Computer Security Symposium 2018
22 - 25 October 2018

Webサイトのセキュリティ・センサス

櫻井　悠次[1]　奥田　哲矢[2]　秋山　満昭[2]　渡邉　卓弥[2]　高田　雄太[2]　須賀　祐治[3]　森　達哉[1]

概要：
これまでに Web サイトのセキュリティに関して，SSL/TLS 設定の不備，DNS 設定，証明書の不備などを個別に分析した研究事例は多数ある．しかしながらこれらの要素を互いに相関付け，総合的に分析した研究事例はなかった．セキュリティは信頼の鎖であり，関連する要素すべてを解析することではじめて全体としての評価が可能となる．本研究は Web サイトを対象とした**セキュリティ・センサス**と称して，Web サイトにかかる様々なセキュリティ要素を総合的に分析した結果を報告する．Web サイトを様々なカテゴリ，用法，機能で分類し，セキュリティ対策の実践あるいは非実践につながる要因の特定を狙いとする．

キーワード：Web サイト，SSL/TLS, DNS, クラウドソーシング

1. はじめに

インターネットバンキング，仮想通貨取引所，オンラインショッピング，SNS 等のサービスに代表されるように，今日の Web サイトは個人の金融資産やプライバシーに関わる重要な情報を管理している．それに伴い，Web サイトが管理する重要な情報の窃取，改竄を目的としたサイバー攻撃も増加している．証券取引やクレジットカード決済等における重要な情報の機密性，完全性を保証する手段として，多くの Web サイトが SSL/TLS を採用している．2018 年現在，大多数の著名な Web サイトが SSL/TLS への対応を完了している．さらに Google 等の主要ブラウザベンダは HTTPS に対応しない Web サイトへの接続時に警告を発する等，HTTPS の普及を加速させる動きが広がっている [1]．一方，これまで SSL/TLS に関する様々な脆弱性 [2–4] が報告されている．すなわち，単純に HTTPS を導入しただけでは Web サイトが安全であるとは言えない．Web サイトの管理者はこれまで発見された脆弱性を理解し，適切にサーバの構築と設定・運用を実施する必要がある．

Web サイトの SSL/TLS 通信の脆弱性への対応状況を大規模に調査する先行研究は多く存在する [5–7]．しかしながら，Web サイトの安全性を考えた時，SSL/TLS の対応状況を調査するだけでは不十分である．SSL/TLS の設定に加え，Web サイトのドメイン名を制御する DNS の設定や，使用しているデジタル証明書の不備などを総合的に分析してはじめて，全体として Web サイトのセキュリティ評価が可能となる．本研究ではこの点に着目し，**Web サイトの安全性を総合的に評価するための基礎データの収集・分析**を行う．また，そのようなデータ収集・分析を「**セキュリティ・センサス**」と称し，Web サイトごとのアクセス状況，カテゴリ，機能ごとにセキュリティ対策の実践，非実践の実態を明らかにすることを狙いとする．

Web サイトのセキュリティ・センサスを実施するために，Alexa Top Sites [8] の人気上位 Web サイト 1,000 件と，ランダムに抽出した Web サイト 1,000 件を対象として以下の項目の調査を行う．

- Web サーバの SSL/TLS 設定状況
- 公開鍵証明書の規定準拠状況
- サーバ証明書の認証レベルによる分類
- DNSCAA への対応の有無
- DNSSEC への対応の有無

さらにセンサスを実施する対象となる Web サイトを以下のような分類軸でグループ化し，それぞれの項目の相関およびセキュリティ対策状況に与える影響を分析する．分類により，どのような特性をもつ Web サイトがセキュリティに注意を払っている傾向があるかがわかる．また，一般にプライバシーポリシーを掲載している Web サイト管理者は収集した個人情報に注意を払っていると考えられるが，その事実が SSL/TLS 設定などの他のセキュリティ対策の実践とどのように相関をもつかを明らかにする．

- Web サイトのカテゴリ

[1] 早稲田大学 基幹理工学研究科
[2] NTT セキュアプラットフォーム研究所
[3] インターネットイニシアティブ

- ログインの可否
- 使用料金の要否
- プライバシーポリシーの有無
- 使用している CDN サービス
- 認証レベル

本研究の貢献を以下に示す．

- Web サイトのセキュリティ・センサスを実施し，SSL/TLS 設定の不備，DNS 設定，証明書の不備に関する総合的な分析を行った．
- Alexa 順位の高い `Search Engine` および `Adult` カテゴリでは SSL/TLS 設定を適切に行っている傾向があることを明らかにした．
- SSL/TLS 設定を適切に行っている Web サイトは以下の特徴を持つ傾向が高いことを明らかにした
  - DNSCAA または DNSSEC に対応している．
  - エラーがない証明書を用いている
  - CDN サービスを用いている
- DNSCAA へ対応している Web サイトではプライバシーポリシーを掲載しており，CDN サービスを使用していない傾向があることを示した．
- DNS 設定にまで注意を払っている Web サイトのセキュリティレベルは高い傾向があることを示した．

本論文の構成は以下の通りである．2 章では本論文の理解に必要な研究背景について，3 章では対象 Web サイトの分類，SSL/TLS 設定の評価および各種データ収集手法について，4 章では調査結果および各分類における分析結果について，5 章では本研究における制約，今後の課題および得られた知見について，6 章では関連研究について報告し，7 章で本論文のまとめを行う．

2. 研究背景

本章では本研究の理解に必要な研究背景について述べる．

2.1 SSL/TLS

Web サイトと通信を行う際, 現在では多くの Web サイトが SSL/TLS プロトコルを用いた HTTPS 通信を行うことでクライアントとサーバ間の通信の暗号化を行い，攻撃者からの盗聴や改竄を防ぐ．しかし，単純に HTTPS 通信を行っていることが通信中の安全を保証しているわけではない．Web サーバが提供する暗号アルゴリズム，署名アルゴリズム，鍵長などが脆弱なものであれば攻撃者が通信の盗聴，改竄を行う恐れがある．その他にも SSL/TLS に関する脆弱性は過去に HeartBleed など，様々なものが発見されている．

2.1.1 HTTPS 関連情報の大規模収集

Censys [9] では，定期的な非常に大規模なインターネットスキャンによって集めた多くの Web サイト，IPv4 ホスト，デジタル証明書の情報を保存している．2018 年 8 月現在では，約 144 万の Web サイト，約 1 億 700 万の IPv4 ホスト，7 億 7500 万ものデジタル証明書の情報が格納されており，詳細なデータを閲覧することができる．具体的には，Web サイトでは，HTTPS 通信における暗号スイートや鍵交換アルゴリズムとその鍵長，署名アルゴリズム等の情報が載っている．加えて，HeartBleed や FREAK などの主要な SSL/TLS 通信の脆弱性の有無に関する情報について確認することが可能である．

2.1.2 デジタル証明書の検証

ZLint は Kumar [5] らによって作成されたデジタル証明書を対象に検査するツールである．デジタル証明書が CA/BF Baseline Requirement と RFC5280 に準拠したものであるかを検査し，準拠していないものにはエラーを出力する．これら 2 つの仕様書は各認証局が順守すべき証明書発行の規定となっているものである．Kumar らは，この ZLint の検査の結果，エラーを多く含む認証局では証明書の不正発行など，証明書の発行プロセスが妥当でない傾向があると示唆している．また，彼らは世界中のデジタル証明書の ZLint による検査結果を Censys にて公開しているので，本研究ではこのデータベースを活用する．

2.1.3 認証レベル

認証局はドメイン管理者の正当性，実在性を確認し，証明書を発行する必要があるが，その認証レベルは 3 段階 (DV, OV, EV 認証) に分かれている．OV 認証では企業の組織情報の審査や実在性の確認が行われ，最高レベルの EV 認証ではそれに加えて第三者機関のデータベース等を参照するなど，さらに厳格なチェックが行われる．しかし，DV 認証ではドメインの管理権限を有しているかの確認しか行われず，無料の証明書発行サービスなども多く存在するため，フィッシングサイト等に利用されることが多い．

2.2 CDN

CDN(Content Delivery Network) とは動画や画像、JavaScript などの大容量のコンテンツを大量に配信するするための負荷分散ネットワークである．近年の Web サイトでは大容量なコンテンツを扱うことが増加しているが，エンドユーザーからのリクエストをオリジンサーバのみで対処していては莫大な負荷に耐え切れず，障害が発生する可能性が増加する．CDN では，事前に様々な場所のコンテンツサーバにコンテンツをキャッシュし，エンドユーザからのリクエストにはそのユーザから物理的に近い，最適な位置のコンテンツサーバが応答する．これにより，応答速度を向上させるだけでなく，オリジンサーバの負荷を軽減することが可能となる．また，多くの CDN サービスでは手軽に SSL/TLS 設定を行うことができるサービスを提供している．そのため，各 CDN サービスにおいて脆弱な SSL/TLS 設定が存在すると，その CDN サービスを使用している多くのユーザ（Web サーバ管理者）に悪影響を

与えてしまう．現在，多くの Web サイトで利用される主要な CDN サービスには様々なものがあるが，本研究では CloudFlare, Akamai, Amazon CloudFront を対象として使用の有無を調査し，他の調査結果とともに分析を行う．

2.3 プライバシーポリシー

現在多くの Web サイトでは会員登録などで個人の名前や生年月日，パスワードの設定を求めるなど，多数の個人情報を取り扱っている．そのような収集した個人情報をどのように扱うのかについてサイトの管理者が決定した方針をプライバシーポリシーという．近年，相次ぐ企業の個人情報漏洩事案によるユーザからの懸念の高まりや，EU 一般データ保護規制（GDPR）等の法規制強化の動きを受けて，各企業は個人情報の取り扱いには注意を払って適切に処理している旨をプライバシーポリシーにて宣言している．

2.4 DNSCAA

DNSCAA はドメイン管理者が，そのドメインの証明書に対して発行できる認証局を指定することができる仕組みである．DNS の CAA レコードに認証局のドメインを指定することでその認証局のみがそのドメインに対して証明書を発行することが許可される．これにより，他の認証局による誤発行を防ぐことができる．この仕組みは 2013 年に策定された RFC6844 にて記述されているが，あまり利用は進んでいなかった．だが，2017 年に CA/Browser Forum が証明書発行の際に各認証局が CAA レコードを確認することを義務化すると発表したため，普及が少しずつ広がっている [10]．Qualys 社による調査では 2018 年 8 月現在，Alexa-Top の 150k のドメインにて CAA レコードを設定しているドメインは 3.7%であった [11]．

2.5 DNSSEC

DNSSEC とは公開鍵基盤 (PKI) と電子署名の仕組みを利用して DNS サーバとの通信の機密性，完全性を保証する仕組みである．DNSSEC を実現するにはドメインの各権威 DNS サーバ，オープンリゾルバ，ドメインの接続先となるサーバが全て DNSSEC に対応している必要があり，普及はあまり進んでいない．しかし，DNS キャッシュポイズニングなどの DNS に関するインシデントは数多く発生しているため，このような DNS 通信を守る取り組みは重要である．

3. Web サイトの分類，SSL/TLS 設定評価および各種データ収集手法

本研究ではまず，Alexa のアクセス数上位 1,000 件の Web サイト（Top 1k）と，アクセス数上位 100 万件の中からランダムで抽出した 1,000 件（Random 1k）を対象として各 Web サイトを SNS サイト，金融サイト等の各カテゴリに分類した．それに加え，対象サイトにおけるログインの可否や使用料金の要否，プライバシーポリシーの有無，使用する CDN サービスなどについても分類を行った．また，SSL/TLS 通信に関するサーバの実装状況，および DNSCAA，DNSSEC 等のセキュリティオプションへの対応状況を調査した．本章では Web サイトの分類手法，SSL/TLS 設定の評価方法および各データの収集手法について述べる．

3.1 Web サイトの分類

本研究では Web サイトを以下の観点で分類した．

(1) Web サイトのカテゴリ（一般的に用いられる 9 つのカテゴリである `Search engine`，`Shopping`，`Social Media`，`Company`，`Underground`，`Adult websites`，`Finance`，`News`，`Education`）
(2) ログインの可否
(3) サイトを利用する上での使用料金の要否
(4) プライバシーポリシーの有無
(5) 使用する CDN サービス
(6) 認証レベル

（1）から（4）の観点にて Web サイトを分類するために，本研究では Amazon Mechanical Turk（以下，MTurk と呼ぶ）によるクラウドソーシングを用いた．MTurk では世界中の登録されたワーカーに写真や動画オブジェクトの識別などの現在でも自動化が難しいタスクを依頼することができる．今回の Web サイトの分類では 25 サイトごとに 1 つのタスクを作成し，1 つのタスクに 3 人のワーカーを割り当てることで多数決判定による分類を実施し，1 人のみによる分類よりも精度を高める．本研究ではグローバルで人気のある Web サイトを含むため，英語で記述されたサイトが多数存在する．したがって，本研究におけるクラウドソーシングでは US 在住のワーカーを対象とした．ワーカーの回答時間が 5 分を切るような短いものや，全てのサイトを同じカテゴリに分類しているといった明らかに適切に分類を行っていない回答は拒否した．Web サイトがワーカーの知らない言語で書かれていた場合，それが Web サイトの分類への精度を左右する可能性があるため，サイトの言語が英語であるかどうかの質問も用意した．

（5）の分類について，本研究では主要な CDN サービスとして知られる，CloudFlare, Amazon CloudFront, Akamai の 3 つについて焦点を当て，分類を行った．Linux のシェル上でドメインの DNS 設定状況が確認できる dig コマンドを使用し，NS レコードに記載されているドメインから対象サイトが利用する CDN サービスを判別した．具体的には，それぞれのドメインの NS レコードに `cloudflare.com`，`awsdns`，`akam.net` が含まれていれば，CloudFlare，Amazon CloudFront，Akamai をそれぞれ CDN サービスとして利用していると判断した．（6）につ

いては Censys のデータベースに既に存在する，対象 Web サイトが使用するデジタル証明書を収集し，それに記載されている認証レベルから分類した．

3.2 SSL/TLS

Web サイトの SSL/TLS 設定状況を調査するために，Censys が提供しているデータを利用する．これは，サーバの SSL/TLS 設定のスキャンを行うツールを使用することで対象 Web サイトのサーバへかかる負荷を考慮したためである．本研究では Censys のデータから取得可能な以下の SSL/TLS 設定に注目し，解析を進める．

- HeartBleed, FREAK, Logjam の脆弱性の有無
- DH 鍵交換，RSA 鍵交換の鍵長
- 署名アルゴリズム
- 最新のプロトコルバージョン

この他にも RC4 の使用や,ROBOT 攻撃等が主要な脆弱性として存在するが，Censys のデータベースからでは確認ができないため，本研究では割愛する．また，Censys のデータベース上の暗号スイートでは Censys 調査用クライアントが当該サーバとのネゴシエーション結果として得られたもののみしか掲載されていないため，今回はその暗号スイートをデータとして用いる．今回，脆弱性を Minimal, Medium, Critical と分類し，それぞれの脆弱性の有無の程度で SSL/TLS 設定を評価が高い順に A, B, C, F といった形で評価するプログラムを実装した [12, 13]．HeartBleed などの非常に脆弱な SSL/TLS 実装をしている Web サイトについてはその旨を強調するため，最低評価を F としている．

その脆弱性の分類を**表 1** に示し，それを用いて Web サイトの SSL/TLS 設定を評価するプログラムのフローチャートを**図 1** に示す．鍵長が 1024 ビットの RSA および DH 鍵交換と署名用途のハッシュ関数 SHA1 は現在では既に危殆化しているため，Critical な脆弱性として分類を行った．フローチャートに示すように，Critical な脆弱性が一つでも見つかれば評価を F にまで下げ，どの脆弱性も有していなければ評価を A にするといった減点方式で評価を行う．Medium の脆弱性を有しているもしくは Minimal の脆弱性を複数有していれば評価を C に，Minimal の脆弱性をひとつ有していれば評価を B とする．

以上のように各 Web サイトの評価を行うプログラムを実行することで大まかな Web サイトの SSL/TLS 実装の状況をみる．Censys のデータベース上において，対象 Web サイトの情報が存在しない場合，表 1 に示す SSL/TLS 設定の評価を行う上で必要な項目が欠損していた場合は評価を行わない．したがって，SSL/TLS 設定評価を行うデータ数は Top 1k では 924, Random 1k では 801 となった．

対象 Web サイトの証明書の不備については，使用しているサーバ証明書，またはその中間証明書の ZLint による検

表 1 SSL/TLS 設定における各脆弱性の分類

脆弱性の分類名	脆弱性
Minimal	Not Support TLS1.2
	DH Key Length < 2048
	Using RSA Key Exchange
Medium	RSA Key Length < 2048
Critical	Including HeartBleed or FREAK or Logjam
	DH Key Length or RSA Key Length < 1024
	Signature Algorithm using SHA1 or MD5

査結果のエラーの有無で判断する．Censys が研究コミュニティ向けに提供しているデータを利用して収集する．

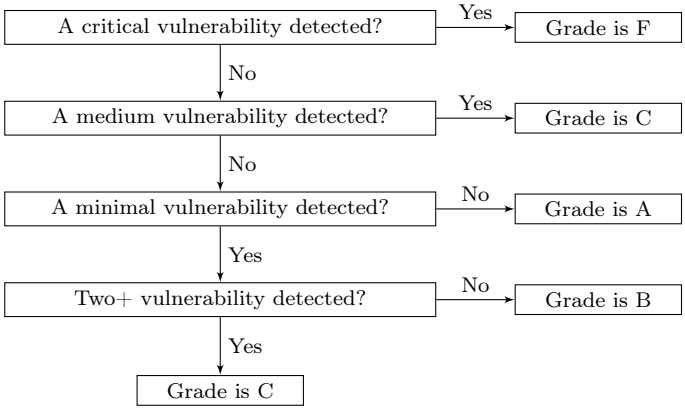

図 1 Web サイトの SSL/TLS 設定の評価手法

3.3 DNSCAA および DNSSEC

Web サイトのドメインが DNSCAA, DNSSEC に対応しているかの有無について，`dig` コマンドを使用してその応答の内容から判断した．具体的には，DNSCAA では `dig caa` ドメイン名 のコマンドを打ち込み，応答内容にそのドメインの CAA レコードが含まれていれば CAA レコードに対応していると判断する．DNSSEC では，`dig +dnssec` ドメイン名 のコマンドを打ち込み，応答内容にリソースレコードの電子署名を表す RRSIG レコードが含まれていれば，DNSSEC に対応していると判断する．

4. 結果

本章では 3 節で示した手法による Web サイトの分類結果，SSL/TLS 設定評価結果，DNSCAA，DNSSEC への対応状況および各分類におけるセキュリティレベルを示す結果を述べる．

4.1 Web サイト分類結果

4.1.1 カテゴリ別の分類結果

MTurk による Web サイトの各カテゴリへの分類結果を**図 2** に示す．`Other` はどのカテゴリにも属さないと判断された Web サイト，N/A は 3 人のワーカーがそれぞれ別々の分類を行い，多数決判定ができなかったことを示す．図に示す通り，Top 1k では 296, Random 1k では 169 も

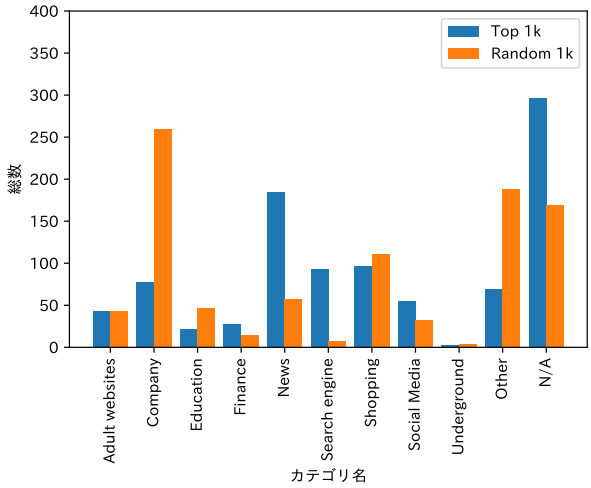

図 2　MTurk による Web サイトのカテゴリ分類結果

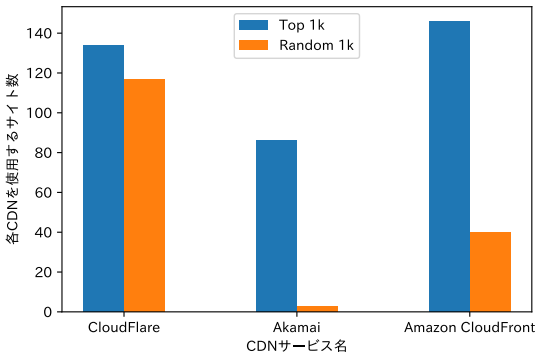

図 4　各 Web サイトが使用する CDN サービス

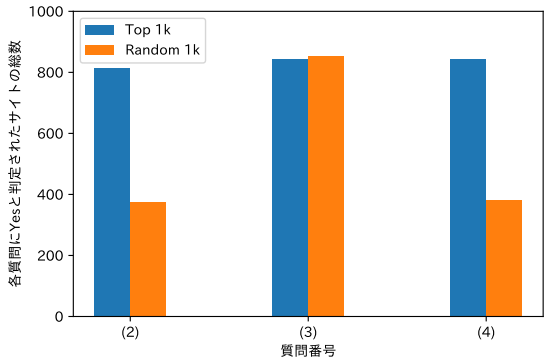

図 3　MTurk による（2）ログインの可否，（3）サイトを利用する上での金銭支払いの有無，（4）プライバシーポリシーの有無による Web サイトの分類結果

のサイトを適切に分類することができなかった．この原因については 5 節にて考察する．Top 1k においては `News`，`Search Engine`，`Shopping` といったカテゴリの Web サイトが多く見られた．`News` では Fox New や CNN，`Search Engine` では Google 社が提供する検索サイト，`Shopping` では Amazon や ebay などの有名な Web サイトが Top 1k に多くあるためと考えられる．Random 1k では `Company` と `Other` が突出して多く見られた．ランダムに抽出した際に `Company` が多く含まれたこと，Random 1k では Alexa の順位が低い Web サイトも多数含まれていたため，見慣れないサイトに対して判断に迷ってしまったことなどが原因として考えられる．

4.1.2　ログインの可否，料金支払いの要否およびプライバシーポリシーの有無による分類結果

次に（2）ログインの可否，（3）Web サイトを利用する上での金銭支払いの要否，（4）プライバシーポリシーの有無による分類結果を**図 3** に示す．グラフの縦軸は各質問において答えが Yes と判定されたもの，つまり，（2）ではログイン可能，（3）では無料で使用可能，（4）ではプライバシーポリシーを掲載している Web サイトの総数を表している．Top 1k では（2）から（4）までどれも 800 前後であるが，Random 1k では（2）と（4）で総数が 400 弱という結果であった．これより，Alexa の上位にある Web サイトほどログイン可能で，かつプライバシーポリシーを適切に掲載している傾向があり，無料で使用できる Web サイトは Alexa の順位によらず多く存在することが分かる．

4.1.3　使用する CDN サービスによる分類結果

（5）の各サイトが使用する主要な 3 つ（CloudFlare, Akamai, Amazon CloudFront）の CDN サービスによる分類結果を**図 4** に示す．図に示す通り，Random 1k に比べ Top 1k では多くの Web サイトでこれらの主要な CDN サービスを使用している．特筆すべきは Akamai を使用する Web サイトが，Top 1k では 86 あるのに対して Random 1k ではわずか 3 しかないことである．Akamai は Alexa 上位の非常に大容量のコンテンツを配信する大規模な事業者が使用する傾向があり，Alexa のランクが高くない，あまり規模が大きくない事業者からはあまり使用されない傾向があることが伺える．Amazon CloudFront についても同様のことが言える．CloudFlare は Top 1k，Random 1k のどちらにおいても 100 以上の Web サイトで使用されており，Alexa の順位によらずに人気があることが分かる．CloudFlare では無料のプランを提供していることから [14]，サービスの価格が影響していると考えられる．

4.1.4　認証レベルによる分類結果

（6）の認証レベルによる分類結果を**表 2** に示す．Top 1k では大規模な企業や組織が運営している Web サイトが多く存在するため OV 認証が多く，Random 1k では規模の小さな事業者や個人が運営する Web サイトが多くあるため，DV 認証が突出して多かったと考えられる．EV 認証を得るには厳格な審査が必要であり，加えて費用も多くかかることから，どちらの場合においても EV 認証を得ている Web サイトは少ない．しかし，Top 1k の方がその割合が高いことから，Alexa の順位が高いほど EV 認証を得ている Web サイトが多い傾向がある．

表 2　認証レベルの分類結果

認証レベル	Top 1k	Random 1k
EV	8.4%	3.7%
OV	58.7%	16.1%
DV	32.8%	80.2%

表 3　SSL/TLS 設定の評価結果

Grade	Top 1k	Random 1k
A	95.2%	93.5%
B	4.43%	2.50%
C	0.22%	0.25%
F	0.11%	3.74%

表 4　DNSCAA, DNSSEC への対応状況

	Top 1k	Random 1k
DNSCAA	10.1%	3.2%
DNSSEC	2.2%	2.9%

4.2 SSL/TLS

Top 1k, Random 1k において, 3.2 節で示した SSL/TLS 設定の評価手法による対象 Web サイトの結果の割合を**表 3** に示す．どちらにおいても 90%以上の Web サイトが A 評価となっている．F 評価の割合は Top 1k では 0.11%, Random 1k では 3.74%あり，Random 1k では Critical な脆弱性を含む Web サイトが多く存在した．また，F 評価となった全ての Web サイトの証明書では SHA1，もしくは MD5 を用いた署名アルゴリズムを用いており，ブラウザの信頼された証明書ストアに入っていない自己署名の証明書を用いていた．

また，対象 Web サイトが用いるサーバ証明書，中間証明書における ZLint の検査結果は，Top 1k では 7.6%，Random 1k では 10.3%がエラーを含んでいるという結果であった．以降の 4.4 節にてこの結果と他の調査結果を用いた分析を行っていく．

4.3 DNSCAA，DNSSEC

Top 1k，Random 1k における DNSCAA,DNSSEC の対応状況を**表 4** に示す．表 4 に示す通り，DNSCAA に関しては Top 1k が 10.1%，Random 1k が 3.2%が対応しており，Alexa の順位が高いほど普及が進んでいることが伺える．DNSSEC に関しては Top 1k が 2.2%，Random 1k が 2.9%が対応しているという状況であり，Web サイト全体として普及はあまり進んでいない．Top 1k において DNSCAA に対応している Web サイトのうち，29.7%は Google 社が提供する検索エンジンの Web サイトであった．これらの Web サイトのドメインに対しては，Google の自社向けの認証局のみが発行許可を与えられていた．

4.4 各分類別のセキュリティレベル

3.1 節の（1）のカテゴリ別に Web サイトを分類した際の，各分類における SSL/TLS 設定の評価結果を**表 5** に示す．表 5 では Top 1k と Random 1k のそれぞれにおいて，集計結果を行方向に正規化したものである．どちらの場合においても多くのカテゴリの 90%以上が A 評価となっており，適切な設定を行なっているとみられる．しかし，`Search Engine` における A 評価の割合は，Top 1k では 96.7%であるのに対して Random 1k では 83.3%と大きく下がっている．Top 1k における `Search Engine` の 72.9%は Google 社が提供する Web サイトであったため，Google 社が適切な設定を行っているとみられる．`Adult` カテゴリの Web サイトに注目してみると，Top 1k ではすべて A 評価であるのに対して Random 1k では 8.6%が F 評価となっている．これより，Alexa 上位の `Adult` カテゴリの Web サイトでは適切な SSL/TLS 設定を行っている傾向があることが伺える．`Finance` ではどちらの場合においても B 評価となったものは 7.7%, 8.3%と存在しているが，F 評価は存在しなかった．したがって，`Finance` のような重要な情報を取り扱う Web サイトでは Alexa の順位によらず最低限の脆弱性対策を行っていると考えられる．Shopping についても，Random 1k にて 1.0%のみ F 評価が存在するが，A 評価の割合はどちらの場合でも 96%を超えており，適切な設定を行っていることが伺える．

表 5　カテゴリ別の SSL/TLS 評価結果

	Grade							
	Top 1k				Random 1k			
Category	A	B	C	F	A	B	C	F
Adult websites	100%	0.0%	0.0%	0.0%	91.4%	0.0%	0.0%	8.6%
Company	95.9%	4.1%	0.0%	0.0%	94.3%	2.6%	0.0%	3.1%
Education	90.5%	9.5%	0.0%	0.0%	94.7%	2.6%	0.0%	2.6%
Finance	92.3%	7.7%	0.0%	0.0%	91.7%	8.3%	0.0%	0.0%
News	95.3%	4.7%	0.0%	0.0%	95.6%	0.0%	0.0%	4.4%
Search Engine	96.7%	3.3%	0.0%	0.0%	83.3%	0.0%	0.0%	16.7%
Shopping	96.8%	3.2%	0.0%	0.0%	96.2%	2.9%	0.0%	1.0%
Social Media	98.1%	1.9%	0.0%	0.0%	96.3%	0.0%	0.0%	3.7%
Underground	100%	0.0%	0.0%	0.0%	75.0%	25.0%	0.0%	0.0%
Other	88.0%	9.6%	2.4%	0.0%	92.2%	3.1%	1.0%	3.6%
N/A	95.7%	4.0%	0.0%	0.4%	94.3%	1.3%	0.0%	4.5%

表 6　(2) ログインの可否, (3) 用料金支払いの要否, (4) プライバシーポリシーの有無, (5) CDN サービスおよび (6) 認証レベルによる分類結果と SSL/TLS 設定評価, DNSCAA,DNSSEC,ZLint の調査結果との関係（それぞれのカイ二乗検定結果の p 値）

		Grade	DNSCAA	DNSSEC	ZLint
(2)	Top 1k	0.10	0.07	0.99	0.64
	Random 1k	0.96	0.95	0.78	0.96
(3)	Top 1k	0.97	0.50	0.18	0.14
	Random 1k	0.81	0.37	0.99	0.17
(4)	Top 1k	0.08	<0.01**	0.41	0.94
	Random 1k	0.17	0.12	0.87	0.84
(5)	Top 1k	0.09	<0.01**	0.12	0.28
	Random 1k	<0.01**	0.01*	0.93	<0.01**
(6)	Top 1k	0.71	<0.01**	0.01*	<0.01**
	Random 1k	<0.01**	<0.01**	0.31	<0.01**

$*$ は p 値が 0.05 未満 0.01 以上，$**$ は p 値が 0.01 未満.

表 7　SSL/TLS 設定評価結果と DNSCAA, DNSSEC, ZLint，および CDN サービスとの関係

	DNSCAA		DNSSEC		ZLint		CDN	
Grade	False	True	False	True	Error	Safe	False	True
A	94.5%	100.0%	94.9%	97.9%	78.5%	96.3%	93.4%	97.6%
B	3.4%	0.0%	3.1%	2.1%	6.4%	3.1%	3.9%	2.4%
C	0.2%	0.0%	0.2%	0.0%	0.6%	0.2%	0.3%	0.0%
F	1.8%	0.0%	1.7%	0.0%	14.5%	0.4%	2.4%	0.0%

各分類結果とこれまでの調査結果を総合的にみるため，カイ二乗検定による分析を行う．3.1 節の（2）から（6）による Web サイトの分類結果と，SSL/TLS 設定評価，DNSCAA，DNSSEC，ZLint との各組み合わせにおいてクロス集計をとった時のカイ二乗検定結果の p 値を計算することで，各分類結果との相関をみる．各 p 値の計算結果を**表 6** に示す．表 6 に示す通り，(2)，(3) による分類では目立った有意差は得ることができなかったが，(4)，(5) における分類では DNSCAA との有意差がみられた．DNSCAA を設定している Web サイトのうち 97%がプライバシーポリシーを掲載しており，DNSCAA を設定している Web サイトほど，プライバシーポリシーを掲載している傾向がみられた．対照的に，DNSCAA を設定している Web サイトのうち，いずれかの CDN サービスを用いているものは僅かに 12.2%であり，そのような Web サイトでは CDN サービスを用いていない傾向にあることが分かった．(7) による分類ではより多くの有意差のある結果が得られ，特に DNSCAA，ZLint の調査結果との有意差は強い．DNSCAA を設定しているのは OV 認証である傾向が強く，DNSCAA を設定している Web サイトのうち，85.2%が OV 認証であった．また，ZLint エラーは意外にも EV 認証のサーバ証明書，またはその中間証明書で発生する傾向にあった。DV 認証のものでは 3.7%，OV 認証では 9.1%のエラーが含まれていたが，EV 認証では 18.2%のエラーを含んでいた．

次に，SSL/TLS 設定評価結果と DNSCAA，DNSSEC，ZLint，および CDN との関係を**表 7** に示す．この表では Top 1k と Random 1k の結果を統合させたクロス集計表を，各項目ごとに列方向で正規化したものである．CDN の項目は CloudFlare，Akamai，Amazon CloudFront のいずれかを用いている Web サイトを対象とした．表 7 に示す通り，DNSCAA または DNSSEC を設定しているもの，ZLint にてエラーと出力されない証明書を用いている，いずれかの CDN サービスを使用している Web サイトの評価は非常に良い傾向にあることが分かる．特に，DNSCAA を設定している Web サイトはその全てが A 評価であり，上述した特徴が強いと考えられる．また，DNSCAA，DNSSEC を設定している Web サイトのうち，98.5%，90.2%が ZLint によるエラーのない証明書を用いていたことから，DNSCAA，DNSSEC は Web サイトのセキュリティレベルを見る上で重要な指標になると言える．

5. 議論

本章では本研究における制約，今後の課題，および分析により得られた知見について述べる．

5.1 SSL/TLS 設定

3.2 節で述べたように，本研究では Censys のデータを用いて対象の Web サイトの SSL/TLS 設定の評価を行った．Censys のデータでは確認できない脆弱性の項目があるため，本研究における評価手法で A と評価されたものが全て安全性が高いとは言えない．実際には RC4 の使用や ROBOT 攻撃，POODLE 攻撃などの脆弱性を有する Web サイトが存在することが示されており [11]，今後は Censys で確認できなかった項目も含めた詳細な調査を行う必要がある．

5.2 Web サイトのカテゴリ別の分類

本研究ではクラウドソーシングによって Web サイトのカテゴリ分類を自動化することでカテゴリに基づく大規模調査を実施できたが，約 2 割強の Web サイトはカテゴリが N/A となり適切に分類を行えなかった．クラウドソーシングではワーカーの技量によって結果が大きく左右されるが，このような結果となった原因として，どのカテゴリに分類すべきか判断に迷うサイトが多くあったことが考えられる．例えば，企業がニュースや教育，ショッピングに関する様々な情報を提供するポータルサイトなどが判断に迷うサイトとして挙げられる．また，N/A となったサイトのうち，Top1 k では 22.2%，Random 1k では 44.4%が英語ではない言語で書かれていた．したがって，Top 1k では上述した判断に迷う Web サイトが多くあったことにより N/A が増加してしまい，言語による分類の判断の難しさへの影響は低かったと考えられる．一方で，Random 1k の外国語で書かれた Web サイトについては，たとえブラウザの機能を用いて翻訳したとしても，見慣れない Web サイトのため多くのものは判断に迷ってしまったと思われる．

複数の評価者による分類の一致度を示す指標である Fleiss' kappa におけるカッパ係数を計算すると，Top 1k では 0.29，Random 1k では 0.33 となり，0.6 以上でないと一致度が高いとは言えないため，このようなクラウドソーシングによる Web サイトの分類は難しいタスクであることが分かる [15]．

DNSCAA や DNSSEC などの対応しているデータ数が少ないものでは，各カテゴリの分類におけるデータ数が少なくなってしまい，精度の高い解析を行うことができなかった．より精度を高めるためにもより多くの Web サイトを，ワーカーによって判断が分かれることのないように分類する手法を考案することが重要である．

5.3 分析から得られた知見

4.4節における分析結果から，DNSCAA またはDNSSECへ対応している Web サイトは，そのセキュリティレベルを見る上で重要な指標となることを示した．DNSCAA は自身のドメインに対する証明書への誤発行を防ぐ仕組みであり，DNSSEC は DNS のセキュリティを確保する仕組みであることから，SSL/TLS 設定との関係は必ずしも高くはない．したがって，今回の分析でこのような結果が得られたことは，DNS 設定にまで注意を払っている Web サイトのセキュリティレベルの高さを示している．

6. 関連研究

本研究の実施した調査に関しては多数の関連研究がある．以下では紙面の都合上，SSL/TLS および DNSCAA に絞って関連研究を示す．

6.1 SSL/TLS

Kumar [5] らは，認証局が発行する証明書が CA/BF Baseline Requirement と RFC5280 に準拠しているかを検査するツールである ZLint を開発し，それを用いてインターネット上で証明書に関する大規模な調査を行った．調査結果によると，ZLint によるエラーを含む証明書は小規模な認証局，または不正発行が多いと指摘される認証局が多い傾向にあることが判明した．また，それらは中間認証局が原因であることが多いことも明らかにした．

6.2 DNSCAA

Scheitle [10] らは，DNSCAA の運用に関して各ドメイン管理者，認証局が適切に運用しているかを調査した結果を示している．調査結果によると，適切に運用していないものが少なからず存在しており，また，CAA レコードの設定ができないドメインが約 44%ほど存在していることも課題として述べている．とはいえ，DNSCAA を設定しているドメインの DNSSEC への対応率は 12%と，一般的な DNSSEC 対応率に比べて高いことからも，当論文では DNSCAA への対応は適切に行った上で，セキュリティのベストプラクティスとして推奨されている．

7. まとめ

本研究では Web サイトのセキュリティ・センサスと称し，Web サイトにおけるセキュリティ対策の実践状況を総合的に分析した．著者らの知る限り，このような総合的な分析に基づく Web サイトセキュリティの研究は初の試みである．この結果，Alexa 順位の高い `Search Engine` や `Adult` カテゴリでは SSL/TLS を適切に設定する傾向が高いことを明らかにした．また，SSL/TLS の設定状況は，DNSCAA または DNSSEC の設定状況，証明書の正当性，CDN サービスの利用有無と高い相関を持つことがわかった．一般に，DNS 設定にまで注意を払っている Web サイトのセキュリティレベルは高い傾向があることも明らかにした．セキュリティ・センサスの内容をさらに拡充すること，また問題が発見された場合の具体的なアクションの示唆も含めたガイドラインの提供・実践は今後の課題である．

参考文献

[1] Google: Moving towards a more secure web (2016).

[2] : The Heartbleed Bug, `http://heartbleed.com/` (2014).

[3] Möller, B., Duong, T. and Krzysztof, K.: This POODLE Bites: Exploiting The SSL 3.0 Fallback, *Security Advisory, Google*, (online), available from ⟨https://www.openssl.org/ bodo/ssl-poodle.pdf⟩ (2014).

[4] : The FREAK Attack, `https://censys.io/blog/freak` (2015).

[5] Kumar, D., Wang, Z., Hyder, M., Dickinson, J., Beck, G., Adrian, D., Mason, J., Durumeric, Z., Halderman, J. A. and Bailey, M.: Tracking Certificate Misissuance in the Wild, *Proc. of IEEE Symposium on Security and Privacy 2018*, pp. 785–798 (2018).

[6] Durumeric, Z., Kasten, J., Bailey, M. and Halderman, J. A.: Analysis of the HTTPS certificate ecosystem, *Proceedings of the 2013 Internet Measurement Conference, IMC 2013, Barcelona, Spain, October 23-25, 2013*, pp. 291–304 (online), DOI: 10.1145/2504730.2504755 (2013).

[7] Aviram, N., Schinzel, S., Somorovsky, J., Heninger, N., Dankel, M., Steube, J., Valenta, L., Adrian, D., Halderman, J. A., Dukhovni, V., Käsper, E., Cohney, S., Engels, S., Paar, C. and Shavitt, Y.: DROWN: Breaking TLS Using SSLv2, *25th USENIX Security Symposium, USENIX Security 16, Austin, TX, USA, August 10-12, 2016.*, pp. 689–706 (2016).

[8] AWS: Alexa Top Sites, `https://aws.amazon.com/alexa-top-sites/`.

[9] Durumeric, Z., Adrian, D., Mirian, A., Bailey, M. and Halderman, J. A.: A Search Engine Backed by Internet-Wide Scanning, *22nd ACM Conference on Computer and Communications Security* (2015).

[10] Scheitle, Q., Chung, T., Hiller, J., Gasser, O., Naab, J., van Rijswijk-Deij, R., Hohlfeld, O., Holz, R., Choffnes, D., Mislove, A. and Carle, G.: A First Look at Certification Authority Authorization (CAA), *SIGCOMM Comput. Commun. Rev.*, Vol. 48, No. 2, pp. 10–23 (2018).

[11] Qualys: SSL Labs SSL Pulse, `https://www.ssllabs.com/ssl-pulse/` (2018).

[12] Qualys: SSL Server Rating Guide, `https://github.com/ssllabs/research/wiki/SSL-Server-Rating-Guide` (2017).

[13] 情報処理推進機構（IPA）：SSL/TLS 暗号設定ガイドライン～安全な Web サイトのために（暗号設定対策編）～, `https://www.ipa.go.jp/security/vuln/ssl_crypt_config.html` (2015).

[14] CloudFlare: Cloudflare Pricing, `https://www.cloudflare.com/plans/` (2018).

[15] Landis, J. R. and Koch, G. G.: The Measurement of Observer Agreement for Categorical Data, *Biometrics*, Vol. 33, No. 1 (1977).