5ZE-01

情報セキュリティの導入教育を目的とした出題型ハッキング競技 CTF の試行実践における解答ログの分析

西村拓海† 中矢誠† 富永浩之†
香川大学† 香川大学† 香川大学†

1. はじめに

ハッキング競技 CTF は，サーバ側に隠された情報を旗(フラッグ)に見立て，ハッカーとしての知識や技能を総動員して，探し出すものである．インターネット上で参加できるチーム対抗の大会が多い．世界各地で開催され，マスコミでも注目されている.日本でも，SECCON[1]が開催され，参加者の裾野も広がっている．

本論では，各問題に対する競技者の解答ログを分析し，問題の分類や難易度設定の妥当性を検討する．

2. 情報セキュリティの導入教育の CTF 大会

本研究室でも，初心者を対象とする情報セキュリティの導入教育として，出題型(ジェパディ型)の CTF の大会イベントを提案している[2]．ハッカーのための本格的な CTF と異なり，ゲーム感覚で楽しみながら，誰でも気軽に参加できる大会を目指す．大会運営サーバ BeeCon を開発し，試行的に運用している[3]．競技者は，チーム単位で取り組む．大会の進捗状況を Web で公開し，観戦者にも広く関心を持ってもらう．大会の後は，講評の時間を設け，復習を促す．

参加が難しい入門者には，サポータ制のように，特定の競技チームへの応援者という役割を与える．応援者は，競技者と協調して取り組める余興ゲームに参加する．余興ゲームは，CTF と連動し，ゲームのポイントが競技の過程や結果に影響を与える．競技者とともにゲームに参加することで，興味や関心を沸かせ，次回以降の CTF への参加を促している．

3. CTF の問題の分類と構築

BeeCon で出題する問題は，情報セキュリティおよび情報リテラシにおける学習内容で分類し，難度に応じて 6 段階のレベルに分けている(図 1)[4].レベル 1 は，初心者が日常的に起こす操作ミスや，知っておくと便利なチップスに関連する．レベル 2 は，不審なデータや安易な操作の危険性を実感させる．レベル 3 は，情報系の新入生が情報処理の仕組みとして理解し，積極的に体験してもらいたい問題である．自分で計算したり，テキストエディタや電卓などの活用が必要である．レベル 4 は，セキュリティに大きく関係してくる．バイナリエディタも必要となる．レベル 5 は，専用のツールやコマンドを利用して，データの特徴を分析する問題である．レベル 6 は，CGI や DBMS など，Web サイトの脆弱性を突く問題である．

一般に，CTF の問題は，解法を明示せず，不親切な出題が多い．そのため，実際には，上位のレベルの問題を，コンテストの開催中に解くことは難しい．そのため，BeeCon では一定時間の後にヒントを提示することにしている．

4. 試行実践の内容

今回の試行実践では，図 1 の分野を網羅するように問題を 20 問用意した(図 2)．しかし，2 つの問題に不備があった．問題番号 8 は運営者の問題の読解ミスにより，振り分けを誤った．本来は 2-2 ではなく，5-2 に所属すべき問題であった．問題番号 12 は解くことができない問題であった

問題の配点を決定するにあたり，図 1 の各段階に基礎点を設定した．段階 1，2 の基礎点，段階 3，4 の基礎点，段階 5，6 の基礎点はそれぞれ 150 点，250 点，350 点である．また解答に必要な知識と手間から問題の難度を A~E の 5 段階で設定した．難度 A, B, C, D, E では，それぞれ基礎点に対し，-100, -50, 0, +50, +100 点とする．この難度の設定により，同じ段階の問題でも必要な知識や手間が多ければ，問題の配点を高くすることができる．

試行実践では，本学の情報系サークルの 1～2 年生 18 名に競技者として参加してもらい，チーム対抗ではなく，個人対抗で問題に取り組んでもらった．競技時間は，90 分とし，協議開始から 60 分後にヒントを開示した．

今回は、競技者の解答ログから問題のレベル設定の妥当性を検討することが目的のため，観戦者は存在せず，余興ゲームも行わなかった．

Answer Log Analysis in Hacking Competition CTF with Jeopardy Style for Introductory Learning about Information Security
†Takumi NISHIMURA, Kagawa University
†Makoto NAKAYA, Kagawa University
†Hiroyuki TOMINAGA, Kagawa University

5.試行実践の結果

競技者の各問題に対する解答ログや解答状況から，閲覧数，着手数，正答数，着手時間，吟味時間を算出した(図 2)．閲覧数は，問題を閲覧した人数，着手数は，問題の解答を一度でも提出した人数，正答数は問題を正答した人数である．着手時間は，問題ページを閲覧してから，最初の解答を提出するまでの時間を平均した時間，吟味時間は，最初の解答を提出してから，正答または最後に誤答の解答を提出するまでの時間を平均した時間である．

最初に，図 2 から，基礎点の違う問題の解答状況を比較し，分析する．基礎点が高いほど，閲覧数，着手数，正答数が少なくなる傾向がある．また，閲覧数と着手数の差も大きくなり，着手時間も長くなる．これらより，レベルが高くなるにつれて，問題を見ただけで諦めたり，問題を見すらしない競技者，解答を提出するまでの時間が増える傾向にある．

次に，同じ基礎点で難度が違う問題の解答状況を比較し，分析する．閲覧数は，同じ基礎点の場合は，ほとんど変わらない．しかし，着手数，正答数，着手数と正答数の差，着手時間については，難度に従い，上記のレベルの場合と同様の傾向が見られた．問題番号 12 だけ例外的に，閲覧数と着手数の差が大きくなっている．この問題は，外部サイトで問題を解けなければ，解答を検討できないため，着手ができなかったと考えられる．

上記の傾向から，今回の試行実践で用意した問題の難易度や配点の設定は妥当であったと見られる．

6.おわりに

情報セキュリティの導入教育の一環として，初心者を対象とする CTF 大会を提案し，大会運営サーバ BeeCon を開発している．CTF は，ジェパディ型で，分野と難易度に応じて，6 段階のレベルを設ける．本論では,試行実践を実施し,競技者の解答ログと解答状況から,問題の難易度や配点の妥当性を検討した．問題の配点が高いにも関わらず，着手数や正答数が多い問題はほとんどなかったことから，問題の難易度や配点の妥当性を確認した．

今後の課題として，各問題の解答提出数のうち，正答，誤答の割合を算出することで，正答数は多いが、正答の割合は低い問題や，正答数は少ないが、正答の割合は高い問題などを割り出し，分析する．また，振り分けミスをなくすため，モデレータによる難易度の吟味を行う．他にも，Web 問題など，外部サイトに依存している問題を BeeCon 上で実施し，ログを収集できるように改良を試みる．

レベル	対象者	出題分野
1	一般の大学生	1 キーボードのキー配置とシフト操作 2 マウスやタブレットの操作 3 Web ページの閲覧や URL 指定の構成 4 Web ブラウザと情報検索エンジンの機能 5 セキュリティ関連の用語
2	理系の高校生	1 様々なユーザ認証とパスワードの重要性 2 圧縮ファイルや実行ファイルのバイナリ 3 マルチメディアのファイル形式の復元再生 4 Web ページの HTML ソースの閲覧 5 オープンな SNS からの情報入手
3	情報系新入生	1 文字化けのテキストと文字コードの変換 2 二進数やビット列の変換と計算 3 簡単な暗号解読やエンコード文字列の復元 4 悪意のある Web ページへのアクセス回避 5 PC の OS のコマンドのコマンド操作
4	意欲的高校生	1 文字列とハッシュ値の変換 2 文字列の検索と正規表現の利用 3 バイナリエディタによるビット列の走査 4 C 言語のプログラムの実行
5	情報系上級生	1 Linux のコマンド操作と簡単なスクリプト 2 バイナリデータの特徴の分析 3 ネットワーク通信のパケットの解析 4 オブジェクト指向言語の実行
6	意欲的新入生	1 クライアント側スクリプトの脆弱性（JS） 2 Web CGI の脆弱性(XSS) 3 DBMS の脆弱性(SQL インジェクション) 4 本格的なフォレンジックス

図 1 問題のレベル設定と出題分野

番号	問題名	分野	難度	配点	閲覧	着手	正答	着手時間	吟味時間
1	かな文字入力	1-1	B	100	18	18	18	10:25	07:43
2	友人からの奇妙なメール	1-2	B	100	18	18	17	13:40	05:48
3	何かが違う	1-3	B	100	18	17	11	06:47	15:09
4	記事を探そう	1-4	C	150	17	12	9	12:38	05:06
5	攻撃手法の名前の検索	1-5	A	50	18	18	18	03:43	03:34
6	セキュリティ用語の検索	1-5	A	50	18	18	17	03:15	02:17
7	パスワードの使いまわし	2-1	C	150	18	18	18	00:50	00:51
8	Only one	2-2	C	150	17	0	0		
9	開けない画像ファイル 初級編	2-3	B	100	17	12	12	03:13	00:04
10	見えないフラグ	2-4	B	100	18	15	13	02:05	13:31
11	Infinity Links	2-4	D	200	15	4	3	16:26	00:00
12	SNS に気をつけろ	2-5	B	100	15	14	0	06:05	26:45
13	メッセージを読め！	3-1	C	250	16	12	11	07:51	04:40
14	計算せよ	3-2	B	200	18	18	17	05:31	00:50
15	暗号!?!?	3-3	C	250	17	7	0	35:17	02:40
16	MD5 can restore!	4-1	B	200	16	13	11	15:08	02:37
17	正規表現を使おう	4-2	B	200	15	1	0	04:25	00:20
18	開けない画像ファイル 上級編	4-3	D	300	10	4	2	33:53	00:03
19	履歴からパスワードを検索	5-1	B	300	17	10	8	18:38	04:18
20	SQL インジェクション	6-3	C	350	12	1	0	01:52	24:57

図 2 試行実践で用意した問題内容と解答状況

参考文献

1) SECCON : SECCON CTF, http://www. seccon. jp/.
2) 中矢誠, 赤木智史, 富永浩之, “情報リテラシとセキュリティの導入教育のための初心者向けのハッキング競技 CTF による大会イベント -大会運営サーバ BeeCon の設計と実装-”, 信学技法, Vol.115, No.223, pp.53-60 (2015).
3) 中矢誠, 富永浩之, “情報リテラシとセキュリティの導入教育のための初心者向けのハッキング競技 CTF による大会イベント - オープン利用のための仮想化の導入と運用方法 -”, 情処研報, Vol.2015-CE-133, No.16, pp.1-8 (2016).
4) 楠目幹, 阿部隆幸, 中矢誠 富永浩之, “情報セキュリティの導入教育のための大会イベント BeeCon におけるハッキング競技 CTF の問題構築”, 情報処理学会 第 79 回全国大会講演論文集, Vol.2017, No.1, pp.739-740 (2017)