PRACTICAL ATTACKS AGAINST GRAPH-BASED CLUSTERING

YIZHENG CHEN, YACIN NADJI, ATHANASIOS KOUNTOURAS, FABIAN MONROSE, ROBERTO PERDISCI, MANOS ANTONAKAKIS, NIKOLAOS VASIOLOGLOU
WHAT ABOUT ADVERSARIAL CLUSTERING?

Community Detection
Spectral Clustering
Hierarchical Clustering
Graph Embedding

Contributions
- First *practical* attempt to attack graph-based modeling techniques
- Cost analysis
- Defenses
INTUITION

Attacker’s Nodes → Inject Noise → Create Islands (Graph Partitions); Join Noisy Clusters (Graph Embeddings)

Attack Success: Different Label
• Dynamic domain generation for Command and Control (C&C)
• Algorithm in the *malware* or on the *server*
• Bipartite Graph $G = (U, V, E)$
  • Hosts ($U$) query NXDOMAINS ($V$)
  • An edge connects a vertex in $U$ and one in $V$
DGA ADVERSARIAL EXAMPLE

Create Islands
- Server-side DGA, push config
- Cost is too high? Size of islands can be bigger, depending on the clustering system hyperparameters
• Clustering needs global features
  • Local features: an image, a spam email, a PDF itself
  • Global features: all objects being clustered; classification features such as historical IPs

• Clustering system hyperparameters
• Graph modeling component
• Reimplementation in real-world telecommunication data
  • Accuracy: 96.08%, and False Positive Rate: 0.9%
  • Discovered 12 new DGAs
Attacker’s Goal: Change G to G’ to evade clustering.

**Minimal Knowledge**: Only G and open-source intelligence. E.g., 395K RE DGA domains from 14 malware families.


Prior Work: Randomly Inject Noise


Goal

- Make additional nodes to join the same cluster
  - Join the same graph partition
  - Have the same node embeddings

\[ G’ = \{U, V \cup V’, E \cup E’\} \]

Choices of \( V’ \) depends on attacker’s knowledge

- \( V’ \) very different from \( V \), such that \( V \cup V’ \): different label

Results

- Effective against spectral clustering, community detection, and node2vec
- Retraining defends against the same type of noise
Anomaly cost in U node degree percentile change.

Example

# of NXDomains: from 2 to 10

host percentile: from 48th to 95th

Results: Low Anomaly Cost

• 9.12% of infected hosts: 69.86th to 88.73th

• 90.88% remain the same: 99.74th to 99.85th
Prior Work: Small Community Phenomenon


Keep 3 nodes from V, 1 edge per remaining node (V'). E.g., server-side DGA

One configuration of the attack: Y nodes from V, X edges per remaining node.

\[ G' = \{U', V', E'\} \]
AGILITY COST

Agility Cost: Decrease in attacker bipartite graph density
- Losing redundancy
- Losing utility
Different configurations:
Y nodes from V, X edges per remaining node

- **Minimal Knowledge**: Success Rate 75.16%
- **Perfect Knowledge**: Guaranteed Success
- **Moderate Knowledge**: Surrogate network dataset should be **smaller** than the original network dataset. (Section 5.3.4, Fig 10)
NODE2VEC (KDD 2016)

Outperforms spectral clustering, DeepWalk (KDD 14), and LINE (WWW 15).

Start: $h_1$, Walk Length: 6, Random Walk Sample =  $\{h_1, q_2, h_2, q_6, h_3, q_7\}$

Neighborhood Size: 3
- $N(h_1) = \{q_2, h_2, q_6\}$, $N(q_2) = \{h_2, q_6, h_3\}$, $N(h_2) = \{q_6, h_3, q_7\}$

Learn embeddings: maximize $P(N(node) \mid node)$, similar to Word2Vec

- $N(h_4) = \{q_8\}$, $N(q_8) = \{h_4\}$, Undersampled
NODE2VEC: SMALL COMMUNITY ATTACK

- Guaranteed successful attack
  - # of shared Hosts = 1
  - Or <= 40 DGA Domains
- Randomness in the middle of the plot
- Fewer guarantees and higher costs than Spectral Clustering.
## AGILITY COST: SMALL COMMUNITY ATTACK

<table>
<thead>
<tr>
<th>Spectral Clustering</th>
<th>Hyperparameter</th>
<th>Minimum Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SVD Rank 35</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>SVD Rank 50</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>SVD Rank 80</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>SVD Rank 100</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>SVD Rank 200</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>SVD Rank 300</td>
<td>0.22</td>
</tr>
<tr>
<td>node2vec</td>
<td>Neighborhood Size 6</td>
<td>0.415</td>
</tr>
</tbody>
</table>

- SVD rank 35: no cost
- SVD rank 200: highest cost
- Higher cost using node2vec than spectral clustering
DEFENSE: TUNING HYPERPARAMETERS

Pick the elbow

![Graph showing Eigenvalue vs Factor Number and Attack Success Rate vs Number of Eigenvalues]
Walk Length 12 has lower validity scores than 20
LIMITATION AND FUTURE WORK

- General Attack, specific implementation
- **Anomaly cost**: node degree change
- **Agility cost**: graph density change

**Injectable/Removable nodes**
- Cost of manipulating nodes
- Cost of manipulating edges
- Bound the cost
SUMMARY

• First *practical* adversarial clustering analysis for graph-based modeling
  • Threat model
  • Targeted Noise Injection Attack
  • Small Community Attack

• Low Attack Cost
  • Hosts do not become more anomalous
  • No graph density decrease in spectral clustering; higher cost in node2vec

• Defenses
  • Retraining with noise
  • Tuning hyperparameters
THANK YOU!

YIZHENG CHEN

YZCHEN@GATECH.EDU