2025年度

修　士　論　文

GCNを用いたCFAの脆弱エリア抽出法

指導教員: 上山 憲昭

立命館大学大学院 情報理工学研究科
情報理工学専攻 博士課程前期課程
計算機科学コース

学生証番号: 6611230074-1

氏名: WANG Tianyu

概要

CFA（Crossfire Attack）は他の DDoS（分散型サービス拒否）攻撃とは異なり，ターゲットがサーバではなくネットワークリンクである点に特徴を有する．このため，検出が困難であり，ネットワークに対して重大な被害をもたらす可能性がある．しかし，従来の防御手法はサーバに対する DDoS 攻撃の検出に依存しているため，リンクのみを標的とする CFA には有効ではない．CFA に対するより効果的な防御として，攻撃発生前にサーバが属するネットワークトポロジを基にした防御を展開する手法が考えられるが，攻撃者が選択するターゲットエリアを事前に予測することは困難であり，防御計画の策定は容易ではない．

本論では，一部のネットワークに存在するトポロジが CFA 攻撃の影響を大きく左右する可能性を明らかにし，CFA 攻撃の影響を評価するための手法を提案する．この手法により，ネットワークトポロジ内の CFA に対して脆弱なエリアを特定することが可能である．さらに，GCN（グラフ畳み込みネットワーク）を用いた学習モデルを提案し，このモデルにより脆弱エリアの特徴を学習するとともに，トポロジ生成モデルである GraphTune を用いて CFA 攻撃に脆弱なネットワークトポロジ内の脆弱エリアを効率的に識別する．本論の評価結果により，提案モデルが各種類のトポロジ構造から脆弱エリアを正確に特定できることが示された．

目 次

第1章　序論

1.1　研究の背景

DDoS（分散型サービス拒否）攻撃とは，大量のデータパケットやリクエストをターゲットサーバに送信し，サーバを過負荷にすることでネットワークサービスを利用不能にする攻撃手法である．ホストに対する特定の DDoS 攻撃を軽減するため，サーバとネットワークの接続部分にファイアウォールを配置して攻撃パケットを制御する，あるいはネットワーク運用者が BGP ルータを使用して攻撃対象のプレフィックスを制限するといった対策が一般的に採用されている [1]．しかし近年，ターゲットエリアへのネットワークリンクを過負荷にすることで，そのエリア内のホスト間通信を不能にする CFA の問題が指摘されている [2]．従来の DDoS 攻撃とは異なり，CFA の特徴は攻撃対象がサーバではなく，その周辺のネットワークリンクである点にある．CFA は，ターゲットサーバエリア外のリンクを過負荷にすることにより，ターゲットエリア（TA）と外部ネットワークとの通信を妨害し，TA 内のホストが外部通信を行えないようにしてサービスを阻害する．CFA では，攻撃対象となるリンクを選定するため，攻撃の準備段階で大量のボットおよびターゲットエリア内の複数のサーバに対して traceroute を実行し，ターゲットエリアと外部ネットワークを接続するリンクを特定する．この探索フェーズの後，ターゲットエリア周辺の複数のデコイサーバが選定され，ボットから送信されるトラフィックがこれらのリンクを通過するように設定される．そして攻撃フェーズでは，大量のボットから送信されたトラフィックが複数のデコイサーバに向けられ，ターゲットリンクを過負荷にすることで攻撃が実行される [2].

1.2　GCN

GCN（Graph Convolutional Networks）は，グラフ構造データを処理するために設計されたニューラルネットワークであり，複数の層で構成され，それぞれの層でグラフ畳み込みを実行することで，ノード近傍の情報を集約・変換する [3]．その主な目的は，ノード（エンティティ）とエッジ（関係）からなるトポロジを活用し，グラフデータの空間的特徴を効率的に抽出することである．GCN は，グラフ内の構造的および関係的情報を効果的に学習できるため，ノード分類やリンク予測などのタスクに適用可能である．この特性を踏まえ，本論では GCN トレーニングモデルを採用しており，詳細は第 5.2 章で述べる．

1.3　研究の目的

本論の目的は，CFA（クロスファイア攻撃）に対して脆弱なエリアを高精度かつ効率的に特定する手法を開発することである．この目的を達成するため，以下の取り組みを行った．

- CFA 攻撃者がトポロジ内でリンクおよびエリアを選定する戦略に着目し，これを定量化する尺度を提案した．具体的には，攻撃者がリンクの混雑度やトポロジの構造的特性を基に選択を行うという仮定に基づき，

CFA に対しての脆弱性を評価する指標を設計した．この尺度は，特定のエリアが CFA 攻撃によってどの程度のリスクを受けるかを数値的に表現し，これによりトポロジ全体の脆弱性分析が可能となる．

- CFA 脆弱性を予測するための GCN モデルを設計し，十分な訓練データを用いてモデルを学習させた．本モデルでは，トポロジ構造内のノードやリンク間の関係を考慮しながら，各エリアの特徴を抽出することを可能にしている．また，提案した脆弱性尺度を用いることで，モデルがトポロジ内の脆弱なエリアをより正確に予測できるようにした．これにより，提案モデルは任意のトポロジにおいて，脆弱なエリアを効率的に特定する能力を備える．
- モデルの汎化性能を向上させるため，GraphTune というトポロジ生成モデルを活用して多様なネットワーク構造を生成し，訓練データとして利用した．このアプローチにより，多様なネットワーク構造に対する適応力を高めた．
- シミュレーションを通じて，提案モデルの性能を定量的に評価して提案モデルの有効性と実用性が示された．

本論は，CFA に対して脆弱なエリアを効率的に特定することで，これらの脆弱エリアに対する効果的な防御策を将来的に設計するための基盤を提供するものである．

第2章　関連研究

既存の CFA の検出および防御手法は，大きく分けて探索フェーズと攻撃フェーズのいずれかを対象とした技術に分類される．探索フェーズにおいては，攻撃者が脆弱なリンクやエリアを特定する能力を妨害することを目的とし，ネットワークトポロジの隠蔽，リンクメトリクスの動的な変更，または偽情報を注入して攻撃者を欺くといった手法が採用されている．一方，攻撃フェーズでは，トラフィック過負荷の影響を軽減することに重点が置かれ，リアルタイムのトラフィック監視，適応型ルーティング，およびトラフィックフィルタリングを通じて重要なリンクの輻輳を防ぐ取り組みが行われている．

さらに，近年の研究では，機械学習技術を活用して CFA の攻撃特性を識別し，防御する手法が注目されている．これらのアプローチでは，攻撃パターンを学習し，潜在的な攻撃対象を予測し，リアルタイムで脅威に適応的に対応するモデルの構築を目指している．ネットワークトラフィックやトポロジ変化の大規模なデータセットを分析することで，機械学習に基づく手法は，CFA の検出および緩和プロセスの精度と効率を向上させることを目的としており，より知的で先進的な防御策の実現に向けた有望な方向性を示している．

2.1　攻撃フェーズのテクニック

攻撃フェーズにおいて CFA の発生を検出するための多くの手法が提案されている．Narayanadoss らは，ターゲットリンクのトラフィック量を測定し，ANN，CNN，LSTM といった深層学習技術を活用して，CFA が発生しているリンクを検出する手法を提案している [4]．また，Xue らは，ルータ間で E2E またはホップツーホップのアクティブ測定を行うことで，CFA が発生しているリンクを検出する手法を提案した [5]．しかし，これらの手法は CFA を実行している攻撃フローを特定することができないため，CFA に対する完全な防御を提供することはできない．

攻撃フェーズにおいて，ソフトウェア定義ネットワーク（SDN）を活用して迂回制御などの手法を用いることで攻撃に対抗する技術も提案されている．例えば，Hyder らは，ONOS Rest API および DNS ポートリダイレクションを利用した意図ベースのトラフィック変更によって CFA 防御を強化するフレームワークを提案している [6] また，Rafique らは，リンク選択，攻撃検出，悪意のあるフローのブロックモジュールを組み込んだ新しい CFA 対策「CFADefense」の設計と実装を提案している [7].

さらに，Aydeger らは，CFA 攻撃に対抗するための SDN ベースの MTD（Moving Target Defense）メカニズムを提案している．評価結果では，経路の変動がターゲットリンクの負荷を効果的に軽減できる一方で，ネットワークサービスに大きな影響を与えないことが示されている [8]．しかし，これらの手法はいずれも攻撃発生後に検出および防御を行うため，CFA を完全に防ぐことはできない．

Y.Cao らは，SDN を利用した Spatio-Temporal Graph Convolutional Networks（ST-GCN）に基づく検出手法を提案している．この手法では，ネットワークをグラフにマッピングし，ネットワーク状態を ST-GCN モデルに入力することで，DDoS 攻撃フローが通過するスイッチやネットワークを横断する経路を特定する [9]．一方，Saunders らは，DDoS 攻撃検出問題をトポロジ情報と統計情報を組み合わせた GCN のエッジ分類問題として抽

象化し，いくつかの DDoS 攻撃を検出・分類するための GCN-IDS モデルを提案している [10]．しかし，CFA を防御する際には，CFA の特有の特徴を考慮する必要がある．

2.2 トポロジ視点

一部の研究では，DDoS 攻撃に対する防御がトポロジの観点から実現可能であると主張し，深層学習を用いた DDoS 攻撃の検出および防御手法を提案している．W. Guo らは，トポロジおよびトラフィック特徴に基づく深層学習手法（GLD-Net）を提案した．この手法では，トポロジ特徴とトラフィック特徴を初めて融合し，グラフニューラルネットワークを用いて高性能な DDoS 攻撃侵入検出を実現している [11]．

また，Liaskos らは，トポロジと検出効率の関係に関する形式的な証明を提供し，この 2 つの関係を定量化する新しいオフライン測定手法を提案している [12]．結果として，新たな評価基準がトポロジにおける検出関係を効果的に表現できることを示し，既存の広く使用されている評価基準ではこの目標を十分に達成できないことを明らかにしている．

L. Guo らは，SDN に基づく検出および防御手法を提案している．ネットワークのボトルネックとなるノードおよびリンクを選定し，仮想トポロジ構造を導入することで，正規のトラフィックの正常な転送に影響を与えることなく，クロスファイア攻撃の検出および緩和を効果的に実現している [13]．しかし，この手法は SDN のグローバルビューおよびリアルタイム計算に依存しているため，複雑または大規模なネットワークでは計算コストが高くなる傾向があり，さらに異なる種類のネットワークトポロジに一般化することが難しいという課題がある．

本論の先行研究では，攻撃者が CFA のターゲットリンクを選定するために多数の traceroute を実行するという事実に着目し，traceroute の発生間隔に基づいて攻撃ホストを検出する手法を提案した [14]．しかし，CFA に関する既存の研究は，準備フェーズおよび攻撃フェーズにおける CFA の検出を主な対象としており，CFA に対して脆弱なエリアを事前に予測し，重点的な設備投資などを通じて事前準備を行う研究は存在しない．

第3章　CFA(クロスファイア攻撃)

3.1　攻撃手法

CFA（クロスファイア攻撃）は，探索フェーズと攻撃フェーズという 2 つの異なるフェーズで実行される．探索フェーズでは，図 1 に示すように，攻撃者は多数のボットホストからターゲットエリア（TA）内の公開サーバや，TA 周辺に設置されたデコイサーバに向けて traceroute パケットを送信する．その後，取得したルーティング情報に基づいて攻撃リンクを選定する．具体的には，traceroute ツールを使用してノードへのルーティング情報を取得し，指定したノードへの経路（つまり，通過するルータのリスト）を取得する．攻撃フェーズでは，攻撃者は多数のボットからデコイサーバに向けて少量のトラフィックを生成し，これにより TA と外部世界の通信を遮断する．

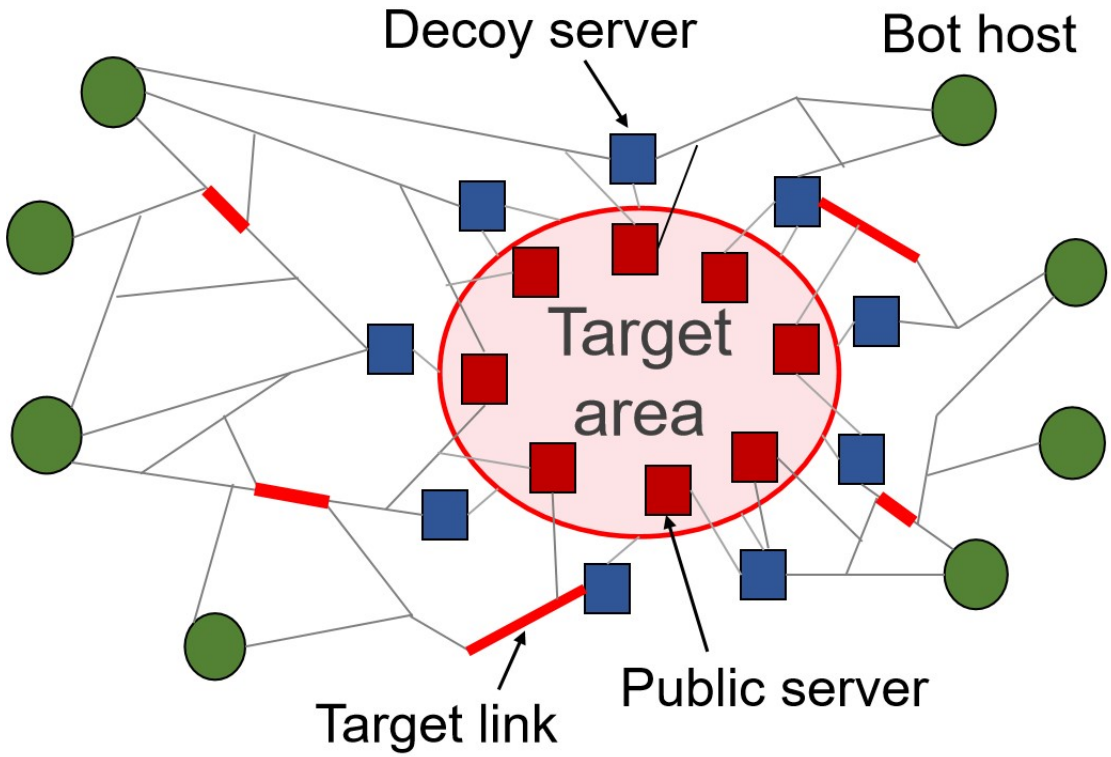


図 1: Crossfire Attack

CFA の攻撃者の流れは以下の通りである．

1. 多数のボットからターゲットエリア（TA）内の多数の公開サーバ（ウェブサーバなど）に対して traceroute を実行し，TA 内のホストに向かう多くのフローが通過する少数のリンク（ターゲットリンク）を特定する．

2. 多数のボットから TA 周辺の多数の公開サーバ（デコイサーバ）に対して traceroute を実行し，攻撃対象のリンクを通過するフローを生成するボットおよびデコイサーバの組み合わせを選定する．

3. 選定したボットおよびデコイサーバの組み合わせに対して，通常の HTTP リクエスト/レスポンスなどの検出困難な少量のトラフィックを生成する．

これにより，CFA の攻撃者はリンクの過負荷を引き起こし，ターゲットエリアと外部ネットワークとの通信を遮断することが可能となる．

3.2 検知の困難さ

CFA（クロスファイア攻撃）は，従来の攻撃とは異なり，その特異な特性により検出および防御が非常に困難である．特に，攻撃の予測や防御が難しい理由として，以下の点が挙げられる．

- CFA のターゲットエリア（TA）は攻撃が開始されるまで特定することができない．このため，既存の防御策では，攻撃を迅速に検出して効果的な保護を提供することが困難であり，攻撃を受ける可能性のあるエリアを事前に防御することが難しい．
- CFA では TA そのものが直接的に攻撃されるわけではなく，TA 近傍に配置されたデコイサーバも異常なトラフィックを検出しない．その結果，TA 内のサーバは攻撃の発生に気付くことができず，通常の異常検知システムでは CFA 攻撃を見逃してしまう可能性が高い．
- CFA では，ターゲットリンクを多数の低速な攻撃フローが通過する．このような低速フローは正規の通信と見分けがつきにくいため，ターゲットリンクに接続されたルータが攻撃フローを識別し，遮断することが非常に困難である．

3.3 CFA の特徴

CFA において，攻撃前の探索フェーズでは，攻撃者が使用するボットとデコイサーバのペアを選定する．この過程で，ボットとターゲットエリア（TA）内のサーバやデコイサーバ間で大量の traceroute が実行される．この選定は，ネットワークリンクが更新される前に完了する必要があるため，短期間のうちに連続的に実行されることが多い．さらに，多くの場合，攻撃者はボット市場（PPI: Pay-Per-Install）を利用しており，コストが使用するボットの数に比例するため，1 つのホストが多数のターゲットサーバやデコイサーバに対して traceroute を実行することが一般的である．

第4章　CFAの影響尺度

CFA の攻撃者は，ターゲットエリア（TA）周辺のデコイサーバ間で多数のボットから少量のトラフィックを転送することによって，TA 周辺に存在するリンクを過負荷にし，TA と他のエリア間のトラフィック分配を妨害する．このため，攻撃者は，TA と外部世界間のトラフィックの大部分が通過するリンクを攻撃対象として選定する．ネットワーク内で複数の隣接ノードが CFA のターゲットエリア x である場合，x を通過し他のエリアと接続する限られたリンクが削除されるとき，x と x 外部間のトラフィックで通信不可能となる割合が大きいほど，エリア x は CFA に対してより脆弱であると考えられる．

したがって，本論では CFA に対する脆弱性を測定する指標として，以下の変数を定義する．

1. $A_n(x)$: n 個の隣接ノードで構成されるエリア x
2. $E_n(x, y)$: 任意の $A_n(x)$ に対して，$A_n(x)$ と他エリアを跨る任意のリンク y
3. $R_n(x, y)$: $A_n(x)$ 以外の任意のノードと，$A_n(x)$ の任意のノードとの間の最短ホップ経路のうち，リンク $E_n(x, y)$ を通るものの割合
4. Max $R_n(x)$: $R_n(x, y)$ の最大値

ただし CFA では $A_n(x)$ に隣接しないリンクを攻撃対象とすることも考えられるが，本論では便宜上，$A_n(x)$ と外部のエリアとの境界に存在するリンクを攻撃対象の候補として考える．

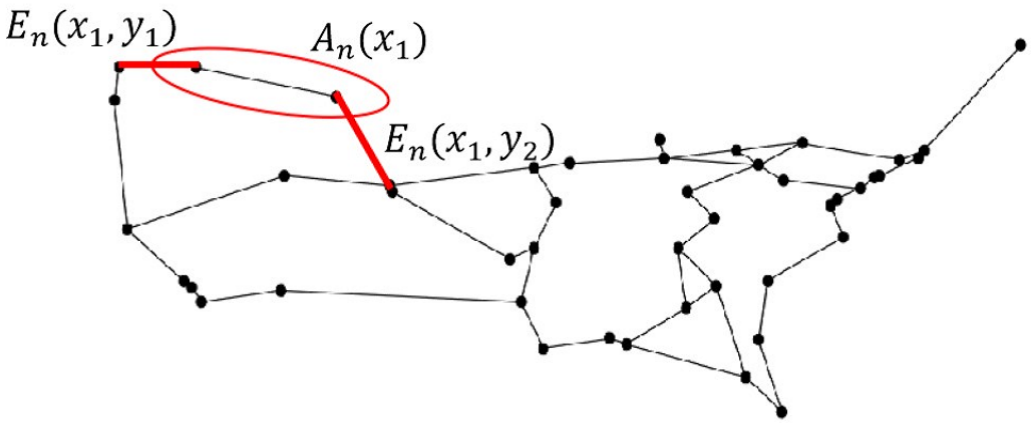
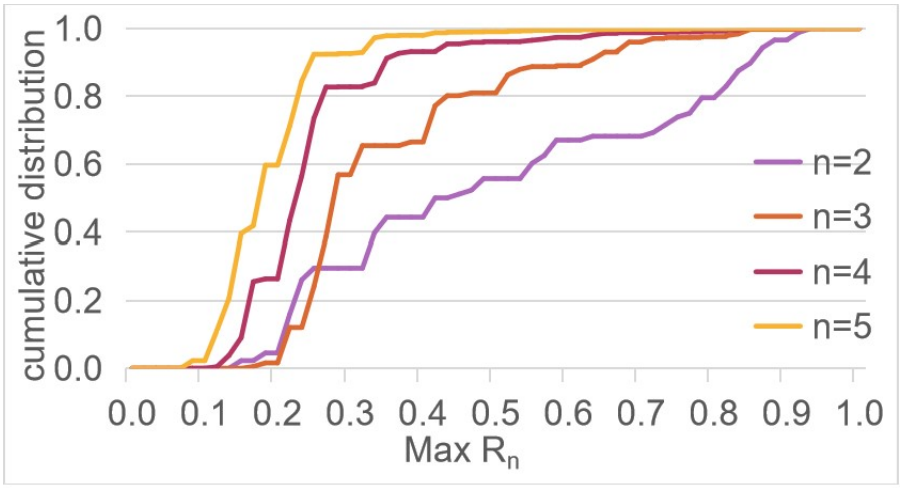


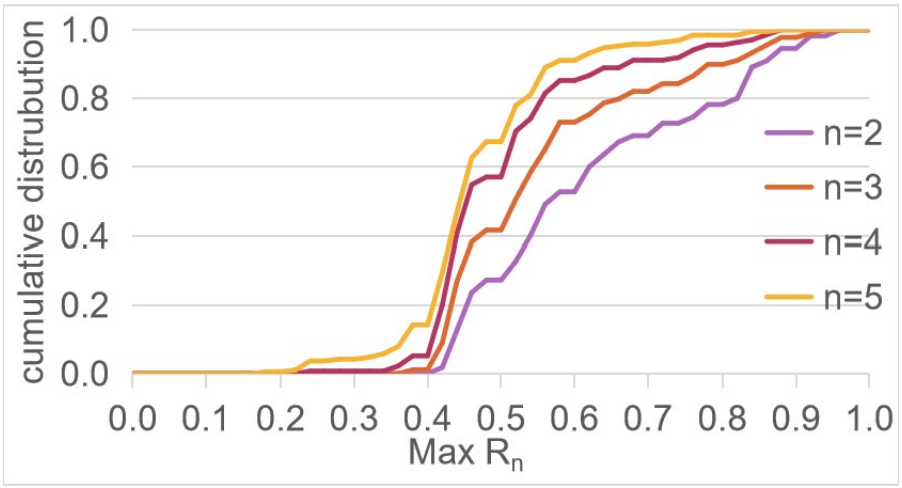
図 2: Example of $A_n(x)$ in At Home Network

図 2 に示すのは，$A_n(x)$ とその 2 つの $E_n(x, y)$ の例である．

Allegiance Telecom および At Home Network の 2 つのトポロジにおいて，それぞれ 2，3，4，5 ノードで構成されるすべてのエリアに対する $Max\ R_n(x)$ の値をカウントし，累積分布を図 3 にプロットした．



(a) Allegiance Telecom



(b) At Home Network

図 3: Cumulative distribution of $Max\ R_n(x)$ in two topologies

注目すべき点として，両トポロジにおいて $Max\ R_n(x)$ が 0.8 を超えるエリアが存在することが挙げられる．これは，CFA 攻撃者がこれらのエリアを標的とする場合，たった 1 本のリンクを切断するだけで，そのエリアと外部ネットワーク間のトラフィックの 80%以上を遮断し，通信を効果的に妨害できることを示している．さらに，$Max\ R_n(x)$ が 0.8 を超えるエリアは，図 4 の赤い円内に示されているように，特有の構造を持つ傾向がある．本論では，この構造を「直列構造」と命名するが，限られた接続パターンを特徴としており，各ノードが両隣のノードとしか接続されていない．この直列構造を持つターゲットエリアでは，エリアに接続するリンクの数は通常 2 本または 3 本に限定され，トラフィックは 1 つの主要なリンクを通じて流れる．n の増加に伴い，$Max\ R_n(x)$ の分布はより値の小さいものに集中する．すなわちターゲットエリアに至るトラフィックが各境界リンクに分散される度合いが大きくなる．そのため CFA 攻撃者はより低いコストで攻撃を達成するためには，より少ないノードを含むエリアを攻撃する必要がある．

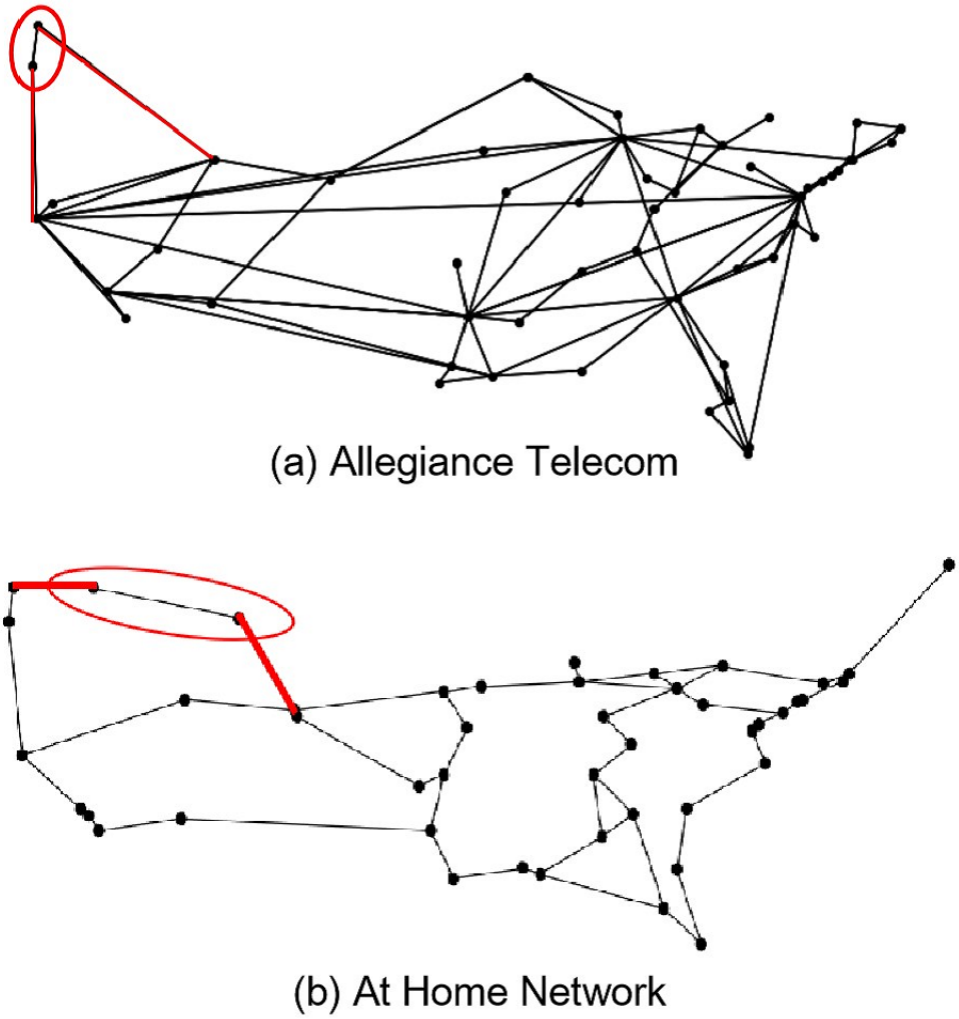


図 4: Series structure in topology

一方で，n が増加するにつれて，$Max\ R_n(x)$ は減少する．これは，ターゲットエリアに含まれるノード数が増加し，直列構造が減少するためである．このことから，直列構造はネットワークトポロジにおける相対的に脆弱な部分であり，CFA の対象になりやすいことがわかる．防御策を開発する際には，直列構造に着目することが重要である．各 $A_n(x)$ には対応する $Max\ R_n(x)$ が存在し，CFA 攻撃者は CFA を実行する際に，このリンクを攻撃対象として選択することで，最も効率的かつ効果的に攻撃を行うことが可能である．ただし，実際の攻撃では，攻撃者は単一のリンクではなく，複数のリンクを選択する可能性が高い．したがって，これまでに定義した変数に以下の定義を追加する．

5. $Max_2\ R_n(x)$: $R_n(x, y)$ の最大値と 2 番目に大きな値との最大値

以下のセクションでは，本論で提案するモデルが $Max_2\ R_n(x)$ をどのように活用し，ネットワークトポロジ内で CFA に対して脆弱なエリアを示す特徴（直列構造を含む）を特定するかについて詳細に説明する．また，これらの特徴に基づいて GCN を用いた深層学習を実行し，与えられたネットワークトポロジ内のすべての脆弱なエリ

アを正確に検出する方法についても示す.

第5章　提案方式

5.1　初期手法と限界

本論はまず，特定のトポロジ構造がエリアの脆弱性に影響を与えるという結論に基づいた予測アルゴリズムを提案した．このアルゴリズムの理論的基盤は，特定のネットワークトポロジ構造が存在することで，特定のエリアの脆弱性が顕著に増加する可能性があるという点である．具体的には，エリアの脆弱性に影響を及ぼすいくつかの既知のトポロジ構造に重みを付与し，それらの重みと事前に設定された閾値に基づいて，特定のネットワークトポロジ内の全てのエリアの脆弱性を予測することを目的としている．この方法の目標は，$Max_2\ R_n(x)$ の値を直接計算することなく，指定されたネットワーク内の脆弱なエリアを迅速に特定することである．

しかし，大量のシミュレーション実験を経た結果，この方法にはいくつかの重要な限界があることが明らかになった．第一に，このアルゴリズムは予測精度を安定的に保証することが困難であるという点である．第二に，ネットワークトポロジ構造が変化した際に予測精度が大幅に変動し，この方法が異なるトポロジ構造に適応しにくいことが示された．さらに，この現象は，特定のエリアが CFA 攻撃を受けやすい原因が，我々が提案した特定のトポロジ構造に由来するだけでなく，現行の方法では正確に捉えられない潜在的な特徴にも関係している可能性を示唆している．予測精度を向上させるために，重みの値を調整し続けることで，多様なネットワークトポロジ構造に適応できる計算式を構築しようと試みたが，この手動によるパラメータ調整は非常に時間がかかるだけでなく，多様なトポロジ構造に対応することは極めて困難であることが判明した．これらの問題は，初期手法が有する限界を明確に示すものである．

以上の課題を踏まえ，本論は研究の重点をグラフニューラルネットワーク（GCN）に基づく深層学習手法へと移した．この新たなアプローチを通じて，複雑なネットワークに隠された潜在的な特徴を抽出し，アルゴリズムの予測精度を向上させることで，初期手法の固有の欠点を克服することを目指している．

5.2　GCN モデル

グラフ畳み込みネットワーク（Graph Convolutional Network, GCN）は，2016 年に Kipf らによって提案されたもので，深層ニューラルネットワークの一種である [3]．GCN はグラフの非ユークリッド的な特性を活用し，データのノード情報を畳み込み的に集約する手法を提供する．これにより，従来の深層学習モデル（CNN や RNN など）がグラフ構造データに対して十分な性能を発揮できないという問題を解決する優れた手段となっている．CFA 攻撃に関する解析をトポロジレベルで行う際，ネットワークトポロジがグラフ構造データの形式で表されることから，GCN はこの問題を効果的に処理できることが期待される．

GCN は，ランダムに初期化されたパラメータ W を使用して未訓練の状態でもグラフ構造から効率的に特徴を抽出できるという強力な特性を持つ．この特性により，GCN はグラフデータの処理において高い可能性を示している．この特性に基づき，本論では，GCN モデルに対して最小限の調整を加え，CFA 攻撃に対するエリアの脆弱性を示す指標である $Max_2\ R_n(x)$ に関連する特徴の学習に重点を置けるようにした．この調整により，GCN の基本的な利点を保持しつつ，CFA 攻撃シナリオの認識におけるモデルの適応性と精度を向上させることに成功した．

本論で提案する GCN モデルは，ネットワークトポロジの構造特性を活用して，CFA 攻撃に脆弱なエリアを識別することを目的としている．モデル設計において，ネットワークトポロジは入力データとして扱われ，ノードとエッジに関する詳細な情報を含む．その中で，ノード属性は入力データの中心的な要素であり，ノードの次数中心性，媒介中心性，近接中心性，クラスタリング係数といった重要な指標が含まれる．これらの属性は，各ノードのネットワーク内での重要性を示すだけでなく，トラフィック分布や潜在的なボトルネックへの影響を明らかにし，モデルに豊富な特徴基盤を提供する．

モデルは，これらの属性をグラフの構造に埋め込み，ネットワークトポロジのグローバルな特性と結合することで，ネットワーク全体の構造や局所的な接続パターンを的確に捉える能力を発揮する．GCN モデルは独自の多層情報集約メカニズムを活用し，各ノードおよびその近傍の情報を再帰的に統合する．このメカニズムにより，ノード属性とトポロジ構造との間に存在する潜在的な関連性を効果的に掘り下げ，従来の分析手法では捉えにくい複雑な関係性を抽出することが可能となる．

さらに，モデルはネットワークトポロジ内の各エリアに「脆弱」か否かを示すラベルを付与することで，CFA 攻撃シナリオにおけるエリアの脆弱性状態を明確にする．ノード属性とトポロジ特性を組み合わせることで，モデルは高次元の特徴表現を生成し，脆弱エリアの識別および特徴付けにおいて高い精度と適応性を実現している．この特性により，モデルは複雑なネットワーク環境への理解能力を向上させるだけでなく，攻撃予測や防御最適化に向けた強力な基盤を提供する．アルゴリズムの詳細は 5.2.2 節に記載している．

5.2.1 生成モデル

提案するモデルが幅広いネットワークトポロジにおいて優れた性能を発揮することを目指している．しかし，既存のネットワークトポロジのみを用いて訓練するだけでは，モデルの広範な適用性を十分に確保することはできない．そのため，2023 年に Watabe らが提案したネットワークトポロジ生成モデル [15] を採用し，多様なトポロジを生成することを決定した．この生成モデル「GraphTune」は，グローバルレベルの構造的特徴の値を条件として調整することが可能であり，異なるネットワーク環境においてモデルを効果的に訓練することを保証する．これにより，モデルの適応性と実用性をさらに向上させることができる．

表 1: Phase 1 generation settings

Parameters	Given Values
Power exponent	2.0, 2.5, 3.0, 3.5, 4.0
Average clustering coefficient	0.10, 0.20
Average path length	3.0, 5.0
Average degree	2.5, 3.0, 3.5
Edge density	0.05, 0.07, 0.09
Diameter	5, 10
Maximum component size	30

本論では，モデルの予測精度および異なるネットワークトポロジへの適用性を段階的に向上させるため，2 段階のモデル訓練スキームを提案した．第 1 段階では，ネットワークトポロジ生成手法の主要なパラメータを調整することで，360 個の異なるネットワークトポロジ構造を生成し，初期モデルの訓練サンプルとした．表 1 のように第 1 段階でのパラメータの設定を示す．この段階の目的は，モデルの基本的な予測能力を検証し，特定のネット

ワークトポロジ特性が予測結果に与える影響を初期的に探索することである．しかし，実験の結果，第 1 段階で訓練されたモデルは，一部のネットワークトポロジ（例えば At Home Network）において予測精度が低いことが判明した．これは，これらのネットワークトポロジの一部のパラメータ（平均次数やエッジ密度など）が表 1 で示される範囲を超えており，その結果，モデルがこれらの構造に対して十分に適応できなかったためである．

表 2: Parameters of 5 topology

Parameters	Allegiance Telecom	At Home Network	CAIS Internet	Verio	Att
Power exponent	2.327	5.810	5.100	2.965	1.731
Average clustering coefficient	0.294	0.051	0.041	0.243	0.102
Average path length	3.321	5.060	5.048	2.508	2.955
Average degree	3.321	2.391	2.378	4.114	3.312
Edge density	0.064	0.053	0.066	0.121	0.360
Diameter	9	13	11	4	5
Maximum component size	53	46	37	35	91

また，表 2 には性能評価の対象となった 5 つの特定のネットワークトポロジのパラメータが示されている．表 2 を確認すると，これらのトポロジの一部のパラメータが一定程度，表 1 で設定された範囲を超えていることがわかる．そのため，これらのトポロジにおいてモデルが精度の高い予測を行えないことは，予想される結果であるといえる．

表 3: Phase 2 generation settings

Parameters	Given Values
Power exponent	2.0, 3.0, 4.0, 5.0
Average clustering coefficient	0.05, 0.10, 0.15, 0.20
Average path length	3.0, 4.0, 5.0
Average degree	2.5, 3.5, 4.5
Edge density	0.05, 0.07, 0.09
Diameter	5, 7, 9, 11
Maximum component size	30, 40

しかし，モデルの精度をさらに向上させ，さまざまな種類のネットワークトポロジにおける適用性を強化するため，第 2 段階では改良型モデルを提案した．表 3 には性能評価の対象となった 5 つの特定のネットワークトポロジのパラメータが示されている．この段階では，一部のパラメータに対してより広範な値の範囲を設定し，それらのパラメータのすべての可能な組み合わせを利用し，2500 個の異なるネットワークトポロジ構造を生成し，改良

モデルの訓練サンプルとした．この手法は，さまざまな構造の変化を効果的にカバーするだけでなく，少数の基本的なトポロジタイプを基に系統的に訓練データを生成するものであり，ランダムにトポロジ構造を選択するわけではない．このようなターゲットを絞ったトポロジ生成手法により，モデルの訓練効率が向上するだけでなく，実際の応用における適応性と汎化能力が大幅に強化された．

5.2.2 モデルの訓練

本論で提案するモデルの訓練には，第 4 章で提案した CFA 攻撃に対するエリアの脆弱性指標である $Max_2\ R_n(x)$ を使用する必要がある．訓練データの作成にあたり，ネットワークトポロジの各エリアに対して脆弱性指標を計算し，設定した閾値を基に「脆弱（Vulnerable）」または「非脆弱（Non-vulnerable）」のラベルを付与する．このラベルは，GCN モデルが学習すべき対象を明確にするものであり，各エリアの脆弱性状態を二値化した形で表現している．ラベル付けの基準として $Max_2\ R_n(x)$ の値を用いることで，特定のリンクやノードがネットワーク全体に与える影響を適切に反映できるようにした．生成されたネットワークトポロジが所望のデータセット形式に適合するように処理するために，与えられたネットワークトポロジ内のすべてのエリアに対して $Max_2\ R_n(x)$ 値を効率的に計算するアルゴリズム 1 を設計した．このアルゴリズムにより，CFA 攻撃に対する脆弱性を正確に反映するデータセットを構築し，信頼性の高いモデル訓練が可能となる．

Algorithm 1 Generate Dataset and Label Vulnerabilities

```
Input: G = (V, E), Threshold
Find all regions consisting of five nodes in G:
for Each region r in G do
    Compute vulnerability score Max_2 R_n(r)
    if Max_2 R_n(r) ≥ Threshold then
        Label region r as "Vulnerable"
    end if
end for
Store regions and labels in Dataset
Return Dataset
```

生成されたデータセットは，GCN の入力要件に適合させるため，追加の処理が必要である．まず，トポロジに自己ループ（Self-Loop）を追加し，隣接行列を正規化することで，モデル訓練時の安定性を確保した．次に，ノードの初期特徴と正規化された隣接行列を組み合わせることで，訓練用の入力特徴を生成した．アルゴリズム 2 の処理を通じて，データセットが GCN に適合する形式となり，効率的かつ安定したモデル訓練が可能となった．

Algorithm 2 Data Loading and Preprocessing

```
Input: G = (V, E), X(Node feature matrix)
A_self ← A + I
A_norm ← D^{-1/2} A_self D^{-1/2}
X_combined ← A_norm · X
Return A_norm, X_combined
```

GCN モデルは，ノードの隣接情報を逐層的に集約することで，グラフの局所的および全体的な特徴を効果的に抽出する仕組みを持つ．各層の出力は，現在のノードの特徴だけでなく，その隣接ノードの特徴にも影響を受け

るため，グラフ全体の構造を反映した学習が可能である．この特性により，GCN は複雑なネットワーク構造に対する予測や解析において優れた性能を発揮する．アルゴリズム 3 に示すのは，提案モデルの構築におけるアルゴリズムであり，具体的な実装プロセスを概観するものである．

Algorithm 3 Build GCN Layers

Input: $A_{\text{norm}}, X_{\text{combined}}, L$
$Z[0] \leftarrow X_{\text{combined}}$
for $l = 1$ **to** L **do**
    $Z[l] \leftarrow \text{ReLU}(A_{\text{norm}} \cdot Z[l-1] \cdot W[l])$
end for
Output: $Z[L]$

深層学習モデルの訓練において効果的であることから，損失関数として CrossEntropyLoss を，最適化アルゴリズムとして Adam を選択した．CrossEntropyLoss は，予測確率と実際のクラスラベルとの差異を測定するため，特に分類タスクに適している．この損失関数は，クラス予測誤差の最小化に焦点を当てることを可能にし，トポロジ内のエリアが CFA 攻撃に対してどの程度脆弱であるかという二値的なケースにおいて適用性が高い．このアプローチにより，高い予測精度が実現される．

一方で，Adam オプティマイザはその適応的な学習率調整機能により選択された．Adam は，AdaGrad と RMSProp という 2 つの一般的な最適化手法の利点を統合しており，各パラメータに個別の学習率を適用し，勾配の一階および二階モーメントに基づいて学習率を調整する．この結果，収束に必要な時間が短縮され，特に勾配が疎である場合やデータにノイズが多い場合において，性能が向上することが期待される．

Algorithm 4 Train GCN Model

Input: $Z[L]$, Y_{true}, *MaxEpoch*, *LearningRate*
Initialize parameters θ randomly
for epoch $= 1$ **to** *MaxEpoch* **do**
    Compute predictions: $Y_{\text{pred}} = \text{Softmax}(Z[L])$
    Compute loss: $\text{Loss} = \text{CrossEntropy}(Y_{\text{pred}}, Y_{\text{true}})$
    Perform backpropagation and update θ:
    $\theta \leftarrow \theta - \textit{LearningRate} \cdot \text{Gradients}$
end for
Output: Return trained θ

アルゴリズム 4 では，訓練部分に対する提案モデル構築アルゴリズムの実装を示す．モデルの訓練において，エポック数を 200 に設定した．これは，十分な訓練時間を確保しつつ，過学習のリスクを抑えるバランスを取ることを目的としている．この設定により，データセット内の本質的なパターンを識別しながら，モデルのパラメータを十分に洗練させるための反復を可能にする．過去の研究において，エポック数が少ない場合，モデルがデータの複雑性を十分に捉えることができず，未学習の状態（アンダーフィッティング）に陥ることが確認された．一方で，エポック数を 200 以上に増やすと，訓練データに対しては良好な性能を示す一方，検証データにおいて性能が低下する過学習の兆候が観察された．このような理由から，エポック数を 200 に設定することで，適切な訓練と汎化性能のバランスを実現した．

第6章　性能評価

6.1　評価条件

評価では，図 4 に示すように，米国の商業用バックボーン ISP ネットワークである At Home Network，Allegiance Telecom，CAIS Internet，Verio，ATT の 5 つのネットワークトポロジと生成モデルからランダムに生成された 250 個トポロジを使用した．このうち，Allegiance Telecom，Verio，ATT は hub & spoke 型ネットワークに分類され，ノードが指数分布パターンを示す構造である．この構造では，中央のハブが複数のスポークを接続しており，CFA においてハブが攻撃されると，ネットワーク全体が遮断され，すべての接続されたスポークに影響を及ぼす可能性がある．

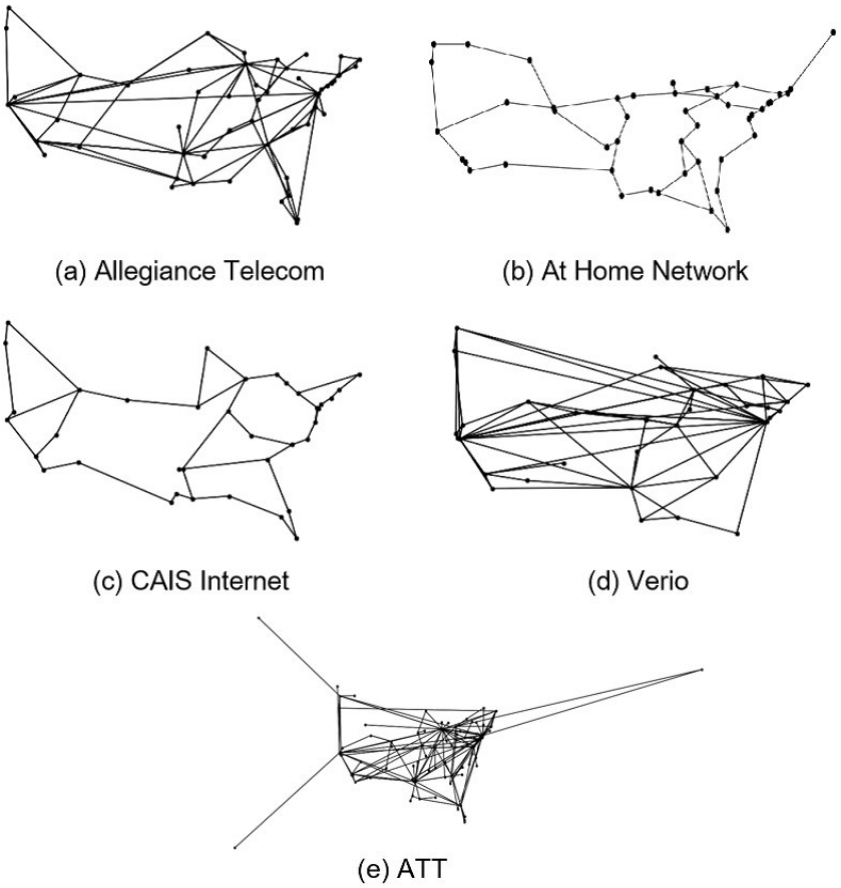


図 4: 5 network topologies

一方，At Home Network と CAIS Internet は ladder ネットワークに分類され，ノードが線形に分布している．このトポロジは，各ノードが隣接するノードと直接通信できるため，冗長性が高い特徴を持つ．

また，評価対象エリア内のノード数 n は 5 に設定している．これは，CFA 攻撃者が通常，ノード数が多すぎるエリアを標的とすることを避けるという理解に基づいている．ノード数が多いエリアを攻撃する場合，攻撃者はより多くのリンクを攻撃し，より大量の攻撃トラフィックを生成する必要があり，その結果，攻撃がより複雑化するためである．

攻撃者の攻撃コストの違いを考慮すると，脆弱エリアの特定は固定的ではない．一部の攻撃者は，攻撃対象として少数のリンクを選択する場合があり，このとき脆弱エリアの $Max_2\ R_n(x)$ は 80%以上になる傾向がある．一方で，より多くのリンクを攻撃対象として選択する場合，$Max_2\ R_n(x)$ の値は 60%程度まで低下することもある．本論では，トポロジ内で $Max_2\ R_n(x)$ が 70%以上のすべてのエリアを脆弱エリアと見なしており，この 70%は設定された閾値である．

モデルは 3 種類の異なる訓練データを使用して訓練を行った．1 つ目は，Allegiance Telecom のネットワークトポロジのみを訓練サンプルとして使用したものであり，残りの 2 つは，生成されたすべてのネットワークトポロジを訓練サンプルとして使用したものである．この点については第 5.2 節で説明されている．これら 3 種類の異なるモデルの予測精度を比較することで，提案モデルの性能を評価した．

6.2　モデル精度

図 5 の青い部分は，Allegiance Telecom を訓練サンプルとして使用したモデルが，Allegiance Telecom，Verio，ATT において非常に高い予測精度（0.9 以上）を示している一方で，At Home Network および CAIS Internet に対する予測精度が非常に低く，0.5 未満であることを示している．この結果は，モデルが hub & spoke ネットワークにおける CFA 攻撃に脆弱なエリアを正確に予測できる一方で，ladder ネットワークのような他の構造に対する予測では期待を満たさないことを示している．この結果は，モデルが特定のトポロジタイプについて効果的に学習し，そのタイプのネットワークトポロジに対する予測能力を向上させることができることを示唆している．この結論に基づき，生成モデルによって生成された 360 個のネットワークトポロジを教師サンプルとして用い，予測モデルを再訓練した．

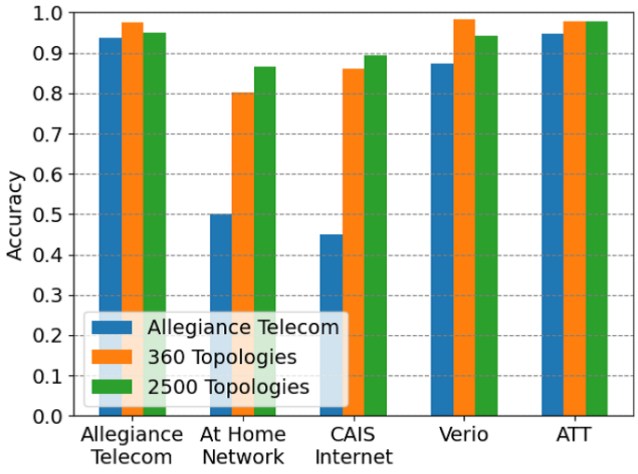


図 5: Accuracy of 3 models in 5 topologies

図 5 のオレンジ色の部分は，360 個のネットワークトポロジを教師サンプルとして使用して訓練したモデルが，5 つの与えられたネットワークトポロジにおいて得られた予測精度を示している．この結果から，モデルの全体的な予測精度が大幅に向上し，すべての予測精度が 0.8 以上であることが確認できる．このことから，本モデルは異なるネットワークにおける CFA 攻撃に脆弱なエリアを正確に予測できると考えられる．

360 個のネットワークトポロジを教師サンプルとして訓練したモデルでは，At Home Network と CAIS Internet における予測精度がやや低下していることがわかる．これは，生成された教師サンプルのネットワークトポロジにおいて選択したすべてのパラメータの範囲が制限されており，At Home Network と CAIS Internet のトポロジにおける平均次数やエッジ密度などの一部のパラメータがその範囲を超えていたためである．このため，サンプルの多様性が不十分であり，360 個のネットワークトポロジを教師サンプルとして訓練したモデルが，これら 2 つのトポロジの脆弱エリアを予測する際に性能が低下した．この問題に対処するため，生成されるトポロジのパラメータ範囲をさらに拡大し，生成されるトポロジの数を従来の 360 個から 2500 個に増やした．このように訓練サンプルの多様性を拡大することで，モデルがさまざまな種類のネットワークトポロジに対して高い予測精度を持つようになると考えられる．

図 5 に示される緑色の部分は，2500 個の生成されたネットワークトポロジを教師サンプルとして訓練したモデルが，5 つの与えられたネットワークトポロジにおいて示した予測精度を表している．全体的に，このモデルは 5 つのネットワークトポロジすべてで予測精度が 0.85 を超え，優れた性能を示している．特に，360 個のトポロジを用いて訓練したモデルと比較して，At Home Network および CAIS Internet という 2 つの ladder ネットワークにおいて，予測精度が大幅に向上している．この改善は，訓練プロセスにおいてサンプルとなるトポロジの多様性を拡張したことに起因しており，モデルが ladder ネットワークを含むさまざまな構造のネットワークトポロジに適応できるようになった結果である．このことにより，モデルの汎化能力および予測精度が向上した．

さらに，Allegiance Telecom および Verio という 2 つの hub & spoke トポロジにおいて，モデルの性能がわずかに低下していることもわかる．これは，訓練サンプルの多様性を増加させることで，複雑なネットワーク構造への適応性が向上する一方で，ある程度の分布の偏りが生じ，比較的単純なネットワークにおける性能が若干低下する可能性を示している．しかしながら，訓練されたモデルは全体として期待に応えるものであり，複雑なネットワーク構造において優れた性能を発揮するとともに，ネットワークトポロジの予測能力を向上させるという設計目標を達成していると評価できる．

6.3 予測精度の累積分布

モデルの予測精度をさらに検証するため，生成モデルを用いて訓練サンプルと同じパラメータ範囲内でランダムに 250 個のネットワークトポロジを生成し，訓練済みモデルを用いてこれらのトポロジを予測した．最終的な予測結果は累積分布図 6 として示されており，横軸はモデルの各トポロジに対する予測精度を，縦軸は予測精度がある値以上のトポロジが全体に占める割合を表している．

図 6 から，予測精度が 0.8 に達した時点で縦軸の値が約 0.8 であることがわかる．これは，全体の 80%のネットワークトポロジにおいて予測精度が 0.8 以上であることを示している．さらに，予測精度が 0.9 に達した時点では，縦軸の値が約 0.2 となっており，20%のネットワークトポロジが 0.9 を超える予測精度を達成していることが示される．これらの結果は，モデルが大部分のネットワークトポロジに対して安定した予測性能を示しており，高精度の範囲で優れたカバー能力を持つことを示唆している．また，予測精度が 0.7 から 0.8 の範囲内で曲線の上昇が急峻になっていることが観察される．これは，モデルの予測性能がこの範囲で急速に向上し，大多数のランダム生成トポロジに適応していることを示している．

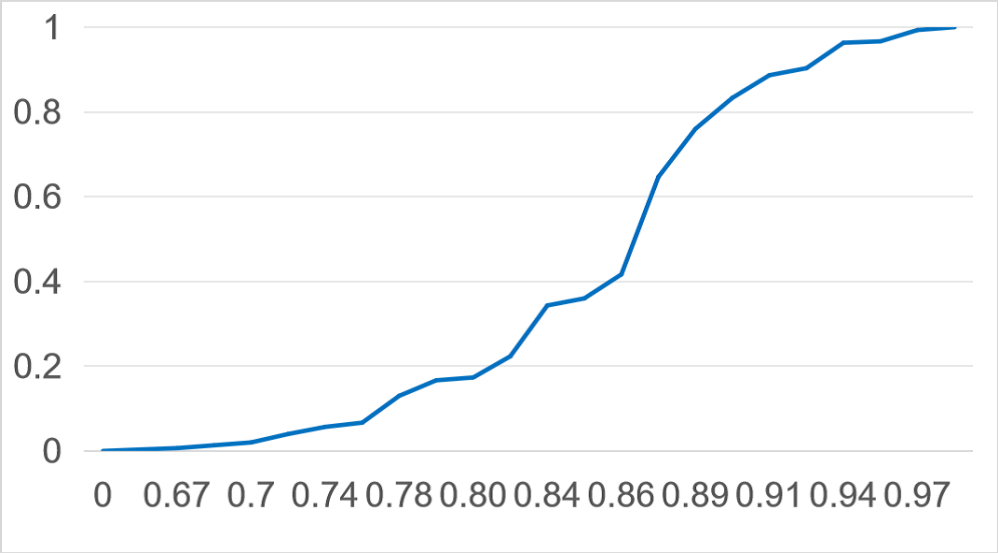


図 6: Cumulative distribution of model prediction accuracy

一方で，予測精度が低いトポロジも少数存在しており，全体の約 5%未満のトポロジで予測精度が 0.7 を下回っている．これは，モデルが特定の構造やパラメータ分布を持つネットワークトポロジに対して適応性が限られている可能性を示している．このような不足は，訓練データの多様性やモデルの極端なトポロジへの適応能力に起因している可能性がある．しかしながら，モデルは大多数のトポロジで優れた性能を示し，予測の信頼性が高いことが確認された．

6.4 検出性能分析

Allegiance Telecom トポロジにおいて，ノード 0，1，2，3，5 で構成されるエリア A がモデルによって脆弱エリアとして予測された．このエリアがネットワークトポロジ内での位置を図 7 に示す．平均次数や平均クラスタ係数といった基本的なネットワーク特性を見る限り，このエリアの構造には顕著な異常は見られず，他のエリアと同程度の特性を持つ．しかしながら，本論の GCN モデルはこのエリアを脆弱エリアとして識別した．この結果は，エリア A が従来のネットワーク解析手法では直接発見することが難しい，より複雑な特徴を持つ可能性を示唆している．

この点をさらに検証するために，表 4 では脆弱なエリア A と，脆弱ではないエリア B および C の主要な特徴値を比較した．明示的な特徴（例えば，エリア内の次数，平均次数，平均クラスタ係数）においては，A と B および C の値はほぼ同等である．しかし，エリア A の $Max_2\ R_n(x)$ 値は顕著に高いことが確認された．この結果は，エリア A の脆弱性の原因が単純なネットワーク構造特性によるものではなく，より高次元的で隠れた特徴に起因している可能性を示している．

具体的に，エリア A の脆弱性は，従来の手法では直接捕捉することが難しい複数の潜在的な特性に起因している可能性がある．例えば，エリア A のトポロジ位置が特定の条件を満たし，トラフィックがいくつかの重要なエッジに集中している場合，これらのエッジが攻撃を受けると，エリア全体が外部ネットワークとの通信を断絶するリスクがある．また，エリア A の隣接構造が他のエリアよりも複雑である可能性があり，この複雑性は隣接ノード間の依存関係に表れている．GCN は層ごとの情報集約を通じて，従来の指標では定量化が難しい動的な相互作

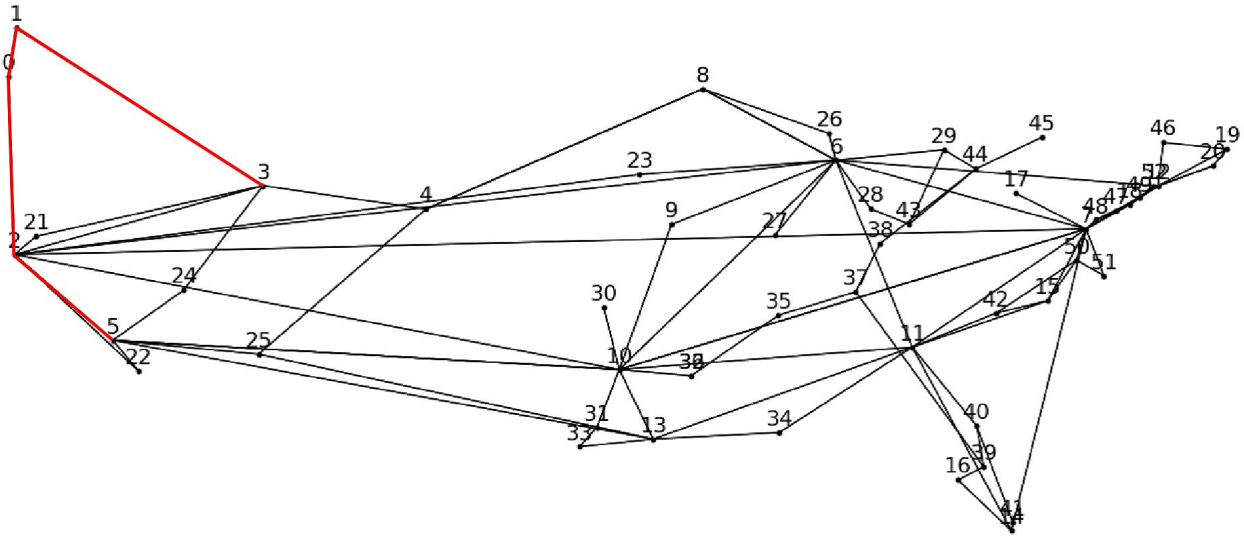


図 7: Vulnerable areas in Allegiance Telecom

用特性を識別することが可能である．さらに，地理的に見ると，このエリアはネットワークトポロジの端部に位置しているように見えるが，実際にはボトルネックリンクに近接している可能性がある．このようなグローバルな特性は，通常，複雑な全体的なグラフ解析を必要とするが，GCN は情報の伝播と集約のプロセスを通じてこれらの依存関係を自動的に捉えることができる．このような潜在特性が相互に作用する結果として，エリア A の $Max_2\ R_n(x)$ 値が顕著に高くなり，CFA 攻撃の潜在的な標的となっていると考えられる．

表 4: 3 Areas in Allegiance Telecom

Area type	Nodes	Degree	Average degree	Betweenness centrality
Vulnerable	0, 1, 2, 3, 5	14	4.8	0.31
Non-vulnerable	1, 3, 5, 10, 24	17	5.0	0.28
Non-vulnerable	9, 10, 11, 14, 41	19	5.4	0.49

この例から，GCN モデルの能力は単に基本的なネットワーク特性の分析にとどまらず，データ内に潜む定量化が難しい隠れた関係性を発掘する点において，従来の手法の欠点を補完することがわかる．この特性は，特に複雑なネットワークの脆弱性分析において重要である．例えば，DDoS 攻撃や CFA の防御において，これらの脆弱エリアを事前に特定できることは，ネットワーク資源の効率的な配置や防御戦略の強化に実際的な意義を持つ．

第7章　まとめ

本論では，CFA に対して脆弱性を示すネットワークトポロジエリアを特定するための GCN ベースの予測モデルの設計を提案した．このモデルは，指定された指標 $Max_2R_n(x)$ に基づいて脆弱なエリアを識別するものである．評価では，hub & spoke 型トポロジである Allegiance Telecom のみを用いて訓練したモデルと，生成モデルによって生成された 360 個異なるネットワークトポロジを用いて訓練したモデルと，生成モデルによって生成された 2500 個異なるネットワークトポロジを用いて訓練したモデルとの性能の性能を比較した．

検証結果により，提案したモデルは異なるネットワーク構造のトポロジにおいて高い予測精度を示し，特定のネットワークトポロジ内で CFA 攻撃に脆弱なエリアを正確に特定できることが確認された．また，ネットワークトポロジ内のエリアにおける潜在的な特徴を発見する能力も実証された．本論の目的は，攻撃が発生する前に効果的かつ詳細な防御策を構築し，ネットワークの安全性と安定性を確保する防御システムを設計することである．

謝辞

本研究を行うにあたり，ご指導をいただいた上山教授に感謝します．また日常，有益な議論をしていただいた研究室の皆様に感謝します．

参考文献

[1] C. Dietzel, M. Wichtlhuber, G. Smaragdakis, and A. Feldmann, Stellar: Network Attack Mitigation using Advanced Blackholing, ACM CoNEXT 2018.

[2] M. Kang, S. B. Lee, and V. D. Gligor, The Crossfire Attack, IEEE SSP 2013.

[3] Max Welling and Thomas N. Kipf, Semi-supervised classification with graph convolutional networks, J. ICLR 2017, 2016.

[4] A. R. Narayanadoss, M. Gurusamy, P. M. Mohan, and T. Truong-Huu, Crossfire Attack Detection using Deep Learning in Software Defined ITS Networks, VTC Spring 2019.

[5] L. Xue, E. W. W. Chan, G. Gu, X. Luo, X. Ma, and T. T. N. Miu, LinkScope: Toward Detecting Target Link Flooding Attacks, IEEE Trans. Information Forensics and Security, 2018.

[6] M. F. Hyder and T. Fatima, Towards Crossfire Distributed Denial of Service Attack Protection Using Intent-Based Moving Target Defense Over Software-Defined Networking, IEEE Access, vol. 9, pp. 112792-112804, 2021.

[7] W. Rafique, X. He, Z. Liu, Y. Sun, and W. Dou, CFADefense: A Security Solution to Detect and Mitigate Crossfire Attacks in Software-Defined IoT-Edge Infrastructure," IEEE HPCC/SmartCity/DSS 2019.

[8] A. Aydeger, N. Saputro, K. Akkaya, and M. Rahman, Mitigating Crossfire Attacks Using SDN-Based Moving Target Defense, IEEE LCN 2016.

[9] Y. Cao, H. Jiang, Y. Deng, J. Wu, P. Zhou and W. Luo, Detecting and Mitigating DDoS Attacks in SDN Using Spatial-Temporal Graph Convolutional Network, IEEE Transactions on Dependable and Secure Computing, vol. 19, no. 6, pp. 3855-3872, 1 Nov.-Dec. 2022.

[10] B. J. Saunders, R. E. de Grande, G. H. S. Carvalho and I. Woungang, "Deep Graph Learning for DDoS Detection and Multi-Class Classification IDS," 2024 IEEE International Conference on Cyber Security and Resilience (CSR), London, United Kingdom, 2024

[11] Wei Guo, Han Qiu, Zimian Liu, Junhu Zhu, and Qingxian Wang, GLD-Net: Deep Learning to Detect DDoS Attack via Topological and Traffic Feature Fusion, Computational Intelligence and Neuroscience, vol. 2022, Article ID 4611331, 2022.

[12] C. Liaskos and S. Ioannidis, Network Topology Effects on the Detectability of Crossfire Attacks, IEEE Transactions on Information Forensics and Security, vol. 13, no. 7, pp. 1682-1695, July 2018.

[13] L. Guo, S. Jing, L. Wei and C. Zhao, "Crossfire Attack Defense Method Based on Virtual Topology," 2023 19th International Conference on Mobility, Sensing and Networking (MSN), Nanjing, China, 2023, pp. 341-350

[14] Manami Nakahara and Noriaki Kamiyama, Detecting Crossfire-Attack Hosts in Search Phase, APNOMS 2022 (Poster Session)

[15] K. Watabe, S. Nakazawa, Y. Sato, S. Tsugawa and K. Nakagawa, "GraphTune: A Learning-Based Graph Generative Model With Tunable Structural Features," in IEEE Transactions on Network Science and Engineering, vol. 10, no. 4, pp. 2226-2238, 1 July-Aug. 2023

外部発表リスト

[1] 王 天嶼, 上山 憲昭, “ クロスファイア攻撃に脆弱なトポロジ上の位置に関する分析 “ , 電子情報通信学会 2023 年ソサイエティ大会, B-6-57, 名古屋, 2023 年 9 月

[2] 王 天嶼, 上山 憲昭, “ クロスファイア攻撃に対して脆弱なエリアの選定法 “ , 電子情報通信学会 ネットワークシステム（NS）研究会, NS2023-195, 沖縄, 2024 年 2 月

[3] 王 天嶼, 上山 憲昭, “クロスファイア攻撃のターゲットエリア選定法“, 電子情報通信学会総合大会, BPO-1-10, 広島, 2024 年 3 月

[4] 王 天嶼, 上山 憲昭, “ GCN に基づくクロスファイア攻撃のターゲットエリア選定法 “ , 2024 年電子情報通信学会ソサイエティ大会, B-11-01, 埼玉, 2024 年 9 月

[5] 王 天嶼, 上山 憲昭, “ GCN に基づくクロスファイア攻撃のエリア選定法 “ , 2024 年電子情報通信学会インターネットアーキテクチャ（IA）研究会, 鹿児島, 2025 年 3 月