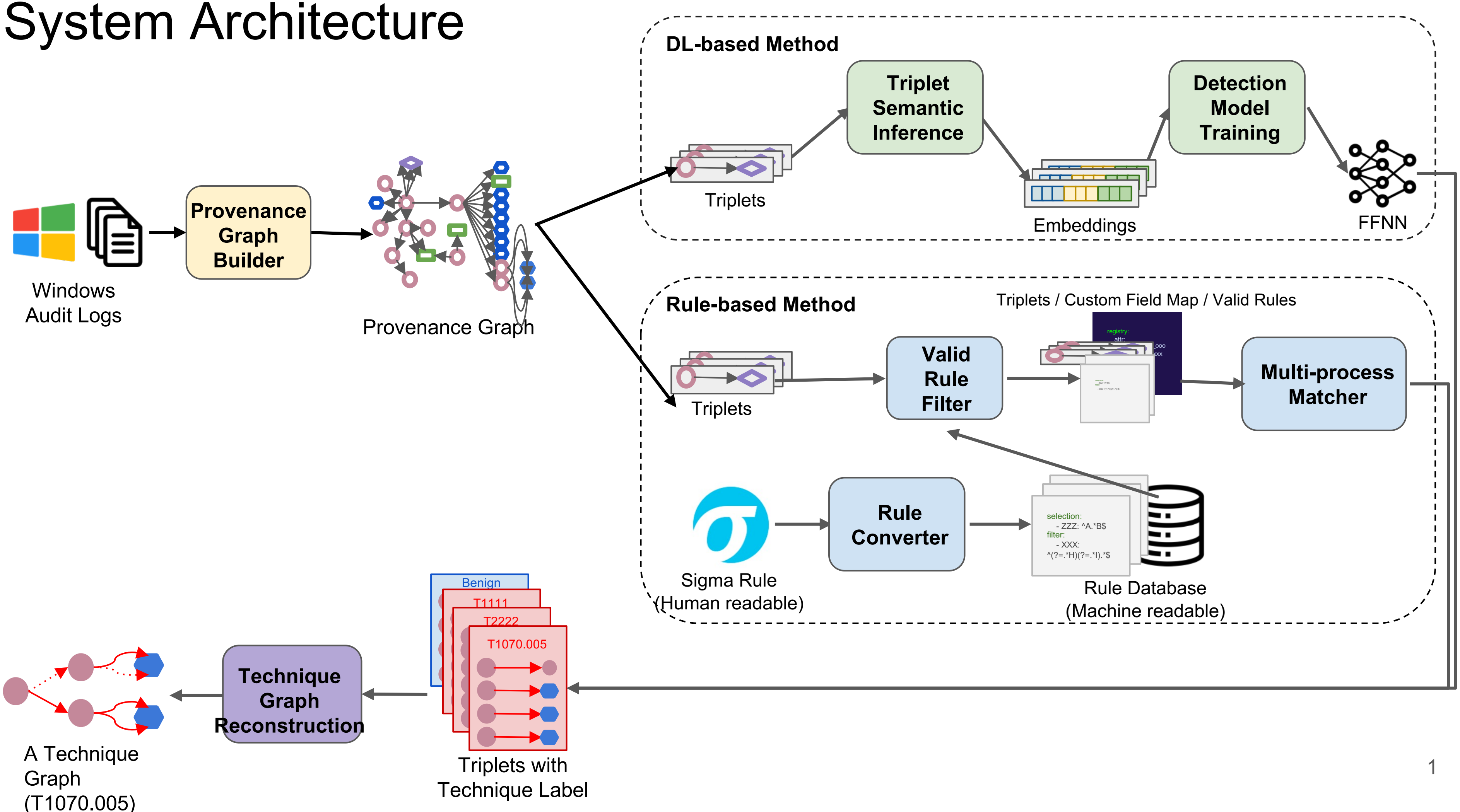
System Architecture



Windows Audit Logs

Provenance Graph Builder

Provenance Graph

DL-based Method

Triplets

Triplet Semantic Inference

Embeddings

Detection Model Training

FFNN

Rule-based Method

Triplets / Custom Field Map / Valid Rules

Triplets

Valid Rule Filter

```
registry:
  attr:
```

Multi-process Matcher

Sigma Rule (Human readable)

Rule Converter

```
selection:
    - ZZZ: ^A.*B$
filter:
    - XXX:
^(?=.*H)(?=.*I).*$
```

Rule Database (Machine readable)

Benign

T1111

T2222

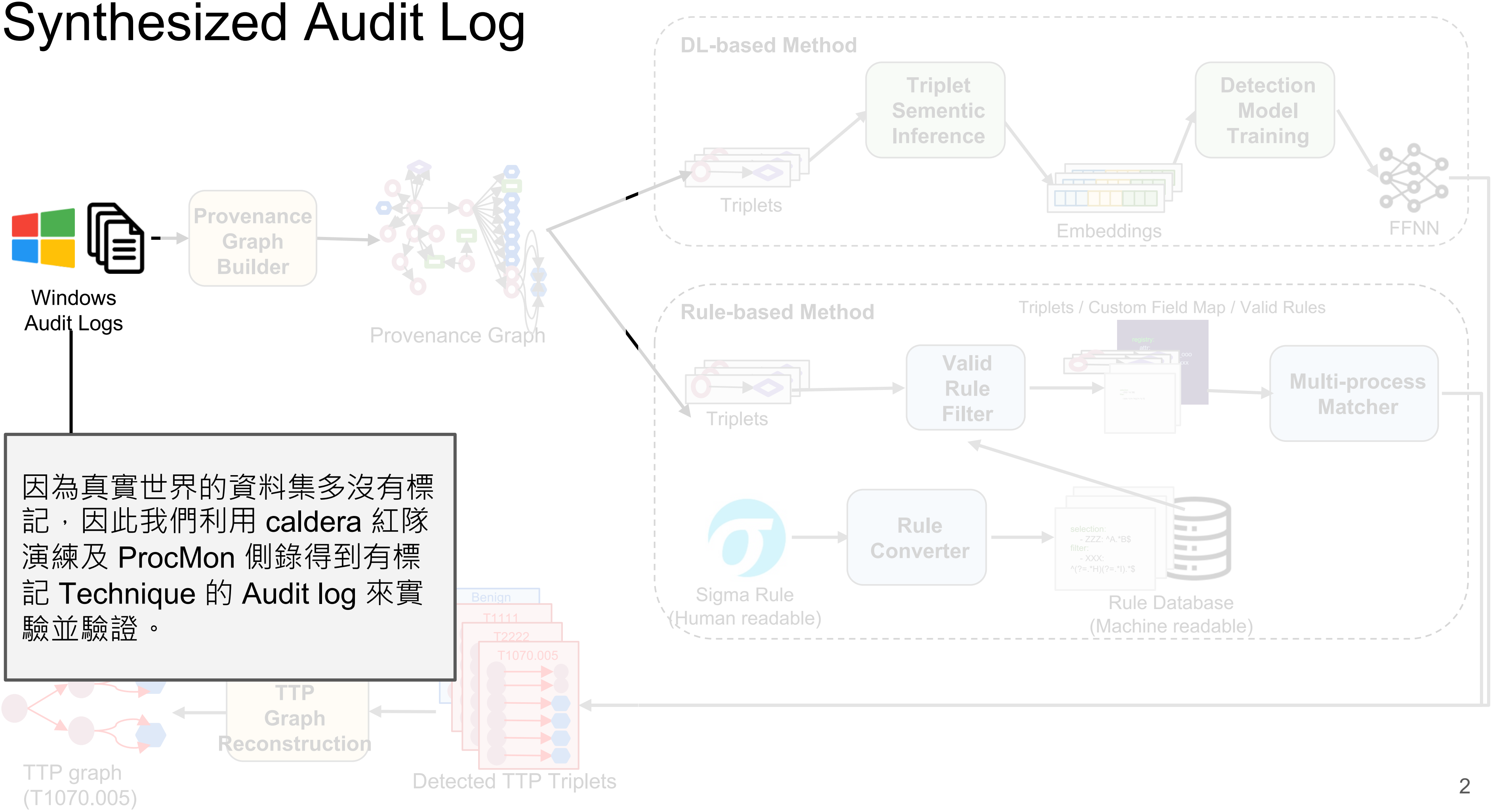
T1070.005

Triplets with Technique Label

Technique Graph Reconstruction

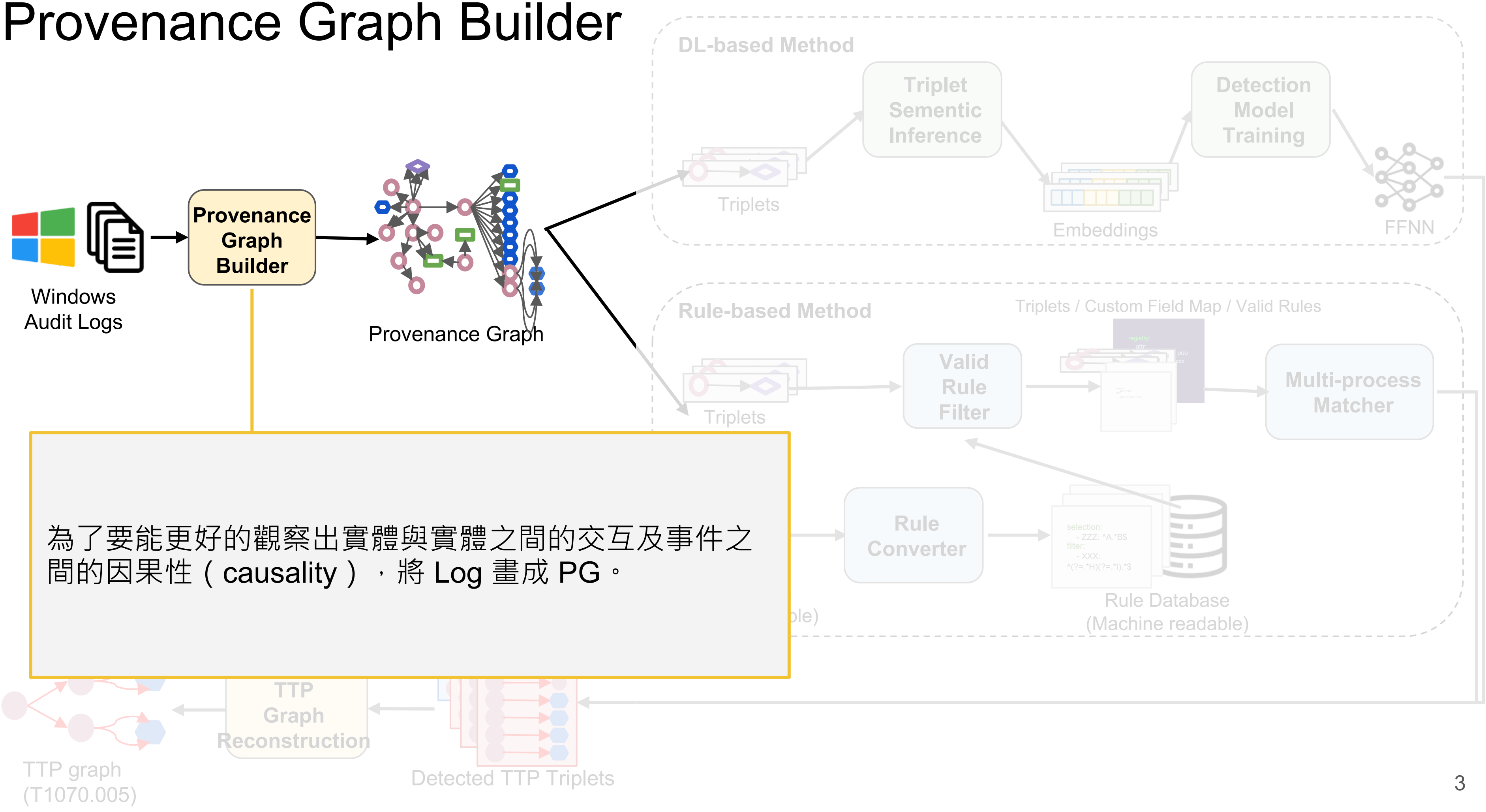
A Technique Graph (T1070.005)

Synthesized Audit Log



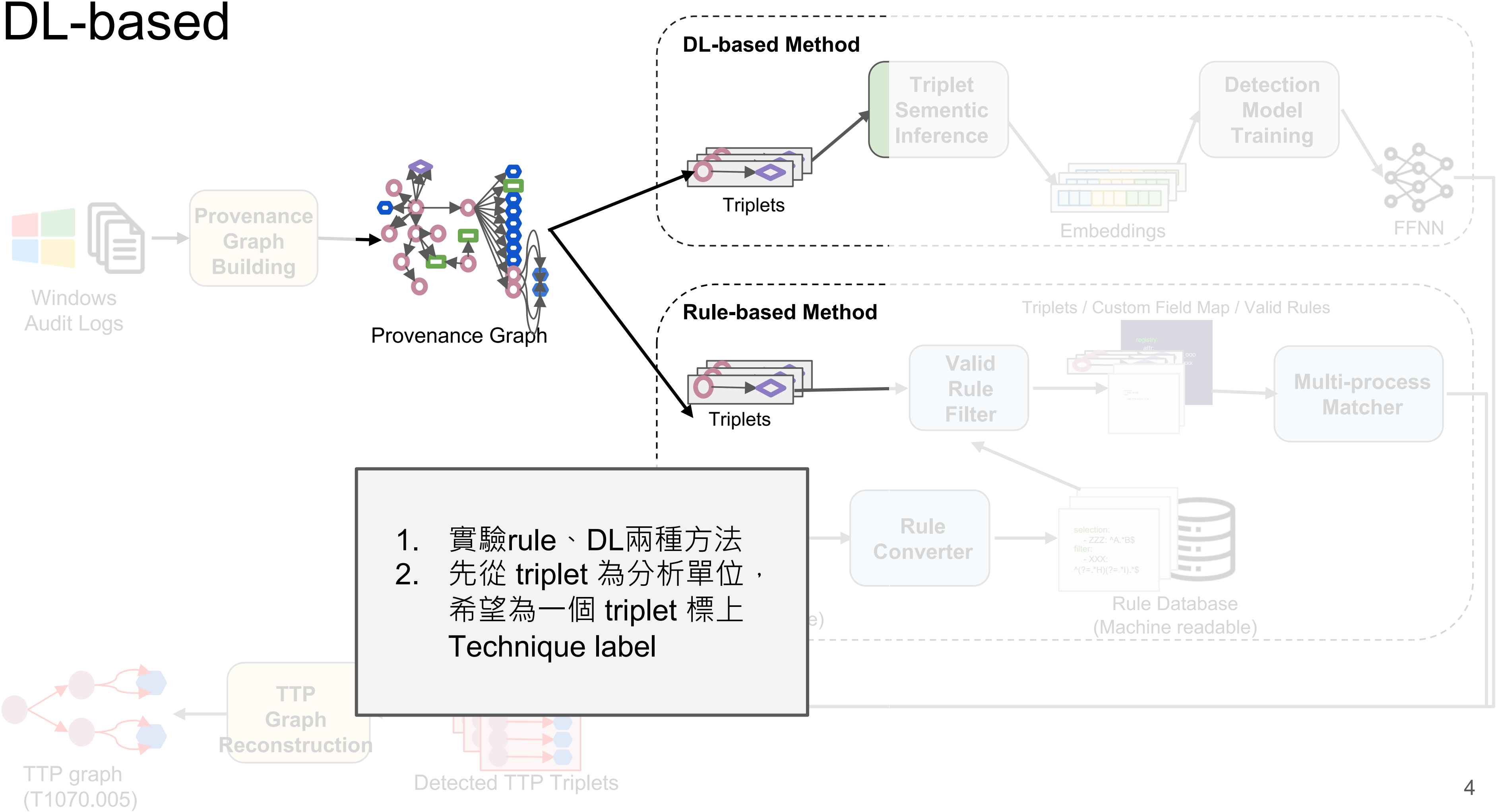
因為真實世界的資料集多沒有標記，因此我們利用 caldera 紅隊演練及 ProcMon 側錄得到有標記 Technique 的 Audit log 來實驗並驗證。

Provenance Graph Builder



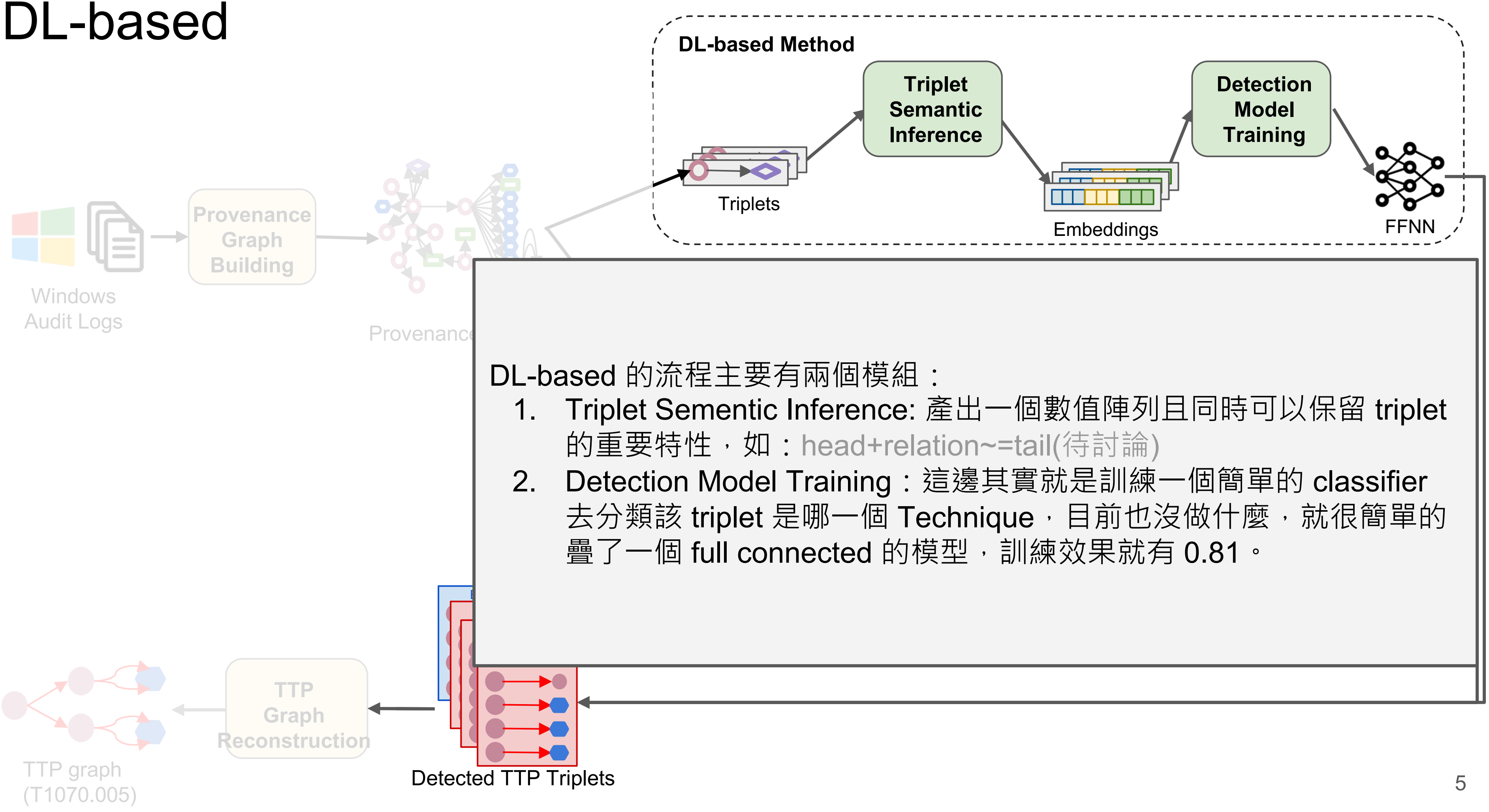
為了要能更好的觀察出實體與實體之間的交互及事件之間的因果性（causality），將 Log 畫成 PG。

DL-based



1. 實驗rule、DL兩種方法
2. 先從 triplet 為分析單位，希望為一個 triplet 標上 Technique label

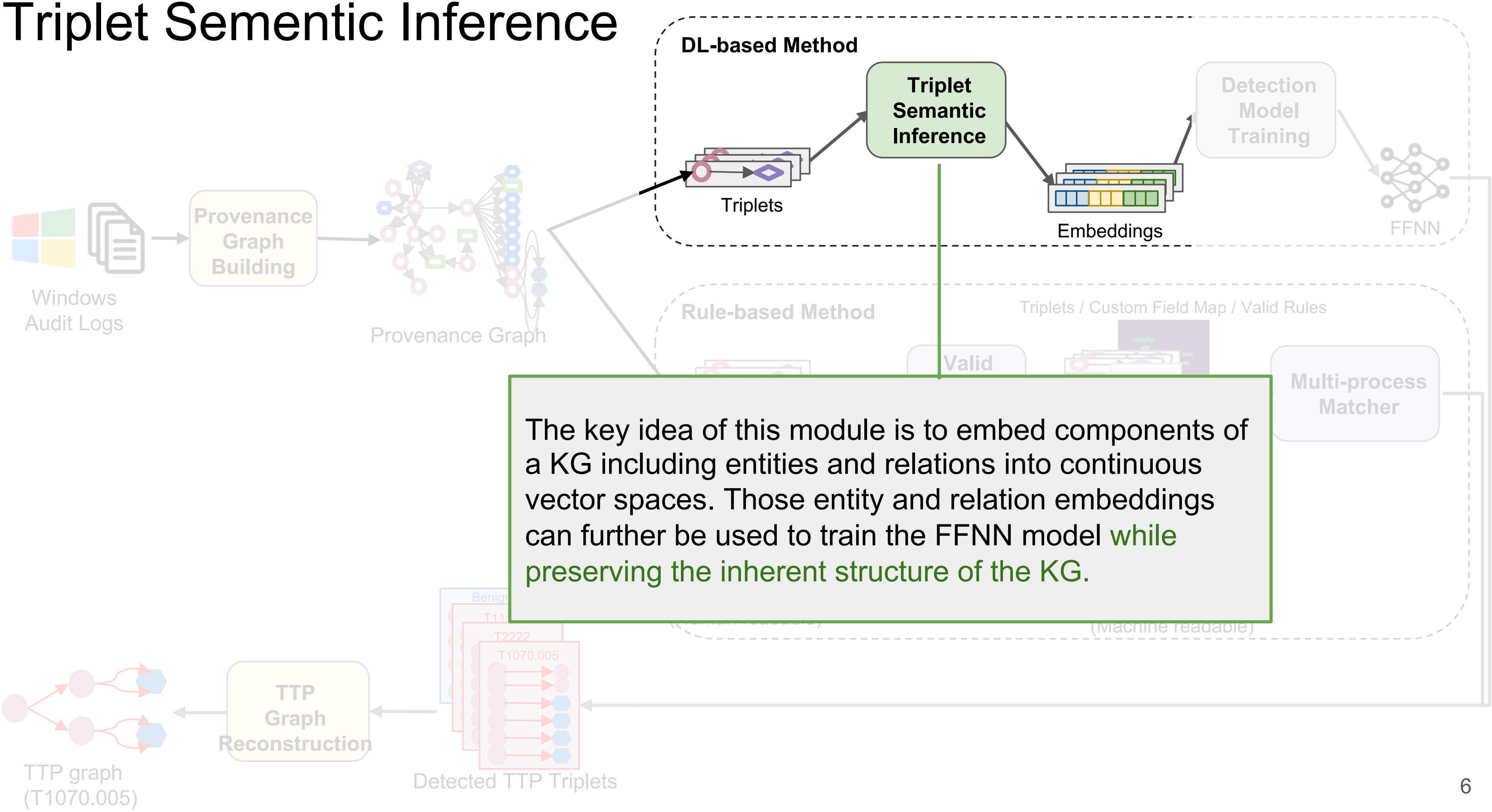
DL-based



DL-based 的流程主要有兩個模組：

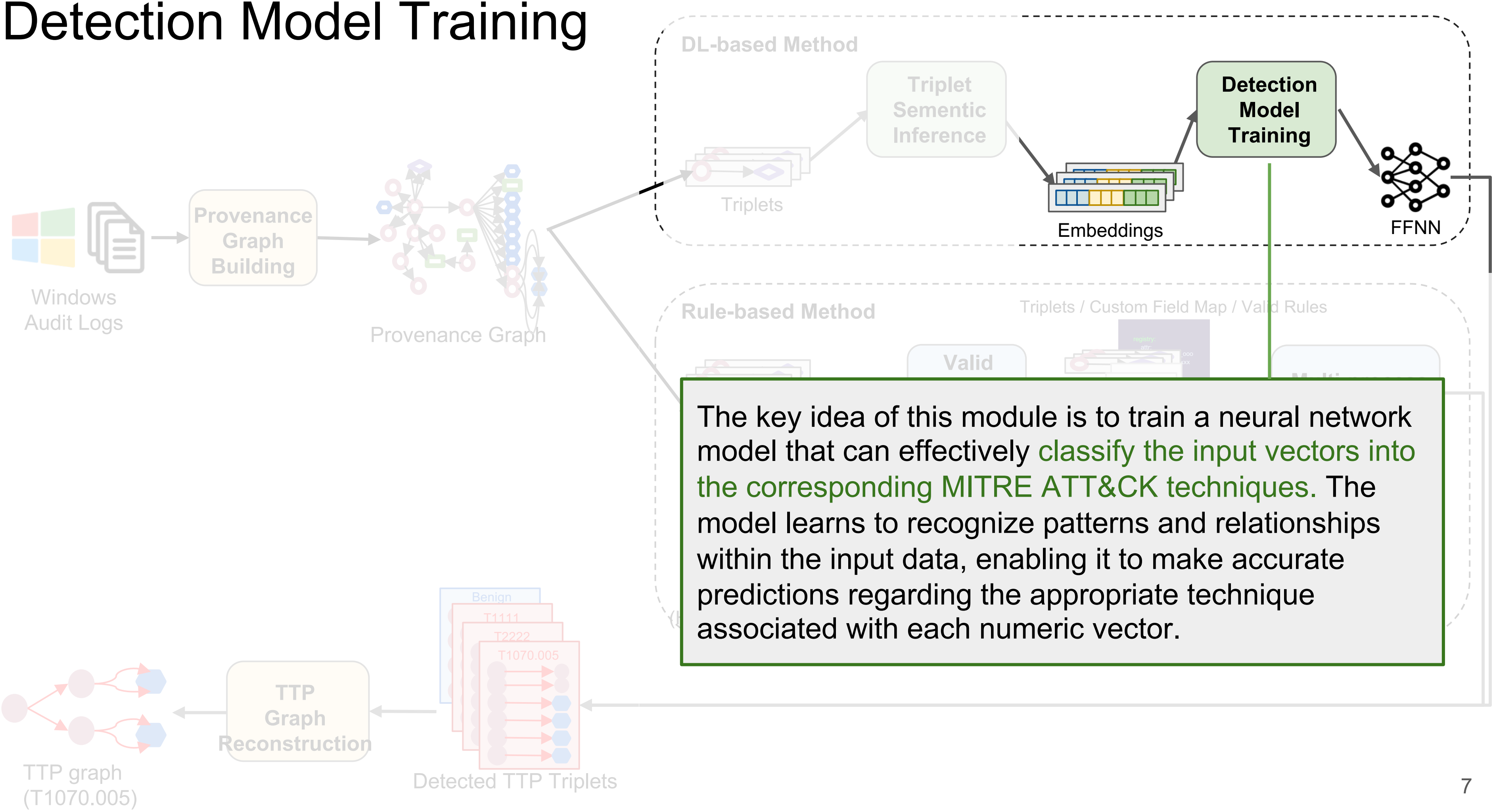
1. Triplet Sementic Inference: 產出一個數值陣列且同時可以保留 triplet 的重要特性，如：head+relation~=tail(待討論)
2. Detection Model Training：這邊其實就是訓練一個簡單的 classifier 去分類該 triplet 是哪一個 Technique，目前也沒做什麼，就很簡單的疊了一個 full connected 的模型，訓練效果就有 0.81。

Triplet Sementic Inference



The key idea of this module is to embed components of a KG including entities and relations into continuous vector spaces. Those entity and relation embeddings can further be used to train the FFNN model while preserving the inherent structure of the KG.

Detection Model Training

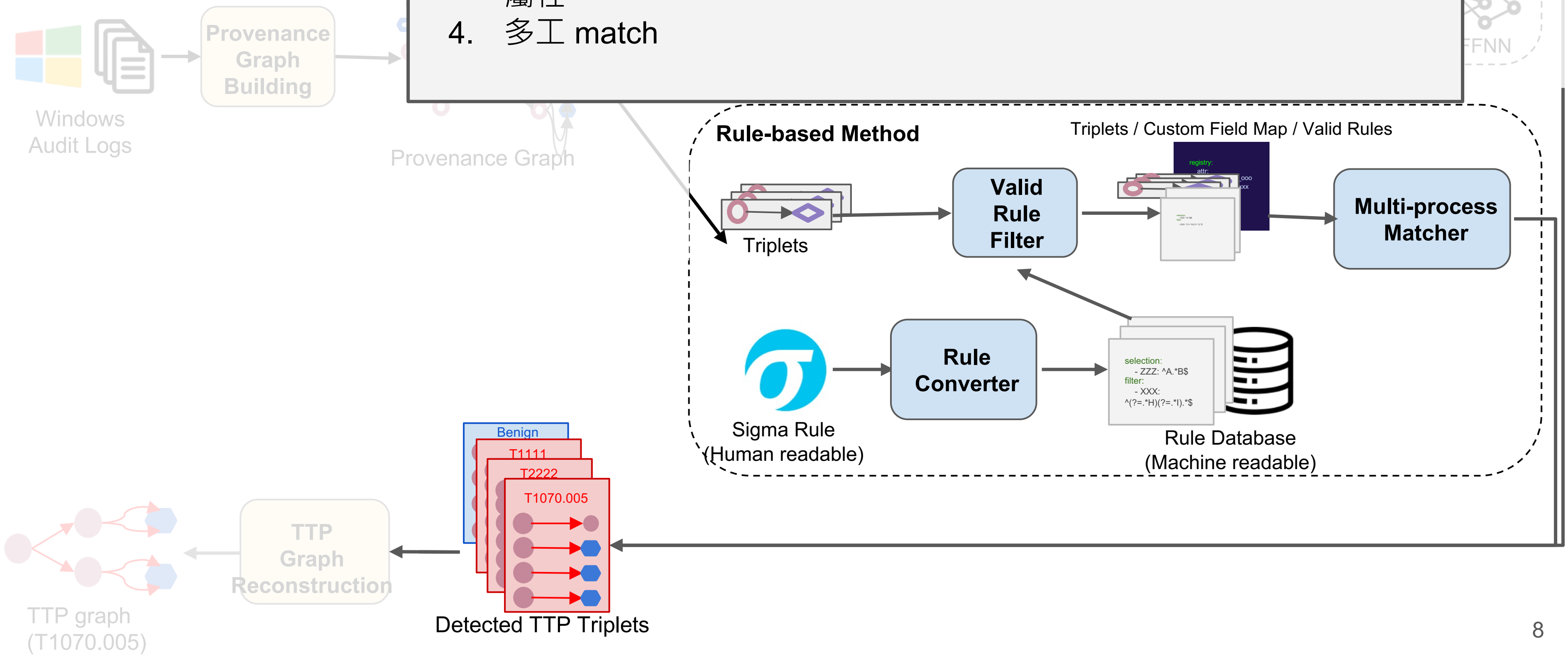


The key idea of this module is to train a neural network model that can effectively classify the input vectors into the corresponding MITRE ATT&CK techniques. The model learns to recognize patterns and relationships within the input data, enabling it to make accurate predictions regarding the appropriate technique associated with each numeric vector.

Rule-based

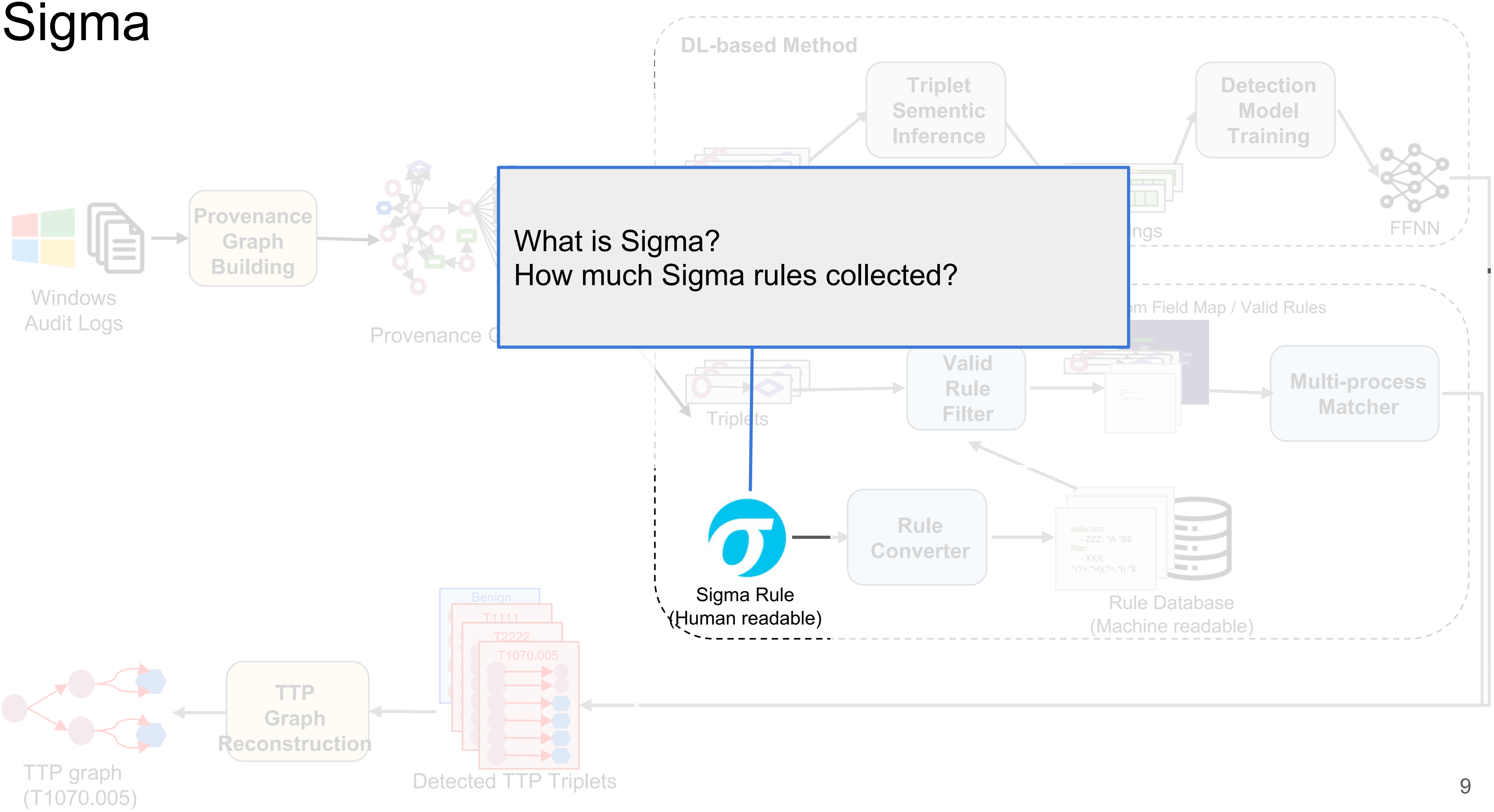
Rule-based 有幾件事要做

1. 蒐集 rule 存起來
2. 挑選出有效的 rule
3. 人工作一個屬性對應表。紀錄 Sigma rule 的屬性要對應到 log 的哪個屬性
4. 多工 match



```
selection:
  - ZZZ: ^A.*B$
filter:
  - XXX:
^(?=.*H)(?=.*I).*$
```

Sigma



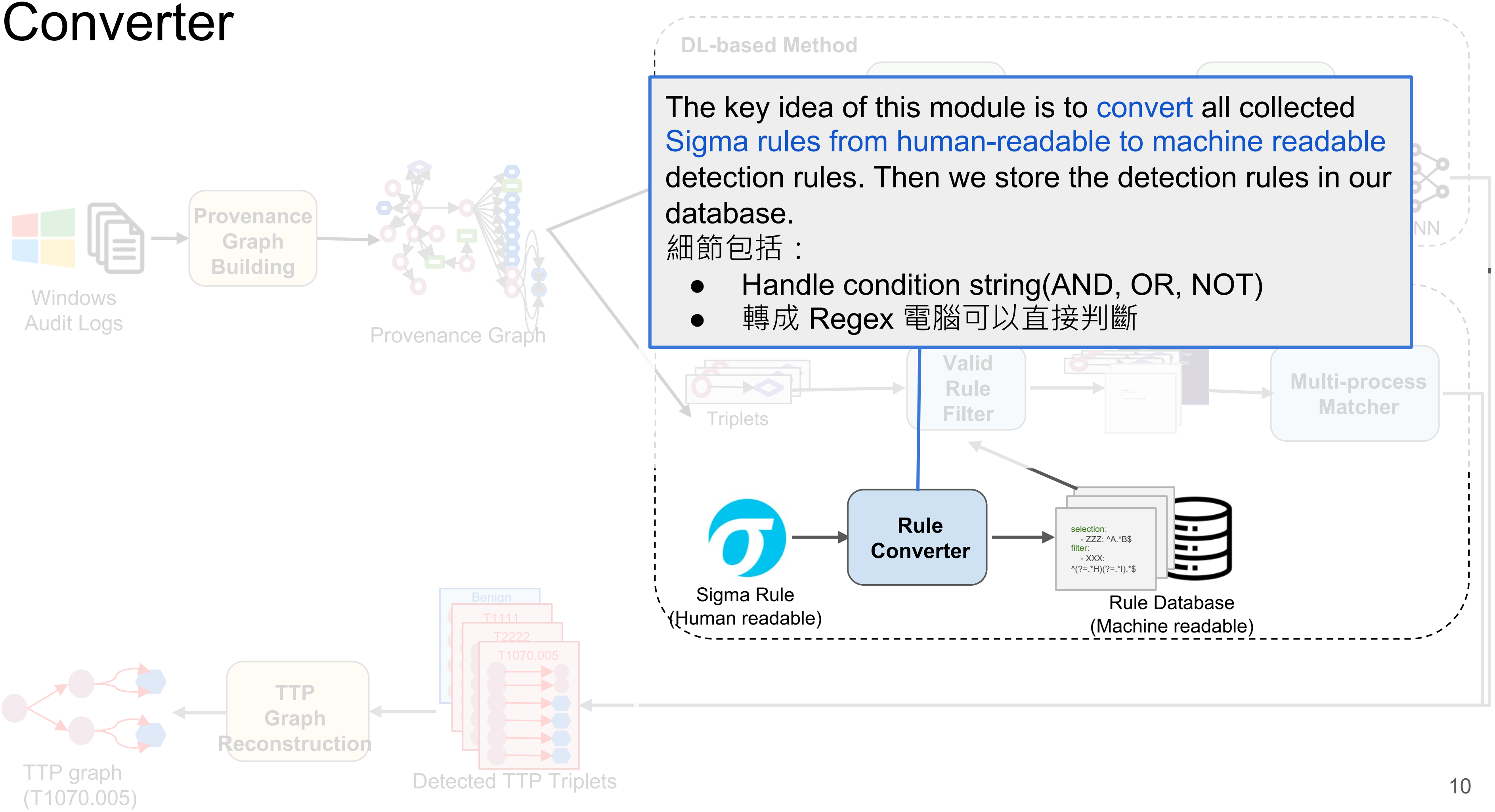
What is Sigma?
How much Sigma rules collected?

Converter

The key idea of this module is to convert all collected Sigma rules from human-readable to machine readable detection rules. Then we store the detection rules in our database.

細節包括：

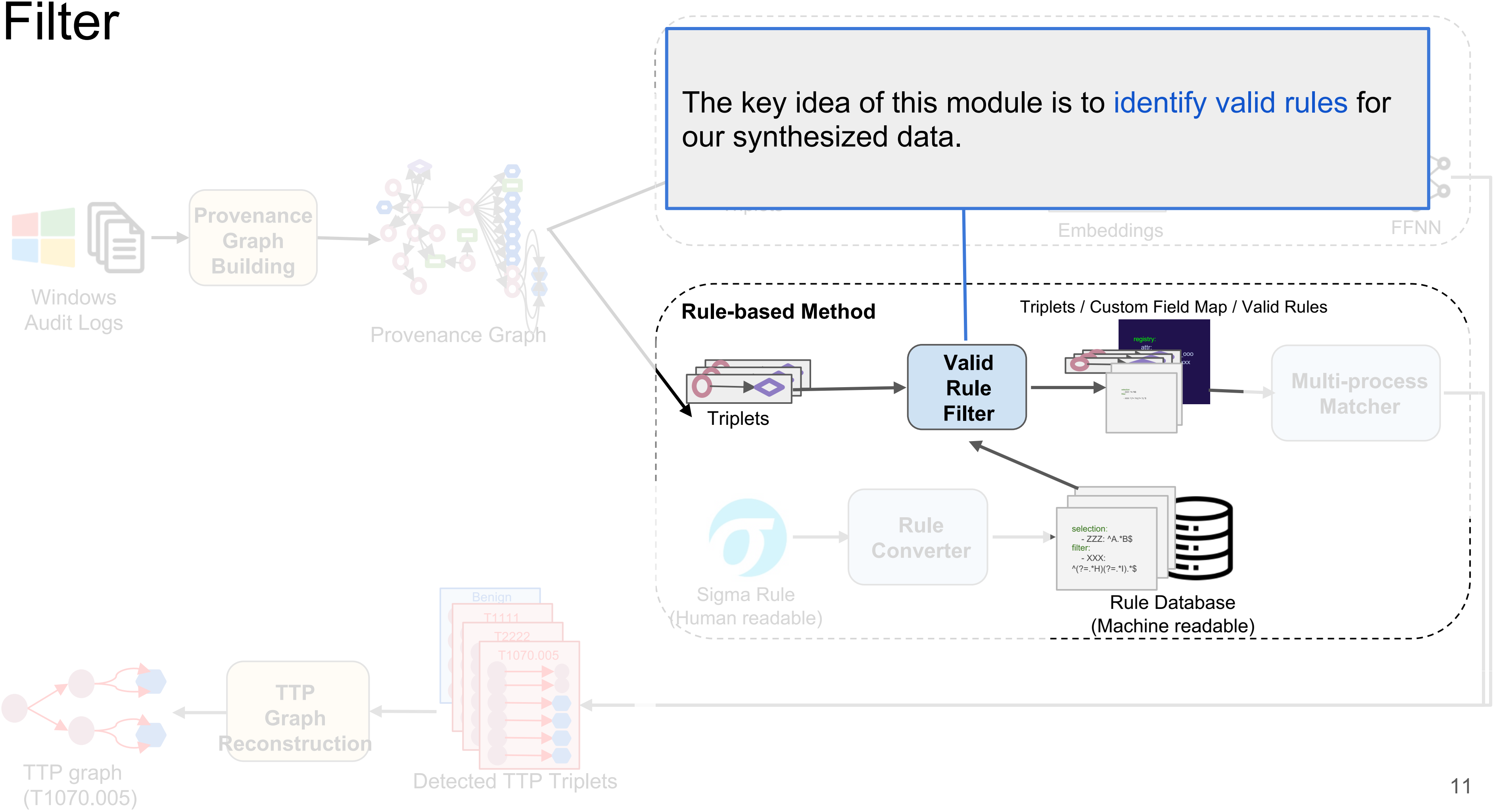
- Handle condition string(AND, OR, NOT)
- 轉成 Regex 電腦可以直接判斷



```
selection:
    - ZZZ: ^A.*B$
filter:
    - XXX:
^(?=.*H)(?=.*I).*$
```

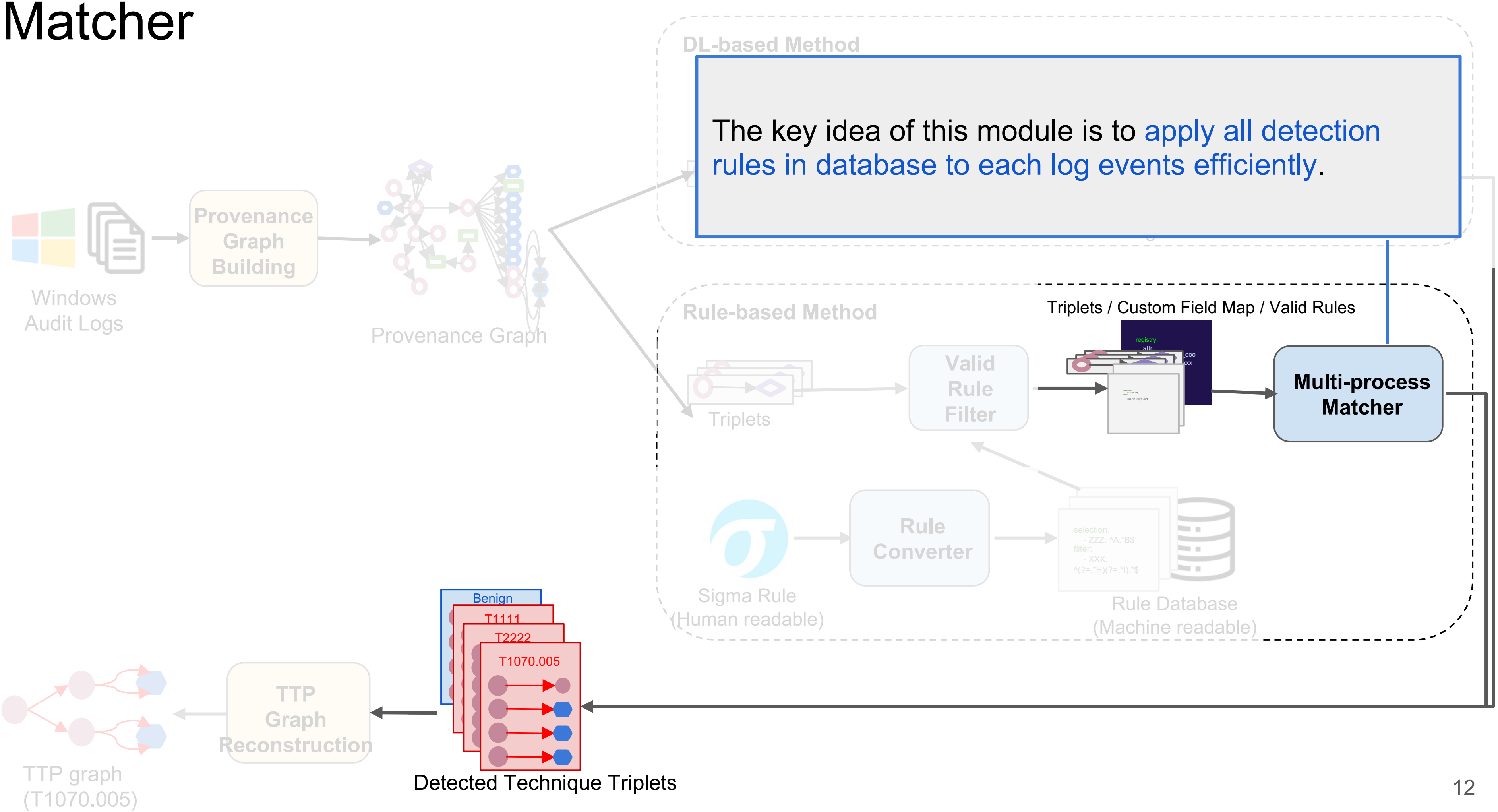
Filter

The key idea of this module is to identify valid rules for our synthesized data.



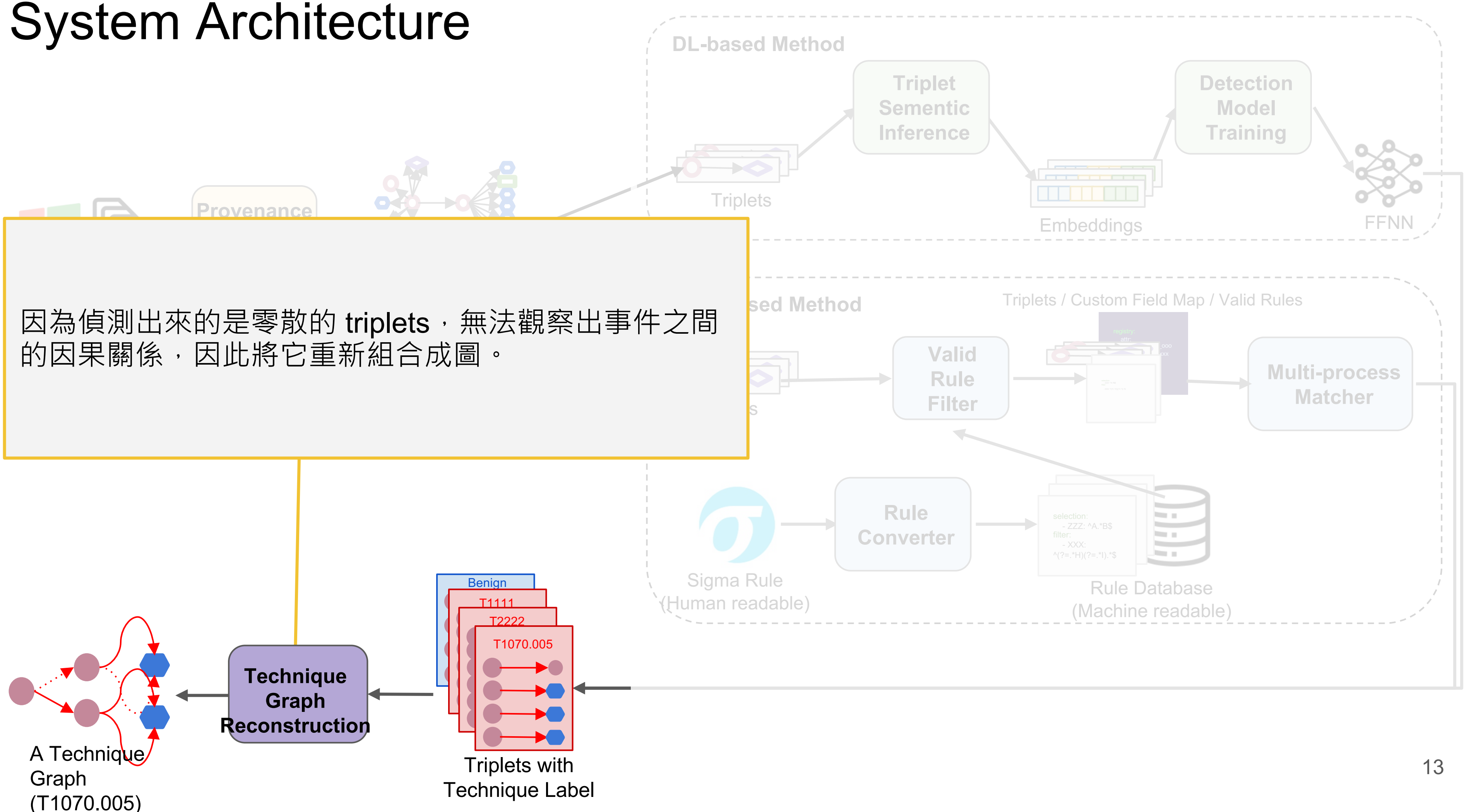
Matcher

The key idea of this module is to apply all detection rules in database to each log events efficiently.



System Architecture

因為偵測出來的是零散的 triplets，無法觀察出事件之間的因果關係，因此將它重新組合成圖。


DL-based Method
Triplet Semantic Inference
Detection Model Training
Triplets
Embeddings
FFNN
Provenance
sed Method
Triplets / Custom Field Map / Valid Rules
Valid Rule Filter
Multi-process Matcher
Rule Converter
selection:
- ZZZ: ^A.*B$
filter:
- XXX:
^(?=.*H)(?=.*I).*$
Sigma Rule
(Human readable)
Rule Database
(Machine readable)
Benign
T1111
T2222
T1070.005
Technique Graph Reconstruction
Triplets with Technique Label
A Technique Graph
(T1070.005)