令和6年度　秋学期　卒業研究3 (BI)
学士論文

題目 GCNを用いたFake型CPAに対する脆弱ノード検出法

指導教員 上山憲昭 教授

立命館大学 情報理工学部 セキュリティ・ネットワークコース

学籍番号 2600200056-2
上野優衣

令和7年1月31日

概要

CDN(content delivery network) に代わるネットワークとして，ICN(information-centric networking) の研究が進められている．しかし，ICN には悪意を持ったユーザが不当なコンテンツをルータにキャッシュさせることで，キャッシュヒット率を低下させるコンテンツポイズニング攻撃 (CPA:content poisoning attack) という脆弱性が存在することが指摘されている．中でも，独自 Fake 型 CPA は検知がしづらく，その対策についての研究は十分になされていない．そこで，本稿では畳み込みグラフニューラルネットワーク (GCN:graph convolusional network) を用いて，Fake コンテンツが注入されている可能性のある脆弱なノードを検出する手法を提案する．さまざまな条件下での GCN の予測精度を測定し，有効な推測を行うことができる教師データ数の閾値を調査する．

目 次

第1章　序論

1.1　研究の背景

現在，コンテンツ配信のプラットフォームとして Content Delivery Network(CDN) が主に使用されている．CDN では，コンテンツをサーバにキャッシュすることで，高速なコンテンツ配信を実現している．近年，スマートフォンなどの高性能な移動型端末の普及により，コンテンツプロバイダによる映画やドラマなどの大容量コンテンツの配信サービスの需要がさらに拡大している．また，スマートフォンの普及に伴う SNS や動画投稿サービスなどのユーザが生成するコンテンツの投稿サービスの普及も，インターネットのトラフィックが増加する原因となっている．CDN を含む現在の Internet Protocol (IP) では，コンテンツ要求を行う際に IP アドレスをもとにパケット転送を行うホスト指向の通信が行われているが，コンテンツ配信がトラフィックの大部分を占める昨今では，このホスト指向の通信方式は非効率的である．

そこで，コンテンツ名を指定してコンテンツを要求することで，従来の IP で必要であった名前解決のためのオーバーヘッドを削減可能かつ，要求パケット (Interest) の転送経路上のルータからもコンテンツの取得を可能とする情報指向ネットワーク (ICN: information-centric networking) の研究が進められている．また，ICN では誰でもコンテンツをアップロードすることができるため，近年のユーザ生成コンテンツの増加という傾向にも適応している．

しかし，悪意を持ったユーザが不当なコンテンツをルータにキャッシュさせることで，キャッシュヒット率を低下させるコンテンツポイズニング攻撃 (CPA: content poisoning attack) が指摘されている [1].CPA によって，ルータのキャッシュ(CS:cache store) に偽りのコンテンツがキャッシュされキャッシュヒット率の低下などを引き起こす問題がある．

CPA には，Fake 型と Corrupted 型という 2 種類の攻撃が存在する [2]．ICN には，コンテンツの要求者が受信コンテンツに対して，コンテンツに対応する公開鍵で生成されたデジタル署名が一致するかを確認することでコンテンツの正当性を確認する方法が存在する．Corrupted 型は，署名とコンテンツが一致しないため．デジタル署名を確認することで検知が可能である [3]．一方で，Fake 型は有効なデジタル署名を持つコンテンツを配信するため，署名による検知は困難である．さらに Fake 型 CPA の中でも．自身で無意味なオリジナルのコンテンツ (Fake コンテンツ) を生成し，Bot からコンテンツを要求させる独自 Fake 型 CPA は，正常なユーザがコンテンツを受信することがないため，検知がさらに困難となる．

1.2 研究の目的

正常なユーザからの要求が発生しない独自 Fake 型 CPA では，コンテンツの正当性を確認する機会がなく検知されづらいため，不当なコンテンツを排除するにはキャッシュされているコンテンツ全てを正当なコンテンツであるか検証する必要がある．しかし，ネットワークの全てのノードのキャッシュを全て検証することは現実的ではない．そこで，本稿では，Fake コンテンツが注入されている可能性のある脆弱なノードを，一部のノードのキャッシュ情報をもとに GCN(graph convolusional neural network) を用いて検出する方式を提案する．

第2章　関連研究

2.1　CPA の分類

CPA には大きく分けて 4 つの型が存在する [2].

- **結託 Corrupted 型:** 実在するコンテンツを騙った署名が一致しない Fake コンテンツを Bot の要求に応じて配信
- **自然 Corrupted 型:** 実在するコンテンツを騙った署名が一致しない Fake コンテンツを正常ユーザからの要求に応じて配信
- **詐称 Fake 型:** 実在するコンテンツのデジタル署名を持った Fake コンテンツを正常なユーザからの要求に応じて配信
- **独自 Fake 型:** 攻撃者が独自に生成した Fake コンテンツを Bot の要求に応じて配信

Corrupted 型 CPA の 2 種類は署名が一致しない Fake コンテンツを配信するため，デジタル署名を検証することで容易に検知が可能である．また，検知した Fake コンテンツを CS から削除するために，検知した不当なコンテンツをルータに通知する [3] などの防御法が研究されている．

詐称 Fake 型 CPA は，公開鍵を管理する認証局 (CA: certification authority) の職員と攻撃者が結託して，実在するコンテンツの署名に使用している公開鍵を攻撃者の公開鍵と書き換えることによって成立する攻撃である．これは公開鍵を一つの機関でのみ管理することに脆弱性がある．これに対して，分散型台帳である IOTA でコンテンツ名を管理することで，詐称 Fake 型 CPA の発生を防御する手法が提案されている [4]．また，詐称 Fake 型 CPA は CA 職員との結託が必要であるため，頻繁に実現させることは難しいと考えられる．

2.2　独自 Fake 型 CPA

独自 Fake 型 CPA では，攻撃者が生成した Fake コンテンツに Interest を転送するのは Bot のみであり，正常ユーザからの要求は発生しない (図 2.1) ため検知されづらい．また，攻撃者が独自に生成したコンテンツそのものに署名を取得しているため，詐称 Fake 型のような CA 職員との結託も不要で実現が容易である．しかし，独自 Fake 型 CPA がネットワークに与える影響については調査が行われているが，実際に独自の Fake コンテンツを検知する手法は考慮されていない．

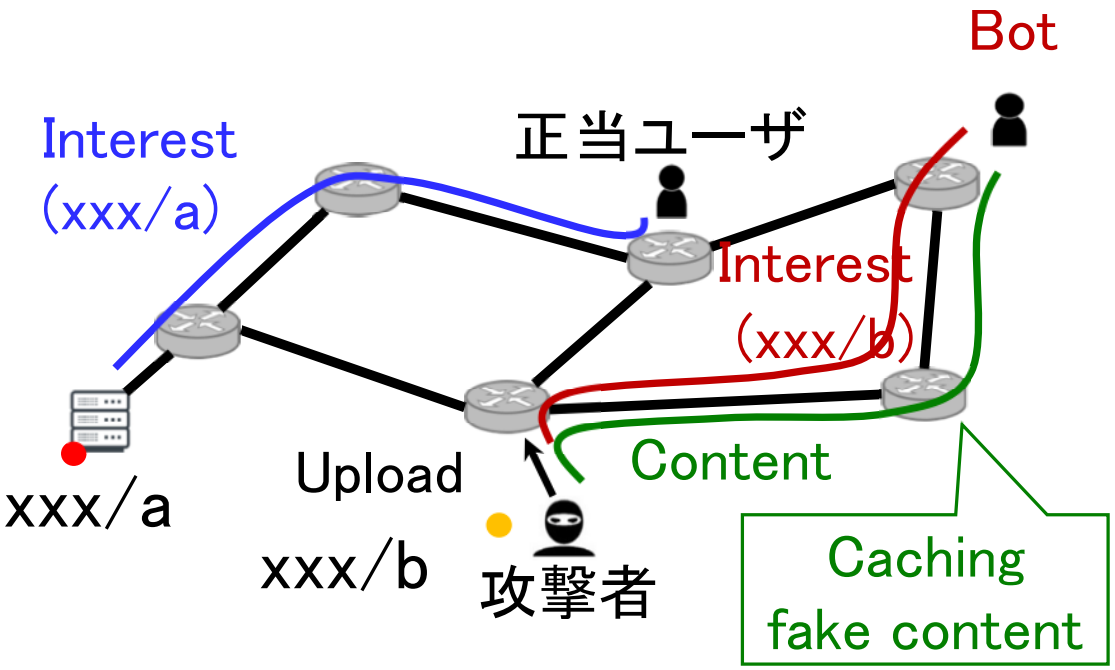


図 2.1: 独自 Fake 型 CPA の Interest 転送の様子

また，独自 Fake 型 CPA の Fake コンテンツは，Bot のみに依存する Interest 発生の特性から，Fake コンテンツのオリジナルが存在するノードまたは Bot が配置されているノードの近くにキャッシュされやすいことが明らかにされている [2]．そのため，ネットワーク内で Fake コンテンツがキャッシュされているノードが判明すれば，周辺ノードに Fake コンテンツのオリジナルか Bot が存在する可能性が高いと考えられる．

本稿では，独自 Fake 型 CPA の攻撃を受けているトポロジに対して，Fake コンテンツが注入される可能性が高い脆弱なノードを検出することで，独自 Fake 型 CPA の防御が特に必要なノードを明らかにする．

第3章　提案方式

3.1　GCN

従来のニューラルネットワークが画像やテキストを学習するモデルであるのに対し，グラフニューラルネットワーク (GNN: graph neural network) はグラフ構造データを学習させる機械学習モデルである．また，畳み込みニューラルネットワーク（CNN: convolutional neural network）は，主に画像認識の分野で活用されてきた機械学習モデル [5] である．一定のストライド（間隔）で入力データとフィルタを畳み込み演算することによって画像の局所的な特徴を抽出する畳み込み層と，特徴の圧縮を行うプーリング層を繰り返すことで，学習を行う手法である．グラフ畳み込みニューラルネットワーク (GCN) は，GNN の中でも代表的な手法のひとつである．CNN が画像の隣接するピクセルの特徴の畳み込みを行う (図 3.1) のに対し，GCN では自身のノードと隣接するノードの特徴量の畳み込みを行う (図 3.2)．GCN には，ノードのクラス分類，ノード間にリンクが存在するかどうかの予測，グラフのクラス分類などさまざまなタスクが存在する [6]．

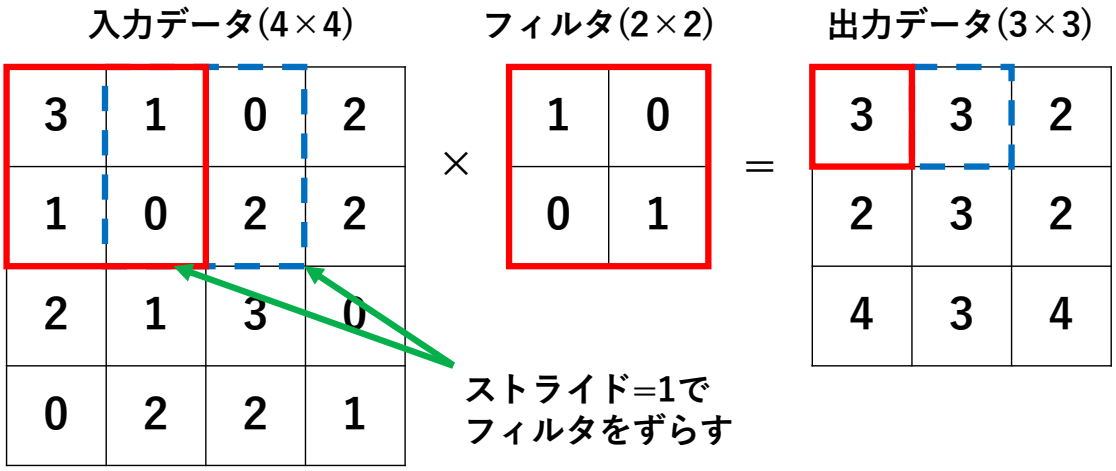


図 3.1: CNN の畳み込みのイメージ

ノードのクラス分類のタスクは，グラフ，全ノードのそれぞれの特徴量，一部のノードが属するクラス (ラベル) が入力として与えられた際に，半教師あり学習を行うことで，ラベルが未判明である残りのノードのラベルを予測するタスクである．GCN では，グラフの接続されたノード同士は同一ラベルを持つ可能性が高いという仮定のもと，以下の伝播規則 [7] に基づいて特徴量を学習層ごとに更新し，最終的な予測ラベルを確定する．

$$H^{(l+1)} = \sigma(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}) \tag{3.1}$$

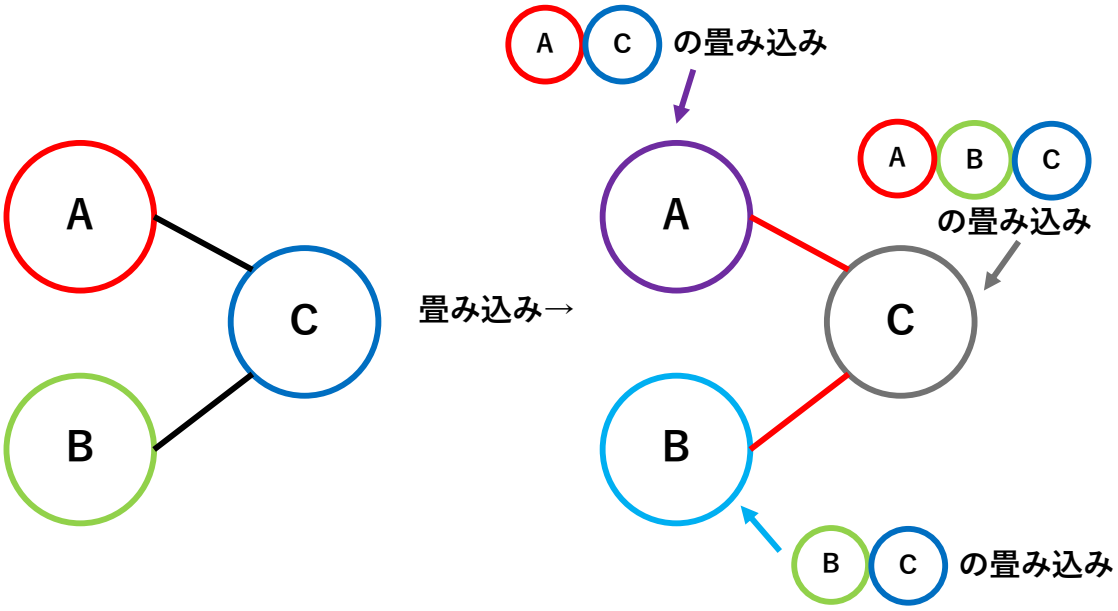


図 3.2: GCN の畳み込みのイメージ

ここで，$\widetilde{A} = A + I_N$ は自己接続が追加されたグラフ G の隣接行列，I_N は単位行列，$\widetilde{D}_{i,i} = \Sigma_j \widetilde{A}_{i,j}$ と $W^{(l)}$ は層固有の学習可能な重み行列であり，$\sigma(\cdot)$ は $\mathrm{ReLU}(\cdot) = \max(0, \cdot)$ などの活性化関数である．$H^{(l)} \in \mathbb{R}^{N \times D}$ は第 l 層の活性化行列であり，$H^{(0)} = X$ となる．ただし X はノードの特徴量行列である．

3.2　GCN で使用する条件

Fake コンテンツが注入される可能性のある脆弱なノードを推測するため，GCN のラベルは，各ノードの CS における Fake コンテンツの平均注入率（AIRF: average injection ratio of fake content）を元に作成する．AIRF を算出するには，各ノードの CS の全てのコンテンツに対して，コンテンツが正常コンテンツであるか Fake コンテンツであるかを確認し，キャッシュされている Fake コンテンツの数を調べる必要がある．そのため，少ない正解ラベルから脆弱なノードを効率よく推測できることが望ましい．また，全ノードの特徴量として，ネットワーク内で人気の高い上位 5 つのコンテンツについて，各ノードにおけるキャッシュヒット率を利用する．

第4章　性能評価

4.1　教師データ作成条件

計算機シミュレーションによって，GCN で使用する特徴量とラベルを作成する．シミュレータには，米国の商用 ISP のバックボーンネットワークトポロジである At Home Network と Allegiance Telecom を使用する．キャッシュ方式は，転送経路上の全てのルータでキャッシュを行う All Cache 方式とし，キャッシュ容量は 100 とする．

正当コンテンツ数は $M = 10,000$ とし，Fake コンテンツ数 F は 32 または 1,024 とする．正常ユーザは $\theta = 0.8$ の Zipf 分布に従い正当コンテンツを要求し，Bot の要求比率 ρ はネットワーク全体の要求比率 θ のうち 0.1，0.3，または 0.5 の割合で Fake コンテンツを要求する．シミュレーション時間は 10,000 秒とし，1,000 秒に 1 回，各ノードの Fake コンテンツの Fake コンテンツの注入率を測定し，測定した 10 回の平均値を AIRF とする．Fake コンテンツのオリジンサーバと Bot は，それぞれ同数を配置する．Fake コンテンツオリジンサーバと Bot の配置数 N_{attack} はそれぞれ 1，2，または 4 箇所とする．配置方式は，次の 3 種類とする．

- **Random 方式:**Fake コンテンツのオリジンサーバと Bot をランダムに配置する．
- **DCf 方式:** 次数中心性 (DC: Degree Centrality) に基づいて，Fake オリジンサーバを X_{fake} 箇所に先行配置する．その後，既に配置されたノードからより遠いノードに Bot を後続配置する [2].
- **DCb 方式:**DC に基づいて，Bot を X_{bot} 箇所に先行配置する．その後，既に配置されたノードからより遠いノードに Fake オリジンサーバを後続配置する．

AIRF を元に，10,000 秒の間に一度でも Fake コンテンツが CS に注入された形跡のあるノード (AIRF > 0) のクラスを Label1 とし，一度も Fake コンテンツが注入されていないノード (AIRF $= 0$) のクラスを Label0 と定義する．

4.2　GCN の学習条件

グラフには，教師データの作成に使用したものと同じ Allegiance Telecom と At Home Network のネットワークトポロジの接続情報を与える．全ノードの特徴量として，計算機シミュレーションによって算出したネットワーク内で人気の高い上位 5 つのコンテンツについての各ノードにおけるキャッシュヒット率を使用する．ラベルは 4.1 節で述べた 2 種を使用する．

また，GCN の作成には，Python のライブラリである PyTorch を使用している．PyTorch ではモデルの重みが実行のたびにランダムに初期化されるため，実行のたびに結果が変化し，予測精度に影響を及ぼす．そこで，seed 値を設定することで，モデルの初期化にばらつきが出ないように固定する．2 種類の seed 値 (42，1,024) を設定して，精度の良いものを研究結果として考察する．

4.3　GCN の性能評価

Fake コンテンツの注入は，Fake オリジンサーバと Bot の最短経路上のノードのみにしか行われないため，Label1 に分類されるノードよりも Label0 に分類されるノードの方が多いと推測される．そのため，本稿では，教師データ数 X_train が偶数の場合は各ラベルが均等な数になるようラベルの教師データを与え，奇数の場合は Label0 が多くなるようにラベルの教師データを与える．

偽陰性率 (FNR:false negative rate) と偽陽性率 (FPR:false positive rate)，精度についての評価を行う. 表 4.1 に示すように，TP(TN) を Label1(0) のノードを正しく判定した数とし，FP(FN) を Label1(0) のノードを誤って判定した数とする．

表 4.1: 陽性陰性の定義

	正解 Label1	正解 Label0
予測 Label1	TP	FP
予測 Label0	FN	TN

FNR は Label1(AIRF $>$ 0) のノードを誤って Label0(AIRF $=$ 0) と判定する割合で，FPR は Label0 のノードを誤って Label1 と判定する割合である．FPR，FNR，精度を以下のように定義する．

$$FPR = \frac{FP}{FP + TN} \tag{4.1}$$

$$FNR\ \frac{FN}{TP + FN} \tag{4.2}$$

$$\text{精度} = \frac{TP + TN}{TP + FP + TN + FN} \tag{4.3}$$

ネットワークトポロジと Fake コンテンツや Bot の配置方式の関係により，Fake コンテンツが広範囲のノードの CS に注入される場合と少数ノードの CS にのみに注入される場合がある．少数のノードの CS に Fake コンテンツが注入されている場合，精度が極端になることが考えられる．そのため，本稿では正解データの Label1 の割合がネットワークトポロジ全体のうち 15%以上となる配置が行われたケースについて比較し，考察を行う．

4.3.1　要求比率による差異

At Home Network において，Fake オリジンサーバと Bot の配置数 N_{attack} を 1 として DCb 方式で配置し，Fake コンテンツ数が $F = 32$ であるときに，Bot の要求比率 ρ を変化させたときの予測精度を比較する．

ρがそれぞれ 0.1，0.3，0.5 の全ての場合において，Label1 となるノードは 8 箇所である．教師データ数を 3 から 6 に変化させたときの，FNR，FPR，精度を表 4.2 に示す．$\rho = 0.5$ の教師データ数が 6 の場合で著しく精度が下がっているが，これは教師データが増えたことによる過学習により引き起こされている可能性がある．

表 4.2: $\rho = 0.1, 0.3, 0.5$ の GCN の予測精度

Bot の要求比率	$\rho = 0.1$			$\rho = 0.3$			$\rho = 0.5$		
教師データ数	FPR	FNR	精度	FPR	FNR	精度	FPR	FNR	精度
3	0.1667	1.0000	0.6977	0.3056	0.1429	0.7209	0.0556	1.0000	0.7907
4	0.3333	0.1667	0.6905	0.5556	0.0000	0.5238	0.2500	0.1667	0.7619
5	0.2000	0.6667	0.7317	0.0857	0.8333	0.8049	0.2286	0.6667	0.7073
6	0.1429	0.4000	0.8250	0.2857	0.0000	0.7500	0.5143	0.0000	0.55

シミュレータで作成した正解データと，それぞれの要求比率において最も精度が良かった場合の予測データを，Label ごとに色付けしてトポロジを出力した図を図 4.3，4.4，4.5 に示す．

4.3.2 Fake コンテンツ数による差異

At Home Network において，Fake オリジンサーバと Bot の配置数 N_{attack} を 1 として DCb 方式で配置し，Bot の要求比率が $\rho = 0.3$ であるときに，Fake コンテンツ数 F を変化させたときの予測精度を比較する．

F が 32，1,024 の場合において，Label1 となるノードはどちらも 8 箇所である．教師データ数を 3 から 6 に変化させたときの，FNR，FPR，精度を表 4.3 に示す．わずかではあるが，$F = 32$ よりも $F = 1024$ の場合の方が精度が良い．Fake コンテンツの総数が増えることで，CS に占める Fake コンテンツの割合が高くなる．そのため，GCN の学習の特徴量に使用している人気コンテンツのキャッシュヒット率に与える影響が大きくなり，よりノードの特徴を正確に表現することができるからであると考えられる．

表 4.3: $F = 32, 1024$ の GCN の予測精度

Fake コンテンツ数	$F = 32$			$F = 1024$		
教師データ数	FPR	FNR	精度	FPR	FNR	精度
3	0.3056	0.1429	0.7209	0.1111	1.0000	0.7442
4	0.5556	0.0000	0.5238	0.5278	0.0000	0.5476
5	0.0857	0.8333	0.8049	0.0000	1.0000	0.8537
6	0.2857	0.0000	0.7500	0.5714	0.2000	0.4750

シミュレータで作成した正解データと，それぞれの Fake コンテンツ数において最も精度が良かった場合の予測データを，Label ごとに色付けしてトポロジを出力した図を図 4.6，4.7 に示す．

4.3.3　攻撃者配置数による差異

At Home Network において，DCb 方式で Fake コンテンツ数が $F = 32$ で，Bot の要求比率 $\rho = 0.3$ であるときに，Fake コンテンツオリジンサーバと Bot の配置箇所 N_{attack} を 1,2,4 箇所に変化させたときの予測精度を比較する．

攻撃者の配置箇所が 1，2，4 のとき，Label1 となるノードはそれぞれ 8，15，17 箇所である．教師データ数を 3 から 6 に変化させたときの，FNR，FPR，精度を表 4.4 に示す．$N_{attack} = 4$ のとき $N_{attack} = 1, 2$ に比べ精度が低くなっている．図 4.1 に，各配置数での AIRF が大きい順に並べたものを示す．攻撃者が 1，2 のときに比べ，攻撃者が 4 の場合は AIRF が 0.08 から 0.12 の間で横ばいになっている数が多い．攻撃者を多数配置することで多くのノードに満遍なく Fake コンテンツがキャッシュされるため，Fake コンテンツをキャッシュさせることで各ノードごとに与える影響が小さくなる．その結果，キャッシュヒット率に与える影響が小さくなり，特徴量行列が不正確になることで学習が適切に行われなかった可能性が考えられる．

表 4.4: $N_{attack} = 1, 2, 4$ の GCN の予測精度

攻撃者配置数	$N_{attack} = 1$			$N_{attack} = 2$			$N_{attack} = 4$		
教師データ数	FPR	FNR	精度	FPR	FNR	精度	FPR	FNR	精度
3	0.3056	0.1429	0.7209	0.0690	0.7857	0.6977	0.2222	0.6250	0.6279
4	0.5556	0.0000	0.5238	0.2414	0.4615	0.6905	0.4074	0.4000	0.5952
5	0.0857	0.8333	0.8049	0.0357	0.8462	0.7073	0.3462	0.7333	0.5122
6	0.2857	0.0000	0.7500	0.3571	0.3333	0.6500	0.3462	0.5000	0.6000

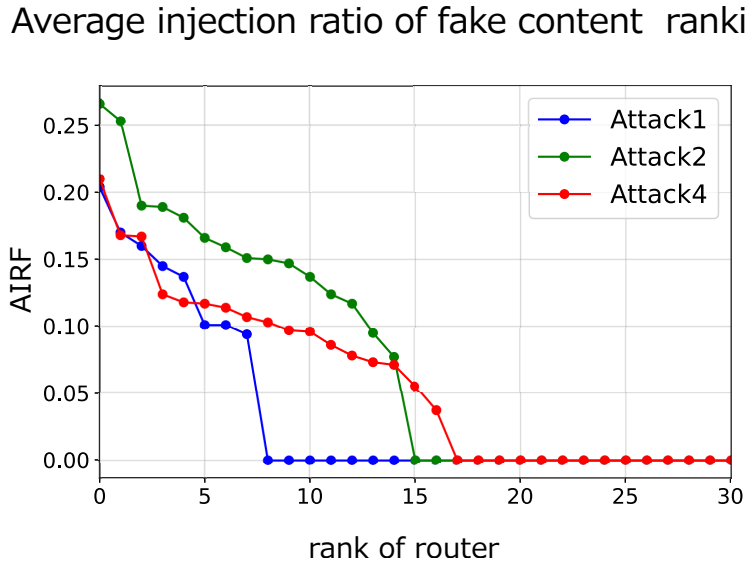


図 4.1: $N_{attack} = 1, 2, 4$ における AIRF ランキング

シミュレータで作成した正解データと，それぞれの N_{attack} において最も精度が良かった場合の予測データを，Label ごとに色付けしてトポロジを出力した図を図 4.8，4.9，4.10 に示す．

4.3.4 攻撃者の配置方式による差異

At Home Network において，Fake オリジンサーバと Bot の配置数 N_{attack} を 1 とし，Fake コンテンツ数が $F = 1024$，Bot の要求比率が $\rho = 0.1$ であるときに，Fake オリジンサーバと Bot の配置方式を変化させたときの予測精度を比較する．

配置方式がそれぞれ DCb 方式，DCf 方式，Random 方式のとき，Label1 となるノードはそれぞれ 8，8，10 箇所である．教師データ数を 3 から 6 に変化させたときの，FNR，FPR，精度を表 4.5 に示す．いずれの場合においても，教師データ数が 5 の場合が概ね良い精度である．4.3.1 項において，教師データ数が増えると過学習が起きる可能性を指摘したが，様々な配置方式においても教師データ数が 6 よりも 5 の場合の方が高い精度であることが確かめられる．本稿の GCN の学習条件において，高精度な予測を行うことができる適切な教師データ数は 5 であることが推測される．

表 4.5: DCb，DCf，Random 方式で配置したときの GCN の予測精度

配置方式	DCb			DCf			Random		
教師データ数	FPR	FNR	精度	FPR	FNR	精度	FPR	FNR	精度
3	0.3611	0.7143	0.5814	0.2222	0.4286	0.7442	0.3529	0.3333	0.6512
4	0.5833	0.1667	0.4762	0.4722	0.1667	0.5714	0.5000	0.1250	0.5714
5	0.2857	0.1667	0.7317	0.2571	0.3333	0.7317	0.3939	0.1250	0.6585
6	0.4286	0.2000	0.6000	0.3143	0.0000	0.7250	0.3333	0.4286	0.6500

シミュレータで作成した正解データと，教師データ数が 5 の場合の予測データを，Label ごとに色付けしてトポロジを出力した図を図 4.11，4.12，4.13 に示す．

4.3.5 ネットワークトポロジによる差異

At Home Network と，Allegiance Telecom の 2 種類のトポロジにおける予測精度を比較する．DCb 方式で Fake オリジンサーバと Bot の配置数 N_{attack} を 1 とし，Fake コンテンツ数が $F = 1024$，Bot の要求比率が $\rho = 0.3$ とする．

At Home Network と Allegiance Telecom において，Label1 となるノードはどちらも 8 箇所である．教師データ数を 3 から 6 に変化させたときの，FNR，FPR，精度を表 4.6 に示す．全体的に Allegiance Telecom の方が精度が高くなっている．これは，4.3.4 項においても指摘した，AIRF が原因だと考えられる．図 4.2 は，各トポロジの AIRF が大きいものに並べたものである．Allegiance Telecom の方が各ノードに注入された Fake コンテンツ数が大きくキャッシュヒット率に与える影響が増加することによって，より正確な特徴量行列が獲得できていると考えられる．

シミュレータで作成した正解データと，それぞれのトポロジにおいて最も精度が良かった場合の予測データを，Label ごとに色付けしてトポロジを出力した図を図 4.14，4.15 に示す．

表 4.6: At Home Network と Allegiance Telecom の GCN の予測精度

トポロジ	At Home Network			Allegiance Telecom		
教師データ数	FPR	FNR	精度	FPR	FNR	精度
3	0.1111	1.0000	0.7442	0.0233	0.4286	0.9200
4	0.5278	0.0000	0.5476	0.4884	0.0000	0.5714
5	0.0000	1.0000	0.8537	0.0238	0.5000	0.9167
6	0.5714	0.2000	0.4750	0.2619	0.2000	0.7447

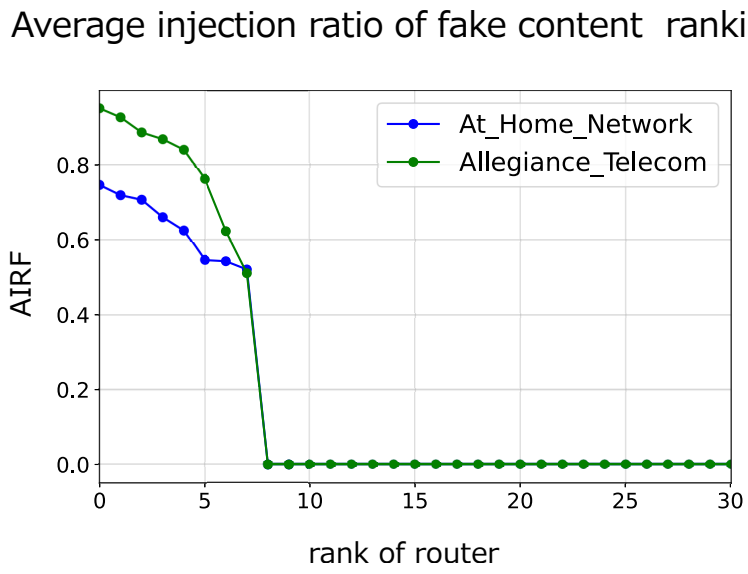
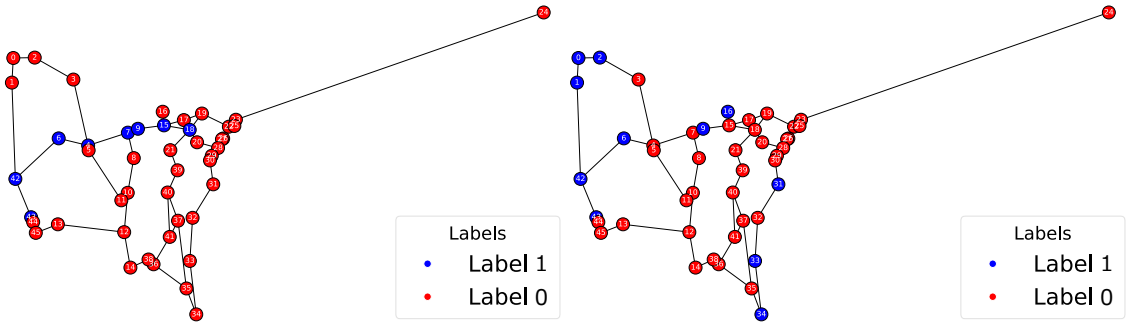


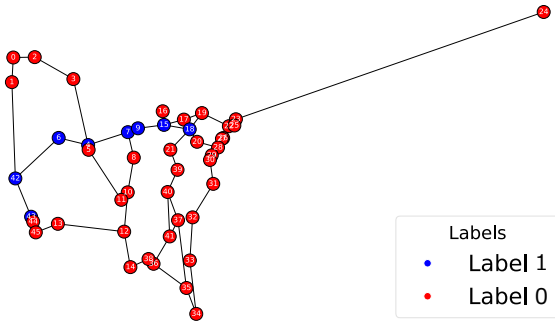
図 4.2: 各トポロジにおける AIRF ランキング



Answer：$\rho = 0.1$　　Predict：$\rho = 0.1$

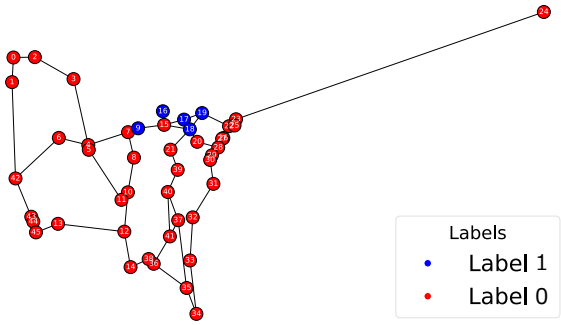
(a) 正解ラベル　　(b) 推測ラベル

図 4.3: $\rho = 0.1$ のときのラベル比較



Answer：$\rho = 0.3$

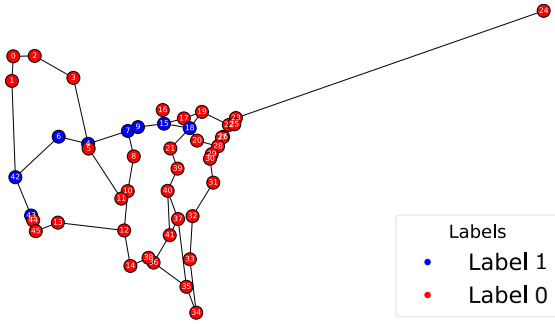
(a) 正解ラベル



Predict：$\rho = 0.3$

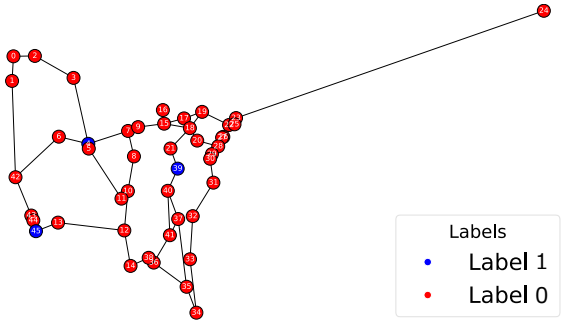
(b) 推測ラベル

図 4.4: $\rho = 0.3$ のときのラベル比較



Answer：$\rho = 0.5$

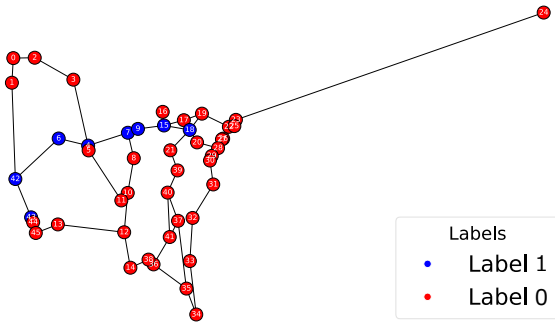
(a) 正解ラベル



Predict：$\rho = 0.5$

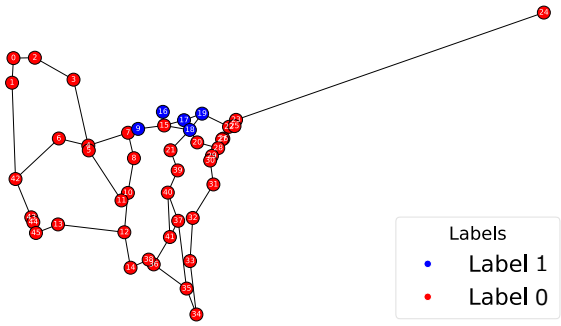
(b) 推測ラベル

図 4.5: $\rho = 0.5$ のときのラベル比較



Answer：$F = 32$

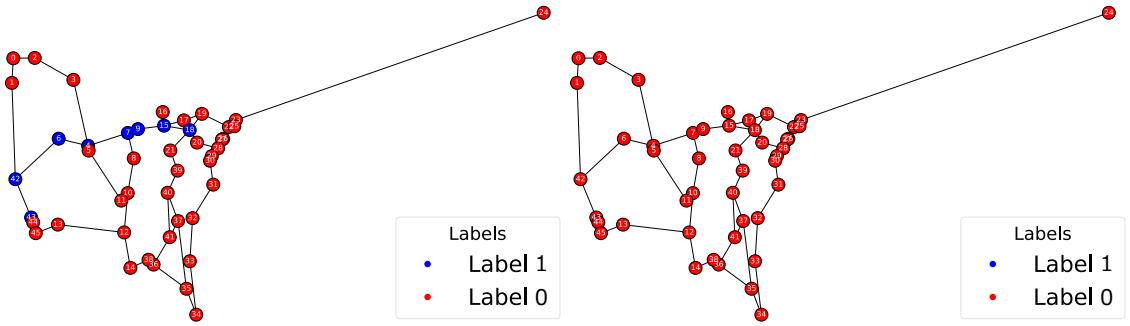
(a) 正解ラベル



Predict：$F = 32$

(b) 推測ラベル

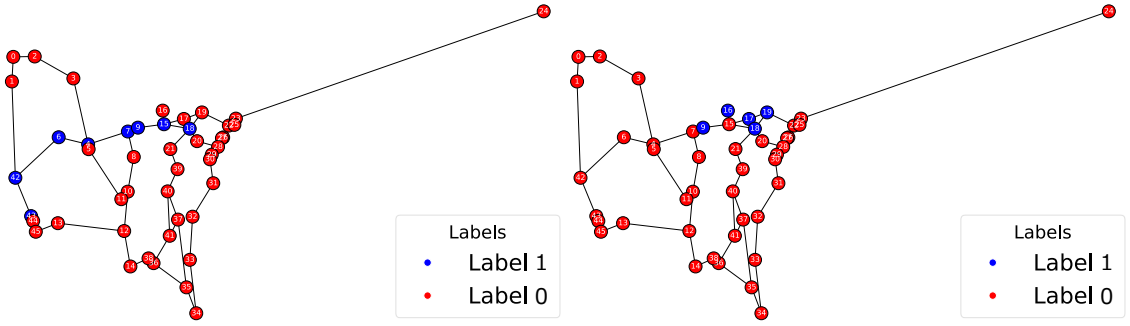
図 4.6: $F = 32$ のときのラベル比較



Answer：$F = 1024$　　Predict：$F = 1024$

(a) 正解ラベル　　(b) 推測ラベル

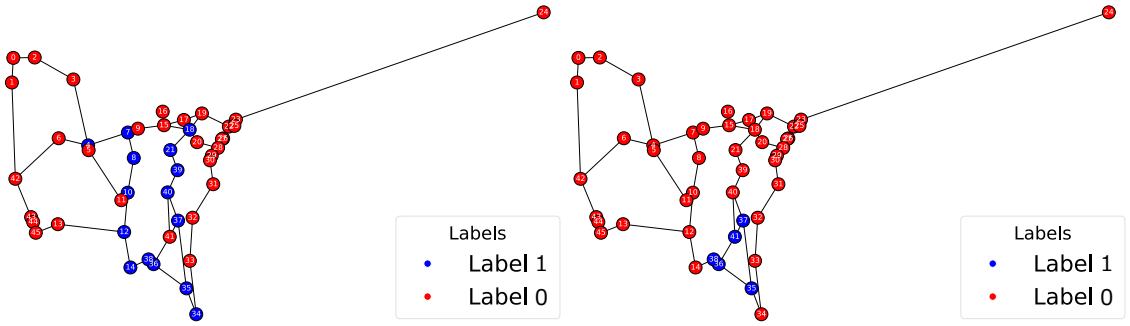
図 4.7: $F = 1024$ のときのラベル比較



Answer：$N_{attack} = 1$　　Predict：$N_{attack} = 1$

(a) 正解ラベル　　(b) 推測ラベル

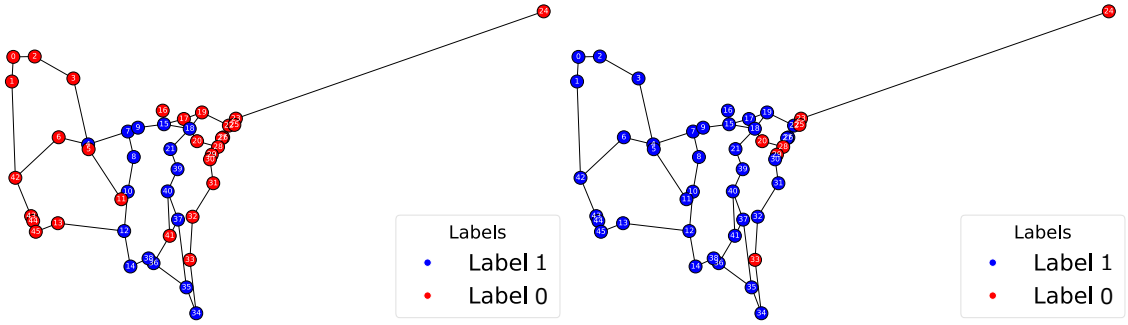
図 4.8: $N_{attack} = 1$ のときのラベル比較



Answer：$N_{attack} = 2$　　Predict：$N_{attack} = 2$

(a) 正解ラベル　　(b) 推測ラベル

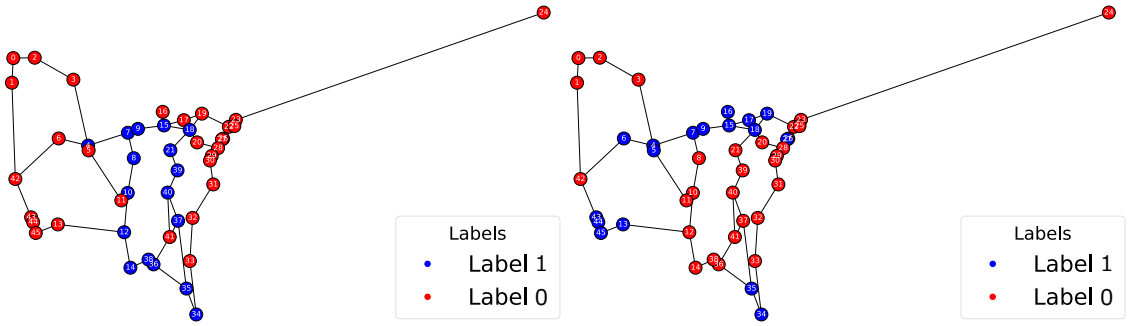
図 4.9: $N_{attack} = 2$ のときのラベル比較



Answer：$N_{attack} = 4$　　Predict：$N_{attack} = 4$

(a) 正解ラベル　　(b) 推測ラベル

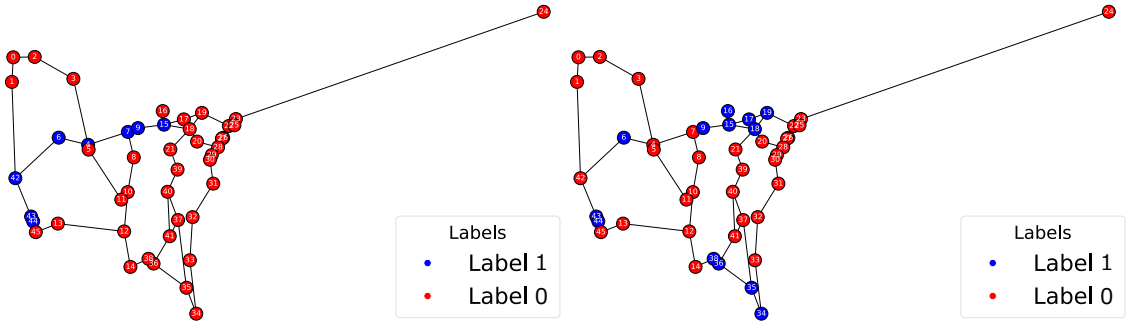
図 4.10: $N_{attack} = 4$ のときのラベル比較



Answer：DCb　　Predict：DCb

(a) 正解ラベル　　(b) 推測ラベル

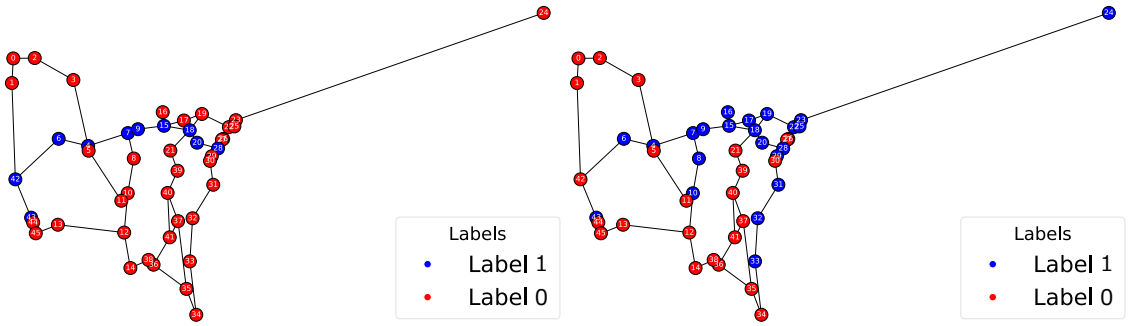
図 4.11: DCb 方式のときのラベル比較



Answer：DCf　　Predict：DCf

(a) 正解ラベル　　(b) 推測ラベル

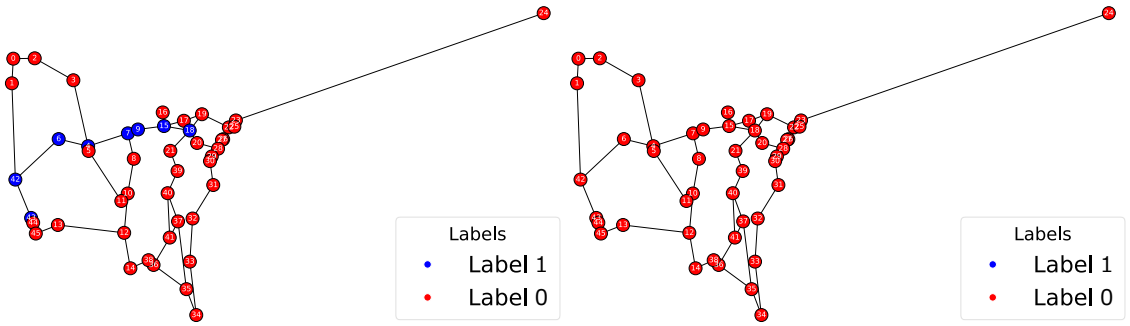
図 4.12: DCf 方式のときのラベル比較



Answer：Random　　Predict：Random

(a) 正解ラベル　　(b) 推測ラベル

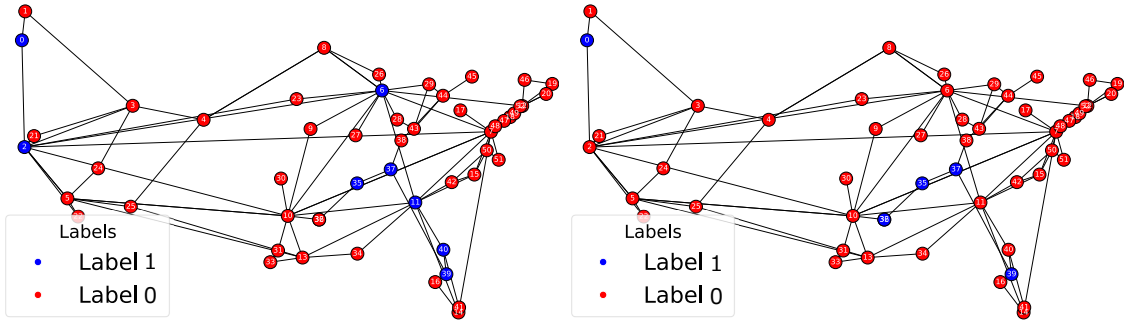
図 4.13: Random 方式のときのラベル比較



Answer：At Home Network　　Predict：At Home Network

(a) 正解ラベル　　(b) 推測ラベル

図 4.14: At Home Network のときのラベル比較



Answer：Allegiance Telecom　　Predict：Allegiance Telecom

(a) 正解ラベル　　(b) 推測ラベル

図 4.15: Allegiance Telecom のときのラベル比較

第5章　まとめ

本稿では，ICN の CPA に対する脆弱性を軽減するための新たなノード検出手法を提案した．従来の CPA の防御法ではデジタル署名の利用が提唱されているが，攻撃者が独自に作成した無意味なコンテンツをキャッシュさせる独自 Fake 型 CPA においては署名は正当であるため署名による防御が不可能であった．また，独自 Fake 型 CPA はユーザからの要求が発生しないため，検知も困難とされていた．この問題に対し，本稿では GCN を用いて Fake コンテンツが注入されやすい脆弱なノードを効率的に特定する手法を提案した．

提案手法では，ノードの AIRF をラベルとして利用し，人気コンテンツのキャッシュヒット率を特徴量として学習を行なった．この学習モデルにより，Fake 型 CPA に対する脆弱ノードを推測することが可能である．異なるトポロジを使用した評価や，攻撃者の配置数などさまざまなパタンでの評価を行うことによって，GCN が特に効率的に予測を行うことができるのは教師データ数が 5 の場合であると結論づけられた．

本稿では，教師データの与え方をランダムで与えていたが，GCN の学習の特性上教師データの与える位置が精度に与える影響は大きいと考えられる．今後は，より GCN の学習精度をあげる教師データの配置法や損失関数の設定を調査する予定である．

謝辞

本研究を行うに当たり，ご指導を頂いた上山教授に感謝します．また日常，有益な議論をして頂いた研究室の皆様に感謝します．

参考文献

[1] T. Nguyen, X. Marchal, G. Doyen, T. Cholez, and R. Cogranne, Content Poisoning in Named Data Networking: Comprehensive Characterization of real Deployment, IM2017, 2017.

[2] N. Kamiyama, and T. Kudo, Investigating Impact of Fake-Type Content Poisoning Attack on NDN, GRASEC 2023.

[3] W. Cui, et al.,“Feedback-Based Content Poisoning Mitigation in Named Data Networking”, IEEE ISCC 2018.

[4] T. Okada and N. Kamiyama, Name Management Using IOTA in ICN, IEEE DAG-DLT 2024, May 2024

[5] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner,Gradient-Based Learning Applied to Document Recognition, Proceedings of the IEEE, Vol. 86, Issue. 11, pp.2278-2324, Nov. 1998.

[6] L. Wu, P. Cui, J. Pei, and L. Zhao. Graph Neural Networks: Foundations, Frontiers, and Applications. Springer, Singapore, 725p.

[7] Kipf, T. N, and Welling, M, Semi-supervised classification with graph convolutional networks. In International Conference on Learning Representations, 2017.

学会発表リスト

- 上野優衣，上山憲昭，”GCN を用いた Fake 型 CPA に対する脆弱ノード検出法”，電子情報通信学会 2025 年総合大会，B-11，東京，2025 年 3 月
- 上野優衣，上山憲昭，”GCN を用いた独自 Fake 型 CPA に対する脆弱ノード検出法”，電子情報通信学会ネットワークシステム研究会, 沖縄, 2025 年 3 月