(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2022年11月17日(17.11.2022)

WIPO | PCT

(10) 国際公開番号
WO 2022/239166 A1

(51) 国際特許分類:
G06F 21/55 (2013.01)

(21) 国際出願番号： PCT/JP2021/018127

(22) 国際出願日： 2021年5月12日(12.05.2021)

(25) 国際出願の言語： 日本語

(26) 国際公開の言語： 日本語

(71) 出願人: 日本電信電話株式会社 (**NIPPON TELEGRAPH AND TELEPHONE CORPORATION**) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目５番１号 Tokyo (JP).

(72) 発明者: 千葉 大紀 (**CHIBA, Daiki**); 〒1808585 東京都武蔵野市緑町３丁目９－１１ ＮＴＴ 知的財産センタ内 Tokyo (JP). 秋山 満昭 (**AKIYAMA, Mitsuaki**); 〒1808585 東京都武蔵野市緑町３丁目９－１１ ＮＴＴ 知的財産センタ内 Tokyo (JP).

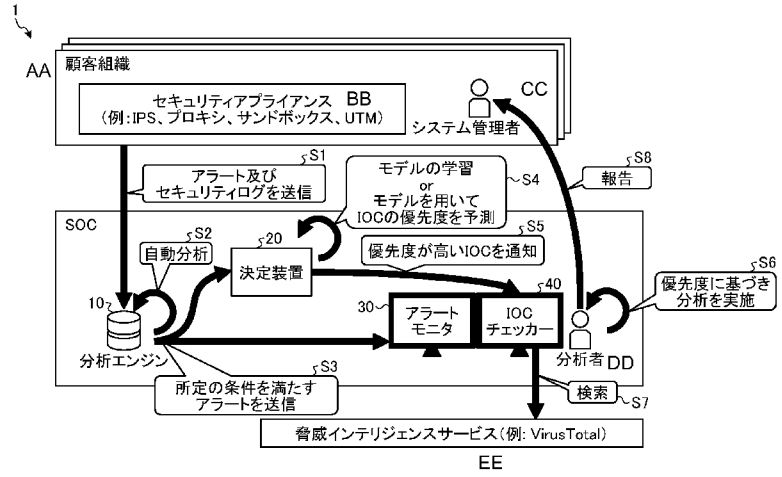
(74) 代理人: 特許業務法人酒井国際特許事務所 (**SAKAI INTERNATIONAL PATENT OFFICE**); 〒1000013 東京都千代田区霞が関３丁目８番１号 虎の門三井ビルディング Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS,

(54) **Title:** EXTRACTION METHOD, EXTRACTION DEVICE, AND EXTRACTION PROGRAM

(54) 発明の名称： 抽出方法、抽出装置及び抽出プログラム



10 Analysis engine
20 Determination device
30 Alert monitor
40 IOC checker
S1 Transmit alert and security log
S2 Automatic analysis
S3 Transmit alert satisfying prescribed conditions
S4 Learn model or use model to predict IOC priority
S5 Perform notification of IOC having high priority
S6 Execute analysis on the basis of priority
S7 Search
S8 Announcement
AA Client organization
BB Security appliance (e.g., IPS, proxy, sandbox, UTM)
CC System manager
DD Analyst
EE Threat intelligence service (e.g. VirusTotal)

(57) **Abstract:** In the present invention, a characteristic information extraction unit (21) acquires observation results, from a prescribed organization, with respect to an indicator of compromise (IOC) included in information related to cyber security. The characteristic information extraction unit (21) creates characteristic information for the IOC, on the basis of information obtained from the observation results acquired by the characteristic information extraction unit (21).

(57) 要約：特徴情報抽出部（２１）は、サイバーセキュリティに関する情報に含まれるＩＯＣに対する所定の組織による観測結果を取得する。特徴情報抽出部（２１）は、特徴情報抽出部（２１）によって取得された観測結果から得られる情報を基に、ＩＯＣの特徴情報を作成する。

WO 2022/239166 A1

［続葉有］

MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類：

一　国際調査報告（条約第21条(3)）

明　細　書

発明の名称 ： 抽出方法、抽出装置及び抽出プログラム

技術分野

[0001] 本発明は、抽出方法、抽出装置及び抽出プログラムに関する。

背景技術

[0002] 企業や組織ではサイバーセキュリティを担保するために、セキュリティ管理や脅威を検知するためのシステムが導入されている。セキュリティオペレーションセンタ（ＳＯＣ：Security Operation Center）は、このようなシステムの運用を行う組織である。ＳＯＣのアナリスト（分析者）は、システムから出力される大量のログやアラートを監視、分析し、必要な対処を行う。

[0003] 一方で、下記の参考文献１、２によると、日々大量に発生するアラートを処理するアナリストが、アラート疲労と呼ばれる状況を起こし、アナリストの燃え尽きにつながることが問題視されている。

参考文献１：S. C. Sundaramurthy, A. G. Bardas, J. Case, X. Ou, M. Wesch, J. McHugh, and S. R. Rajagopalan, “A human capital model for mitigating security analyst burnout,” Proc. SOUPS, 2015.

参考文献２：Ponemon Institute, “Improving the Effectiveness of the Security Operations Center,” 2019.

[0004] また、上記の問題を解決するために必要なことは、よりすぐれた自動化を実現してアナリストの稼働を削減することである。実際、参考文献３によると、多くのＳＯＣのマネージャは、ＳＯＣコンポーネントの自動化レベルが不十分であることが現在のＳＯＣ組織における最重要課題であるととらえている。

参考文献３：F. B. Kokulu, A. Soneji, T. Bao, Y. Shoshitaishvili, Z. Zhao, A. Doupe, and G.-J. Ahn, “Matched and Misma

tched SOCs: A Qualitative Study on Security Operations Center Issues," Proc. ACM CCS, 2019.

[0005] これに対し、例えば、セキュリティに関する各アラートの異常スコアや悪性度スコアを過去のアラートから推定することで、真に悪性なアラートと誤検知となる本来非悪性のアラートとを識別する技術が提案されている（例えば、非特許文献１から５を参照）。

[0006] また、セキュリティに関する各アラートと最も関連性の高い情報を抽出してくることで、アナリストのその後のプロセスをサポートする技術が知られている（例えば、非特許文献６から８を参照）。

先行技術文献

非特許文献

[0007] 非特許文献1：W. U. Hassan, S. Guo, D. Li, Z. Chen, K. Jee, Z. Li, and A. Bates, "NoDoze: Combatting Threat Alert Fatigue with Automated Provenance Triage," Proc. NDSS, 2019.

非特許文献2：A. Oprea, Z. Li, R. Norris, and K. Bowers, "MADE: Security Analytics for Enterprise Threat Detection," Proc. ACSAC, 2018.

非特許文献3：K. A. Roundy, A. Tamersoy, M. Spertus, M. Hart, D. Kats, M. Dell'Amico, and R. Scott, "Smoke Detector: Cross-Product Intrusion Detection With Weak Indicators," Proc. ACSAC, 2017.

非特許文献4：Y. Liu, M. Zhang, D. Li, K. Jee, Z. Li, Z. Wu, J. Rhee, and P. Mittal, "Towards a Timely Causality Analysis for Enterprise Security," Proc. NDSS, 2018.

非特許文献5：P. Najafi, A. Muhle, W. Punter, F. Cheng, and C. Meinel, "MalRank: a measure of maliciousness in SIEM-based knowledge graphs," Proc. ACSAC, 2019.

非特許文献6：C. Zhong, J. Yen, P. Liu, and R. F. Erbacher,

“Automate Cybersecurity Data Triage by Leveraging Human Analysts’ Cognitive Process,” Proc. IEEE IDS, 2016.

非特許文献7：C. Zhong, T. Lin, P. Liu, J. Yen, and K. Chen, “A cyber security data triage operation retrieval system,” Comput. Secur., vol.76, pp.12-31, 2018.

非特許文献8：S. T. Chen, Y. Han, D. H. Chau, C. Gates, M. Hart, and K. A. Roundy, “Predicting Cyber Threats with Virtual Security Products,” Proc. ACSAC, 2017.

非特許文献9：P. Vadrevu, B. Rahbarinia, R. Perdisci, K. Li, and M. Antonakakis, “Measuring and Detecting Malware Downloads in Live Network Traffic,” Proc. ESORICS, 2013.

非特許文献10：M. Antonakakis, R. Perdisci, D. Dagon, W. Lee, and N. Feamster, “Building a dynamic reputation system for DNS,” Proc. USENIX Security Symposium, 2010.

非特許文献11：L. Bilge, E. Kirda, C. Kruegel, and M. Balduzzi, “EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis,” Proc. NDSS, 2011.

非特許文献12：R. Perdisci, I. Corona, and G. Giacinto, “Early Detection of Malicious Flux Networks via Large-Scale Passive DNS Traffic Analysis,” IEEE Trans. Dependable Secure Comput., vol.9, no. 5, pp.714-726, Sep. 2012.

非特許文献13：S. Alrwais, X. Liao, X. Mi, P. Wang, X. Wang, F. Qian, R. Beyah, and D. McCoy, “Under the Shadow of Sunshine: Understanding and Detecting Bulletproof Hosting on Legitimate Service Provider Networks,” Proc. IEEE S&P, 2017.

非特許文献14：B. Rahbarinia, R. Perdisci, and M. Antonakakis, “Segugio: Efficient behavior-based tracking of malware-control domains in large ISP networks,” Proc. IEEE/IFIP DSN, 2015.

発明の概要

発明が解決しようとする課題

[0008] しかしながら、従来の技術には、ＩＯＣの調査の優先度を決定するために有用な特徴情報が得られない場合があるという問題がある。

[0009] 例えば、上記の先行技術文献に記載された技術では、ＩＯＣが異常又は悪性であるか否かを決定するのに必要な特徴情報が採用されている。一方で、ＩＯＣが異常又は悪性であることと、当該ＩＯＣが分析者による詳細な調査が必要であるかどうかは独立である。

課題を解決するための手段

[0010] 上述した課題を解決し、目的を達成するために、抽出方法は、抽出装置によって実行される抽出方法であって、サイバーセキュリティに関する情報に含まれるＩＯＣ（Indicator of Compromise）に対する所定の組織による観測結果を取得する取得工程と、前記取得工程によって取得された観測結果から得られる情報を基に、ＩＯＣの特徴情報を作成する作成工程と、を含むことを特徴とする。

発明の効果

[0011] 本発明によれば、ＩＯＣの調査の優先度を決定するために有用な特徴情報を得ることができる。

図面の簡単な説明

[0012] [図1]図１は、セキュリティシステムについて説明する図である。

[図2]図２は、アラートモニタの画面の例を示す図である。

[図3]図３は、ＩＯＣチェッカーの画面の例を示す図である。

[図4]図４は、第１の実施形態に係る決定装置の構成例を示す図である。

[図5]図５は、リクエストの期間の例を示す図である。

[図6]図６は、学習処理の流れを示すフローチャートである。

[図7]図７は、特徴情報を抽出する処理の流れを示すフローチャートである。

[図8]図８は、予測処理の流れを示すフローチャートである。

[図9]図９は、決定プログラムを実行するコンピュータの一例を示す図である。

発明を実施するための形態

[0013] 以下に、本願に係る抽出方法、抽出装置及び抽出プログラムの実施形態を図面に基づいて詳細に説明する。なお、本発明は、以下に説明する実施形態により限定されるものではない。なお、本実施形態では、決定装置が抽出装置として機能する。

[0014] ［第１の実施形態の構成］

まず、図１を用いて、第１の実施形態に係る決定装置を含むセキュリティシステムについて説明する。図１は、セキュリティシステムについて説明する図である。

[0015] セキュリティシステム１は、顧客組織のセキュリティプライアンスにおいて生じた所定の情報を基に、分析エンジンによる自動分析、又は分析者による分析が行われる。

[0016] セキュリティプライアンスは、例えば侵入防御システム（ＩＰＳ：Intrusion Prevention System）、プロキシ、サンドボックス、統合脅威管理（ＵＴＭ：Unified Threat Management）等である。

[0017] ＳＯＣでは、セキュリティアプライアンスから取得したセキュリティに関する情報をリアルタイムに分析する。例えば、セキュリティに関する情報にはセキュリティログ及びアラートが含まれる。

[0018] 図１の例では、ＳＯＣは大規模なＭＳＳＰ（Managed Security Service Provider）で提供されるアウトソースＳＯＣとして利用されている。一方で、本実施形態は、インハウスＳＯＣにも適用可能である。

[0019] アウトソースＳＯＣとインハウスＳＯＣは組織的な違いはあるが、全体的なワークフローは類似している。このため、スケールメリットを十分に発揮できる規模の組織のインハウスＳＯＣであれば、本実施形態の効果が得られやすい。

[0020] セキュリティシステム１における処理の流れを説明する。図１に示すよう

に、まず、顧客組織のセキュリティアプライアンスは、アラート及びセキュリティログをＳＯＣの分析エンジン１０に送信する（ステップＳ１）。

[0021] 以下、アラートについて処理を行う場合の例を説明する。セキュリティシステム１は、セキュリティログについても、アラートと同様に処理することができる。

[0022] 分析エンジン１０は自動分析を行う（ステップＳ２）。分析エンジン１０は、既知の悪意のある特性や、事前に定義されたルールやブラックリストに基づいて分析を行うことで、アラートに対応する。

[0023] 分析エンジン１０は、ＳＯＡＲ（Security Orchestration, Automation, and Response）と呼ばれる機能を用いて分析を行ってもよい。

[0024] 分析エンジン１０は、所定の条件を満たすアラートを決定装置２０、アラートモニタ３０又はＩＯＣチェッカー４０に送信する（ステップＳ３）。

[0025] このとき、図２に示すように、アラートモニタ３０は、アラートに関する情報を表示する。図２は、アラートモニタの画面の例を示す図である。

[0026] 例えば、アラートモニタ３０には、アラートの原因となったイベントの日付（Date）、顧客名（Customer）、アラートを送信したデバイス（Device）、アラートの名前（Alert Name）、アラートのトリガーとなった状況の概要等が表示される。

[0027] また、図３に示すように、ＩＯＣチェッカー４０は、アラートに含まれるＩＯＣ（Indicator of Compromise）に関する情報を表示する。図３は、ＩＯＣチェッカーの画面の例を示す図である。

[0028] 例えば、ＩＯＣには、ドメイン名、ＩＰアドレス、ＵＲＬ、ファイルハッシュ値等が含まれる。

[0029] 図３に示すように、例えば、ＩＯＣチェッカー４０は、ＳＯＣでの調査状況（Status）、ＩＯＣの悪性度に関するＳＯＣの直近の判断（ＳＯＣ Last Decision）、そのＩＯＣの最新の脅威インテリジェンス結果（Detection in TI）等を表示する。

[0030] 例えば、分析者は、アラートモニタ３０及びＩＯＣチェッカー４０といっ

た、ＩＯＣ評価専用のツールを使って、分析エンジン１０によって処理できなかったアラートについてＩＯＣのトリアージ（評価）を行う。

[0031] ＳＯＣの分析者は日頃のＳＯＣワークフローで大量のアラートを処理する。そこで、決定装置２０は、優先度が高いＩＯＣを決定し、分析者に通知する。これにより、ＳＯＣにおいて複数の分析者が同じＩＯＣを手動で評価することを防止することができる。

[0032] また、決定装置２０によれば、優先度が高いＩＯＣを優先して分析することができるため、分析者の稼働量に対する効果を向上させることができる。

[0033] 決定装置２０は、モデルの学習、又はモデルを用いてＩＯＣの優先度の予測を行う（ステップＳ４）。そして、決定装置２０は、予測結果に基づき優先度が高いＩＯＣを決定し、当該決定したＩＯＣを通知する（ステップＳ５）。

[0034] 例えば、決定装置２０は、決定したＩＯＣを、ＩＯＣチェッカー４０を介して分析者に通知する。

[0035] 分析者は、通知された優先度に基づき分析を実施する（ステップＳ６）。また、分析者は、分析の過程で脅威インテリジェンスサービス（例えば、VirusTotal（https://www.virustotal.com/））の検索を行ってもよい（ステップＳ７）。

[0036] 脅威インテリジェンスサービスの中には、脅威のレベルや悪性度に関するスコアを提供するものがある。しかしながら、本来そのようなスコアは必ずしも分析者の次のアクションを決定するものではない。

[0037] 例えば、既にパッチが展開されている脆弱性を利用した攻撃に関連したＩＯＣは、悪性であるスコアが高いかもしれないが、顧客組織を守るという観点では差し迫った脅威ではない。

[0038] このように、ＳＯＣにおけるアラート分析は単純ではないため、アラート分析を完全自動化することは難しく、分析者による判断が必要になる場合がある。

[0039] このため、決定装置２０による優先度が高いＩＯＣの決定は、分析者の判

断の時間を確保し、また各ＩＯＣの調査稼働を削減するために有用であるということができる。

[0040] 分析者は、最終的に分析対象のアラート及び当該アラートに含まれたＩＯＣが悪性か非悪性かを判断し、さらに顧客への報告が必要かどうかを判断し、顧客への報告が必要な場合、顧客組織のシステム管理者等に報告を行う（ステップＳ８）。

[0041] 例えば、分析者があるＩＯＣの評価を完了すると、その結果に基づいて、分析エンジン１０におけるアラートのトリガーされる条件を変更することができる。

[0042] 例えば、分析者による評価で明らかに悪性のＩＯＣが特定された場合、カスタムブラックリスト又はカスタムシグネチャとして当該ＩＯＣを分析エンジン１０において使用することができる。

[0043] その場合、同一ＩＯＣを含むログをＳＯＣの他の顧客でも自動的に検知できるようにすることができる。また、評価で誤検知や脅威レベルが少ないＩＯＣが特定された場合、アラートをトリガーするＳＩＥＭロジックが変更され、同じ誤検知アラートが再び発生するのを防ぐことができ、分析者の稼働削減につながる。

[0044] 以降、決定装置２０が優先度の高いＩＯＣを決定する処理について、決定装置２０の構成とともに詳細に説明する。

[0045] 図４は、第１の実施形態に係る決定装置の構成例を示す図である。図４に示すように、決定装置２０は、特徴情報抽出部２１、ラベル付与部２２、学習部２３、予測部２４及びモデル情報２５を有する。

[0046] 決定装置２０は、機械学習手法によるモデルの学習処理、及び学習済みのモデルを使った予測処理を行う。

[0047] 学習処理では、特徴情報抽出部２１、ラベル付与部２２及び学習部２３が用いられる。また、予測処理では、特徴情報抽出部２１及び予測部２４が用いられる。

[0048] 特徴情報抽出部２１は、サイバーセキュリティに関する情報に含まれるＩ

ＯＣから特徴情報を抽出する。例えば、サイバーセキュリティに関する情報は、分析エンジン１０から取得するアラートである。

[0049] 特徴情報抽出部２１は、分析エンジン１０から入手した過去のアラートに含まれるＩＯＣから、当該ＩＯＣの特性を特徴づける情報（以後、特徴情報）を抽出する。

[0050] 特徴情報は、ＩＯＣに含まれるドメイン名、ＩＰアドレス、ＵＲＬ、ファイルハッシュ値等であってもよい。

[0051] 例えば、特徴情報抽出部２１は、あらかじめ定められた一定の日数の間に発生したアラートから特徴情報を抽出する。

[0052] ここで、特徴情報抽出部２１による特徴情報の抽出方法を詳細に説明する。特徴情報抽出部２１は、取得部及び作成部を有する抽出装置として機能する。

[0053] 取得部は、サイバーセキュリティに関する情報に含まれるＩＯＣに対する所定の組織による観測結果を取得する。作成部は、取得した観測結果から得られる情報を基に、ＩＯＣの特徴情報を作成する。

[0054] 特徴情報抽出部２１は、脅威インテリジェンスサービスによる観測結果（項目１、２、３）、又はインターネット等のネットワークにおける観測結果（項目４、５）を基に特徴情報を作成する。

[0055] 各項目の特徴情報について説明する。まず、項目１、２及び３の特徴情報は、各ＩＯＣに関連して既に脅威インテリジェンスサービスによって観測された脅威の特性に着目した特徴情報である。

[0056] 特徴情報抽出部２１は、ＩＯＣに関連する事項の脅威インテリジェンスサービスによる検知状況を取得する。特徴情報抽出部２１は、検知状況を基に特徴情報を作成する。

[0057] 脅威インテリジェンスサービスは、顧客組織が用意したものであってもよいし、外部組織が提供しているものであってもよい。例えば、脅威インテリジェンスサービスは、VirusTotalのようなドメイン名やＩＰアドレスやＵＲＬやファイルハッシュ値に関する脅威情報を取得できるサービスである。

[0058] 項目１、２、３に包含される特徴情報は、例えば２８個である。以下、項目Ｘのように表記される項目、及び項目Ｘ－Ｙのように表記される項目の特徴情報は、項目Ｘに包含される特徴情報であるものとする。

[0059]（項目１）

特徴情報抽出部２１は、ドメイン名のＩＯＣがあった場合、脅威インテリジェンスサービスを参照し、（１）当該ドメイン名を含む検知ＵＲＬ、（２）当該ドメイン名宛に通信した検知ファイル、（３）当該ドメイン名からダウンロードされた検知ファイル、（４）当該ドメイン名を言及している検知ファイルの４つの事項の数をそれぞれカウントし、特徴情報とする。

[0060] これにより、特徴情報抽出部２１は、例えば４個の特徴情報を得る。項目１の特徴情報によれば、当該ＩＯＣが既知の脅威と関連しているかどうかを識別することができる。

[0061] ここで、（１）の当該ドメイン名を含む検知ＵＲＬを、ドメイン名部分が共通するＵＲＬのうち、脅威インテリジェンスサービス上の任意の検知エンジンのうち少なくとも１つ以上で検知したものと定義する。

[0062] また、（２）の当該ドメイン名宛に通信した検知ファイルを、サンドボックス化された環境での実行や解析を通じて当該ドメイン名宛に通信することが確認されたファイルのうち、脅威インテリジェンスサービス上の任意の検知エンジンのうち少なくとも１つ以上で検知したものと定義する。

[0063] また、（３）の当該ドメイン名からダウンロードされた検知ファイルを、当該ドメイン名から取得されたファイルのうち、脅威インテリジェンスサービス上の任意の検知エンジンのうち少なくとも１つ以上で検知したものと定義する。

[0064] また、（４）の当該ドメイン名を言及している検知ファイルを、当該ドメイン名の文字列を内部に含むファイルのうち、脅威インテリジェンスサービス上の任意の検知エンジンのうち少なくとも１つ以上で検知したものと定義する。

[0065]（項目２）

特徴情報抽出部２１は、ドメイン名のＩＯＣがあった場合、脅威インテリジェンスサービスを参照し、（１）当該ドメイン名を含む非検知ＵＲＬ、（２）当該ドメイン名宛に通信した非検知ファイル、（３）当該ドメイン名からダウンロードされた非検知ファイル、（４）当該ドメイン名を言及している非検知ファイルの数の４つの事項の数をそれぞれカウントし、特徴情報とする。

[0066] 項目２の特徴情報は、項目１の「検知された」という部分を「検知されなかった」と置き換えた特徴情報に相当する。検知されなかったＵＲＬやファイルとは、脅威インテリジェンスサービスによって検査されたが、いずれの検知エンジンによっても悪意のあるもの又は不審なものとしては検出されなかったことを意味する。

[0067] これにより、特徴情報抽出部２１は、例えば４個の特徴情報を得る。項目２の特徴情報によれば、当該ＩＯＣが良性又は正規なものかどうかを識別することができる。

[0068]（項目３）

特徴情報抽出部２１は、項目１の（１）から（４）の各事項について、脅威インテリジェンスサービスに存在する複数の検知エンジンのうち、何個の検知エンジンが検知したかという検知数の情報を収集する。

[0069] さらに、特徴情報抽出部２１は、４種類の検知数について、５つの統計量（平均値、最小値、最大値、標準偏差、分散）を計算して、合計２０個の特徴情報を作成する。

[0070] このように、特徴情報抽出部２１は、観測結果から得られる情報、及び情報から計算される統計量を基に、特徴情報を作成する。

[0071] 項目３の特徴情報によれば、検知されたＵＲＬやファイルが、より多くの検知エンジンで検出されているメジャーな脅威なのか、それとも少数の検知エンジンでしか検知されていないマイナーな脅威なのかを区別することができる。

[0072] 項目４及び５の、各ＩＯＣに関連するネットワーク内で観測される通信の

特性に着目した特徴情報について説明する。ネットワークは、例えばインターネットである。

[0073] 具体的には、特徴情報抽出部２１は、各ＩＯＣがあるネットワークでどの程度参照されたのかという情報を取得するために、Passive ＤＮＳ（Domain Name System）データベースを利用する。

[0074] Passive ＤＮＳデータベースとは、任意のキャッシュＤＮＳサーバや権威ＤＮＳサーバにおいてその通信を観測し、実際にやりとりされたＤＮＳメッセージから、ドメイン名とＩＰアドレスの対応関係やその履歴を記録したデータベースである。

[0075] Passive ＤＮＳデータベースは、顧客組織が用意したものであってもよいし、外部組織が提供しているものであってもよい。

[0076] 特徴情報抽出部２１は、ネットワーク内で観測される通信特性に関する特徴情報として、項目４及び５に包含される５つの項目、合計１４７個の特徴情報を抽出する。

[0077]（項目４）

特徴情報抽出部２１は、ＩＯＣに紐付くドメイン名に対応するＤＮＳ（Domain Name System）レコードを観測結果として取得し、ＤＮＳレコードの情報の変更回数を基に、例えば７個の特徴情報を作成する。

[0078] 例えば、特徴情報抽出部２１は、ドメイン名のＩＯＣがあった場合、Passive ＤＮＳデータベースを参照し、ドメイン名に対応する７種類のＤＮＳリソースレコード（Ａ、ＡＡＡＡ、ＣＮＡＭＥ、ＭＸ、ＮＳ、ＳＯＡ、ＴＸＴ）ごとに、過去のある時点から現在までのリソースレコードの変更回数を特徴情報としてカウントする。

[0079] 例えば、図５の上側の表に示すように、「example.com」のＡレコード（ＩＰｖ４アドレス）が過去に「192.0.2.1」で、その後「203.0.113.1」になったことが観測されている場合、特徴情報抽出部２１は、ドメイン名「example.com」に相当するＩＯＣの変更回数を１とカウントする。

[0080] 項目４の特徴情報によれば、ＤＮＳレコードそのものが頻繁に変更されて

いるドメイン名と、安定的に利用されているドメイン名とを区別することができる。

[0081] なお、ドメイン名以外のＩＯＣの場合、特徴情報抽出部２１は、当該ＩＯＣをドメイン名に紐付けた上で上記の特徴情報を抽出すればよい。

[0082] 例えば、特徴情報抽出部２１は、ＩＯＣがＵＲＬ「https://www.example.com」である場合、当該ＩＯＣをドメイン名部分「www.example.com」と紐付けてカウントを行う。

[0083] また、ＩＯＣがＩＰアドレスの場合、特徴情報抽出部２１は、ＤＮＳの逆引きレコードを参照して対応するドメイン名を入手するか、あるいは、Passive ＤＮＳデータベースを使うことで当該ＩＰアドレスに紐付いていたドメイン名を抽出できる。

[0084] さらに、ＩＯＣがファイルハッシュ値の場合、特徴情報抽出部２１は、脅威インテリジェンスサービスを参照して、当該ファイルが通信した先あるいは当該ファイルをダウンロードした元のドメイン名を抽出できる。

[0085]（項目５）

特徴情報抽出部２１は、ＩＯＣに紐付くドメイン名に対応するＤＮＳ（Domain Name System）レコードを観測結果として取得し、ＤＮＳレコードの利用回数及び利用期間を基に、例えば１４０個の特徴情報を作成する。

[0086]（項目５－１）

特徴情報抽出部２１は、過去のＤＮＳクエリ数の平均値、最小値、最大値、標準偏差、分散に基づく、例えば３５個の特徴情報を作成する。

[0087] 例えば、ドメイン名のＩＯＣがあった場合、特徴情報抽出部２１は、まず項目４と同様に、Passive ＤＮＳデータベースを参照し、ドメイン名に対応する７種類のＤＮＳリソースレコード（Ａ、ＡＡＡＡ、ＣＮＡＭＥ、ＭＸ、ＮＳ、ＳＯＡ、ＴＸＴ）の中から、各組み合わせのＤＮＳクエリ回数をカウントする。

[0088] ここで、ＤＮＳクエリ回数を、Passive ＤＮＳデータベースにおいて、リソースレコードの組み合わせ（例えば「example.com」、Ａレコード、「192.

0.2.1」）が観測された回数と定義する。

[0089] 図５の上側には、「example.com」の２つの過去のＤＮＳ　Ａレコードと、それらのＤＮＳクエリ回数（それぞれ５，０００回と１５，０００回）の例が示されている。

[0090] 次に、特徴情報抽出部２１は、７種類のリソースレコードについて、５つの統計量（平均値、最小値、最大値、標準偏差、分散）を計算し、合計で３５個の特徴情報を作成する。

[0091] 項目５－１の特徴情報によれば、ドメイン名にアクセスしているインターネットユーザーの数の傾向を反映させることができる。

[0092]（項目５－２）

特徴情報抽出部２１は、最初のＤＮＳクエリからの経過日数の平均値、最小値、最大値、標準偏差、分散に基づく、例えば３５個の特徴情報を作成する。

[0093] 例えば、ドメイン名のＩＯＣがあった場合、特徴情報抽出部２１は、まず項目４と同様に、Passive　ＤＮＳデータベースを参照し、ドメイン名に対応する７種類のＤＮＳリソースレコード（Ａ、ＡＡＡＡ、ＣＮＡＭＥ、ＭＸ、ＮＳ、ＳＯＡ、ＴＸＴ）ごとに、各組み合わせの最初のＤＮＳクエリが行われた日付を抽出する。

[0094] 次に、特徴情報抽出部２１は、各々の日付について、当該日付から特徴情報を抽出する当日までの経過日数を算出する。

[0095] 図５に「example.com」に対する２つのレコードとその最初のＤＮＳクエリが観測された日付の例を示す。

[0096] 例えば、図５に示すように、特徴情報を抽出するのが２０２０年６月１日とした場合、特徴情報抽出部２１は、１つ目のレコードについては２０１９－１０－３１から２０２０－０６－０１までの日数、２つ目のレコードについては２０２０－０１－２４から２０２０－０６－０１までの日数をカウントする。

[0097] 項目５－２の特徴情報によれば、各レコードの使用開始時期を中心に、ド

メイン名に関するＤＮＳトレンドを反映させることができる。

[0098]（項目５－３）

特徴情報抽出部２１は、最後のＤＮＳクエリからの経過日数の平均値、最小値、最大値、標準偏差、分散に基づく、３５個の特徴情報を作成する。

[0099] 例えば、ドメイン名のＩＯＣがあった場合、特徴情報抽出部２１は、項目５－２の「最初のＤＮＳクエリ」を「最後のＤＮＳクエリ」に変更して特徴情報を抽出する。

[0100] 例えば、図５に示すように、特徴情報を抽出するのが２０２０年６月１日とした場合、特徴情報抽出部２１は、１つ目のレコードについては２０２０－０１－２３から２０２０－０６－０１まで、２つ目のレコードについては２０２０－０４－０１から２０２０－０６－０１までの日数をカウントする。

[0101] その後、特徴情報抽出部２１は、７種類のリソースレコード毎にカウントした日数の５つの統計量（平均値、最小値、最大値、標準偏差、分散）を計算し、合計で３５個の特徴情報を作成する。

[0102] 項目５－３の特徴情報によれば、各レコードの使用が停止した時期に着目して、ドメイン名に関するＤＮＳトレンドを反映させることができる。

[0103]（項目５－４）

特徴情報抽出部２１は、ＤＮＳクエリが存在していた期間の平均値、最小値、最大値、標準偏差、分散に基づく。３５個の特徴情報を作成する。

[0104] 例えば、ドメイン名のＩＯＣがあった場合、特徴情報抽出部２１は、ドメイン名に対応する７種類のＤＮＳリソースレコード（Ａ、ＡＡＡＡ、ＣＮＡＭＥ、ＭＸ、ＮＳ、ＳＯＡ、ＴＸＴ）ごとに、項目５－２と同様に最初のＤＮＳクエリの日付を、項目５－３と同様に最後のＤＮＳクエリの日付を入手する。

[0105] 図５の例では、特徴情報抽出部２１は、１つ目のレコードについては２０１９－１０－３１から２０２０－０１－２３までの日数、２つ目のレコードについては２０２０－０１－２４から２０２０－０４－０１までの日数をカ

ウントする。

[0106] その後、特徴情報抽出部２１は、７種類のリソースレコード毎にカウントした日数の５つの統計量（平均値、最小値、最大値、標準偏差、分散）を計算し、合計で３５個の特徴情報を作成する。

[0107] 項目５－４の特徴情報によれば、各レコードがどのくらいの期間使用されているかに注目して、ドメイン名に関するＤＮＳトレンドを反映させることができる。

[0108] ラベル付与部２２は、ＩＯＣのそれぞれについて、関連するアラートの対応に要した稼働量の実績に応じたラベルを付与する。

[0109] ここでは、ラベルは優先度が高いか否かを表す二値データであるものとする。例えば、ラベル付与部２２は、過去に分析者の稼働を多く消費したＩＯＣについては、優先度が高いことを示すラベルを付与し、そうでないものについては優先度が高くないことを示すラベルを付与する。

[0110] なお、従来技術（例えば、非特許文献４から８に記載の技術）では、ＩＯＣが悪性のもの（又は悪意があるもの）であるか否かを示すラベルが付与されていた。一方で、本実施形態では、分析者の稼働量に基づきラベルが付与される。

[0111] ラベル付与部２２は、ＩＯＣのうち、関連するアラートに対して一定期間内に発生した手動調査の回数が所定値以上であるＩＯＣについて、優先度が高いことを示すラベルを付与し、手動調査の回数が所定値未満であるＩＯＣについて、優先度が高くないことを示すラベルを付与する。

[0112] 以降の説明では、優先度が高いことを示すラベルを「優先」、優先度が高くないことを示すラベルを「非優先」と表記する。

[0113] 学習部２３は、特徴情報抽出部２１によって抽出された特徴情報及びラベル付与部２２によって付与されたラベルを組み合わせた学習データを用いて、ＩＯＣの特徴情報からラベルを出力するモデルの学習を行う。

[0114] 学習部２３は、教師あり機械学習により、モデルの作成及び更新を行う。モデル情報２５は、モデルを構築するためのパラメータ等を含む情報である

。学習部２３は、モデル情報２５の作成及び更新を行う。

[0115] 学習部２３は、既知の任意の教師あり機械学習のアルゴリズムを採用することが可能である。本実施形態では、学習部２３は標準的なロジスティック回帰を採用するものとする。

[0116] ロジスティック回帰は、スケーラブルで高速なので、ＳＯＣ環境のように多くの顧客からの大量のアラートに含まれるＩＯＣを予測するのに適している。

[0117] また、ロジスティック回帰は解釈可能性が高いことが知られている。ロジスティック回帰の出力は、その性質上、入力されたＩＯＣが優先される確率として解釈でき、さらに各ＩＯＣに対応する特徴情報のうちどの特徴が結果に貢献しているかを示すことができる。このように、ロジスティック回帰には解釈可能性が高いという利点がある。

[0118] ここでは、学習部２３は、特にＬ１正則化つきロジスティック回帰を利用するものとする。

[0119] まず、学習部２３は、特徴情報抽出部２１によって抽出された特徴情報を表すベクトルｘが与えられたとき、（１）式に示すラベルの条件付き確率ｙを、（２）式のようにモデル化する。

[0120] [数1]

$$y \in \{0(non-priority),\ 1(priority)\} \quad \cdots(1)$$

[0121] [数2]

$$p(y=1|\mathbf{x};\theta) = \sigma(\theta^{\mathrm{T}}\mathbf{x}) = 1/(1+\exp(-\theta^{\mathrm{T}}\mathbf{x})) \quad \cdots(2)$$

[0122] ここで、θはロジスティック回帰モデルのパラメータである。また、σはシグモイド関数である。また、ｘの全ての特徴は、［０，１］の範囲に正規化されるものとする。

[0123] 学習部２３は、正則化の度合いを決定するハイパーパラメータλを導入した（４）式の目的関数を最小化する際のパラメータθを求めるために、（３）式に示すｎ個のラベル付きの学習用データの集合を使用する。

[0124]

[数3]

$$\{(x_i, y_i)\}_{i=1}^{n} \quad \cdots (3)$$

[0125] [数4]

$$\min_{\theta} \sum_{i=1}^{n} -\log p(y_i|x_i;\theta) + \lambda\|\theta\|_1 \quad \cdots (4)$$

[0126] （４）式のうち、Ｌ１正則化部分λ｜｜θ｜｜$_1$は、目的関数にペナルティを加えており、有意に寄与しない特徴情報を識別して削減する効果がある。

[0127] このような特徴量の削減は必要以上に学習データに合わせてしまうオーバーフィッティングの防止に寄与するだけでなく、メモリ使用量の削減や、ＳＯＣアナリストに提示する結果をより簡潔で解釈しやすいものにする効果がある。

[0128] 予測部２４は、学習部２３による学習が行われたモデルを用いて、ＩＯＣの特徴情報からラベルを予測する。

[0129] 予測部２４は、学習部２３によって学習が行われたモデルを利用して、新たにリアルタイムで発生したアラートに含まれるＩＯＣと対応する特徴情報を入力し、どのＩＯＣが将来的に分析者の稼働を多く消費することになるのかを予測する。

[0130] 例えば、予測部２４は、モデル情報２５を基に構築したロジスティック回帰モデルを使って予測を行う。

[0131] 例えば、予測部２４は、分析者が対象のＩＯＣをＰ日以内にＫ回以上手動で分析する確率を予測することである（ただし、Ｐは整数）。

[0132] 予測部２４は、学習部２３によって決定されたパラメータθを用いて、ＩＯＣに対応する特徴情報のベクトルｘが「優先」である確率ｐ求め、予測するラベル＾ｙ（ｙの直上に＾）を（５）式で定義する。

[0133] [数5]

$$\hat{y} = \begin{cases} 1, & \text{if } p(y|\mathbf{x};\theta) \geq 0.5 \\ 0, & \text{otherwise} \end{cases} \quad \cdots (5)$$

[0134] 決定装置２０は、予測部２４によって予測されたラベルを基に、ＳＯＣの

分析者による繰り返しの調査につながると考えられるＩＯＣ、すなわち「優先」ラベルが予測されたＩＯＣを、確率ｐが高い順に出力し、分析者に提示する。

[0135] このとき、分析者は、決定装置２０によって提示された情報を利用して、調査対象の優先順位付けを行い、効率的にトリアージや詳細分析を行うことができる。

[0136] ＳＯＣの分析者は、ＩＯＣに対してどのようなアクションを取るべきかを可能な限り決定し記録することが求められる。

[0137] 本実施形態によれば、分析者は優先度が高いＩＯＣを調査し、その結果を分析エンジン１０に反映させることができる。それによって、分析エンジン１０は同じＩＯＣを含むアラートを自動処理できるようになるため、分析者が毎回当該ＩＯＣを手動で調査することを回避し、ＳＯＣ全体としての稼働量の削減を図ることができる。

[0138] 例えば、分析者は優先度が高いと決定されたＩＯＣを調査し、その結果を基に分析エンジン１０に当該ＩＯＣを自動分析させるようにする。これにより、当該ＩＯＣは他の分析者に受け渡されることがなくなるため、稼働量が削減される。

[0139] なお、決定装置２０は、学習処理をオフラインで定期的（例えば１日に１回）に再実行し、モデル情報２５を更新する。決定装置２０は、図５に示す特徴情報抽出時点前後の所定の期間のデータを利用して学習処理を行う。例えば、決定装置２０は、特徴抽出時点までのＦ日間と特徴情報抽出時点からのＬ日間を合わせたＦ＋Ｌ日間のデータを利用して学習処理を行う（ただし、Ｆ及びＬは整数）。

[0140] 一方、決定装置２０が、顧客組織からのアラートに含まれるＩＯＣをリアルタイムに処理する際、すなわち予測処理を行う際には、当該ＩＯＣに対して、過去Ｆ日間分のデータを利用して特徴情報を抽出する。

[0141] そして、決定装置２０は、抽出した特徴情報から、未来のＰ日間に分析者によるＫ回以上の手動調査が実施される確率ｐを計算する。

[0142] 決定装置２０は、上記の予測処理をリアルタイムに受信するＩＯＣごとに繰り返す。その結果、分析者が優先的に調査するべきＩＯＣのリストが、図３のようにＩＯＣチェッカー４０の画面に表示され継続的に更新される。

[0143] ［第１の実施形態の処理］

図６は、学習処理の流れを示すフローチャートである。図６に示すように、まず、決定装置２０は、過去のアラートの入力を受け付ける（ステップＳ１０１）。

[0144] 次に、決定装置２０は、入力されたアラートに含まれるＩＯＣから特徴情報を抽出する（ステップＳ１０２）。続いて、決定装置２０は、各ＩＯＣに対する分析者の稼働量に基づいて優先度に関する正解ラベルを付与する（ステップＳ１０３）。

[0145] そして、決定装置２０は、正解ラベルを用いて、特徴情報から優先度に関するラベルを出力するモデルを学習する（ステップＳ１０４）。

[0146] 図７は、特徴情報を抽出する処理の流れを示すフローチャートである。図７の処理は、図６のステップＳ１０２に相当する。

[0147] まず、図７に示すように、決定装置２０は、ＩＯＣの観測結果を取得する（ステップＳ１０２ａ）。

[0148] 次に、決定装置２０は、脅威インテリジェンスサービスによる検知状況に基づく特徴情報を作成する（項目１、２、３）（ステップＳ１０２ｂ）。さらに、決定装置２０は、ＩＯＣに紐付くドメイン名に対応するＤＮＳレコードに基づく特徴情報を作成する（項目４、５）（ステップＳ１０２ｃ）。

[0149] 図８は、予測処理の流れを示すフローチャートである。図８に示すように、まず、決定装置２０は、直近のアラートの入力を受け付ける（ステップＳ２０１）。

[0150] 次に、決定装置２０は、入力されたアラートに含まれるＩＯＣから特徴情報を抽出する（ステップＳ２０２）。続いて、決定装置２０は、各ＩＯＣに対する分析者の稼働量に基づいて正解ラベルを抽出する（ステップＳ２０３）。

[0151] そして、決定装置２０は、特徴情報を学習済みのモデルに入力して、優先度に関するラベルを予測する（ステップＳ２０４）。

[0152] 決定装置２０は、予測したラベルに基づき優先度の高いＩＯＣをＳＯＣの分析者に通知することができる。

[0153] ［第１の実施形態の効果］

これまで説明してきたように、特徴情報抽出部２１は、サイバーセキュリティに関する情報に含まれるＩＯＣに対する所定の組織による観測結果を取得する。特徴情報抽出部２１は、特徴情報抽出部２１によって取得された観測結果から得られる情報を基に、ＩＯＣの特徴情報を作成する。

[0154] これにより、ＩＯＣの調査の優先度を決定するために有用な特徴情報を得ることができる。

[0155] 特徴情報抽出部２１は、ＩＯＣに関連する事項の脅威インテリジェンスサービスによる検知状況を取得する。特徴情報抽出部２１は、検知状況を基に、特徴情報を作成する。

[0156] これにより、特徴情報抽出部２１は、ＩＯＣが、悪性であるか、良性であるか、又はＩＯＣの脅威の度合いを特徴情報に反映させることができる。

[0157] 特徴情報抽出部２１は、ＩＯＣに紐付くドメイン名に対応するＤＮＳレコードを観測結果として取得する。特徴情報抽出部２１は、ＤＮＳレコードの情報の変更回数を基に、特徴情報を作成する。

[0158] これにより、特徴情報抽出部２１は、ＤＮＳレコードそのものが頻繁に変更されているドメイン名と、安定的に利用されているドメイン名とを区別することができる。

[0159] 特徴情報抽出部２１は、ＩＯＣに紐付くドメイン名に対応するＤＮＳレコードを観測結果として取得する。特徴情報抽出部２１は、ＤＮＳレコードの利用回数及び利用期間を基に、特徴情報を作成する。

[0160] これにより、特徴情報抽出部２１は、ドメイン名に関するＤＮＳトレンドを特徴情報に反映させることができる。

[0161] 特徴情報抽出部２１は、観測結果から得られる情報、及び情報から計算さ

れる統計量を基に、特徴情報を作成する。

[0162] これにより、特徴情報抽出部２１は、限られた情報からさらに多くの特徴情報を得ることができる。

[0163] ［システム構成等］

また、図示した各装置の各構成要素は機能概念的なものであり、必ずしも物理的に図示のように構成されていることを要しない。すなわち、各装置の分散及び統合の具体的形態は図示のものに限られず、その全部又は一部を、各種の負荷や使用状況等に応じて、任意の単位で機能的又は物理的に分散又は統合して構成することができる。さらに、各装置にて行われる各処理機能は、その全部又は任意の一部が、ＣＰＵ（Central Processing Unit）及び当該ＣＰＵにて解析実行されるプログラムにて実現され、あるいは、ワイヤードロジックによるハードウェアとして実現され得る。なお、プログラムは、ＣＰＵだけでなく、ＧＰＵ等の他のプロセッサによって実行されてもよい。

[0164] また、本実施形態において説明した各処理のうち、自動的に行われるものとして説明した処理の全部又は一部を手動的に行うこともでき、あるいは、手動的に行われるものとして説明した処理の全部又は一部を公知の方法で自動的に行うこともできる。この他、上記文書中や図面中で示した処理手順、制御手順、具体的名称、各種のデータやパラメータを含む情報については、特記する場合を除いて任意に変更することができる。

[0165] ［プログラム］

一実施形態として、決定装置２０は、パッケージソフトウェアやオンラインソフトウェアとして上記の決定処理を実行する決定プログラムを所望のコンピュータにインストールさせることによって実装できる。例えば、上記の決定プログラムを情報処理装置に実行させることにより、情報処理装置を決定装置２０として機能させることができる。ここで言う情報処理装置には、デスクトップ型又はノート型のパーソナルコンピュータが含まれる。また、その他にも、情報処理装置にはスマートフォン、携帯電話機やＰＨＳ（Perso

nal Handyphone System）等の移動体通信端末、さらには、ＰＤＡ（Personal Digital Assistant）等のスレート端末等がその範疇に含まれる。

[0166] また、決定装置２０は、ユーザが使用する端末装置をクライアントとし、当該クライアントに上記の決定処理に関するサービスを提供する決定サーバ装置として実装することもできる。例えば、決定サーバ装置は、セキュリティに関するアラートを入力とし、優先度の高いＩＯＣを出力とする決定サービスを提供するサーバ装置として実装される。この場合、決定サーバ装置は、Ｗｅｂサーバとして実装することとしてもよいし、アウトソーシングによって上記の決定処理に関するサービスを提供するクラウドとして実装することとしてもかまわない。

[0167] 図９は、決定プログラムを実行するコンピュータの一例を示す図である。コンピュータ１０００は、例えば、メモリ１０１０、ＣＰＵ１０２０を有する。また、コンピュータ１０００は、ハードディスクドライブインタフェース１０３０、ディスクドライブインタフェース１０４０、シリアルポートインタフェース１０５０、ビデオアダプタ１０６０、ネットワークインタフェース１０７０を有する。これらの各部は、バス１０８０によって接続される。

[0168] メモリ１０１０は、ＲＯＭ（Read Only Memory）１０１１及びＲＡＭ（Random Access Memory）１０１２を含む。ＲＯＭ１０１１は、例えば、ＢＩＯＳ（Basic Input Output System）等のブートプログラムを記憶する。ハードディスクドライブインタフェース１０３０は、ハードディスクドライブ１０９０に接続される。ディスクドライブインタフェース１０４０は、ディスクドライブ１１００に接続される。例えば磁気ディスクや光ディスク等の着脱可能な記憶媒体が、ディスクドライブ１１００に挿入される。シリアルポートインタフェース１０５０は、例えばマウス１１１０、キーボード１１２０に接続される。ビデオアダプタ１０６０は、例えばディスプレイ１１３０に接続される。

[0169] ハードディスクドライブ１０９０は、例えば、ＯＳ１０９１、アプリケー

ションプログラム１０９２、プログラムモジュール１０９３、プログラムデータ１０９４を記憶する。すなわち、決定装置２０の各処理を規定するプログラムは、コンピュータにより実行可能なコードが記述されたプログラムモジュール１０９３として実装される。プログラムモジュール１０９３は、例えばハードディスクドライブ１０９０に記憶される。例えば、決定装置２０における機能構成と同様の処理を実行するためのプログラムモジュール１０９３が、ハードディスクドライブ１０９０に記憶される。なお、ハードディスクドライブ１０９０は、ＳＳＤ（Solid State Drive）により代替されてもよい。

[0170] また、上述した実施形態の処理で用いられる設定データは、プログラムデータ１０９４として、例えばメモリ１０１０やハードディスクドライブ１０９０に記憶される。そして、ＣＰＵ１０２０は、メモリ１０１０やハードディスクドライブ１０９０に記憶されたプログラムモジュール１０９３やプログラムデータ１０９４を必要に応じてＲＡＭ１０１２に読み出して、上述した実施形態の処理を実行する。

[0171] なお、プログラムモジュール１０９３やプログラムデータ１０９４は、ハードディスクドライブ１０９０に記憶される場合に限らず、例えば着脱可能な記憶媒体に記憶され、ディスクドライブ１１００等を介してＣＰＵ１０２０によって読み出されてもよい。あるいは、プログラムモジュール１０９３及びプログラムデータ１０９４は、ネットワーク（ＬＡＮ（Local Area Network）、ＷＡＮ（Wide Area Network）等）を介して接続された他のコンピュータに記憶されてもよい。そして、プログラムモジュール１０９３及びプログラムデータ１０９４は、他のコンピュータから、ネットワークインタフェース１０７０を介してＣＰＵ１０２０によって読み出されてもよい。

符号の説明

[0172] １　セキュリティシステム

１０　分析エンジン

２０　決定装置

２１　特徴情報抽出部

２２　ラベル付与部

２３　学習部

２４　予測部

２５　モデル情報

３０　アラートモニタ

４０　ＩＯＣチェッカー

請求の範囲

[請求項1] 抽出装置によって実行される抽出方法であって、

サイバーセキュリティに関する情報に含まれるＩＯＣ（Indicator of Compromise）に対する所定の組織による観測結果を取得する取得工程と、

前記取得工程によって取得された観測結果から得られる情報を基に、ＩＯＣの特徴情報を作成する作成工程と、

を含むことを特徴とする抽出方法。

[請求項2] 前記取得工程は、前記ＩＯＣに関連する事項の脅威インテリジェンスサービスによる検知状況を取得し、

前記作成工程は、前記検知状況を基に、前記特徴情報を作成することを特徴とする請求項１に記載の抽出方法。

[請求項3] 前記取得工程は、前記ＩＯＣに紐付くドメイン名に対応するＤＮＳ（Domain Name System）レコードを前記観測結果として取得し、

前記作成工程は、前記ＤＮＳレコードの情報の変更回数を基に、前記特徴情報を作成することを特徴とする請求項１に記載の抽出方法。

[請求項4] 前記取得工程は、前記ＩＯＣに紐付くドメイン名に対応するＤＮＳ（Domain Name System）レコードを前記観測結果として取得し、

前記作成工程は、前記ＤＮＳレコードの利用回数及び利用期間を基に、前記特徴情報を作成することを特徴とする請求項１に記載の抽出方法。

[請求項5] 前記作成工程は、前記観測結果から得られる情報、及び前記情報から計算される統計量を基に、前記特徴情報を作成することを特徴とする請求項１から４のいずれか１項に記載の抽出方法。

[請求項6] サイバーセキュリティに関する情報に含まれるＩＯＣ（Indicator of Compromise）に対する所定の組織による観測結果を取得する取得部と、

前記取得部によって取得された観測結果から得られる情報を基に、

ＩＯＣの特徴情報を作成する作成部と、

を有することを特徴とする抽出装置。

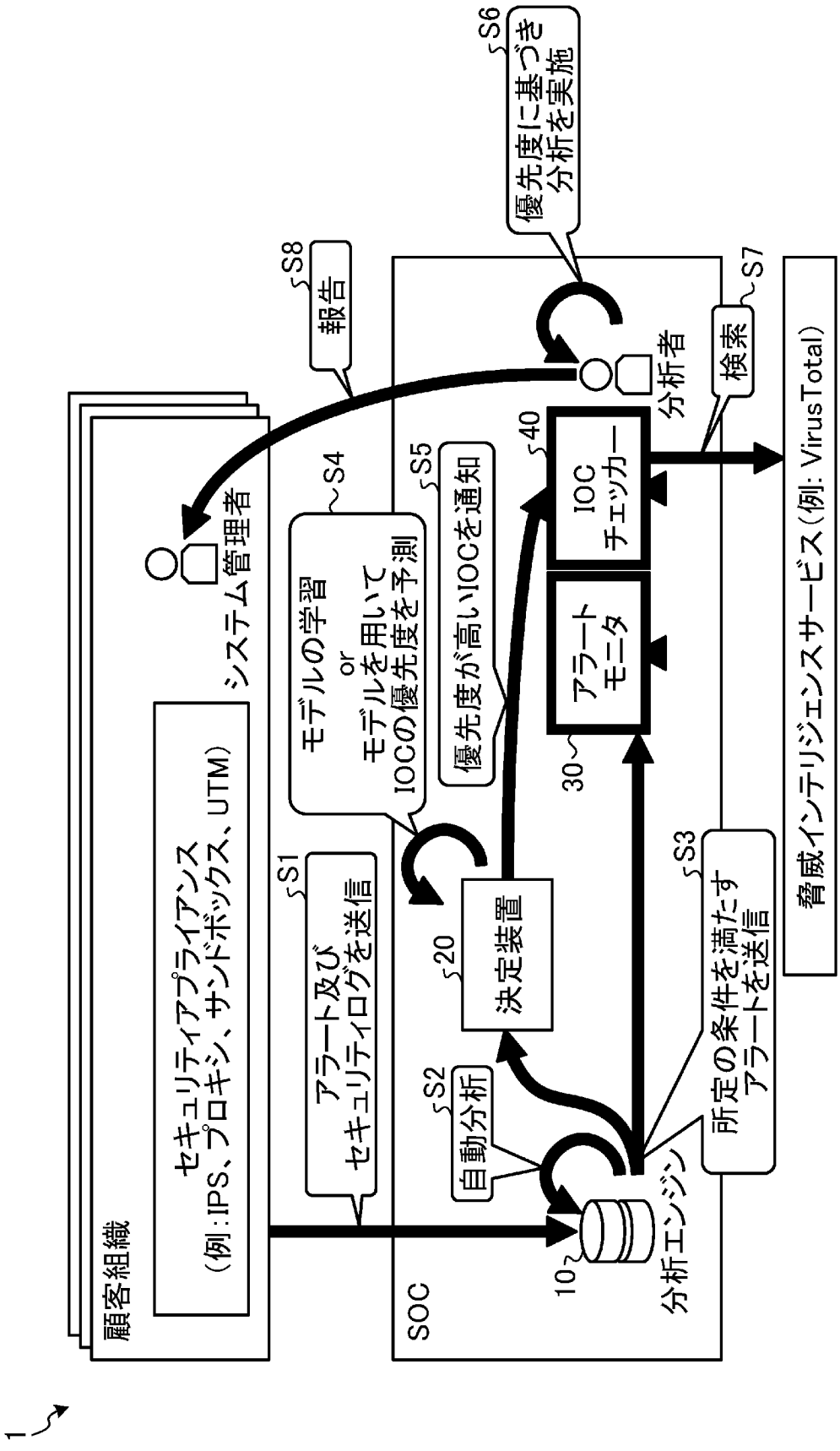
[請求項7] コンピュータに、

サイバーセキュリティに関する情報に含まれるＩＯＣ（Indicator of Compromise）に対する所定の組織による観測結果を取得する取得手順と、

前記取得手順によって取得された観測結果から得られる情報を基に、ＩＯＣの特徴情報を作成する作成手順と、

を実行させることを特徴とする抽出プログラム。

[図1]


1
顧客組織
セキュリティアプライアンス
（例：IPS、プロキシ、サンドボックス、UTM）
システム管理者
SOC
分析エンジン
10
アラート及び
セキュリティログを送信
S1
自動分析
S2
決定装置
20
所定の条件を満たす
アラートを送信
S3
モデルの学習
or
モデルを用いて
IOCの優先度を予測
S4
優先度が高いIOCを通知
S5
アラート
モニタ
30
IOC
チェッカー
40
優先度に基づき
分析を実施
S6
検索
S7
報告
S8
分析者
脅威インテリジェンスサービス（例：VirusTotal）

[図2]

30

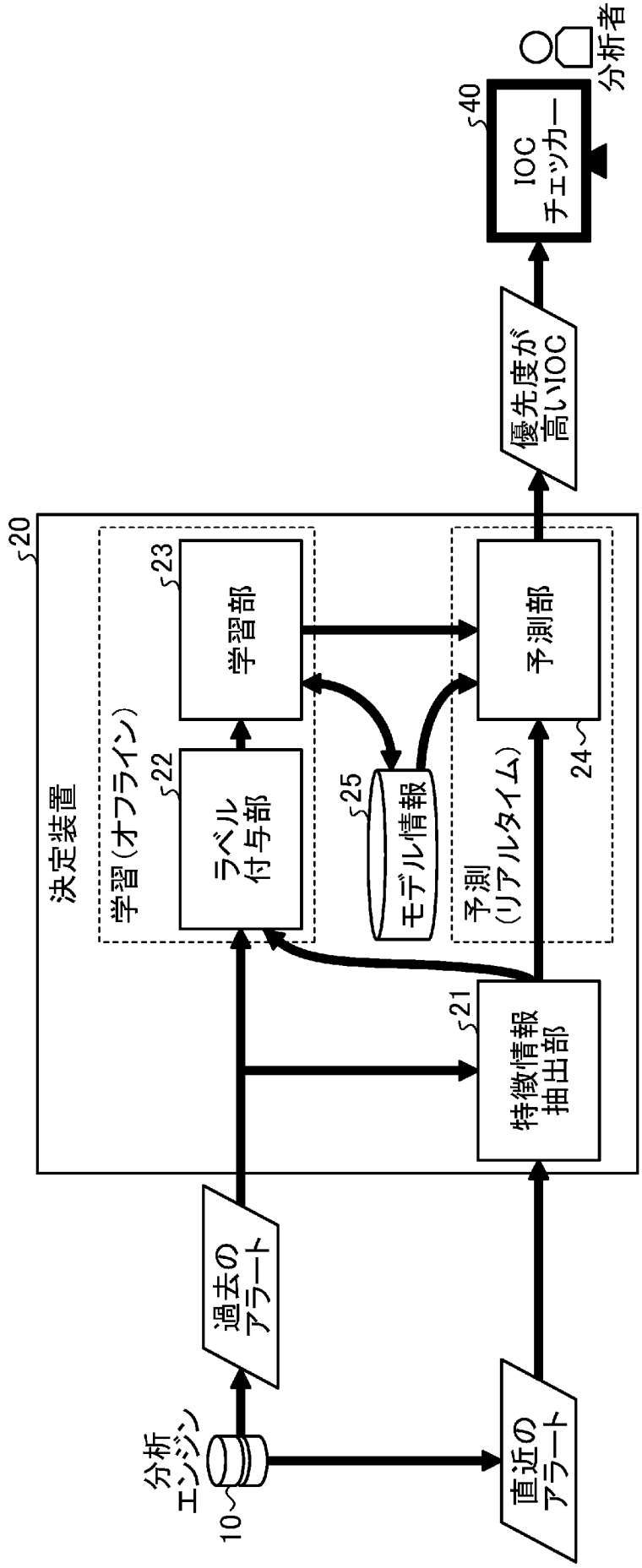
Data	Customer	Device	Alert Name	···
2020-06-01 10:04:00	Foo Corp.	Sandbox	Malware Callback	···
2020-06-01 10:05:30	Bar Inc.	Proxy	Phishing Website	···
2020-06-01 10:06:50	Baz Ltd.	UTM	Suspicious DNS Query	···
···	···	···	···	···

[図3]

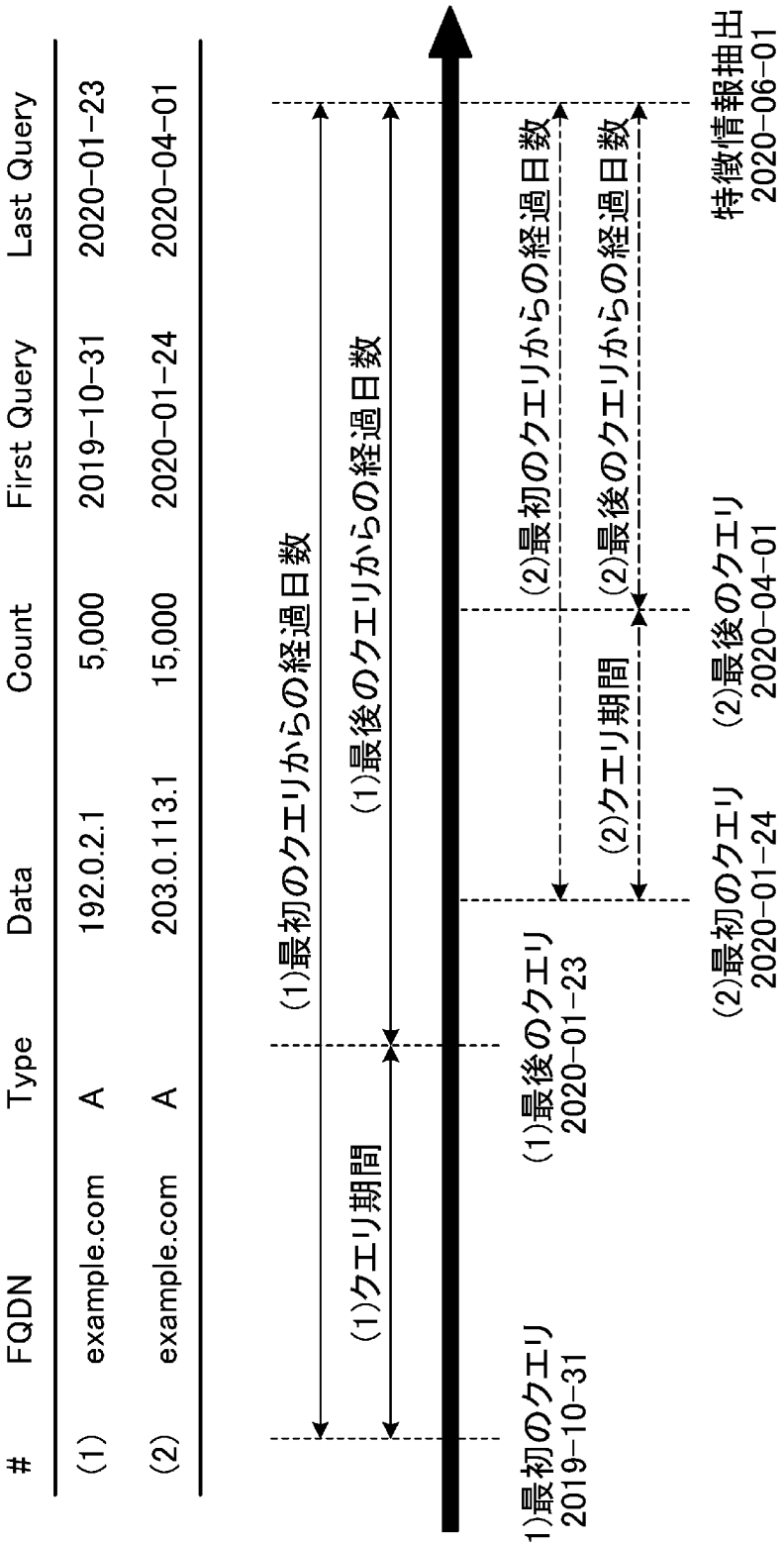
40

Domain Name	Status	SOC Last Decision	# Detection in TI	···
malware.example	Completed	Report Sent	20/80	···
phishing.example	New	Undecided	10/80	···
example.com	Pending	Undecided	2/80	···
···	···	···	···	···

[図4]

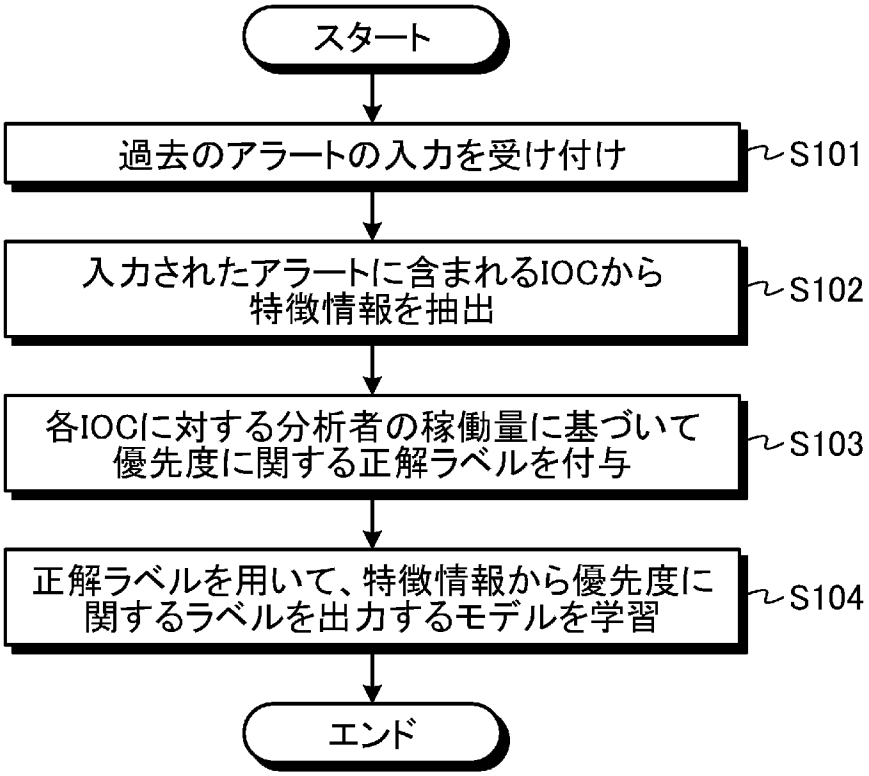

分析エンジン
10
過去のアラート
直近のアラート
決定装置
20
学習（オフライン）
ラベル付与部
22
学習部
23
モデル情報
25
予測（リアルタイム）
24
予測部
特徴情報抽出部
21
優先度が高いIOC
IOCチェッカー
40
分析者

[図5]

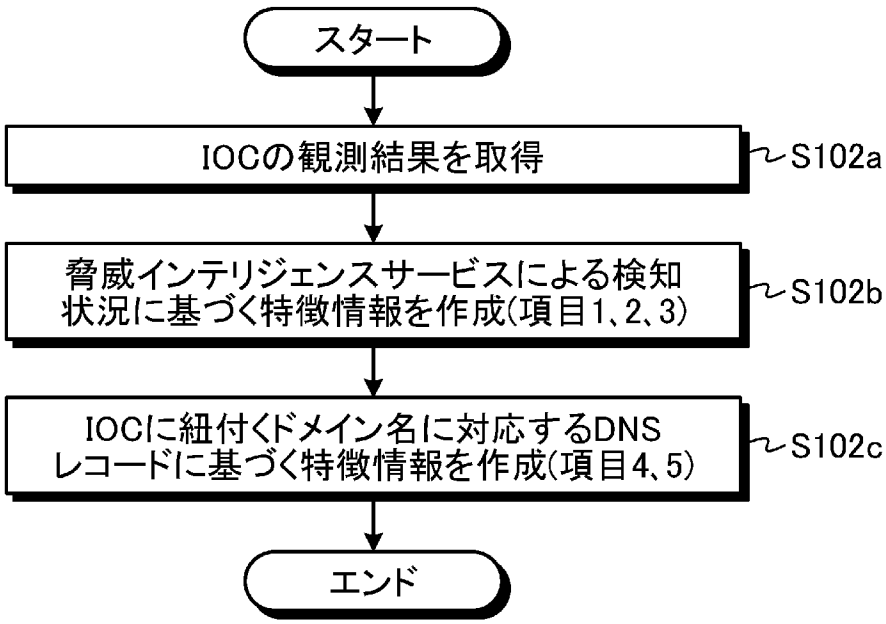


#	FQDN	Type	Data	Count	First Query	Last Query
(1)	example.com	A	192.0.2.1	5,000	2019-10-31	2020-01-23
(2)	example.com	A	203.0.113.1	15,000	2020-01-24	2020-04-01

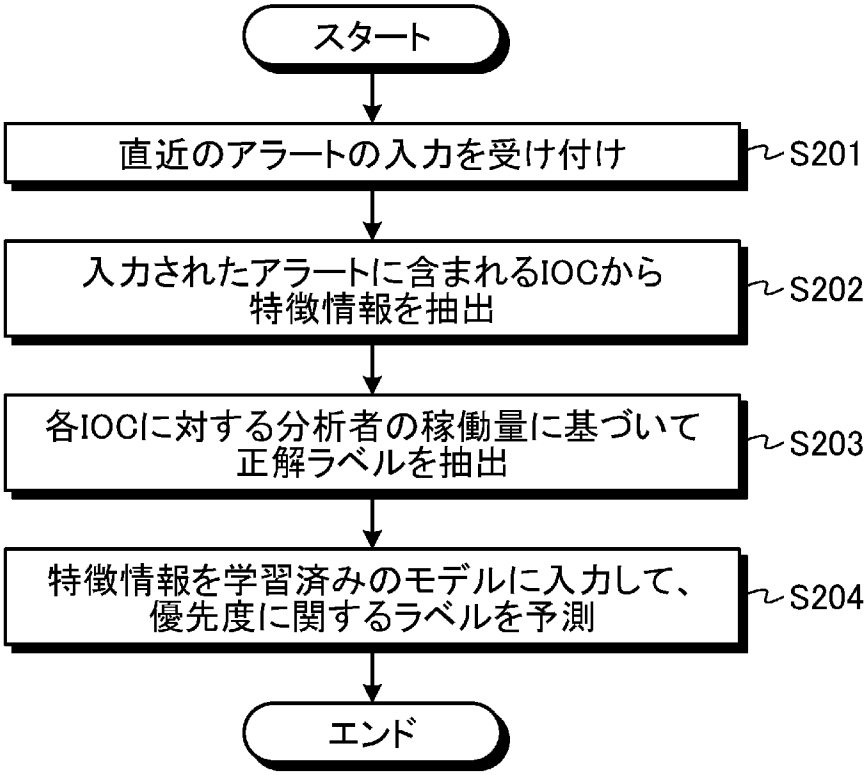
[図6]


スタート
過去のアラートの入力を受け付け
S101
入力されたアラートに含まれるIOCから
特徴情報を抽出
S102
各IOCに対する分析者の稼働量に基づいて
優先度に関する正解ラベルを付与
S103
正解ラベルを用いて、特徴情報から優先度に
関するラベルを出力するモデルを学習
S104
エンド

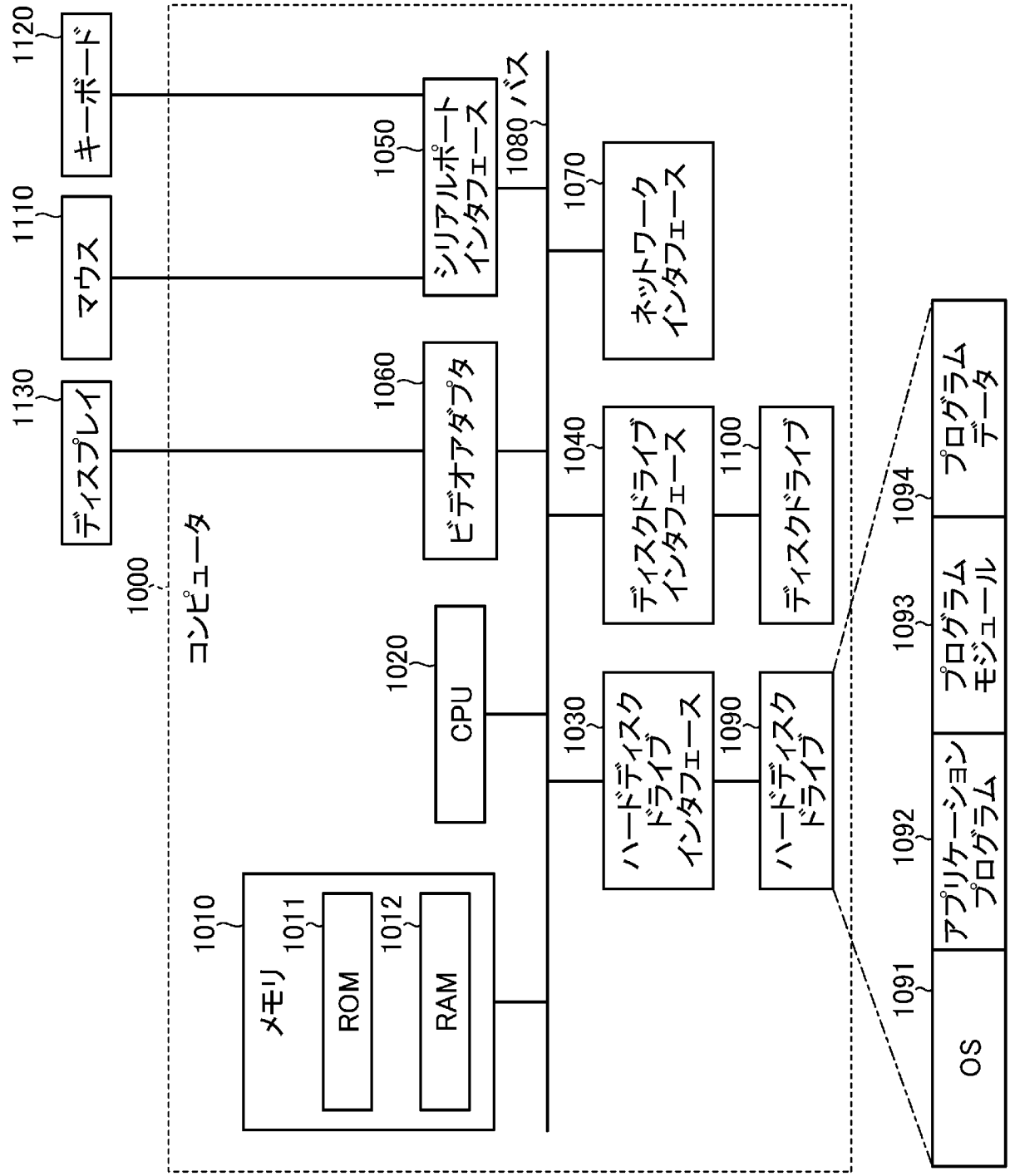
[図7]


スタート
IOCの観測結果を取得
S102a
脅威インテリジェンスサービスによる検知
状況に基づく特徴情報を作成(項目1、2、3)
S102b
IOCに紐付くドメイン名に対応するDNS
レコードに基づく特徴情報を作成(項目4、5)
S102c
エンド

[図8]


スタート
直近のアラートの入力を受け付け
S201
入力されたアラートに含まれるIOCから
特徴情報を抽出
S202
各IOCに対する分析者の稼働量に基づいて
正解ラベルを抽出
S203
特徴情報を学習済みのモデルに入力して、
優先度に関するラベルを予測
S204
エンド

[図9]


1000
コンピュータ
1010
メモリ
1011
ROM
1012
RAM
1020
CPU
1030
ハードディスクドライブインタフェース
1040
ディスクドライブインタフェース
1050
シリアルポートインタフェース
1060
ビデオアダプタ
1070
ネットワークインタフェース
1080
バス
1090
ハードディスクドライブ
1091
OS
1092
アプリケーションプログラム
1093
プログラムモジュール
1094
プログラムデータ
1100
ディスクドライブ
1110
マウス
1120
キーボード
1130
ディスプレイ

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2021/018127

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl. G06F21/55(2013.01)i
FI: G06F21/55320

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
Int.Cl. G06F21/55

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan	1922-1996
Published unexamined utility model applications of Japan	1971-2021
Registered utility model specifications of Japan	1996-2021
Published registered utility model applications of Japan	1994-2021

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	芝原 俊樹, 潜在的な重要アラートの推定によるインシデント特定の効率化, 情報処理学会 シンポジウム コンピュータセキュリティシンポジウム 2019 [online], 14 October 2019, pp. 1092-1 pp. 1094-1097, 3. 提案システム, (SHIBAHARA, Toshiki, Efficient incident detection by predicting potential important alerts), non-official translation (Symposium of Information Processing Society of Japan Computer Security Symposium 2019, 3. Proposal system)	1, 5-7 2-4
A	WO 2018/235252 A1 (NEC CORP.) 27 December 2018 (2018-12-27)	1-7
A	WO 2016/117132 A1 (UBIC KK) 28 July 2016 (2016-07-28)	1-7

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search 15 July 2021	Date of mailing of the international search report 27 July 2021
Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2021/018127

WO 2018/235252 A1	27 December 2018	US 2020/0184072 A1
WO 2016/117132 A1	28 July 2016	(Family: none)
JP 2018-521430 A	02 August 2018	WO 2017/014823 A2 CN 107835982 A

国際調査報告

国際出願番号
PCT/JP2021/018127

A. 発明の属する分野の分類（国際特許分類（ＩＰＣ））

G06F 21/55(2013.01)i
FI: G06F21/55 320

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（ＩＰＣ））

G06F21/55

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	１９２２－１９９６年
日本国公開実用新案公報	１９７１－２０２１年
日本国実用新案登録公報	１９９６－２０２１年
日本国登録実用新案公報	１９９４－２０２１年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献のカテゴリー*	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
X	芝原 俊樹 , 潜在的な重要アラートの推定によるインシデント特定の効率化, 情報処理学会 シンポジウム コンピュータセキュリティシンポジウム ２０１９ [ｏｎｌｉｎｅ] , 2019.10.14, pp.1092-1099 pp.1094-1097「３．提案システム」	1,5-7
A		2-4
A	WO 2018/235252 A1 (日本電気株式会社) 27.12.2018 (2018-12-27)	1-7
A	WO 2016/117132 A1 (株式会社ＵＢＩＣ) 28.07.2016 (2016-07-28)	1-7
A	JP 2018-521430 A (ハサン・シェド・カムラン) 02.08.2018 (2018-08-02)	1-7

☐ Ｃ欄の続きにも文献が列挙されている。

☑ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー
"A" 特に関連のある文献ではなく、一般的技術水準を示すもの
"E" 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
"L" 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
"O" 口頭による開示、使用、展示等に言及する文献
"P" 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
"T" 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの
"X" 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
"Y" 特に関連のある文献であって、当該文献と他の１以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
"&" 同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
15.07.2021	27.07.2021

名称及びあて先
日本国特許庁(ISA/JP)
〒100-8915
日本国
東京都千代田区霞が関三丁目４番３号

権限のある職員（特許庁審査官）
岸野　徹 5S 3983
電話番号 03-3581-1101 内線 3546

国際調査報告
パテントファミリーに関する情報

国際出願番号
PCT/JP2021/018127

引用文献	公表日	パテントファミリー文献	公表日
WO 2018/235252 A1	27.12.2018	US 2020/0184072 A1	
WO 2016/117132 A1	28.07.2016	（ファミリーなし）	
JP 2018-521430 A	02.08.2018	WO 2017/014823 A2	
		CN 107835982 A	