Vol.2019-MPS-123 No.44
Vol.2019-BIO-58 No.44
2019/6/19

入門レベル CTF 問題解析による入門者向け教材開発

田中雅浩[†1]　松田健[†1]　園田道夫[†2]

概要：CTF(Capture The Flag)は，ゲーミフィケーションの手法を取り入れた，情報セキュリティ分野の知識を競う競技であり，世界各国で競技が開催されている．CTF はセキュリティ技術を学ぶ学習コンテンツとしても有用であるが，入門者向けの問題であってもある程度の専門知識が要求される場合もあるため，初学者の中にはハードルが高く感じる者も存在する．本研究では，問題文とその問題に対する CTF 経験者の正解 Flag 取得までの課程が分かるデータを収集し，それらの中から技術用語を中心とする単語の共起頻度を調べることで，分野ごとに必要な知識や技術が具体的にどのようなものか，重用語の関連性から調査することで，CTF 入門者に有用な情報が抽出できるかどうか検討する．

キーワード：CTF，共起頻度

Teaching material development for beginners by introductory level CTF problem analysis

MASAHIRO TANAKA[†1]　TAKESHI MATSUDA[†1]
MICHIO SONODA[†3]

Abstract: CTF (Capture The Flag) is a competition that competes in the information security field, incorporating gamification methods, and competitions are held around the world. Although CTF are also useful as learning content for learning security technologies, some beginners may find that the hurdles are high, as some problems for beginners may require some specialized knowledge. In this study, we collect the data which the problem sentence and shows the process of obtaining the correct Flag of the expert CTF player for the problem, and examine the co-occurrence frequency of the word mainly on the technical term from them, and examine what kind of knowledge and technology are necessary for each field concretely, and whether it is possible to extract the information which is useful for the CTF beginners by investigating from the relevance of the important word.

Keywords: CTF, Co-occurrence frequency

1. はじめに

本研究は，Capture the Flag (以下，CTF という)と呼ばれる，情報セキュリティに関する技術を駆使しながら，サーバー上に設置されたフラグ情報を取り合うゲーミフィケーションを取り入れた教育コンテンツの問題文とその解き方を解析することを目的としている．

海外では 20 年以上前から CTF の大会が開催されており，大会の開催数は図 1 のように年々増加している．

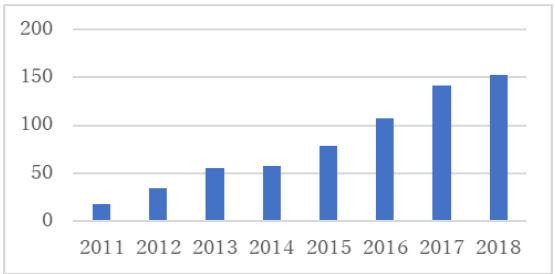

図 1　CTF 大会の開催数遷移

日本では SECCON が開催する SECCON CTF が最も有名で，世界中の技術者が参加し，技術力を競い合っている．しかしながら，このような大きい大会は予選であっても出題される問題の難易度が高く，入門者の中にはハードルが高く感じる者も存在する．

CTF について，過去に出題された問題をまとめたり初心者向けの解説をしたりする Web サイトは複数存在するが，CTF の問題そのものをまとめたり，問題を解くための手法に関する関連付けをしたりするものは見当たらない．

本研究では，CTF に出題された問題文と問題を解くための手法に関するキーワードを収集し，その共起頻度に着目することで問題を出題ジャンルとは異なる観点から再分類することで，CTF 入門者に有用な学習方法の提供の仕方について考察する．

2. CTF の問題について

CTF には，主にクイズ形式と攻防戦形式の 2 つの形式があるが，本稿ではクイズ形式の CTF のみ扱うこととする．

CTF のジャンルには，バイナリを解析する Reversing (Binary)，HDD やメモリなどのイメージファイルを解析する Forensics，プログラムの脆弱性を攻撃する Pwn，Web 上の脆弱性を攻撃する Web，ネットワークパケットなどを解析する Network，暗号を復号する Crypto，様々なデータに隠されたフラグを見つけ出す Misc などがある．

†1 長崎県立大学
University of Nagasaki
†2 情報通信研究機構
National Institute of Information and Communications Technology

ⓒ2019 Information Processing Society of Japan

Vol.2019-MPS-123 No.44
Vol.2019-BIO-58 No.44
2019/6/19

3. 例題

2018 年に開催された SECCON Beginners CTF 2018 で実際に出題された問題の中から 1 問紹介する．

Crypto の問題で，以下の 3 つの暗号が渡される．

・Gur svefg cneg bs gur synt vf: pgs4o{a0zber
・Lzw kwugfv hsjl gx lzw xdsy ak: _uDskk!usd_u
・{ʎɥdɐɹƃ0ʇdʎɹ :sı ƃɐlɟ ǝɥʇ ɟo ʇɹɐd pɹıɥʇ ǝɥ⊥

それぞれ ROT13，ROT18，180 度回転させることで復号できる．

・The first part of the flag is: ctf4b{n0more
・The second part of the flag is: _cLass!cal_c
・The third part of the flag is: Rypt0graphy}

この 3 つを繋げることでフラグとなる．

ctf4b{n0more_cLass!cal_cRypt0graphy}

4. 解析手法

本研究では，インターネット上に公開されている CTF の解法(以下，writeup という)100 件から問題を解くための手法に関するキーワードを収集し，その共起頻度に着目することで問題を出題ジャンルとは異なる観点から再分類を行い，CTF 入門者に有用な学習方法の提供の仕方について考察する．

5. 解析結果

5.1 形態素解析

CTF に相応しい単語が並んでいるように見えるものの，これだけではあまり有用な情報とは言えない．

1	実行,24
2	アドレス,22
3	パスワード,18
4	admin,15
5	アクセス,14
6	スクリプト,13
7	条件,12
8	セット,12
9	ログイン,11
10	平文,11

図 2　形態素解析の上位単語 10 個

5.2 共起頻度(2 単語)

関係性のある単語の組み合わせが並んでおり，strings や file など，問題を解く上で最初に実行することが多いコマンドが見られた．

1	アドレス+実行,4
2	実行+gdb,3
3	平文+暗号文,3
4	格納+関数,3
5	乱数+seed,3
6	アドレス+リーク,3
7	実行+strings,2
8	実行+objdump,2
9	fileコマンド+出力,2
10	binwalk+foremost,2

図 3　2 単語の共起頻度の上位組み合わせ 10 個

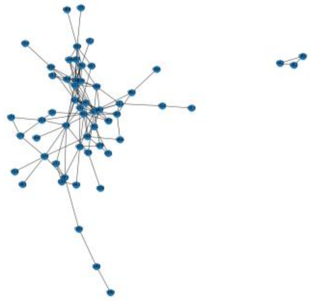

図 4　2 単語の共起頻度の関係図

5.3 共起頻度(3 単語)

2 単語のときよりも具体性が増しており，CTF において有用なツールや，問題を探るような単語の組み合わせが多く見られた．

1	アクセス+ログイン+admin,3
2	libc+リーク+アドレス,3
3	条件+入力文字列+実行,2
4	gdb+実行+objdump,2
5	binwalk+foremost+実行,2
6	セット+アクセス+POST,2
7	shell+実行ファイル+実行,2
8	radare2+実行ファイル+実行,2
9	暗号文+平文+推測,2
10	ログイン+Password+admin,2

図 5　3 単語の共起頻度の上位組み合わせ 10 個

また，3 単語の共起頻度をもとに関係図を作成したところ，図 6 のような結果が得られた．

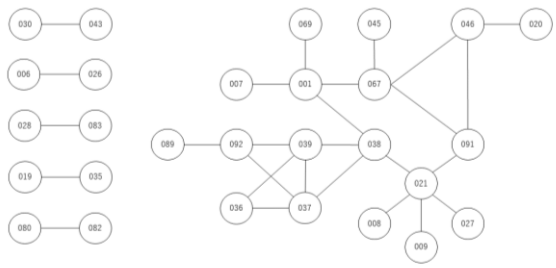

図 6　3 単語の共起頻度の関係図

6. 考察

5.3 節より，図 6 の関係図に紐づいている問題を確認すると，右側のグループが Reversing(Binary)や Web に分類されていた．入門レベルの問題であれば，共起頻度の上位にある組み合わせを試すだけでフラグを得られることもあるため，それらを CTF 入門者に紹介するだけでも一定の効果が得られると推測される．

今後の課題として，サンプル数を増やして解析すること，英語の writeup でも本稿と同様の手法で解析し，日本語の writeup との差を調査すること，などが挙げられる．

7. 参考

[1]SECCON, “SECCON 2018”, https://2018.seccon.jp/, (参照 2019-05-14)
[2]CTFtime team, ”CTFtime.org”, https://ctftime.org/, (参照 2019-05-14)

ⓒ2019 Information Processing Society of Japan