

CDNのキャッシュサーバを騙ったDDoS攻撃 のZスコア法と二段階検知による防御法

立命館大学

谷口和也 上山憲昭

研究背景(CDNの普及)

■ CDN:

インターネット上でコンテンツを高速かつ効率的に配信するための分散型ネットワーク

■ CDNのシェア率: 年々増加傾向

■ CDNの市場規模の成長予測:

2024年: 199.6億ドル → 2029年: 424.6億ドルまで成長

■ CDN普及の原因

- オンラインコンテンツの需要の増加
- モバイルインターネットの普及
- グローバルなアクセス需要
- CDNのセキュリティ機能を利用可能

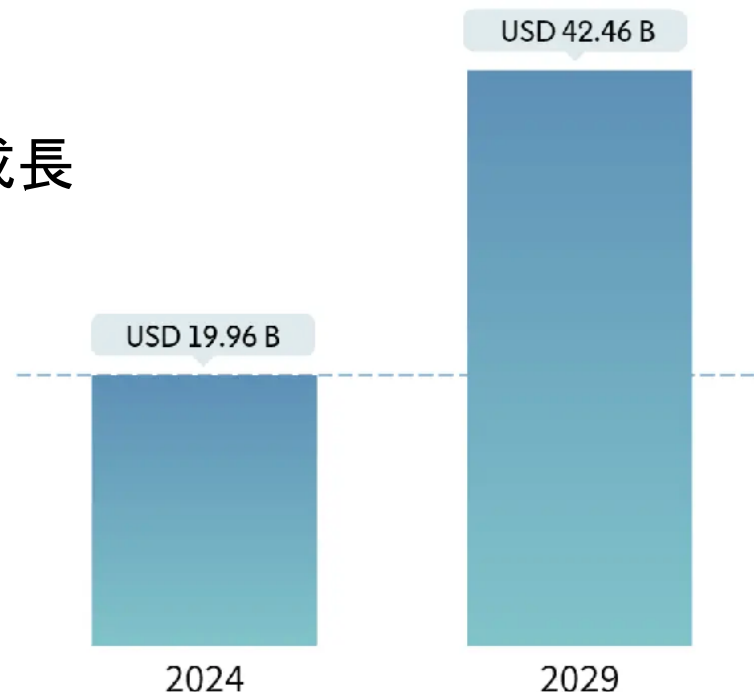


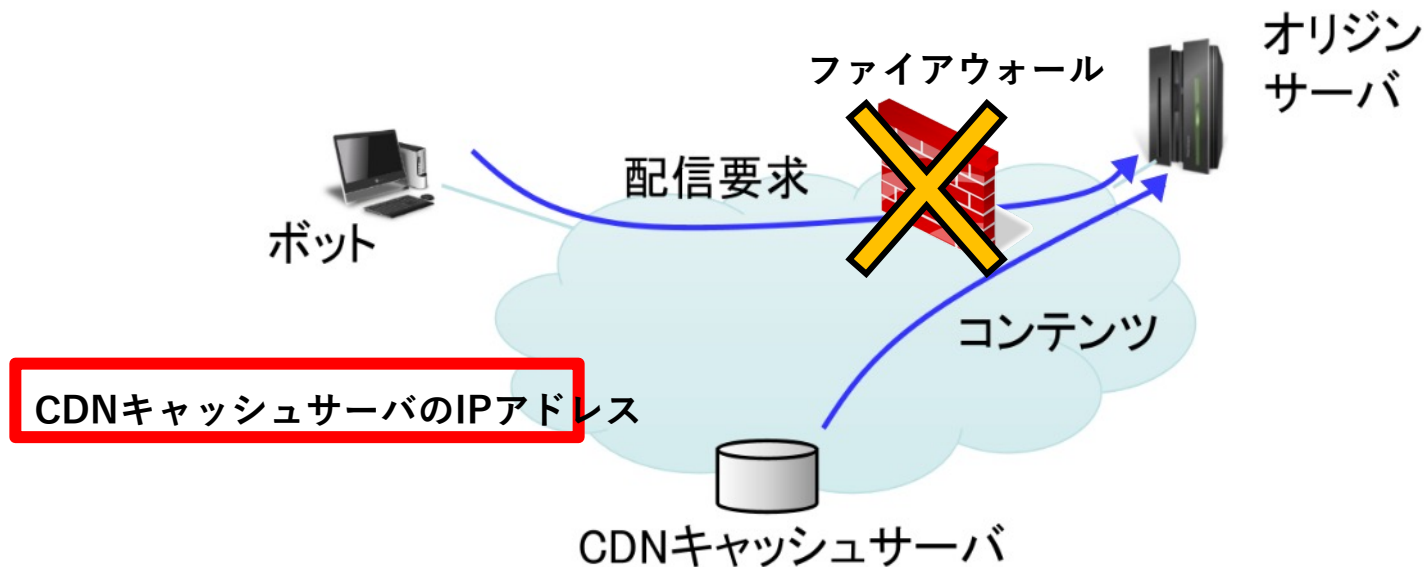
図: コンテンツ配信ネットワークの市場規模と市場規模株式分析 - 成長傾向と成長傾向予測 (2024~2029 年)(Source:mordorintelligence)

研究背景(DDoS攻撃の頻繁化)

- DDoS攻撃の頻繁化
 - 攻撃手法の拡散
 - IoTデバイスの脆弱性
- OS(Origin Server)への直接攻撃の可能性
 - OSのIPアドレスを取得する方式
 - DNSレコードの悪用: 特定のドメインに関連するIPアドレスを取

研究課題

- 攻撃者が**標的サーバのIPアドレスを特定**
 - DNSを用いずに、特定したIPアドレス宛に直接パケットを送信
 - ⇒ **DDoS攻撃が成立**
 - CDNキャッシュサーバ以外から到着したパケットは**ファイアウォールで遮断**
- 攻撃者が**CDNキャッシュサーバのIPアドレスを偽アドレスとして偽った場合**
 - ⇒ DDoSパケットを送ると**ファイアウォールで検知不可**



DNSの名前解決処理のログの活用

■ 検知方法

- CSからの正常な問い合わせ時:
 - コンテンツプロバイダのDNSサーバに名前解決のログが残る
- ボットからのOSのIPアドレスを直接用いた要求:
 - DNSの名前解決を用いないため名前解決の履歴が残らない

■ DNSの名前解決処理のログの確認により, 攻撃を検知

- 問い合わせが有り: 配信処理を実施
- 問い合わせが無し: DDoS攻撃と判断し,アクセスを棄却

⇒ **全ての要求に対し, DNSのログを確認するとオリジンサーバの負荷が増大**

研究目的

■ 要求発生パターンの違い

- 正常な配信要求: キャッシュミス時に発生
- DDoS攻撃: 短い時間間隔で膨大な数が連続して発生

⇒ 到着間隔に閾値を設定し, 攻撃を検知

問題点: 単一閾値を用いると見逃し, 誤検知が多い

■ Zスコアの活用

- リアルタイムで外れ値検知が可能
- 動的に閾値を修正可能

二段階検知

- 課題の攻撃方式に対し、DNSの名前解決ログの確認により検知

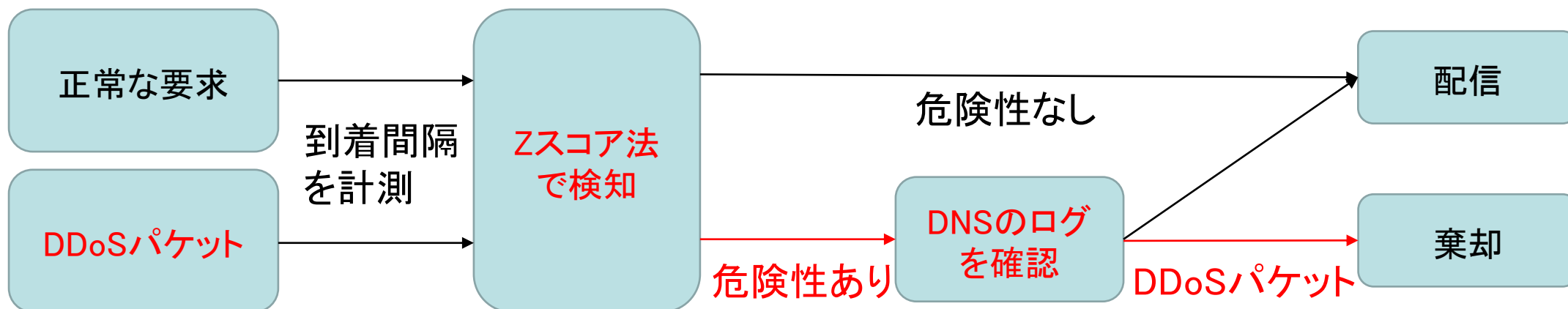
⇒全要求に対し確認を行うと処理負荷の懸念

- オリジンサーバの負荷を軽減する方法を検討

⇒二段階検知を実施

- 要求発生パターンの違いから危険性のある要求を到着間隔より検知

⇒閾値を動的に変更するため、Zスコアを導入



Zスコア法

- データの外れ値を検知するアルゴリズム
- メカニズム
 - 過去のデータより平均を算出
 - 平均から外れ値を検知
 - 外れ値を平均に基づいて編集しデータとして記録
- 3つのパラメタ
 - ラグ(L): 平均値を作成する過去のデータの個数
 - 閾値(η): 信号を検知する際の感度
 - 影響(α): 値を更新する際の過去のデータの影響度

$$S_i = \begin{cases} 1, & \text{if } E_c - \mu_{i-1} > \eta\sigma_{i-1} \\ -1, & \text{if } E_c - \mu_{i-1} < -\eta\sigma_{i-1} \\ 0, & \text{otherwise} \end{cases}$$

$$E_i = \begin{cases} E_c, & \text{if } S_i = 0 \\ \alpha \times E_c + (1 - \alpha) \times E_{i-1}, & \text{otherwise} \end{cases}$$

$$\mu_i = \text{mean}(E_{i-L+1}, E_{i-L+2}, \dots, E_i)$$

$$\sigma_i = \text{std}(E_{i-L+1}, E_{i-L+2}, \dots, E_i)$$

Zスコア法の適応

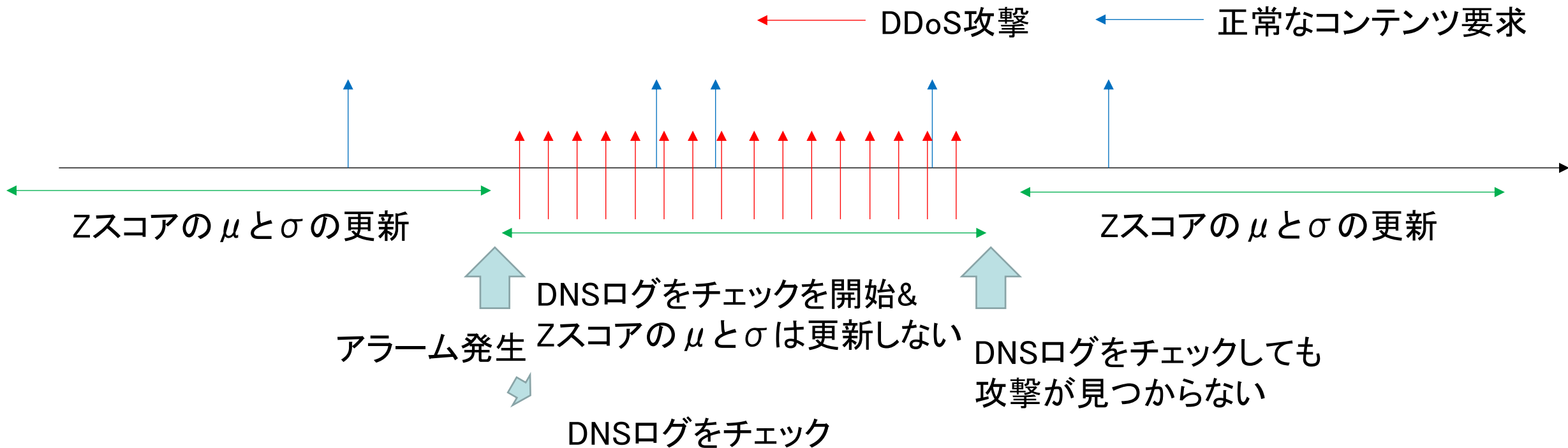
■ Zスコア導入での検知法

■ 到着間隔の逆数をデータとして活用

→ 到着間隔の違いよりDDoSパケットが発生したことを判別

■ 攻撃発生中の誤ったデータを平均に反映しない

→ 正常な要求のみが起こっている時とは異なった到着間隔のパターンを検知



オーダ表記での評価

■ DNSログの検索処理の時間計算量

- 線形探索であると想定
- エントリ数 n に対して最悪時間計算量は $O(n)$

■ Zスコアアルゴリズムの時間計算量

- $Z = (X - \mu) / \sigma$ (Z はZスコア、 X は入力値、 μ (ミュー)は平均、 σ (シグマ)は標準偏差)
- 単純な式であるため、平均値と標準偏差を事前に計算することで瞬時に結果を出力可能
- **Zスコアアルゴリズムの時間計算量は入力数に依存しない**
- 時間計算量は $O(1)$

⇒ **Zスコア法とDNSのログ検索の計算量の違いにより処理コストの低減が可能**

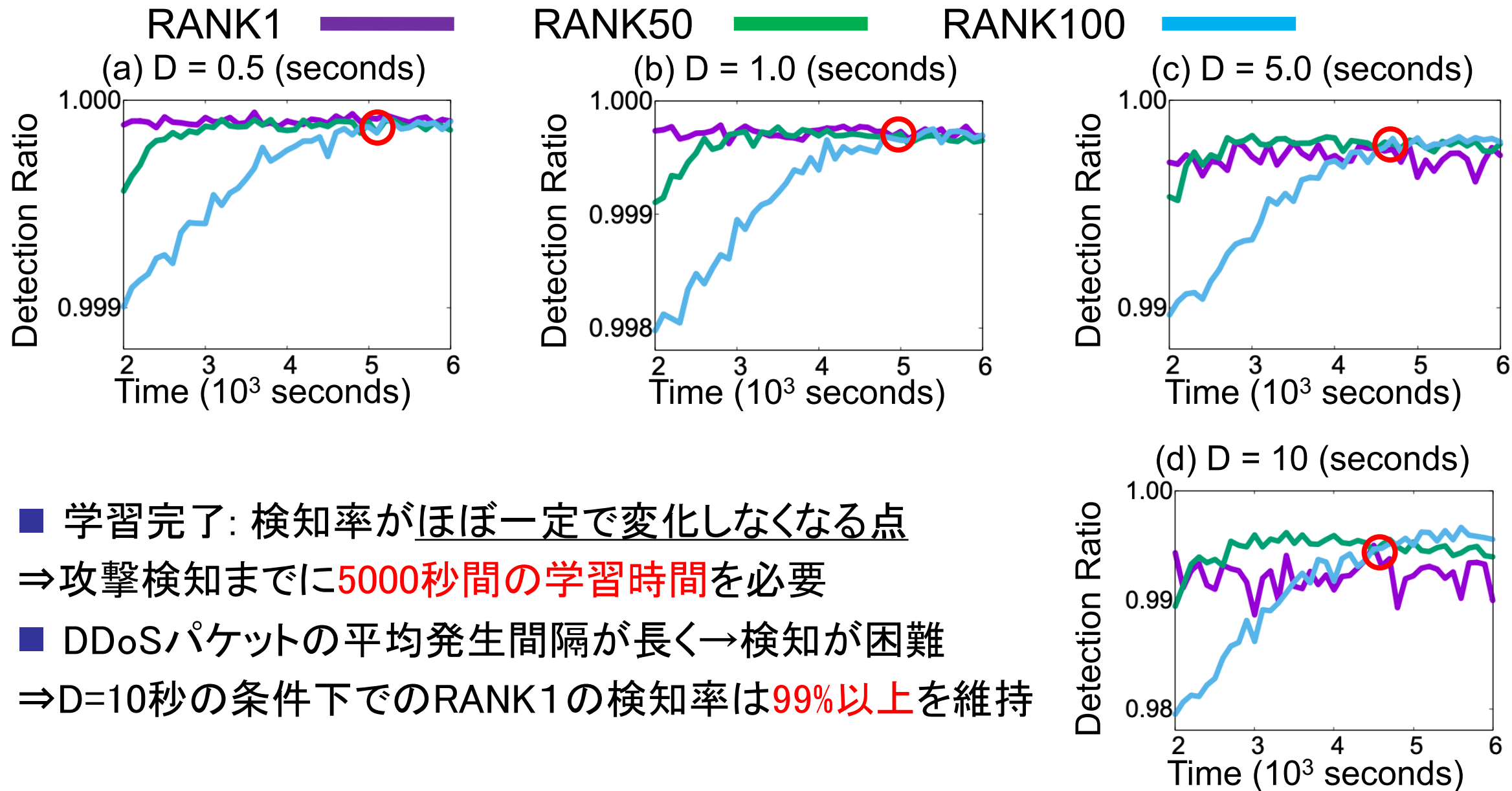
シミュレーション条件

- Zスコアのパラメタ: ラグ(L): 10, 閾値(η): 4.0, 影響(α): 0.5
- コンテンツ要求のシミュレータの設定

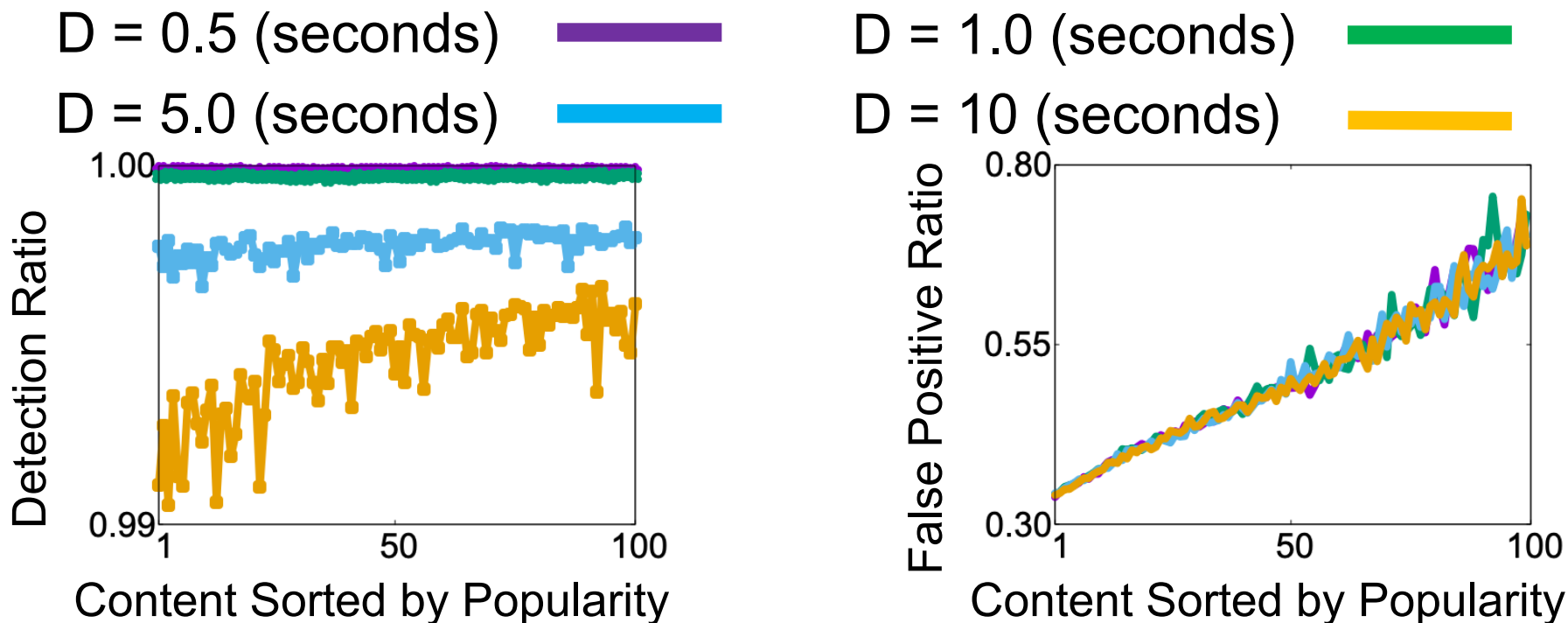
項目	設定した値
シミュレーション時間	10000秒
コンテンツの種類	100個
キャッシュ容量	10個
要求の平均発生回数	1回
攻撃を受ける期間	5000秒経過後3000秒間

- 評価項目
 - 検知率の時間変化
 - 人気度の異なるコンテンツに対する評価
 - DDoSパケットの到着レートに対する評価

結果(検知率の時間変化)



結果(人気度の異なるコンテンツに対する評価)



■ 到着間隔に着目しているため、高人気コンテンツの方が検知困難と推測

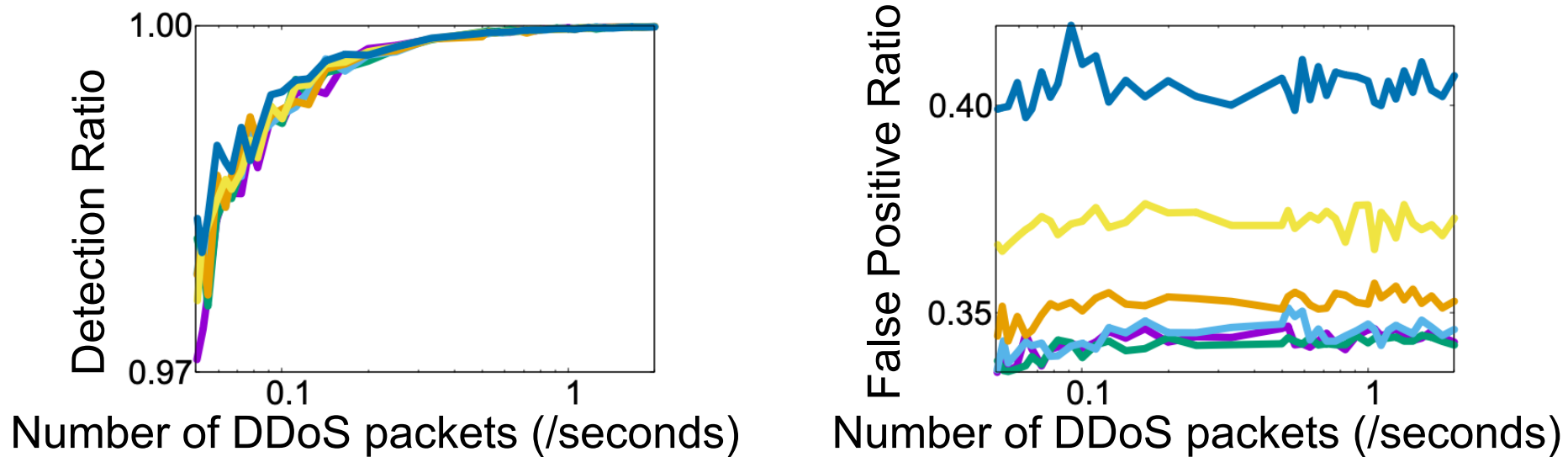
⇒人気度に関わらず**0.99以上**の高い検知率を記録

■ 誤検知率: 30～80%

⇒DNSのログを確認することで、**CSからの正常な要求パケットを棄却することは防げる**

結果(DDoSパケットの到着レートに対する評価①)

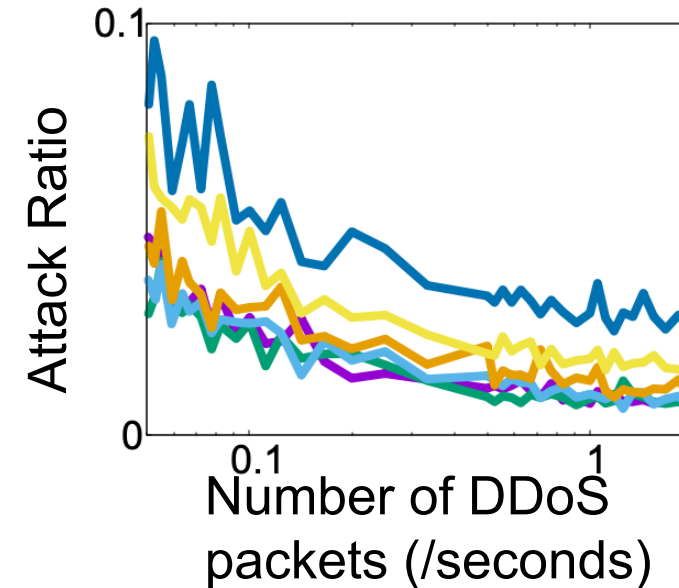
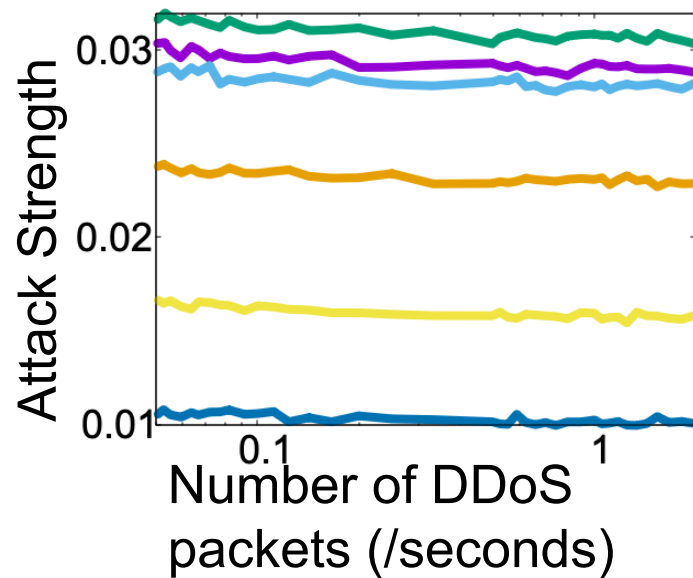
RANK1 RANK2 RANK3
RANK5 RANK10 RANK20



- 検知率: DDoSパケットの到着レートが高い方が検知が容易
⇒ 検知レートの増加に伴い検知率は増加し, 検知率は**ほぼ100%**
 - 誤検知率: DDoSの発生レートに依存しない
 - 攻撃対象コンテンツの人気度により値が異なる
- ⇒ ランダムで発生する要求の中には一定数の短い間隔の要求が存在

結果(DDoSパケットの到着レートに対する評価②)

RANK1 RANK2 RANK3
RANK5 RANK10 RANK20



■ 攻撃強度: 単位時間あたりの配信回数

⇒DDoSパケットを棄却→キャッシュミスの発生間隔を反映した結果に

■ 攻撃率: 配信した要求に含まれるDDoSパケットの割合

■ 配信要求中にDDoSパケットが含まれる割合が低い

⇒DDoSパケットの検知の見逃しは少なく, ボリューム攻撃の危険性を大幅に低下

まとめ

- CDNキャッシュサーバのIPアドレスを騙ったOSへの直接攻撃を想定

⇒ファイアウォールでの検知が困難

- Zスコア法を用いて動的にDNSログを確認する必要がある要求を絞る検知方式を提案

- 正常パケットとDDoSパケットの発生パターンの違いから発生する到着間隔の差異に着目

⇒Zスコア法とDNSのログ検索の計算量の違いにより処理コストの低減が可能

- 提案方式によるDDoSパケットの検知精度は高いことを確認

- 今後の方針

- 送信元IPアドレスの種類に着目することで、危険性のあるCDNキャッシュサーバを絞りこむ方式を実現

⇒多大なCDNキャッシュサーバ全てに提案方式を適応する必要がなく、検知を行うコストを低減可能と予想