Vol.2020-ICS-200 No.8
2020/9/16

マルチエージェント技術に基づく シミュレーション機構を用いた通信先自動匿名化手法の試作

杉山 慶多[1,a)]　福田 直樹[2,b)]

概要：本論文では，ネットワーク仮想化に対するマルチエージェント技術の応用の一つとして，マルチエージェント強化学習を利用した，シミュレーション機構を用いた通信先自動匿名化手法の試作について述べる．具体的なシナリオの例として，小規模な大学のネットワークの通信の通信先を匿名化することを考え，ネットワークの利用者団体の情報学部と事務のそれぞれが我々の通信先自動匿名化機構を使用して匿名化を行う状況を想定し，オンライン学習で学習を進める場合の学習の収束の速度を確認し，シミュレーションによるオフライン学習を繰り返して学習を進める場合についての議論を加える．

1. はじめに

ゲーム理論やマルチエージェント強化学習などのマルチエージェント技術を用いて，セキュリティリソースを効果的に配分する課題を扱う研究は，これまでにいくつもなされてきており，具体的には，密猟対策における資源配分 [1], [2], [3] や，その分野で成功しているモデルのサイバーセキュリティにおける資源配分への応用 [4], [5], [6] などがある．

我々もこれまでに，ネットワークの通信の通信先を匿名化する課題を，ゲーム理論やマルチエージェント強化学習などのマルチエージェント技術を用いて解決することを目指してきた．通信先の匿名性を満たすことの必要性については，通信先匿名化手法の一つである U-TRI [7] を提案している Wang らもその論文内で指摘している．なぜなら，セキュリティを保護するためにネットワーク層の IPsec などで適切に暗号化がなされていても，Wiretap [8] やスイッチの脆弱性 [9], [10] を利用したパケットの盗聴により，データリンク層の物理アドレスについて情報を攻撃者に収集されてしまうと，どのエンドホストが起動しているかの調査や重要なエンドホストの物理アドレスの特定がなされてしまうためである．このような攻撃は，機密性の高い情報を取り扱う企業の企業内ネットワークシステムにとって大きな脅威となる可能性がある．

Wang らの U-TRI はそれを目的とするアプローチの一つとして提案されているものである．U-TRI は，Software Defined Network の技術を用いて，Moving Target Defense [11] のアイディアに基づき通信に用いる物理アドレスを一定間隔で書き換えることで，匿名性を満たす手法である [7].

我々のこれまでの研究 [12], [13], [14], [15] では，U-TRI を拡張して，対象のネットワークを盗聴やトラフィック分析などの方法で攻撃をしてくる敵対者を想定し，そのような敵対者をモデルに組み込むことで，効果的な通信先の匿名化リソースの配分を行うことに取り組んでいた．しかし，次のようなボトルネックが存在していた．第一に，通信先匿名化手法の一つである U-TRI [7] を提案している Wang らもその論文内で指摘しているように，盗聴やトラフィック分析などの攻撃は受動的であるため，その検出がとても困難であり，利用者側がどこがどのタイミングで盗聴されるのかを予想して決めなくてはならなかったことがある．第二に，能動的な攻撃を確認できたとしても，実際の攻撃を複数回受けないことには，敵対者の正確なモデルの構築が困難であることがある．

そこで，本論文では，攻撃をしてくる敵対者について知らなくても，攻撃は受けるものとして最大限に通信先を匿名化をするという考えを基として通信先の匿名化をすることを考える．ただしその場合，ネットワークの利用者団体（チーム・部署など）が限られたネットワークであれば，利用者団体のネットワーク管理者が全体を見てパラメータを調節すれば良いが，ネットワークの利用者団体が複数存在するようなネットワークでは，とりわけ，パブリッククラ

[1] 静岡大学大学院総合科学技術研究科情報学専攻
Department of Informatics, Graduate School of Integrated Science and Technology, Shizuoka University
[2] 静岡大学学術院情報学領域
College of Informatics, Academic Institute, Shizuoka University
[a)] sugiyama.keita.15@shizuoka.ac.jp
[b)] fukuta@inf.shizuoka.ac.jp

ⓒ 2020 Information Processing Society of Japan

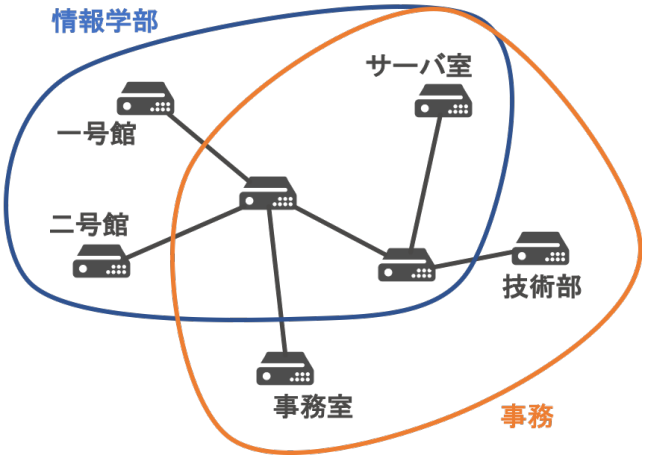

図 1 対象とするネットワークの概観

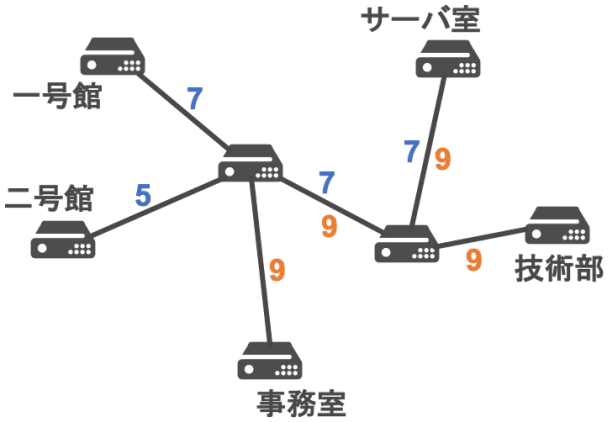

図 2 それぞれの利用者団体の要求する匿名化のレベル

ウドのように利害関係が独立している利用者団体が同じネットワークを物理的に共有している場合には，他の利用者団体の業務を妨害しないように通信先の匿名化のためのリソースを配分することが必要となる可能性がある．本論文では，そのような可能性がある場合に対し，マルチエージェント強化学習を利用して，他の利用者団体の業務を妨害しない程度に自動的に通信先の匿名化のためのリソースを配分して対応する手法の試作について述べる．

2. 具体的なシナリオの例

具体的なシナリオの例として，小規模な大学のネットワークの通信の通信先を匿名化することを考える．対象とするネットワークの概観を，図 1 に示す．このネットワークの利用者団体は，主に情報学部と事務の 2 団体である．

情報学部は，講義や研究で一号館および二号館からサーバ室のサーバに接続することがあり，事務は，業務で事務室および技術部からサーバ室のサーバに接続することがある．サーバ室に接続するには，いくつかの L2 スイッチを経由する場合があり，それらの L2 スイッチは，情報学部と事務が共同で管理している．

今回，情報学部は，講義や研究で使用する部分ネットワーク（青線で囲われている範囲）の，事務は，業務で使用する部分ネットワーク（橙線で囲われている範囲）の通信の通信先を匿名化することを考えており，それぞれが独立して我々の通信先自動匿名化機構を使用して，通信の通信先を匿名化することに決めた．

我々の通信先自動匿名化機構は，すべてのリンクの通信の通信先を同程度のレベルで匿名化するのでなく，パラメータとしてリンクごとの匿名化のレベル（0 から 9）を受け取り，それらを基に通信の通信先を匿名化する．情報学部と事務のネットワーク管理者は，自団体の業務の内容と照らして，図 2 のように匿名化のレベルをパラメータとして設定した．

情報学部は，研究室が集中している一号館からサーバへの通信については，通信の内容の多くが研究に関わる重要なものであると判断し，匿名化のレベルを比較的高く設定しており，講義室が集中している二号館からサーバへの通信については，通信の内容は機密性がそれほど高くないと判断し，匿名化のレベルを比較的低く設定している．事務は，事務室からサーバへの通信については，通信の内容に大学や学生に関する機密性が非常に高い情報が含まれると判断し，匿名化のレベルを最大に設定しており，技術部からサーバ室への通信についても，通信の内容にセキュリティに関する機密性の高い情報が含まれると判断し，匿名化のレベルを最大に設定している．

我々の通信先自動匿名化機構は，これらの提示された通信の通信先匿名化要件を，次の節で示すような制約の中で最大限満たすように努める．

3. 制約

我々の通信先自動匿名化機構は，以下のような制約の中で通信先の匿名化を行う．

(1) 通信先の匿名化とネットワークのパフォーマンスは，トレードオフの関係にある．
(2) 他のネットワーク利用者団体の業務を著しく妨害していると判断される場合には，ペナルティが課せられる．
(3) ただし，他のネットワーク利用者団体が運用しているシステムや発生するトラフィックについては，あらかじめ完全な情報を知ることはできず，ペナルティを受ける中で予測しなくてはならない．

制約 1 で述べているトレードオフの関係は，以下の式 1 で表される．

$$p = 1 - \alpha k^2, \quad p \in [0,1], \quad k \in [0,1] \tag{1}$$

ここで，p はネットワークのパフォーマンスを，k は通信先の匿名化のレベルを，α は運用するシステムのトラフィックの性質によって決定される係数を表している．実際に，U-TRI では，hid と呼ばれるアドレスの更新間隔を 165 秒から 105 秒，45 秒と短くすると，Web ページのダウン

ⓒ 2020 Information Processing Society of Japan

ロード時間が 0.028 秒から 0.033 秒，0.041 秒と長くなっていることが実験により示されている [7].

制約 2 における他のネットワーク利用者団体の業務を著しく妨害していると判断される場合とは，自身のネットワーク利用者団体の通信先の匿名化のレベル k_{own} の設定が，他のネットワーク利用者団体 i の業務におけるネットワークのパフォーマンス p_i を，閾値 $\theta \in [0,1]$ より下げてしまう原因となっている状況である．

制約 3 は，他のネットワーク利用者団体 i が運用するシステムのトラフィックの性質を表す α_i の値を，あらかじめ知ることができないことを意味している．

上記のような制約の中で，我々の通信先自動匿名化機構は，次の節で述べるアプローチに基づき，自動的に適切な通信先の匿名化を行う．

4. 本研究のアプローチ

我々の通信先自動匿名化機構は，他のネットワーク利用者団体が運用しているシステムや発生するトラフィックに関する情報を信念として，マルチエージェント強化学習を行うことで自動的に適切な通信先の匿名化を行う．

強化学習は，類似の状態に対する行動価値関数を同時に更新することで，学習の収束を早める手法の一つである QS-Learning [16] を使用する（我々の過去の研究 [17] では，エージェントのパトロール行動の獲得に QS-Learning を用いることで，その効果を確認している）．今回は，マルコフ決定過程における状態に現在の通信先の匿名化のレベルの設定状況を，行動に通信先の匿名化のレベルの上げ下げを対応させることとする．対象とするリンクの数を N，各リンクに設定できる匿名化のレベルの集合を $L \in \{0,1,\ldots,9\}$ とすると，状態の集合 S は，L^N で与えられる．

本論文では，この学習設定に基づきオンライン学習をするアプローチについてその学習の収束の速度を確認し，それを踏まえて，シミュレーションによるオフライン学習を繰り返すアプローチについて議論を加える．

4.1 オンライン学習によるアプローチ

オンライン学習によるアプローチでは，各ネットワーク利用者団体のネットワーク管理者から入力として受け取った通信先匿名化要件を基に，実際のネットワークでペナルティを受けながら探索をして学習を進めていく．

図 3 は，情報学部のネットワーク管理者による入力を JSON 形式にしたものの例である．

target-links は，通信先の匿名化を行う対象となるリンクをリスト化したものであり，requirement-levels は，各対象に対する通信先の匿名化のレベルの要求をリスト化したものである．この場合は，リンク a にレベル 7 を，リンク b にレベル 5 を要求している．

```
{
  "target-links": ["a", "b", "d", "e"],
  "requirement-levels": [7, 5, 7, 7]
}
```

図 3 オンライン学習の場合の入力例

```
[
  ...
  {
    "state": [7, 5, 7, 7],
    "action": [0, 0, -1, 0]
  },
  {
    "state": [7, 5, 7, 8],
    "action": [0, 0, 0, -1]
  },
  {
    "state": [7, 5, 7, 9],
    "action": [0, 0, 0, -1]
  },
  {
    "state": [7, 5, 8, 0],
    "action": [0, 0, -1, 0]
  },
  ...
]
```

図 4 現在の状態における最適と思われる行動の出力例

試作した通信先自動匿名化機構は，通信先の匿名化の方法として，現在の状態（通信先の匿名化のレベルの設定状況）における最適と思われる行動（通信先の匿名化のレベルの上げ下げ）を出力し，今回の学習結果として，Q-Table を出力する．図 4 は，情報学部が我々の通信先自動匿名化機構を利用した場合の，現在の状態における最適と思われる行動を JSON 形式で出力したものの例である．

一つ目の state および action は，現在の通信先の匿名化のレベルの設定状況が $[7,5,7,7]$ である場合には，3 番目のリンクの通信先の匿名化のレベルを一つ下げて，図 5 のように $[7,5,6,7]$ に設定を変更するべきであることを表している．他の state および action も同様に読み取ることができる．

図 6 は，情報学部が我々の通信先自動匿名化機構を利用した場合の，Q-Table を JSON 形式で出力したものの例である．

一つ目の state および action，value は，通信先の匿名化のレベルの設定状況が $[7,5,7,7]$ である場合の，3 番目のリンクの通信先の匿名化のレベルを一つ下げる行動の評価値は，0.7 であることを表しており，二つ目の state および action，value は，通信先の匿名化のレベルの設定状況が $[7,5,7,7]$ である場合の，4 番目のリンクの通信先の匿名化のレベルを一つ下げる行動の評価値は，0.1 である

ⓒ 2020 Information Processing Society of Japan

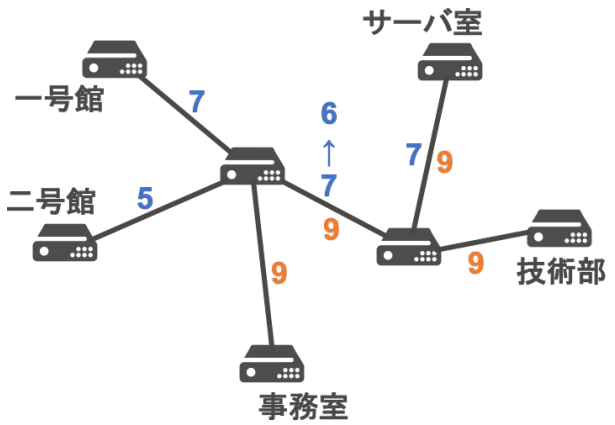

図 5 通信先の匿名化のレベルが変化する例

```
[
  ...
  {
    "state": [7, 5, 7, 7],
    "action": [0, 0, -1, 0],
    "value": 0.7
  },
  {
    "state": [7, 5, 7, 7],
    "action": [0, 0, 0, -1],
    "value": 0.1
  },
  {
    "state": [7, 5, 7, 7],
    "action": [0, 0, 0, 0],
    "value": 0.05
  },
  {
    "state": [7, 5, 7, 7],
    "action": [1, 0, 0, 0],
    "value": 0.02
  },
  ...
]
```

図 6 Q-Table の出力例

ことを表している．このことから，4 番目のリンクの通信先の匿名化のレベルを一つ下げるよりも，3 番目のリンクの通信先の匿名化のレベルを一つ下げる方が好ましいことがわかる．他の state および action，value も同様に読み取ることができる．

4.2 シミュレーションによる学習の収束の様子の確認

シミュレーションで学習の収束の様子を確認したところ，情報学部の通信先自動匿名化機構の学習の収束の様子は図 7 のように，事務の通信先自動匿名化機構の学習の収束の様子は図 8 のようになった．

1 エピソードあたりの学習のステップ数は 1000 ステップである．なお，シミュレータの実装言語は Scala 2.12.12 である．この結果から，オンライン学習で学習を進めると，今回のような小規模なネットワークでも，学習を収束させるまでに多くのペナルティを受けてしまう可能性があることがわかる．一方で，十分に学習が進んだ状態からであれば，その後は大きなペナルティはほとんどなく学習を継続できていると考えられる．

5. 議論

前節のシミュレーション結果から，オンライン学習以外の方法で学習を進めることを検討する必要がある．ここでは，シミュレーションによるオフライン学習を繰り返すアプローチについて考える．

5.1 シミュレーションによるオフライン学習を繰り返すアプローチ

シミュレーションによるオフライン学習を繰り返すアプローチでは，実際のネットワークでペナルティを受けたところで，ペナルティを受けた時の状況を基にシミュレーションによる追加の学習を行い，その結果を実際のネットワークに反映することを繰り返すことで学習を進めていく．

図 9 は，情報学部のネットワーク管理者による入力を JSON 形式にしたものの例である．

target-links および requirement-levels は，オンライン学習と同様のものである．オンライン学習の場合の入力と違う点は，過去にペナルティを受けたときに実際に設定されていた通信先の匿名化のレベルを，penalty-lines として指定する必要がある点である．この penalty-lines を基に，他のネットワークの利用者団体の運用しているシステムや発生するトラフィックに関する信念が更新され学習が進められる．シミュレーションするネットワークの設定と，前回の学習結果（Q-Table）は，別ファイルで入力される．

ただし，他のネットワークの利用者団体の運用しているシステムや発生するトラフィックに関する信念が状態として増えてしまうことが，学習の収束を遅らせてしまう要因になる可能性がある．この課題を解決する目的で，マルチエージェントな状況下で他エージェントに関する信念について効果的に学習する手法として提案されている，Bayesian Action Decoder [18] を QS-Learning と組み合わせて利用することの検討は価値があるかもしれない．

6. おわりに

本論文では，ネットワーク仮想化に対するマルチエージェント技術の応用の一つとして，マルチエージェント強化学習を利用した，シミュレーション機構を用いた通信先自動匿名化手法の試作について述べた．

我々のこれまでの研究では，ネットワークにおける通信の通信先を効果的に匿名化する目的で，通信先の匿名化のためのリソースを配分する課題のモデルに，対象のネットワークを盗聴やトラフィック分析などの方法で攻撃をしてくる敵対者を組み込んでいた．一方，今回我々が試作した

ⓒ 2020 Information Processing Society of Japan

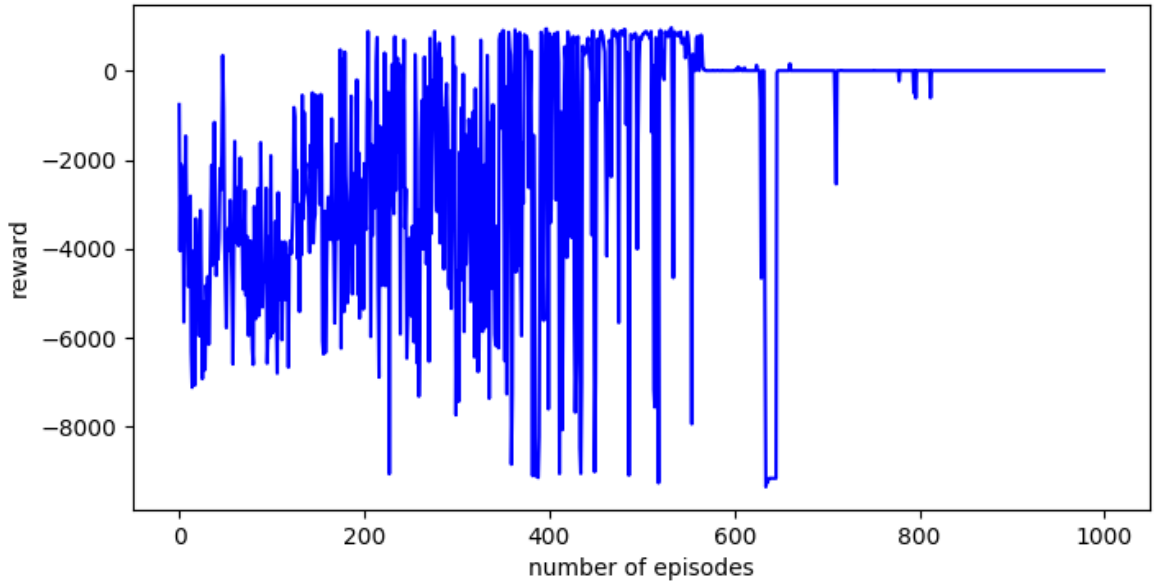

図 7 情報学部の通信先自動匿名化機構のオンライン学習による学習の収束の様子

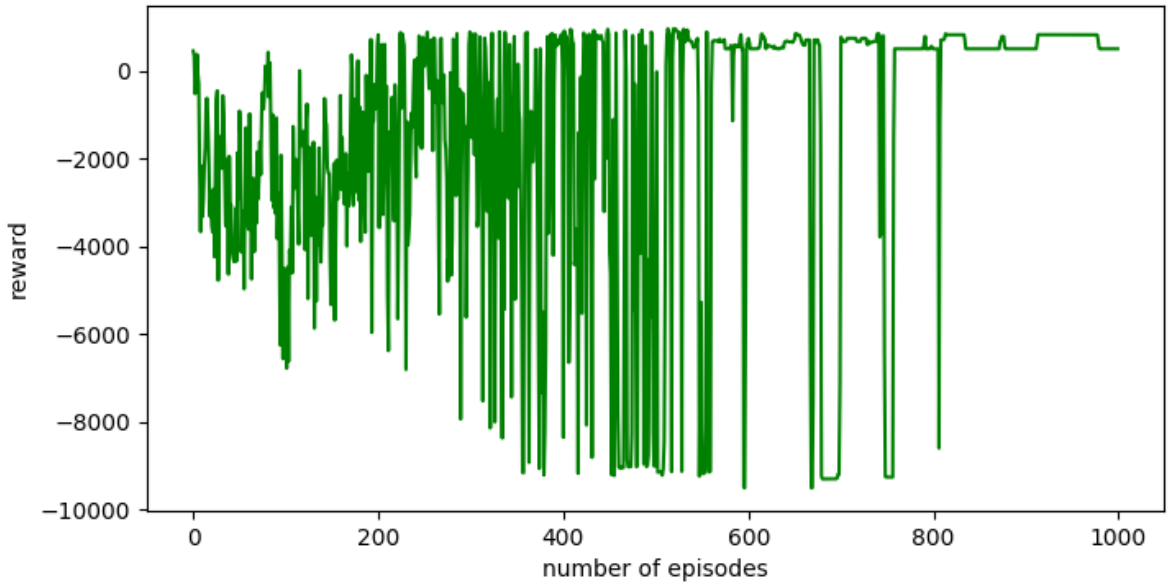

図 8 事務の通信先自動匿名化機構のオンライン学習による学習の収束の様子

通信先自動匿名化機構では，攻撃をしてくる敵対者について知らなくても，攻撃は受けるものとして最大限に通信先を匿名化をするという考えを基として通信先の匿名化を行うものとした．これにより，盗聴やトラフィック分析などの方法で攻撃をしてくる敵対者をモデルに組み込んだ場合の，匿名化リソースの配分の自動的な準最適化の課題から，同じネットワークインフラを複数の利用者団体が共有して利用することで相互にリソースを取り合っている状況で，他の利用者団体を敵対者としてモデルに組み込んだ場合の，匿名化リソースの配分の自動的な準最適化の課題へと変化した．このリソース配分の課題を解く中で，他エージェントに関する情報がペナルティとして間接的にアドホックに得られるような状況での，エージェントの行動の獲得を狙った．

本論文では，具体的なシナリオの例として，小規模な大学のネットワークの通信の通信先を匿名化することを考え，ネットワークの利用者団体の情報学部と事務のそれぞれが我々の通信先自動匿名化機構を使用して匿名化を行う状況を想定し，オンライン学習で学習を進める場合の学習の収束の速度を確認し，シミュレーションによるオフライ

ⓒ 2020 Information Processing Society of Japan

```
{
  "target-links": ["a", "b", "d", "e"],
  "requirement-levels": [7, 5, 7, 7],
  "penalty-lines": [
    [8, 5, 7, 8],
    [8, 6, 8, 7],
    [7, 4, 9, 7]
  ]
}
```

図 9 シミュレーションによるオフライン学習を繰り返す場合の入力例

ン学習を繰り返して学習を進める場合についての議論を加えた．

今回は利用者団体が必ずしも敵対関係にあるとは言えないネットワークを想定したが，パブリッククラウドで仮想プライベートクラウドを構築するような場合には，非常にレアケースかもしれないが，同じネットワークインフラを敵対関係にある利用者団体が共有する可能性がある．そのようなネットワークを想定する場合には，ペナルティの獲得状況を分析することで，トラフィック分析のように他の利用者団体が運用しているシステムに関する情報を獲得できてしまう可能性があることを考慮する必要がある．一方で，利用者団体が非常に親密な関係にあるようなネットワークでは，より積極的なエージェント間のティーチングにより，学習を高速に行える可能性がある．そのような状況への対処は，今後の課題とする．

参考文献

[1] Kar, D., Ford, B., Gholami, S., Fang, F., Plumptre, A., Tambe, M., Driciru, M., Wanyama, F., Rwetsiba, A., Nsubaga, M. and Mabonga, J.: Cloudy with a Chance of Poaching: Adversary Behavior Modeling and Forecasting with Real-World Poaching Data, *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems*, pp. 159–167 (2017).

[2] Gholami, S., Yadav, A., Tran-Thanh, L., Dilkina, B. and Tambe, M.: Don'T Put All Your Strategies in One Basket: Playing Green Security Games with Imperfect Prior Knowledge, *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 395–403 (2019).

[3] Wang, K., Perrault, A., Mate, A. and Tambe, M.: Scalable Game-Focused Learning of Adversary Models: Data-to-Decisions in Network Security Games, *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 1449–1457 (2020).

[4] Sengupta, S., Vadlamudi, S. G., Kambhampati, S., Doupé, A., Zhao, Z., Taguinod, M. and Ahn, G.-J.: A Game Theoretic Approach to Strategy Generation for Moving Target Defense in Web Applications, *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems*, pp. 178–186 (2017).

[5] Nguyen, T., P. Wellman, M. and Singh, S.: A Stackelberg Game Model for Botnet Data Exfiltration, *Proceedings of the 8th Conference on Decision and Game Theory for Security*, pp. 151–170 (2017).

[6] Schlenker, A., Thakoor, O., Xu, H., Fang, F., Tambe, M., Tran-Thanh, L., Vayanos, P. and Vorobeychik, Y.: Deceiving Cyber Adversaries: A Game Theoretic Approach, *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 892–900 (2018).

[7] Wang, Y., Chen, Q., Yi, J. and Guo, J.: U-TRI: Unlinkability Through Random Identifier for SDN Network, *Proceedings of the 2017 Workshop on Moving Target Defense*, pp. 3–15 (2017).

[8] Oggier, F. and Hassibi, B.: The Secrecy Capacity of the MIMO Wiretap Channel, *IEEE Transactions on Information Theory*, pp. 4961–4972 (2011).

[9] Chi, P.-W., Kuo, C.-T., Jing-Wei, G. and Lei, C.-L.: How to detect a compromised SDN switch, *Proceedings of the 2015 1st IEEE Conference on Network Softwarization*, pp. 1–6 (2015).

[10] Chiu, Y.-C. and Lin, P.-C.: Rapid detection of disobedient forwarding on compromised OpenFlow switches, *Proceedings of 2017 International Conference on Computing, Networking and Communications*, pp. 672–677 (2017).

[11] Jajodia, S., Ghosh, A. K., Swarup, V., Wang, C. and Wang, X. S.: *Moving Target Defense: Creating Asymmetric Uncertainty for Cyber Threats*, Springer Publishing Company, Incorporated (2011).

[12] Sugiyama, K. and Fukuta, N.: Toward a SSG-based SDN for Unlinkability of Enterprise IoT Network, Technical report, JAWS: Joint Agent Workshops & Symposium 2018 (2018).

[13] Sugiyama, K. and Fukuta, N.: An Autonomous Cooperative Randomization Approach to Prevent Attacks Based on Traffic Trends in the Communication Destination Anonymization Problem, Technical report, The 33th Annual Conference of the Japanese Society for Artificial Intelligence (2019).

[14] Sugiyama, K. and Fukuta, N.: A Mixed Stackelberg Security Game Model for Security Resources to be Anonymized, Technical report, The 4th International Workshop on Smart Simulation and Modelling for Complex Systems (2019).

[15] Sugiyama, K. and Fukuta, N.: A Multiagent Learning Approach for Distributed Control of Address Randomization in Communication Destination Anonymization, *Proceedings of IEEE/IIAI International Congress on Applied Information Technology* (2019).

[16] Ariel Rosenfeld, Matthew E. Taylor, S. K.: Leveraging Human Knowledge in Tabular Reinforcement Learning: A Study of Human Subjects, *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pp. 3823–3830 (2017).

[17] Sugiyama, K. and Fukuta, N.: A QS-Learning-based Patrol Behavior Learning for Avoiding Illegal Disposals, *Proceedings of 7th IIAI International Congress on Advanced Applied Informatics*, pp. 576–581 (2018).

[18] Foerster, J., Song, F., Hughes, E., Burch, N., Dunning, I., Whiteson, S., Botvinick, M. and Bowling, M.: Bayesian Action Decoder for Deep Multi-Agent Reinforcement Learning, *Proceedings of the 36th International Conference on Machine Learning*, pp. 1942–1951 (2019).

ⓒ 2020 Information Processing Society of Japan