情報処理学会論文誌 Vol.60 No.12 2147–2156 (Dec. 2019)

Man-In-The-Browser攻撃対策を実現する人間・銀行サーバ間のセキュア通信プロトコル

向平 浩貴[1,a)] 神農 泰圭[1] 土屋 貴史[1] 大木 哲史[1] 高橋 健太[2] 尾形 わかは[3] 西垣 正勝[1]

受付日 2019年3月12日, 採録日 2019年9月11日

概要：近年，インターネットバンキングにおいて，マルウェアによるMan-In-The-Browser（MITB）攻撃を用いた不正送金被害が社会問題となっている．MITB攻撃では，マルウェアによりユーザには改ざんした画面を表示しながら裏で不正送金が行われる．MITB攻撃対策として様々な方式が提案されてきており，より高度なMITB攻撃に対しても対策が進んできている．しかし，防御側の技術の進歩にともない，攻撃者も攻撃技術や攻撃手法を多様化させてくると考えられる．そして，究極的にはMachine to Machine（M2M）通信を行う限り，防御側は多様化する攻撃者に対し遅れをとることとなり，対策が困難になることが考えられる．そこで我々は，より高度なMITB攻撃を対策するためには，通信のエンドポイントに人間が存在する必要があると考え，Human to Machine（H2M）セキュア通信によりMITB攻撃対策を実現することを提案する．人間は低い計算機能力しか有していないため，計算の困難性のギャップを用いた従来の暗号等はH2M通信では用いることができないが，その代わりに，人間は機械にはない高い認知能力を有している．そこで，この人間の有する高い認知能力を用いてH2Mセキュア通信を実現する．我々は，“人間には解答が容易であるが，機械には解答が困難である”という性質を持つ問題の一例としてCAPTCHAを用いることで，H2Mセキュア通信チャネルを構築し，MITB攻撃対策を実現するプロトコルを提案する．さらに，公開鍵暗号における識別不可能性に関する安全性定義（IND-CCA安全性）を拡張させ，CAPTCHAにおける識別不可能性に関する安全性定義（IND-C-CCA安全性）を定義する．そして，プロトコルの安全性をIND-C-CCA安全性に帰着させることにより，用いるCAPTCHAがIND-C-CCA安全性を満たすならば，提案プロトコルはMITB攻撃に対し安全性を有することを証明する．

キーワード：Man-In-The-Browser攻撃，CAPTCHA，安全性証明，インターネットバンキング

Secure Communication Protocol between Humans and a Bank Server Secure against Man-In-The-Browser Attack

Koki Mukaihira[1,a)] Yasuyoshi Jinno[1] Takashi Tsuchiya[1] Tetsushi Ohki[1] Kenta Takahashi[2] Wakaha Ogata[3] Masakatsu Nishigaki[1]

Received: March 12, 2019, Accepted: September 11, 2019

Abstract: In recent years, illegal remittance using Man-In-The-Browser (MITB) attack by malware in Internet banking has become a social problem. In MITB attack, malware carry out illegal remittance while displaying a tampered screen to users. Various schemes have been proposed as countermeasures against MITB attack, and countermeasures have been progressed against more sophisticated MITB attacks. However, it is concerned that attackers will diversify attack techniques and methods along with the advances in defense-side technologies. Ultimately, as long as communication is carried out by Machine to Machine (M2M), the defender may be lagging behind diversifying attackers, and it may be difficult to take measures. In view of this, believing that it is necessary for humans to exist at the communication endpoint in order to deal with more sophisticated MITB attacks, we propose Human to Machine (H2M) secure communication for countermeasures against MITB attack. In H2M communication, however, we cannot use conventional cryptography etc. using gaps in computational difficulty, since humans have only low computational power. Instead, humans have a high cognitive ability not found in machines, that we adopt to achieve a H2M secure communication channel. More precisely, we adopt CAPTCHA that has the property “it is easy for humans to answer correctly, but difficult for machines.” Based on such a H2M secure communication channel, we propose a secure banking protocol that can be a countermeasure against MITB attack. Furthermore, we extend the security definition of indistinguishability for public key encryption (IND-CCA security), and define the security definition of indistinguishability for CAPTCHA (IND-C-CCA security). Finally, we prove that the protocol is secure against MITB attacks as long as CAPTCHA satisfies IND-C-CCA security, by providing a reduction from the security of the protocol to IND-C-CCA security.

Keywords: Man-In-The-Browser attack, CAPTCHA, security proof, internet banking

1. はじめに

近年，インターネットバンキングにおける不正送金の被害は減少傾向にあるものの，その被害額は約 3 億 7,200 万円と依然として多くの被害が報告されている [1]．不正送金の手口には，メールによって偽の Web サイトへ誘導するフィッシング等の様々な攻撃手法が存在するが，その 1 つとしてマルウェアによる Man-In-The-Browser 攻撃（以降，MITB 攻撃とする）があげられる．MITB 攻撃とは，ユーザの PC に侵入したマルウェアがブラウザを乗っ取り，ブラウザ上の表示内容を書き換えることで，ユーザには正常な取引が行われているように表示しつつ，裏で不正送金を行う攻撃である．

MITB 攻撃は，文献 [2] で攻撃シナリオの違いによって “ID 盗取型 MITB 攻撃” と “取引内容改ざん型 MITB 攻撃” の 2 種類に分類されている．ID 盗取型 MITB 攻撃は，ユーザの PC に感染したマルウェアが，ユーザのログイン時にログイン画面を改ざんし，金融機関が本来要求しない情報（乱数表のすべての情報等）を要求する偽の画面を表示することで認証情報を盗取する攻撃である．一方，取引内容改ざん型 MITB 攻撃は，ユーザの PC に感染したマルウェアが，ユーザにより入力される送金情報（振込先，金額）およびサーバから送信される確認情報をユーザの PC 内で改ざんすることで，攻撃者の望む任意の取引内容をサーバに受理させる攻撃である．本論文では，取引内容改ざん型 MITB 攻撃に焦点を当て，対策を検討する．

既存の対策手法 [3], [4] では，実際に送金操作を行う端末のほかにトークンデバイス等の端末を用い，実際に送金を行う経路とは別のセキュアな経路を確保することにより，取引内容改ざん型 MITB 攻撃を対策するというアプローチをとっている．しかし，防御側の技術の進歩にともない，攻撃者も攻撃技術や攻撃手法を多様化させてくることが考えられる．これにより，これまで安全とされていた端末にまでマルウェアの侵入が可能となり，複数の経路を用いるというアプローチが取引内容改ざん型 MITB 対策として有効ではなくなることが考えられる．そして，究極的には，Machine to Machine（M2M）による通信を行う限り，防御側は高度化する攻撃者に対し遅れをとり，取引内容改ざん型 MITB 攻撃対策が困難になることが考えられる．

そこで我々は，取引内容改ざん型 MITB 攻撃に対応するためには通信のエンドポイントに人間が存在する必要があると考え，Human to Machine（H2M）セキュア通信による対策を提案する．そして，単一の端末のみを用い，かつその端末がマルウェアに乗っ取られる可能性があるという状況下においても取引内容改ざん型 MITB 攻撃に対する安全性を保つことができる送金要求プロトコルを提案する．

H2M セキュア通信の実現には，複雑な計算を要する従来の暗号方式が利用できないという課題が存在する．従来の暗号方式では，一方向の計算は容易であるが，その逆計算は困難であるといった計算の困難性のギャップを利用している．しかし，機械とは異なり，人間は低い計算能力しか有していないため，計算の困難性のギャップを用いた従来の暗号等は H2M 通信では用いることができない．その代わりに，人間は機械にはない高い認知能力を有しているため，この人間の有する高い認知能力を用いてマルウェア（機械）との間に認知能力のギャップを生み出し，H2M セキュア通信を実現する．我々は，“人間には解答が容易であるが，機械には解答が困難である” といった性質を持つ問題を HS・AIuS（Human-Solvable・AI-unSolvable）問題と定義し，これを用いることで，人間と機械の間に存在する認知能力のギャップを利用した H2M セキュア通信チャネルを構築する．

本論文では，HS・AIuS 問題の一例として CAPTCHA を用いる．そして，公開鍵暗号における識別不可能性に関する安全性定義（IND-CCA 安全性）を拡張し，CAPTCHA に対する安全性として識別不可能性（IND-C-CCA 安全性）を定義する．そのうえで，提案プロトコルの取引内容改ざん型 MITB 攻撃に対する安全性を CAPTCHA の IND-C-CCA 安全性に帰着させることにより，用いる CAPTCHA が IND-C-CCA 安全性を満たすならば，提案プロトコルが取引内容改ざん型 MITB 攻撃に対し安全性を有することを証明する．

本論文の貢献は以下の 3 点である．

A) CAPTCHA に対する最初の暗号学的な安全性定義を示す．今後，送金要求プロトコルで用いる CAPTCHA を具体的に設計する際の指標の 1 つとして，これを用いることができる．

B) CAPTCHA を用いた送金要求プロトコルのモデルを定義し，さらに送金要求プロトコルが持つべき要件として，取引内容改ざん型 MITB 攻撃に対する安全性を定義する．これらは，送金要求プロトコルを具体的に設計していくにあたっての指標として用いることができる．

C) A) で定義した安全性を満たす CAPTCHA の存在のもとで，B) で定義した取引内容改ざん型 MITB 攻撃に対する安全性を暗号学的に証明可能な送金要求プロトコルの構成方法を示す．したがって，A) で定義した安全性を持つ CAPTCHA を構成しさえすれば，B) で定義した安全性を持つ送金要求プロトコルが実現可能である．

[1] 静岡大学
Shizuoka University, Hamamatsu, Shizuoka 432–8011, Japan
[2] 株式会社日立製作所
Hitachi Ltd., Yokohama, Kanagawa 244–0817, Japan
[3] 東京工業大学
Tokyo Institute of Technology, Meguro, Tokyo 152–8552, Japan
[a)] mukaihira.koki@shizuoka.ac.jp

本論文の趣旨は，今後の MITB 攻撃対策の一助として，送金要求プロトコルに適用可能な CAPTCHA の構成に関する方針を示すことである．近年の AI 技術の発達により，高い機械耐性と人間の解読容易性を兼ね備える CAPTCHA を実現することが難しくなっている．文献 [5] で提案されている非現実画像 CAPTCHA のように，人間の有する高度な認知能力を活用して CAPTCHA を強化する試みが続けられているが，提案プロトコルが要求する安全性を満たす CAPTCHA の構築には至っていない．そのため，提案プロトコルを実際のインターネットバンキングシステムに組み込むことは現時点では困難であることに注意されたい．

なお，取引内容改ざん型 MITB 攻撃は，マルウェアが人間である攻撃者と連携して攻撃を実施する場合も想定されるが，本論文においては，取引内容改ざん型 MITB 攻撃対策の先駆けとして，マルウェア単体で攻撃が実施される場合のみを想定する．

海外では，完全自動化された取引内容改ざん型 MITB 攻撃による大規模な不正送金事例が報告されている [6]．文献 [6] の事例では，人手を介することなくマルウェアが自動的に口座からある一定の金額を吸い上げるシステムを構築することで，数カ月の間に約 60 億から 2,000 億円の不正送金被害をもたらした．このように，完全自動化された取引内容改ざん型 MITB 攻撃は，効率良く広範囲にわたり攻撃を実施することが可能であり，将来増加が予想される脅威であるといえる．マルウェア単体による完全自動化された取引内容改ざん型 MITB 攻撃を防ぐことができれば，人手を介さない限り攻撃を成功させることが困難となるため，攻撃者のコストを増大させることができ，結果的に攻撃のペースが抑制されることが期待できる．そのため，マルウェア単体による取引内容改ざん型 MITB 攻撃への対策は十分意味のあるものと考えられる．

2. 準備

2.1 インターネットバンキングにおける一般的な送金要求プロトコルのモデル

本節では，本論文の前提となるインターネットバンキングにおける一般的な送金要求プロトコルのモデルを，文献 [2] の示すモデルを用いて示す．文献 [2] における送金要求プロトコルのモデルは以下の要素から構成される．

ユーザ：インターネットバンキングを利用する顧客である．送金処理を実行する際には，金融機関が提供する送金要求プロトコルに従い PC を操作し，ブラウザを利用する．ユーザは人間であり，低い計算能力・記憶能力しか有していないが，高い認知能力を有するものとする．

ブラウザ：ユーザがインターネットバンキングを利用する際に用いる PC にインストールされたブラウザである．ブラウザは PC の有する高い計算能力・記憶能力を利用できる．また，MITB 攻撃においてマルウェアはブラウザに感染し，不正を行う．

サーバ：インターネットバンキングを提供する金融機関のサーバである．サーバは高い計算能力・記憶能力を有しているが，低い認知能力しか有していないものとする．また，MITB 攻撃においてマルウェアはサーバに感染することはなく，サーバ自身も不正を行うことはないものとする．

一般的な送金要求プロトコル [2] は以下の手順に従い動作する．

① ユーザは，送金情報 x をブラウザに入力する．
② ブラウザは，x をサーバへ送信する．
③ サーバは，用いるプロトコルに従って送金情報 x に対応した確認情報 y を求めてブラウザへ送信する．
④ ブラウザは，y をユーザへ提示する．
⑤ ユーザは，y が自身が ① で入力した x に対応する確認情報であることを確認し，正しければブラウザに Q $(= \mathit{TRUE})$ を入力する．そうでないならばブラウザに Q $(= \mathit{FALSE})$ を入力する．
⑥ ブラウザは，Q をサーバへ送信する．
⑦ サーバは，Q を受信し，$Q = \mathit{TRUE}$ ならば x を受理する．$Q = \mathit{FALSE}$ ならば送金中止処理を行う．

2.2 取引内容改ざん型 MITB 攻撃

本論文では，取引内容改ざん型 MITB 攻撃のモデルとして，文献 [2] の示すモデルを用いる．一般的な送金要求プロトコルに対する取引内容改ざん型 MITB 攻撃は，以下の手順で行われる．

① ユーザは，送金情報 x をブラウザに入力する．
② マルウェアに感染したブラウザは，ユーザが入力した送金情報 x を任意の送金情報 x' に改ざんし，サーバへ送信する．
③ サーバは，x' に対応した確認情報 y' をブラウザへ送信する．
④ マルウェアに感染したブラウザは，y' を x に対する確認情報 y に改ざんを行い，ユーザへ提示する．
⑤ ユーザは，y が，自身が ① で入力した x に対応する確認情報であることを確認し，正しければブラウザに Q $(= \mathit{TRUE})$ を入力する．
⑥ マルウェアに感染したブラウザは，Q をサーバへ送信する．
⑦ サーバは，Q $(= \mathit{TRUE})$ を受信し，x' を受理する．

2.3 CAPTCHA

CAPTCHA（Completely Automated Public Turing test to tell Computers and Humans Apart）[7] は，“人間には解答が容易であるが，機械には解答が困難である” という性質を持つ問題をユーザに出題し，解答できたユーザを人間，解答できなかったユーザを機械だと判定するセキュリティ技術である．現在，多くの Web サービス提供サイト

では，自動プログラム（マルウェア）によるメールアカウントの不正取得，ブログや掲示板へのスパムコメントの書き込み，パスワードリスト攻撃といった不正行為を防ぐためにCAPTCHAが用いられている．具体的なものとしては，歪曲やノイズが付加された文字列画像を提示し，ユーザがその文字を判読できるか否かを試す文字列判読型CAPTCHA等が典型的な手法として広く用いられている．しかし，近年，マルウェアが用いるOCR（文字判読技術）等の攻撃手法の高度化が進み，文字列判読型CAPTCHAをはじめとする多くのCAPTCHAがマルウェアによって解読されることが報告されている．

近年のCAPTCHA研究においては，文献[5]の非現実画像CAPTCHAのように，人間の有する高度な認知能力に着目することで機械耐性を向上させ，かつ人間には解答が容易である問題を生成する方式の提案が行われている．非現実画像CAPTCHAでは，複数の3Dモデルのうち，“不自然な重なり”を与えた非現実モデルを1体だけ配置した1枚の画像をCAPTCHAとして出題し，画像中から非現実モデルを選択できたユーザを人間と判断する．非現実画像CAPTCHAでは，人間の持つ高度な認知能力として“視覚的形式知からの逸脱”を用い，人間と機械の間に存在する不自然な重なりを認識する能力のギャップを利用することにより，人間と機械を切り分けることを提案している．そして，問題画像の一部を切り取り，“画像中に複数のモデルから生成されているモデルが存在するか否か”を検出するめり込み検出攻撃と総当たり攻撃に対して十分な機械耐性を有していることを示している．

本論文におけるCAPTCHAとは，文献[5]の非現実画像CAPTCHAのように，人間の持つ高度な認知能力を活用し，“人間には解答が容易であるが，機械には解答が困難である”ということを達成すると考えられるCAPTCHAを想定するものとする．

3. 関連研究

取引内容改ざん型MITB攻撃対策としては，サーバが受信した取引が，意図したとおりの取引であることをユーザ本人が認証する取引認証が有効とされている．既存の取引認証手法は，専用端末を用いるものと用いないものに分類される．専用端末を用いる取引認証手法としては，文献[8]が提案されている．文献[8]の手法は，インターネットバンキングでの送金時，専用端末をPCと接続しておくと，サーバが受信した送金情報が専用端末のディスプレイに表示され，ユーザはその表示内容と自身が入力した送金情報が一致するかを確認し，問題がなければ送金処理を実行するというものである．一方，専用端末を用いない取引認証手法としては，文献[9]が提案されている．文献[9]の手法は以下のとおりに行われる．ユーザから受信した送金情報とOne-Time Password（OTP）を埋め込んだ特殊な文字画像をサーバで生成し，それを確認情報としてユーザに表示する．ユーザは送金情報と確認情報の一致を確認し，確認情報をスマートフォンで撮影する．スマートフォンは内蔵されたセキュアSIM内で確認情報からOTPを取得し，これを送金の際の認証情報として用いる．セキュアSIM内で確認情報からOTPを取り出す操作を行うことで，マルウェアがスマートフォンに感染し，PCに感染したマルウェアと連携したとしても，取引内容改ざん型MITB攻撃を防ぐことができるとしている．

いずれもインターネットバンキングを利用する端末とは別の端末を用い，複数の経路を確保することにより取引内容改ざん型MITB攻撃を対策している．著者らの調べた限りでは，単一経路のみで取引内容改ざん型MITB攻撃の対策を実現し，さらに暗号学的に安全性証明を行った既存研究は存在しない．したがって，この点に関して本研究に優位性および独自性があると考えられる．

一方，MITB攻撃対策のためにCAPTCHAをプロトコルの構成要素として利用する事例については，著者らの調べた限りでは存在しない．したがって，CAPTCHAの適用先という観点からも本研究に優位性および独自性があると考えられる．

4. CAPTCHAのモデル化

4.1 計算困難性

人間，機械それぞれにとっての計算の複雑度Complexity for Human，Complexity for AIを定義し，**図1**のように4つのクラスにより問題を分類することにより，問題の計算困難性について定義する．図1の4つのクラスは，人間，機械それぞれが実時間で計算可能か否かにより分けられ，人間が実時間で計算可能であることをHS（Human-Solvable），人間が実時間で計算困難であることをHuS（Human-unSolvable），機械が実時間で計算可能であることをAIS（AI-Solvable），機械が実時間で計算困難で

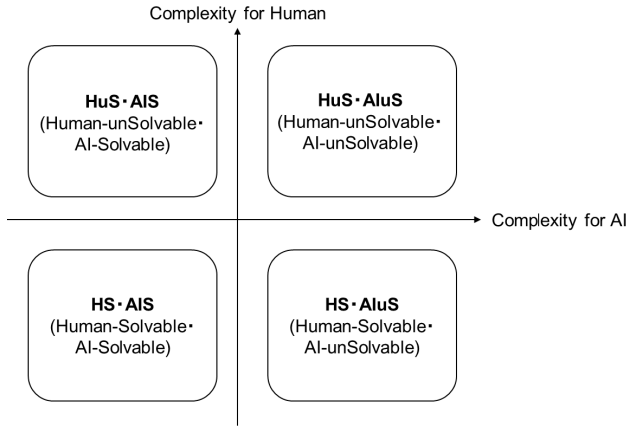

図1 人間と機械に対する問題のクラス

Fig. 1 Problem classes defined by humans and machines solvability.

あることを AIuS（AI-unSolvable）とし，これらを組み合わせることにより 4 つのクラスを表現する．各クラスは以下のとおりである．

HuS・AIS：人間が実時間で計算困難であり，かつ機械が実時間で計算可能な問題のクラス．（例：20 桁の素因数分解）

HS・AIS：人間，機械ともに実時間で計算可能な問題のクラス．（例：1 桁の足し算）

HuS・AIuS：人間，機械ともに実時間で計算困難な問題のクラス．（例：1000 桁の素因数分解）

HS・AIuS：人間が実時間で計算可能であり，かつ機械が実時間で計算困難な問題のクラス．（例：文献 [5] で提案されているような CAPTCHA）

以降では，HS・AIuS クラスに属する問題を HS・AIuS 問題と呼ぶ．

4.2 定式化

本論文における CAPTCHA を文献 [10] をもとに定式化する．

まず，CAPTCHA の問題集合を $\boldsymbol{C}$ ($\boldsymbol{C} := \{\boldsymbol{C_k}\}_{k\in\mathbb{N}}$, $\boldsymbol{C_k} \subseteq \{0,1\}^{l(k)}$)，CAPTCHA の解集合を $\boldsymbol{M}$ ($\boldsymbol{M} := \{\boldsymbol{M_k}\}_{k\in\mathbb{N}}$, $\boldsymbol{M_k} \subseteq \{0,1\}^{k}$) とする．ここで，$l(k)$ は k に関する多項式であり，$\{0,1\}^{l(k)}$ はビット長が $l(k)$ のビット列の集合を表す．続いて，CAPTCHA を以下に示す CAPTCHA 化アルゴリズム C_Enc，解写像 C_Dec の 2 つの組で表し，$Capt := (C\_Enc, C\_Dec)$ と定義する．

CAPTCHA 化アルゴリズム $\boldsymbol{C\_Enc(m)}$：解 m ($\in \boldsymbol{M_k}$) を入力とし，m に対応する（$C\_Dec(c) = m$ となる）問題 c ($\in \boldsymbol{C_k}$) を出力する確率的多項式時間（PPT）アルゴリズムである．また，C_Enc は単射であるとする．

解写像 $\boldsymbol{C\_Dec(c)}$：問題 $c = C\_Enc(m)$ ($\in \boldsymbol{C_k}$) を，解 m にマッピング（$\boldsymbol{C_k} \to \boldsymbol{M_k}$）する写像である．また，高い認知能力を有している人間は C_Dec を実時間で計算可能（HS）であるが，機械は実時間で計算困難（AIuS）であるとする．

4.3 安全性定義：IND-C-CCA 安全性

本論文では，CAPTCHA として HS かつ AIuS であるものを想定するが，ここでは，AIuS よりさらに強い性質である，CAPTCHA の「選択 CAPTCHA 問題攻撃に対する識別不可能」を考える．

選択 CAPTCHA 問題攻撃に対する識別不可能は，公開鍵暗号における選択暗号文攻撃に対する識別不可能と類似した概念である．選択 CAPTCHA 問題攻撃とは，攻撃者が CAPTCHA の問題 c に対応する解 m を入手できる条件を想定しており，識別不可能とはチャレンジャが 2 つの解のうちどちらを CAPTCHA 化したかを攻撃者が識別できないことを意味する．攻撃者 B とチャレンジャによって行われる IND-C-CCA ゲームを，以下のように定義する（**図 2** 参照）．

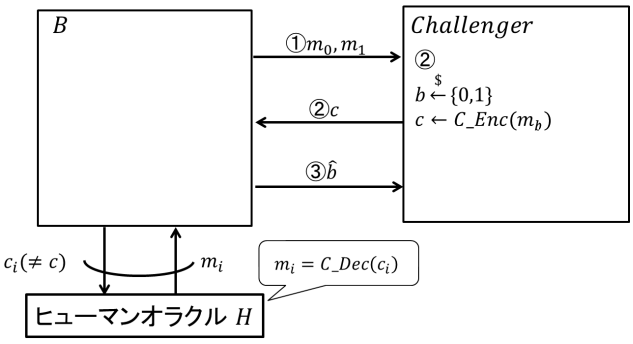

図 2 IND-C-CCA ゲーム
Fig. 2 IND-C-CCA game.

① B は 2 つの解 m_0, m_1 ($m_0, m_1 \in \boldsymbol{M_k}$) を任意に選び，これらをチャレンジャに渡す．

② チャレンジャは m_0，m_1 のうちどちらか 1 つをランダムに選択し，それを m_b とし，問題 $c = C\_Enc(m_b)$ を作成し，これをチャレンジとして B に渡す．

③ B は $\hat{b}$ を出力する．このとき，$\hat{b} = b$ であれば攻撃者の勝ちとする．

このゲームにおいて，B は任意のタイミングでヒューマンオラクル H を利用することができる．ヒューマンオラクルは，ランダムオラクルを拡張した概念であり，文献 [10] のように CAPTCHA を暗号学的にモデル化する際に広く用いられている．ヒューマンオラクルは，人間が高い認知能力を利用することをモデル化しており，“CAPTCHA の問題をクエリするとその解を返す” という動作をする．ただし，B が ② で受け取ったチャレンジ c を H にクエリすることは禁止されており，チャレンジ c をクエリすると B は IND-C-CCA ゲームに敗北する．上記の IND-C-CCA ゲームにおける B のアドバンテージを

$$Adv_B^{IND-C-CCA} = \left|\Pr[\hat{b} = b] - \frac{1}{2}\right|$$

と定義し，いかなる B に対しても $Adv_B^{IND-C-CCA}$ が無視できるとき，CAPTCHA は IND-C-CCA 安全であるという．

CAPTCHA が IND-C-CCA 安全性を満たすということは，PPT アルゴリズムである攻撃者が CAPTCHA の問題から解のいかなる部分情報も得られないことを保証する．

5. CAPTCHA を用いた送金要求プロトコルのモデル

5.1 定式化

我々は，HS・AIuS 問題である CAPTCHA を用いた送金要求プロトコルのモデルとして，2.1 節の送金要求プロトコルを一般化した以下のモデルを考える．

① ユーザは，送金情報 x をブラウザを介してサーバに送

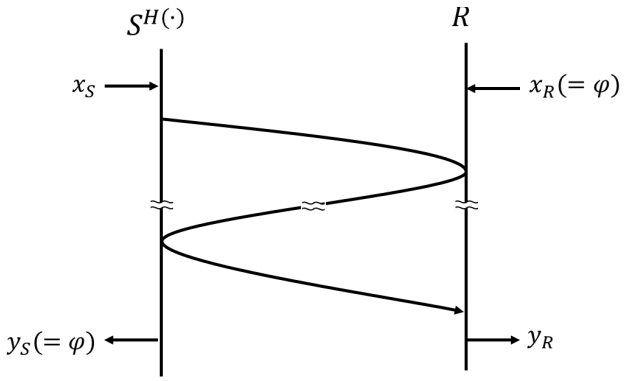

図 3 2 者間プロトコルとして定式化した送金要求プロトコル
Fig. 3 Money-transfer request protocol formulated as two-party protocol.

信する．

② ユーザとサーバはブラウザを介して何らかの通信を行う（ただし，ユーザの行う処理は，すべて HS・AIuS 問題または CAPTCHA を解くことに限るものとする）．

③ サーバはこれまで受信した情報をもとに，送金情報 x を受理するか却下するかを決定する．

MITB 攻撃では，攻撃者は，ブラウザがユーザまたはサーバに情報を転送している ①，② において何らかの改ざんを行うことができる．

上記のプロトコルにおいて，ブラウザは，ユーザやサーバから送られてきたデータをもう一方に受け渡す役割しかしていないため，プロトコルの上では単なる通信路と見なすことができる．したがって，プロトコルはユーザ・サーバ間の 2 者間プロトコルとして定式化することができ，ユーザを送信者 $S^{H(\cdot)}$，サーバを受信者 R とすると，**図 3** のように表すことができる．ここで，$S^{H(\cdot)}$ はヒューマンオラクル $H(\cdot)$ を有する送信者 S を示す．プロトコルを行う 2 者 $S^{H(\cdot)}$，R に対する入力をそれぞれ x_S，x_R とすると，送金要求プロトコルにおいては x_S は送金情報であり，また，サーバは特に入力を持たないため x_R は φ となる．また，$S^{H(\cdot)}$，R に対する出力をそれぞれ y_S，y_R とすると，y_S は φ であり，y_R は R が送金情報を受理した場合は送金情報，送金中止の場合は $\perp$ である．

このように 2 者間プロトコルとして定式化することにより，MITB 攻撃を，ヒューマンオラクルを用いることができる攻撃者とサーバの間で行われるゲームとしてモデル化することができる．これにより，プロトコルの安全性を，公開鍵暗号における安全性定義等と同様に一般的な形で定義することができる（次節）．

CAPTCHA を用いた送金要求プロトコルの要件として，完全性（Completeness）と健全性（Soundness）を定義する．

完全性（Completeness）：人間とサーバが提案プロトコルを実行したとき，以下を満たす場合，プロトコルは完全性を満たすという．ここで，$\langle S^{H(\cdot)}(x_S), R(x_R)\rangle = (y_S, y_R)$

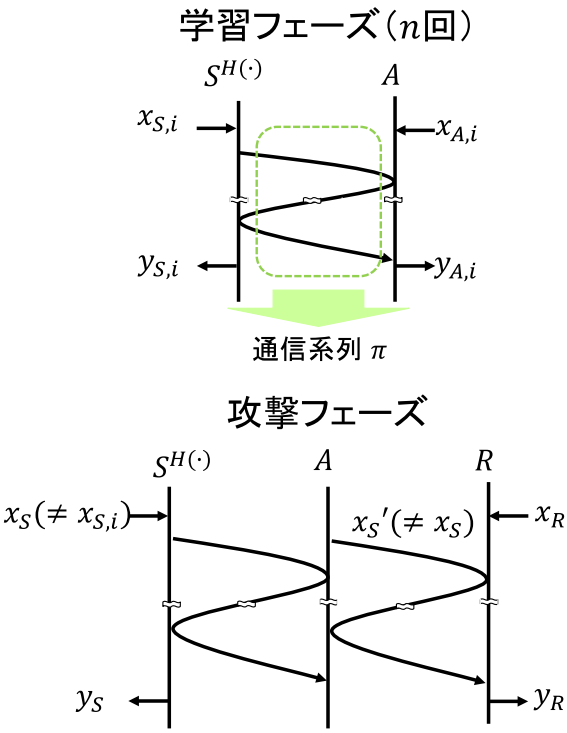

図 4 SUB-MIM ゲーム
Fig. 4 SUB-MIM game.

は，入力 x_S を持つ $S^{H(\cdot)}$ と入力 x_R を持つ R がプロトコルを実行した結果，$S^{H(\cdot)}$ が y_S を得て，R が y_R を得ることを示す．また，ε はネグリジブルな値であり，人間がヒューマンエラーにより CAPTCHA の解答に失敗してしまう確率を示す．

$$Pr\left[x_S = y_R \,|\, \langle S^{H(\cdot)}(x_S), R(x_R)\rangle = (y_S, y_R)\right] \geq 1 - \varepsilon$$

健全性（Soundness）：5.2 節で示す SUB-MIM 安全性を満たす場合，プロトコルは健全性を満たすという．

5.2 MITB 攻撃に対する安全性定義：SUB-MIM 安全性

CAPTCHA を用いた送金要求プロトコルに対する SUB-MIM 攻撃者を A とする．A は学習時および攻撃時に，高度な認知能力を有する人間 $S^{H(\cdot)}$ を送金要求プロトコルのユーザとして使役することができる PPT アルゴリズムである．なお，A が $S^{H(\cdot)}$ を使役することを $A^{S^{H(\cdot)}}$ と表すものとする．ここで，本論文ではマルウェアによる完全自動化された MITB 攻撃を対策の対象としているため，A は学習フェーズおよび攻撃フェーズにおいて，$H(\cdot)$ を直接使役することはできず，S を介してのみ使役することができるということに留意されたい．SUB-MIM（Substitution-Man In the Middle）安全性は，**図 4** に示す SUB-MIM ゲームにより定義される．SUB-MIM ゲームは，以下に示す学習フェーズと攻撃フェーズから構成される．

学習フェーズ：$S^{H(\cdot)}$，A はそれぞれ任意の入力 $x_{S,i}$，$x_{A,i}$ $(1 \leq i \leq n)$ をとり，プロトコルを n 回実行し，通信系列

π を得る．

攻撃フェーズ：$A^{S^{H(\cdot)}}$，R はプロトコルを実行する．ただし，$S^{H(\cdot)}$ への入力 x_S は n 回の学習フェーズで入力した $x_{S,i}$ とは異なる値である．$S^{H(\cdot)}$ から x_S を受信した A は，x_S に対し任意の改ざんを行い，$x_S{}'$ として R に送信する．ただし，A は x_S に対し何らかの改ざんを行わなければならない（$x_S' \neq x_S$ でなければならない）．A は $S^{H(\cdot)}$ を使役しながら R とプロトコルを実行し，最終的に R は y_R を出力する．SUB-MIM ゲームにおける A の勝利条件は，$S^{H(\cdot)}$ への入力 x_S とは異なる y_R を R に受理させることである．したがって，SUB-MIM ゲームにおいて $x_S \neq y_R$ かつ $y_R \neq \perp$ となるとき，A の勝利となり，それ以外の場合負けとなる．

SUB-MIM ゲームにおける A のアドバンテージを

$$Adv_A^{SUB-MIM} = [A \text{ wins SUB-MIM game}]$$

と定義し，いかなるアルゴリズム A に対しても $Adv_A^{SUB-MIM}$ が無視できるとき，プロトコルは SUB-MIM 安全を満たすという．

6. 提案プロトコルおよび安全性証明

6.1 提案プロトコル

本論文で提案するユーザ・サーバ間の H2M セキュア通信プロトコルの概要を**図 5** に示す．同プロトコルは以下の手順に従い動作する．$Capt = (C\_Enc, C\_Dec)$ を CAPTCHA とする．なお，図 5 中の $\boldsymbol{R}$ は乱数集合であり，$\boldsymbol{R}$ は十分大きいものとする．

① ユーザは，送金情報 x をブラウザに入力する．
② ブラウザは，x をサーバへ送信する．
③ サーバは，乱数 $r \in R$ を生成し，ブラウザから受信した送金情報を y として，$C\_Enc(y|r)$ を実行し CAPTCHA 問題（確認情報に相当する）c を生成する．
④ サーバは，c をブラウザへ送信する．
⑤ ブラウザは，c をユーザへ提示する．
⑥ ユーザは，c を解き（$C\_Dec(c)$ を計算し），y と r を得る．y を確認し，自身が ① で入力した x に対する確認情報である $(y = x)$ ならばブラウザに Q $(= r)$ を入力する．そうでない $(y \neq x)$ ならばブラウザに Q $(= \perp)$ を入力する．
⑦ ブラウザは，Q をサーバへ送信する．
⑧ サーバは，Q を受信し，$Q = r$ ならば x を受理する．$Q \neq r$ または $Q = \perp$ ならば送金中止処理を行う．

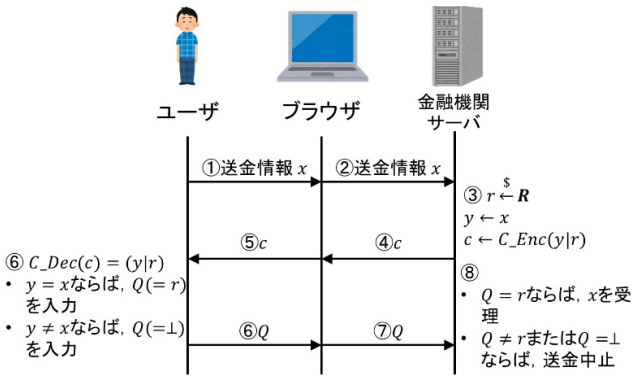

図 5 提案プロトコルの概要
Fig. 5 Overview of proposed protocol.

6.2 安全性証明

提案プロトコルの取引内容改ざん型 MITB 攻撃に対する安全性証明として，IND-C-CCA 安全な CAPTCHA を用いる提案プロトコルは SUB-MIM 安全を満たすことを証明する．

定理 1

CAPTCHA $Capt = (C\_Enc, C\_Dec)$ が IND-C-CCA 安全ならば，その CAPTCHA を用いる提案プロトコルは SUB-MIM 安全を満たす．

定理 1 の証明

定理 1 の対偶をとり，以下の（1）を証明する．

（1） 提案プロトコルの SUB-MIM 安全性を破る攻撃者 A が存在するならば，CAPTCHA の IND-C-CCA 安全性を破る攻撃者 B^A が存在する．

B^A を**図 6** のように構成する．A は提案プロトコルの SUB-MIM 安全性を無視できない確率で破ることができるアルゴリズムであり，B^A は A に対し入力を与え，A からの出力を得ることで，A を内部的に用いながらチャレンジャとの間で IND-C-CCA ゲームを行うアルゴリズムである．また，B^A はヒューマンオラクル H を用いることができる．H はクエリされた CAPTCHA 問題 c に対し解写像 C_Dec を実行することで CAPTCHA の解 m を出力する．B^A は以下の手順でチャレンジャと IND-C-CCA ゲームを行う．

① B は S として A と SUB-MIM ゲームの学習フェーズを実行する．$C\_Dec(c)$ の計算には H を用いることで，π を生成することができる．
② B は A に対し，任意に選んだ送金情報 x を入力する．
③ A は x に対し任意の改ざんを行い，x' として出力する．ただし，SUB-MIM ゲームの定義より，$x' \neq x$ である．
④ B は異なる 2 つの乱数 r_0，r_1 を生成し，2 つの解 m_0 $(= x'|r_0)$，m_1 $(= x'|r_1)$ を生成する．これは，図 5 の ③ におけるサーバがブラウザから受信した送金情報と生成した乱数を連結するという動作に相当する．
⑤ B は m_0，m_1 をチャレンジャに渡す．
⑥ チャレンジャは m_0，m_1 のうち 1 つを選択し，それを m_b とし，$c(m_b) \leftarrow C\_Enc(m_b)$ を計算する．
⑦ チャレンジャは $c(m_b)$ を IND-C-CCA ゲームにおけるチャレンジとして B に入力する．B は $c(m_b)$ を A に入力する．
⑧ A は $c(m_b)$ に対し任意の改ざんを行い，$c(m_b{}')$ として

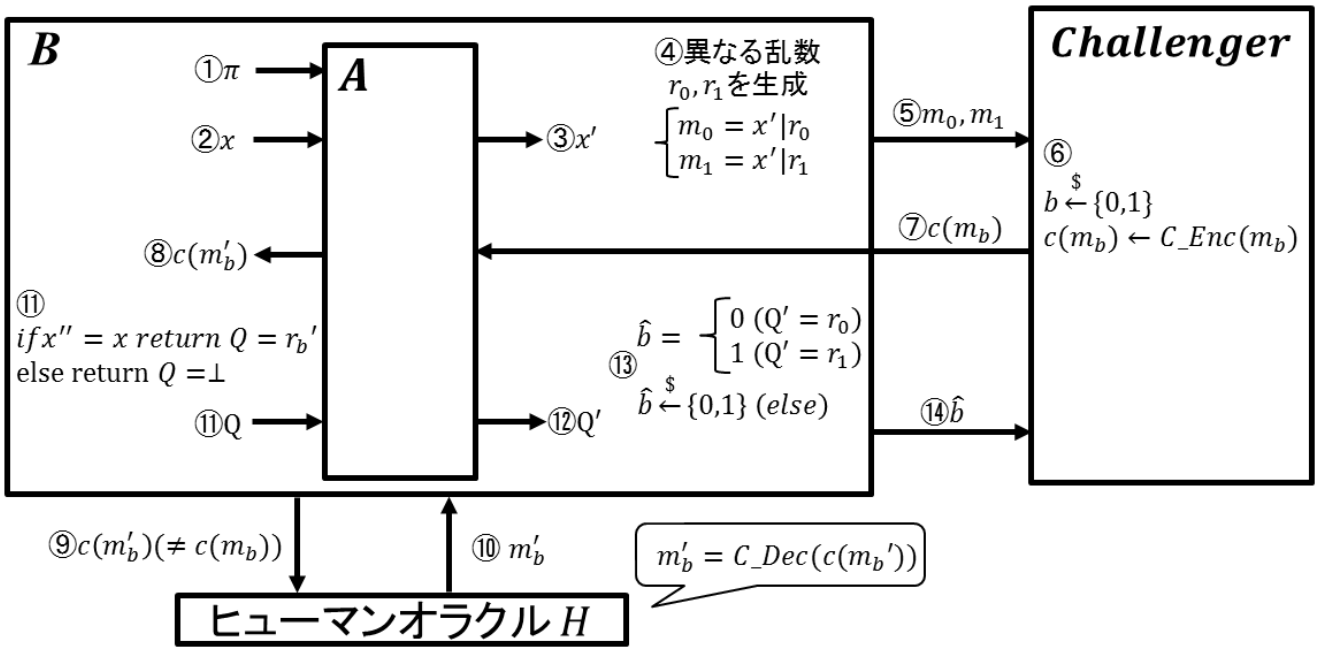

図 6 B^A の構成
Fig. 6 Construction of B^A.

出力する．

⑨ B は $c(m_b')$ を H にクエリする．ただし，$c(m_b') = c(m_b)$ の場合，C_Enc が単射であることから，$m_b' = m_b = x' \,|\, r_b \neq x \,|\, r_b$ となるため，B は H にクエリすることなく，$Q = \perp$ を A に入力する．

⑩ H は $m_b' = C\_Dec(m_b')$ を計算し，m_b' $(= x'' \,|\, r_b')$ を B に渡す．

⑪ B は m_b' を確認し，m_b' の中の x'' が自身が ② において A に対し入力した x と一致しているならば $Q = r'_b$，そうでないならば $Q = \perp$ を A に対し入力する．これは，図 5 の ⑥ におけるユーザの動作部分に相当する．

⑫ A は Q' を出力する．

⑬ B は Q' を確認し，$Q' = r_0$ の場合 $\hat{b} = 0$，$Q' = r_1$ の場合 $\hat{b} = 1$，それ以外の場合ランダムに 0，1 を選択し $\hat{b}$ として出力する．

⑭ B は $\hat{b}$ をチャレンジャに渡す．

B によってシミュレートされた A の環境は，SUB-MIM ゲームにおける A の環境と同一である．A の SUB-MIM ゲームにおける勝利条件は，B が ④ で選択した r_0，r_1 のうち，チャレンジャが ⑥ で選択したビット b に対応している r_b を⑫ で Q' として出力することである．A が SUB-MIM ゲームに勝利するとき，B は Q' を用いることでチャレンジャが選択した b を知ることができ，これを $\hat{b}$ としてチャレンジャに渡すことで IND-C-CCA ゲームに勝利することができる．したがって，A が SUB-MIM ゲームに勝利するとき，B が IND-C-CCA ゲームに勝利することは自明である．

続いて，A が SUB-MIM ゲームに負けたとき B が IND-C-CCA ゲームに勝利する条件について述べる．A が SUB-MIM ゲームに負けるという事象は，⑫ で出力した Q' が r_b ではないもう一方 (r_{1-b}) だった場合と，r_b でも r_{1-b} でもない値であった場合の 2 通りが考えられる．⑫ で出力した Q' が r_b ではないもう一方 (r_{1-b}) だった場合，B はそのことを知りえないため，⑬ で $\hat{b} = 1 - b$ を選択するしかなく，結果的に IND-C-CCA ゲームに負けてしまう．一方，⑫ で出力した Q' が r_b でも r_{1-b} でもない値であった場合，r_0，r_1 を知っている B は Q' が正しくない値であることを知ることができる．そして，コイントスにより 0，1 をランダムに選択し $\hat{b}$ として出力することができる．このとき，B は 1/2 の確率で IND-C-CCA ゲームに勝利することができる．以上より，A が SUB-MIM ゲームに負けたとき，B が IND-C-CCA ゲームに勝利する条件は，$Q' \neq r_{1-b}$ かつ B がコイントスにより IND-C-CCA ゲームに勝利する場合である．

A が SUB-MIM ゲームに勝利する確率を $Pr[A\ wins] = \varepsilon_A$，$A$ が SUB-MIM ゲームに負ける確率を $Pr[A\ loses] = 1 - \varepsilon_A$，乱数の値域を R，⑫ における A の出力を $Q' \in R$ とすると，B が IND-C-CCA ゲームに勝利する確率 $Pr[B\ wins]$ は以下のとおりである．

$$
\begin{aligned}
&Pr[B\ wins] \\
&= Pr[A\ wins] \\
&\quad + Pr[A\ loses \wedge Q' \neq r_{1-b} \wedge B\ wins\ by\ coin\ toss] \\
&= \varepsilon_A + \Pr[\mathrm{A}\ loses] \cdot \Pr[Q' \neq r_{1-b} \,|\, \mathrm{A}\ loses] \\
&\quad \cdot Pr[B\ wins\ by\ coin\ toss \,|\, A\ loses \wedge Q' \neq r_{1-b}] \\
&= \varepsilon_A + (1 - \varepsilon_A) \cdot \frac{|R| - 2}{|R| - 1} \cdot \frac{1}{2}
\end{aligned}
$$

$$= \varepsilon_A \cdot \frac{|R|}{2(|R|-1)} + \frac{|R|-2}{|R|-1} \cdot \frac{1}{2}$$

$$Adv_B^{IND-C-CCA} = \left| Pr[B\ wins] - \frac{1}{2} \right| = \left| \varepsilon_A \cdot \frac{|R|}{2(|R|-1)} - \frac{1}{2(|R|-1)} \right|$$

以上より，乱数空間 $|R|$ が十分大きく，SUB-MIM ゲームにおける A のアドバンテージ $\varepsilon_A = Adv_A^{SUB-MIM}$ が無視できない確率であるとき，IND-C-CCA ゲームにおける B のアドバンテージ $Adv_B^{IND-C-CCA}$ もまた同様に無視できないといえる．したがって，(1) は真であることが証明でき，定理 1 は証明された．

7. おわりに

本論文では，“人間には解答が容易であるが，機械には解答が困難である” という性質を持つ問題の一例として CAPTCHA を用いることにより，取引内容改ざん型 MITB 攻撃に対し安全性を保つことができる H2M セキュア通信プロトコルを構成した．また，CAPTCHA を暗号学的なモデルとして定式化し，CAPTCHA の安全性として識別不可能に関する安全性（IND-C-CCA 安全性）を定義した．さらに，提案プロトコルの取引内容改ざん型 MITB 攻撃に対する安全性として，SUB-MIM 安全性を定義した．最後に，提案プロトコルの安全性証明として，“CAPTCHA が IND-C-CCA 安全ならば，その CAPTCHA を用いる提案プロトコルは SUB-MIM 安全を満たす” ことを示した．

今後は，提案プロトコルにおける CAPTCHA の要件である IND-C-CCA 安全性を満たすような CAPTCHA について検討し，提案プロトコルの実現可能性について検討していきたい．

参考文献

[1] 平成 30 年上半期におけるサイバー空間をめぐる脅威の情勢等について，入手先 ⟨https://www.npa.go.jp/publications/statistics/cybersecurity/data/H30_kami_cyber_jousei.pdf⟩（参照 2019-06-24）．

[2] 鈴木雅貴，中山靖司，古原和邦：インターネット・バンキングに対する Man-in-the-Browser 攻撃への対策「取引認証」の安全性評価，金融研究，Vol.32, No.3, pp.51–76 (2013).

[3] MITB 攻撃によるネットバンキングの不正送金リスクを低減 | 飛天ジャパン，入手先 ⟨https://ftsafe.co.jp/solutions/ocra_mitb/⟩（参照 2019-06-24）．

[4] Man-in-the-Browser の脅威と根本的な解決策，入手先 ⟨http://www.risec.aist.go.jp/files/events/2014/0313-ja/risec-sympo2014-takagi.pdf⟩（参照 2019-06-24）．

[5] 藤田真浩：視覚的形式知を利用した 3DCG 画像 CAPTCHA の研究，静岡大学，博士論文 (2017).

[6] Marcus, D. and Sherstobitoff, R.: Dissecting operation high roller. McAfee, White paper (2012).

[7] The Official CAPTCHA Site, 入手先 ⟨http://www.captcha.net⟩（参照 2019-06-24）．

[8] Weigold, T., Kramp, T., Hermann, R., Höring, F., Buhler, P. and Baentsch, M.: The Zurich Trusted Information Channel – An Efficient Defence against Man-in-the-Middle and Malicious Software Attacks, *Trusted Computing-Challenges and Applications*, pp.75–91 (2008).

[9] 祢宜知孝，森 拓海，平野貴人，小関義博，松田 規，河内清人，米田 健：セキュア SIM を搭載したスマートフォンを利用したトランザクション署名手法の提案，研究報告コンピュータセキュリティ（CSEC），Vol.2015-CSEC-71, No.11, pp.1–8 (2015).

[10] Kumarasubramanian, A., Ostrovsky, R., Pandey, O. and Wadia, A.: Cryptography using captcha puzzles, *Public-Key Cryptography–PKC 2013*, pp.89–106, Springer (2013).

向平 浩貴

2018 年静岡大学情報学部情報科学科卒業．現在，同大学院修士課程．情報セキュリティに関する研究に従事．

神農 泰圭

2016 年静岡大学情報学部情報科学科卒業．2018 年同大学院修士課程修了．在学中，情報セキュリティに関する研究に従事．

土屋 貴史

2015 年静岡大学情報学部情報科学科卒業．2017 年同大学院修士課程修了．同年日本電気株式会社入社．在学中，情報セキュリティに関する研究に従事．

大木 哲史 （正会員）

2002 年早稲田大学理工学部電子情報通信学科卒業．2004 年同大学院理工学研究科電子・情報通信学専攻修士課程修了．2010 年早稲田大学理工学術院情報・ネットワーク専攻博士（工学）取得．2010 年早稲田大学理工学総合研究所次席研究員，2013 年産業技術総合研究所特別研究員を経て，2017 年より静岡大学大学院総合科学技術研究科講師．情報セキュリティ全般，特に個人認証を中心としたネットワークセキュリティに関する研究に従事．電子情報通信学会会員．

高橋 健太 （正会員）

2000年東京大学大学院修士課程修了．2012年同大学院情報理工学系研究科博士課程修了．博士（情報理工学）．2000年（株）日立製作所入社．以来，バイオメトリクス，暗号技術および情報セキュリティの研究開発に従事し，現在ユニットリーダ主任研究員．2015年より東京大学大学院客員准教授．2008年情報処理学会論文賞，2011年SISAP Best paper，2012年IEEE BTAS Best reviewed paper，2015年情報処理学会長尾真記念特別賞，2016年ドコモ・モバイル・サイエンス賞優秀賞，関東地方発明表彰発明奨励賞等受賞．電子情報通信学会会員．

尾形 わかは

1989年東京工業大学理学部物理学科卒業，1991年同大学院総合理工学研究科修士課程修了，1994年同大学院理工学研究科博士後期課程修了．1995年兵庫県立姫路工業大学工学部助手，2000年東京工業大学理財工学研究センター助教授．同大学院イノベーションマネジメント研究科准教授，教授を経て，2016年4月より同大学工学院教授．博士（工学）．情報セキュリティ，主に暗号理論・暗号プロトコルの研究に従事．

西垣 正勝 （正会員）

1990年静岡大学工学部光電機械工学科卒業．1995年同大学院博士課程修了．日本学術振興会特別研究員（PD）を経て，1996年静岡大学情報学部助手．同講師，助教授の後，2010年より同創造科学技術大学院教授．博士（工学）．情報セキュリティ全般，特にヒューマニクスセキュリティ，メディアセキュリティ，ネットワークセキュリティ等に関する研究に従事．2013～2014年情報処理学会コンピュータセキュリティ研究会主査．2015～2016年電子情報通信学会バイオメトリクス研究専門委員会委員長．2016年より日本セキュリティマネジメント学会編集部会長．2019年より情報処理学会情報環境領域委員長．本会フェロー．