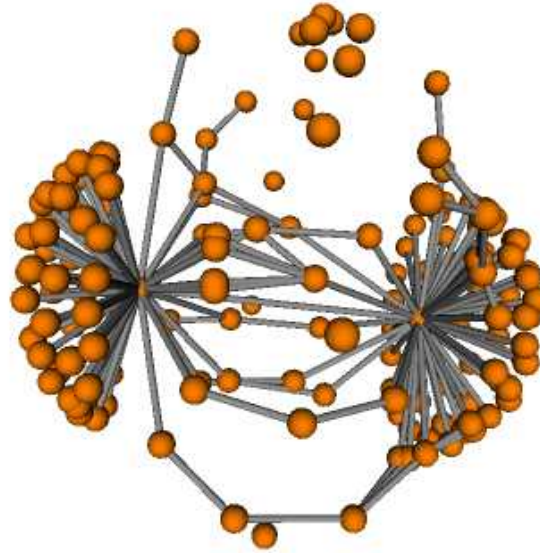


# Computer Forensics



Diffusion and Graph Spectral Methods for  
network forensic analysis

(by W. Wang and T.E. Daniels)

---

# Introduction

- Evidence Graph Model
- Graph spectral methods
- Heat diffusion methods
- Personal conclusions
- Questions

# Evidence graph model

- Why using graphs
  - Computer networks can be represented as mathematical graphs
  - Communication between different hosts is a subgraph of the whole computer network
  - Graphs are well defined. There are a lot of techniques to read data out of a graph
    - structure
    - pattern matching (extraction subgraphs)
    - connectivity

# Evidence graph model

- Difficulty finding patterns
  - Graphs has no natural order for nodes and edges. Correspondence must be used
  - There are structural variations. Number of nodes and edges can change

# Evidence graph model

- extensible graph model used
  - integrates evidence from heterogeneous sources
  - captures entities, events and functional states of the attack scenario for analysis
- $G = (N, E, L_N, L_E)$ 
  - $N$  : Nodes (Vertices)
  - $E$ : Edges
  - $L_N$ : Set of labels for attributes of nodes
  - $L_E$ : Set of labels for attributes of edges

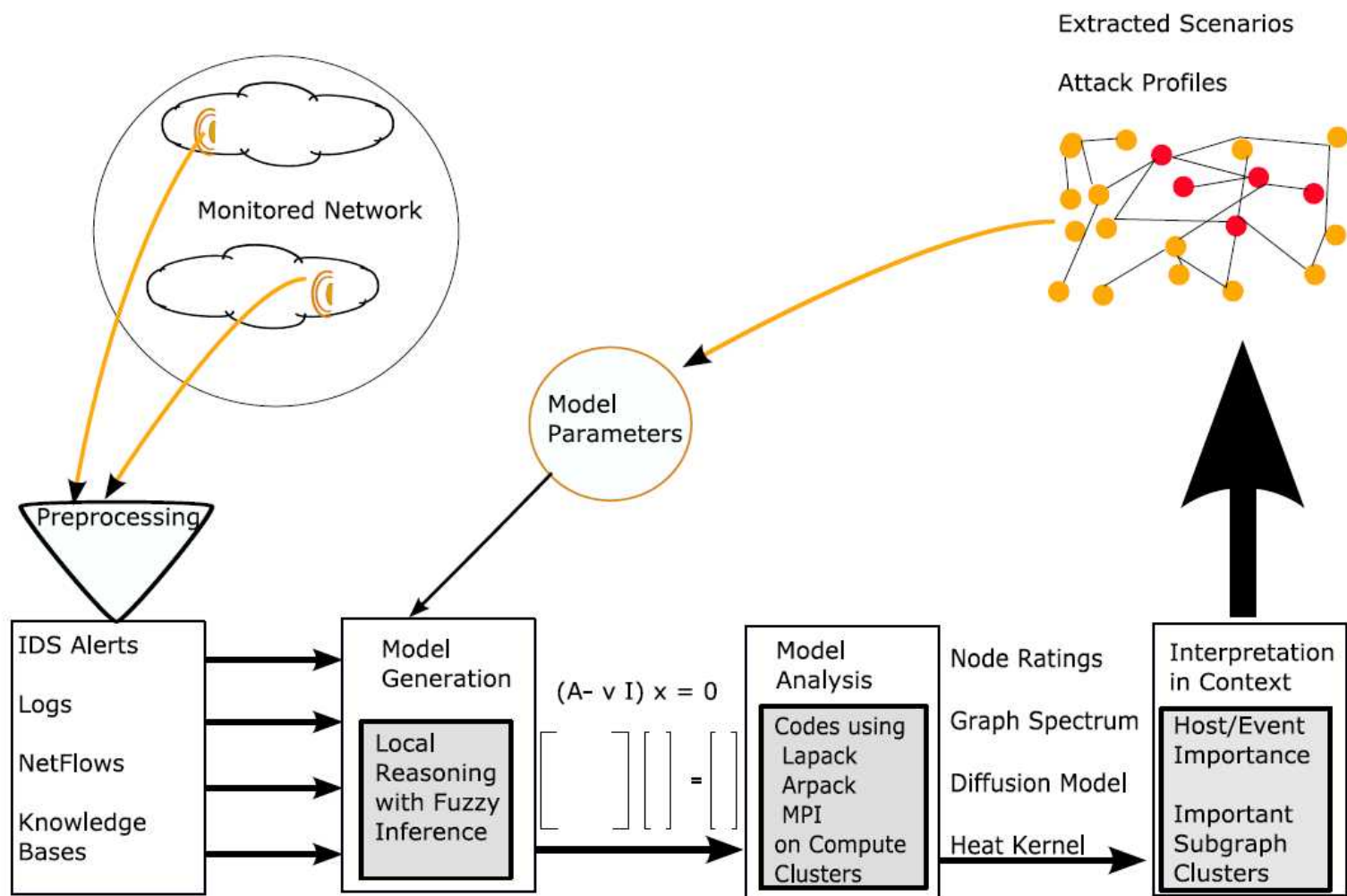
# Evidence graph model

- each node represents a host of interest
  - The state of the node describes the possible role of hosts in the attack scenario, such as
    - Attacker
    - Victim
    - Stepping Stone
    - Affiliated

# Evidence graph model

- each edge represents a piece of observed evidence
  - The priority score of the edge is a product out of:
    - Weight: represents the impact of the attack
    - Relevancy: Belief that the attack would achieve its expected impact.
    - Context importance: used to relate the significance of evidence with value of the host involved.

# Evidence graph model



# Attack extraction methods

- Having the evidence model we can apply different technique to
  - extract the attacks scenario.
  - profile the attack case profiling
- Following methods are explained:
  - Spectral Graph Method
  - Diffusion Method

# Spectral Method

- The Laplacian spectrum provides permutation invariant graph characteristics
- It's used to characterize the evidence graph, extracting two types of information
  - 1. Spectral features representing important nodes (Key player in the attack scenario)
  - 2. natural clusters of highly correlated suspicious nodes (extended attack group)

# Laplacian Spectrum

- The Laplacian representation of Graph  $G$ 
  - $w(u, v)$  represents the weight of all edges between  $u$  and  $v$
  - $d_u$  is the degree of the node  $u$ :  $d_u = \sum_v w(u, v)$

$$\hat{L}(u, v) = \begin{cases} 1 - \frac{w(u, v)}{d_u}, & \text{if } u = v \wedge d_u \neq 0 \\ \frac{-w(u, v)}{\sqrt{d_u \cdot d_v}}, & \text{if } u, v \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases}$$

# Laplacian Spectrum

- The Laplacian spectrum
  - obtained by the eigendecomposition of the Laplacian representation

$$\hat{L} = \Phi \Lambda \Phi^T$$

- Diagonal matrix of eigenvalues

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{|V|})$$

- matrix composed with eigenvectors

$$\Phi = (\Phi_1, \Phi_2, \dots, \Phi_{|V|})$$

# Laplacian Spectrum

- Properties of the Laplacian

- eigenvalues are positive and the smallest eigenvalue is zero

$$0 \leq \lambda_1 < \lambda_2 < \lambda_3 < \dots < \lambda_{|V|}$$

- multiplicity with  $\lambda_1$  gives the number of connected components of the graph

- Eigenvector associated with  $\lambda_2$  returns a vector which can be used for clustering nodes (Fiedler vector)

# Laplacian Spectrum

- Conclusions for evidence graphs
  - The eigenvalues are invariant permutations of the Laplacian
  - therefore they are stable to any “noise”
  - Laplacian spectrum characterizes following graph properties:
    - connectivity of the nodes
    - diameter and path length distribution

# (Heat) Diffusion Method

Second approach to transform the evidence graph analysis into steady state energy diffusion problems.

$$\Delta^2 T = \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} = \frac{c}{k} \frac{\partial T}{\partial t} - Q(x, y, z)$$

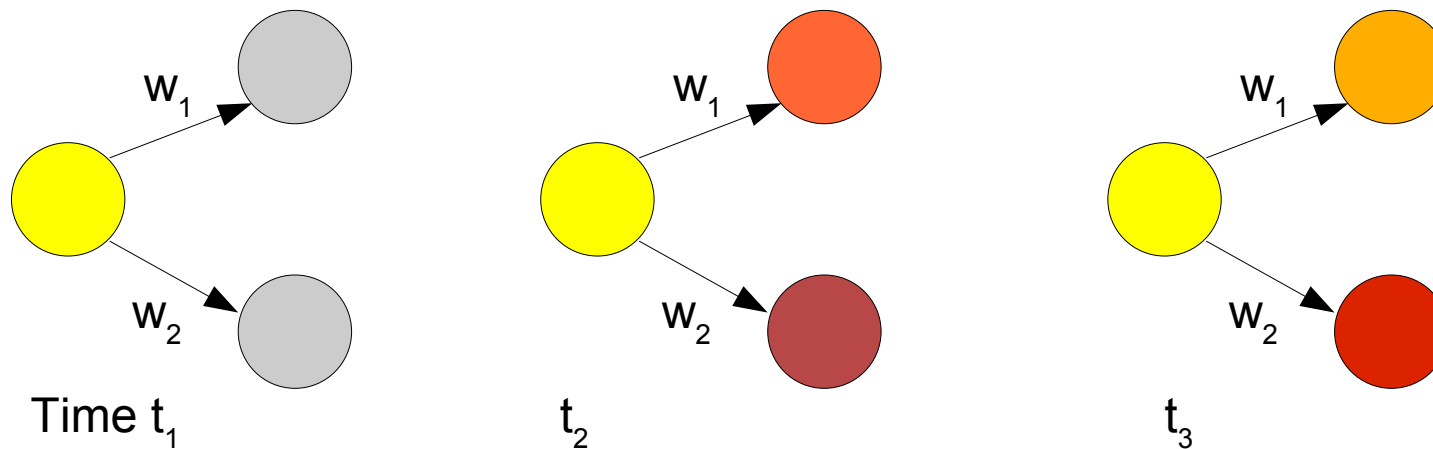
$\Delta$  is the continuous Laplacian operator

$Q$  represents the effects of an internal source of heat

$c$  and  $k$  are constants representing the heat capacity and conductivity of the material

# (Heat) Diffusion Method

- Using the graph  $G=(V, E)$ 
  - Each node is considered as an independent physical item
  - diffusion of heat can occur across the edges in  $E$



# (Heat) Diffusion Methods

- Why using this model
  - The attack process can be regarded as the propagation of suspicion
  - Each entity and piece of evidence are regarded as the container or carrier of certain amount of suspicion.
    - Heat is the suspicion
    - Temperature indicates the level of suspicion

# Diffusion vs. Spectral

- It's closely related to the spectral methods
  - invariants of the spectral methods are linked to the heat kernel
- The diffusion method can be solved by more direct numerical methods
  - extraction of eigenvectors and eigenvalues are not required
  - used to handle massive directed evidence graphs

# Personal conclusions

- Graph spectral methods
  - ++
    - It's are a good way for pattern matching.
    - pretty stable even if the evidence graph changes.
  - --
    - for large evidence graphs not so efficient for attack scenario extraction.

# Personal conclusions

- Heat diffusion methods
  - ++
    - A simple concept based on physical construct
    - Good for thresholding attack scenarios and single nodes

---

# Personal conclusions

- Main task of a project team will be finding the correct measurements to build the graph.
  - finding the parameters out of logged data
- Fine adjustment of parameters is not simple

---

# Questions?



Thank you for your interest.