República Federativa do Brasil
Ministério da Economia
Instituto Nacional da Propriedade Industrial

(21) BR 112019013603-7 A2

BR112019013603A2

(22) Data do Depósito: 22/01/2018

(43) Data da Publicação Nacional: 07/01/2020

(54) Título: APRENDIZADO CONTÍNUO PARA DETECÇÃO DE INTRUSÃO

(51) Int. Cl.: G06F 21/55; H04L 29/06.

(30) Prioridade Unionista: 30/01/2017 US 15/419,933.

(71) Depositante(es): MICROSOFT TECHNOLOGY LICENSING, LLC.

(72) Inventor(es): PENGCHENG LUO; REEVES HOPPE BRIGGS; NAVEED AHMAD.

(86) Pedido PCT: PCT US2018014606 de 22/01/2018

(87) Publicação PCT: WO 2018/140335 de 02/08/2018

(85) Data da Fase Nacional: 28/06/2019

(57) Resumo: A presente invenção refere-se a balancear os sinais observados utilizados para treinar modelos de detecção de intrusão de rede que permite uma alocação mais precisa de recursos de computação para defender a rede de indivíduos maliciosos. Os modelos são treinados contra dados ao vivo definidos dentro de uma janela rolante e dados históricos para detectar características definidas pelo usuário nos dados. Ataques automatizados asseguram que vários tipos de ataques estão sempre presentes na janela de treinamento de rolante. O conjunto de modelos é constantemente treinado para determinar qual modelo colocar em produção para alertar os analistas de intrusões, e/ou para automaticamente desenvolver contramedidas. Os modelos são continuamente atualizados conforme as características são redefinidas e conforme os dados na janela rolante mudam, e o conteúdo da janela rolante é balanceado para prover dados suficientes de cada tipo observado pelo qual treinar os modelos. Quando balanceando o conjunto de dados, sinais de baixa população são sobrepostos sobre sinais de alta população para balancear os seus números relativos.

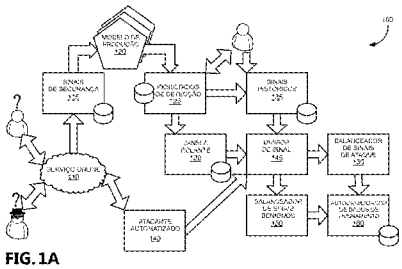

FIG. 1A

Relatório Descritivo da Patente de Invenção para **"APRENDIZADO CONTÍNUO PARA DETECÇÃO DE INTRUSÃO"**.

ANTECEDENTES

[001] As redes de computadores estão sob constante ameaça de indivíduos maliciosos que buscam acesso não autorizado para os sistemas hospedados nas mesmas. As táticas utilizadas por indivíduos maliciosos para atacar as redes e as táticas utilizadas por administradores de rede para defender contra-ataques estão constantemente desenvolvendo à luz uns dos outros; novas façanhas são adicionadas ao arsenal de indivíduos maliciosos e façanhas ineficientes são abandonadas. Implementar contramedidas, no entanto, é frequentemente reativa, em que os administradores de rede devem aguardar para identificar a façanha mais recente antes de desenvolver uma contramedida e determinar quando parar de desenvolver uma contramedida quando a façanha correspondente não é mais utilizada. Identificar e bloquear corretamente as últimas façanhas é frequentemente desafiador para os administradores de rede, especialmente quando uma façanha não está ainda expandida ou ataques a uma pequena população de serviços oferecidos na rede.

SUMÁRIO

[002] Este sumário está provido para introduzir uma seleção de conceitos em uma forma simplificada que estão adicionalmente abaixo descritos na seção de Descrição Detalhada. Este sumário não pretende identificar todas as características chave ou características essenciais do assunto reivindicado, nem este pretende ser como uma ajuda na determinação do escopo do assunto reivindicado.

[003] Sistemas, métodos, e dispositivos de armazenamento de computador incluindo instruções estão aqui providos para prover aprendizado contínuo para detecção de intrusão. Múltiplos modelos de aprendizado de máquina são constantemente retreinados sobre os si-

nais de rede com base em sinais reunidos das máquinas e dispositivos dentro da rede que representam ataques ou comportamento benigno. Uma Janela rolante é utilizada para reunir os sinais, de modo que os modelos utilizam os dados mais atualizados para identificar ataques, e os modelos são continuamente promovidos e rebaixados para guardar a rede conforme as suas habilidades para precisamente detectar ataques mudam de fase em resposta à composição dos dados mais recentes. Modelos distribuídos em redes de produção ativa proveem suas detecções quase em tempo real para analistas de segurança, os quais proveem retorno sobre a precisão dos modelos (por exemplo, intrusões erradas / falsos negativos, falsos positivos, intrusões mal identificadas) para adicionalmente refinar como os modelos são treinados.

[004] Para aperfeiçoar a confiabilidade do conjunto de dados de treinamento utilizados para constantemente retreinar e refinar os modelos de detecção, e por meio disto aperfeiçoar os modelos, os sinais de ataque são balanceados para resolver a sua escassez comparados com um sinal benigno e com relação a tipos de ataque específicos. Os sinais benignos são sobrepostos com sinais de ataque de vários tipos de outras máquinas para prover um conjunto de treinamento balanceado para treinar e refinar os modelos. Entre os sinais de ataque no conjunto de treinamento balanceado, os sinais de vários tipos de ataque são também balanceados para assegurar que o modelo está igualmente treinado sobre todos os tipos de ataque. As características dos sinais são dinamicamente extraídas através de configuração baseada em texto, assim aperfeiçoando a flexibilidade dos modelos para responder a diferentes conjuntos de características indicativas de um ataque sobre a rede.

[005] Em vários aspectos, os ataques são simulados por um atacante interno conhecido para aumentar a prontidão da rede e gerar

sinais de ataque adicionais. Similarmente, sinais de ataque historicamente significativos são utilizados em alguns aspectos de modo que mesmos se sinais de ataque de um tipo específico não foram observados na janela rolante, estes sinais são apresentados para os modelos.

[006] Provendo modelos de detecção de intrusão de aprendizado contínuo para uma rede, as funcionalidades dos dispositivos e software na rede são aperfeiçoadas. Novas formas de ataques são identificadas mais rápido e mais confiavelmente, assim resolvendo o problema centrado em computador de como aperfeiçoar a segurança da rede. Além disso, os recursos de computação não são desperdiçados na tentativa de detectar formas de ataque depreciadas assim reduzindo os recursos de processamento utilizados para salvaguardar a rede de indivíduos maliciosos.

[007] Exemplos estão implementados como um processo de computador, a sistema de computação, ou como um artigo de manufatura tal como um dispositivo, produto de programa de computador, ou meio legível por computador. De acordo com um aspecto, o produto de programa de computador é um meio de armazenamento legível por um sistema de computador e codificar um programa de computador que compreende instruções para executar um processo de computador.

[008] Os detalhes de um ou mais aspectos estão apresentados nos desenhos de acompanhantes e na descrição abaixo. Outras características e vantagens serão aparentes de uma leitura da descrição detalhada seguinte uma revisão dos desenhos associados. Deve ser compreendido que a descrição detalhada seguinte é explanatória somente e não é restritiva das reivindicações.

BREVE DESCRIÇÃO DOS DESENHOS

[009] Os desenhos acompanhantes, os quais estão incorporados e constituem uma parte desta descrição, ilustram vários aspectos. Nos

desenhos:

Figura 1A é um sistema de segurança exemplar com o qual a presente descrição pode ser praticada;

Figura 1B é um sistema de treinamento e seleção de modelo exemplar para utilização com o sistema de segurança exemplar da Figura 1A com o qual a presente descrição pode ser praticada;

Figura 2 é um fluxograma que mostra estágios gerais envolvidos em um método exemplar para desenvolver um conjunto de dados de treinamento pelo qual treinar modelos preditivos para utilização em proteger um serviço online;

Figura 3 é um fluxograma que mostra estágios gerais envolvidos em um método exemplar para treinar e selecionar modelos preditivos para utilização em proteger um serviço online; e

Figura 4 é um diagrama de blocos que ilustra componentes físicos exemplares de um dispositivo de computação.

DESCRIÇÃO DETALHADA

[0010] A descrição detalhada seguinte refere-se aos desenhos acompanhantes. Sempre que possível, os mesmos números de referência são utilizados nos desenhos e a descrição seguinte refere-se aos mesmos ou similares elementos. Apesar de exemplos poderem ser descritos, modificações, adaptações e outras implementações são possíveis. Por exemplo, substituições, adições, ou modificações podem ser feitas nos elementos ilustrados nos desenhos, e os métodos aqui descritos podem ser modificados substituindo, reordenando, ou adicionando estágios aos métodos descritos. Consequentemente, a seguinte descrição detalhada não é limitante, mas, ao invés, o escopo apropriado é definido pelas reivindicações anexas. Exemplos podem tomar a forma de uma implementação em hardware, ou uma implementação inteiramente em software, ou uma implementação que combina aspectos de software e hardware. A descrição detalhada seguin-

te, portanto, não deve ser tomada em um sentido limitante.

[0011] Sistemas, métodos, e dispositivos de armazenamento legíveis por computador que inclui instruções para prover uma segurança de rede aperfeiçoada através de continuamente aprendendo modelos de detecção de intrusão. Continuamente provendo modelos de detecção de intrusão de aprendizado para uma rede, as funcionalidades dos dispositivos e software na rede são aperfeiçoadas. Novas formas de ataques são identificadas mais rapidamente e mais confiavelmente, assim resolvendo o problema centrado em computador de como aperfeiçoar a segurança da rede. Além disso, recursos de computação não são desperdiçados na tentativa de detectar formas de ataque depreciadas, assim reduzindo os recursos de processamento utilizados para salvaguardar a rede de indivíduos maliciosos.

[0012] A Figura 1A é um sistema de segurança exemplar 100 com o qual a presente descrição pode ser praticada. Como mostrado na Figura 1A, um serviço online 110 está conectado por vários usuários – os quais podem ser benignos ou maliciosos – e o sistema de segurança 100. O serviço online 110 representa um conjunto em rede de dispositivos de computação, tal como um centro de dados de nuvem, que provê serviços de "nuvem" para vários usuários, incluindo, mas não limitado a: Infraestrutura como um Serviço (IaaS), onde o usuário provê o sistema de operação e o software que executam nos dispositivos do serviço online 110; Plataforma como um Serviço (PaaS), onde o usuário provê o software e o serviço online 110 provê o sistema de operação e os dispositivos; ou Software como um Serviço (SaaS), onde o serviço online 110 provê tanto o sistema de operação quanto o software para executar nos dispositivos para os usuários. Os usuários os quais buscam acessar o serviço online 110 podem ser usuários legítimos ou indivíduos maliciosos, os quais exploram as vulnerabilidades de segurança para invadir o serviço online 110 para executar pro-

cessos não autorizados e/ou recuperar dados do serviço online 110 sem autorização legítima.

[0013] Para determinar se os usuários são benignos ou maliciosos ou se os dispositivos estão seguros (não enviando sinais maliciosos) ou comprometidos (enviando sinais maliciosos), vários sinais de segurança 115 do serviço online 110 são reunidos e alimentados para os modelos de produção 120 para produzir resultados de detecção 125 que indicam se uma dada seção é maliciosa ou benigna. Os sinais de segurança 115 incluem registos de eventos, rastreios de rede, comandos de sistema, e similares, os quais são analisados pelos modelos de produção 120 por características e seus valores de característica determinados através de treinamento dos modelos de produção 120 para serem indicativos de comportamento malicioso ou benigno. Para o propósito da presente divulgação, sinais de segurança específicos 115 são referidos como sendo "maliciosos" ou "benignos" com base nas ações no serviço online 110 associadas com gerar os sinais de segurança específicos 115. Também, como aqui utilizado, o termo "característica" é um atributo numérico derivado de um ou mais sinais de entrada relativos a uma característica ou comportamento observado na rede os quais são aceitos por um "modelo", o qual é um algoritmo que aceita um conjunto de características (também referido como características de modelo) definido por um analista para converter os valores das características em uma pontuação preditiva ou confiança de se as características indicam uma atividade maliciosa ou benigna.

[0014] Os sinais de segurança 115 são providos para os modelos de produção 120 para extrair várias características dos sinais de segurança 115 para as quais o modelo de produção 120 foram treinados para identificar atividades maliciosas no serviço online 110. Um sinal de segurança 115 é uma coleção de um ou mais eventos relativos que ocorrem em dispositivos dentro do serviço online 110, e podem incluir

diversas características (por exemplo, porta utilizada, endereço de IP conectado, identidade / tipo de dispositivo do qual o sinal é recebido, usuário, ação tomada) das quais um subconjunto é extraído para exame por um dado modelo de produção 120 para determinar se o sinal de segurança 115 é benigno ou malicioso. As características de um ou mais sinais de segurança 115 são combinadas em um vetor de característica para análise, e em vários aspectos, as características podem ser pontuadas para prover uma análise numérica daquela característica para entrada nos modelos de produção 120.

[0015] Por exemplo, um dado endereço IP (Protocolo de Internet) pode ser pontuado com base em sua frequência de utilização, onde uma utilização mais frequente do dado endereço IP durante a janela rolante 130 mudará o valor apresentado para o modelo de produção 120 comparado menor utilização frequente. Contrariamente, se um arquivo sigiloso for acessado, uma ação proibida é tomada, um endereço de IP de lista negra é comunicado com, etc., uma pontuação binária que indica que a condição perigosa ocorreu pode ser provido para o modelo de produção 120 no sinal de segurança 115. O modelo de produção 120 não se baseia em listas brancas ou listas negras, e seu treinamento é relativo às características observadas nos sinais de segurança 115 é discutido em maiores detalhes com relação às Figuras 1B, 2, e 3, as quais podem aprender ao longo do tempo sem a direção de uma lista negra ou lista branca as características que são indicativas de uma intrusão no serviço online 110.

[0016] Para um dado sinal de segurança 115, a determinação pelos modelos de produção 120 especifica se o sinal de segurança 115 em questão é benigno ou malicioso. Estes resultados de detecção 125 estão associados com os sinais de segurança 115 para os identificá-los como maliciosos ou benignos. Em alguns aspectos, pelo menos alguns destes resultados de detecção 125 (por exemplo, os resultados

de detecção maliciosos 125) estão providos para um usuário analista, o qual pode atuar sobre os resultados de detecção 125 para desenvolver contramedidas contra um usuário ou ataque malicioso ou determinar que os resultados de detecção 125 garantam uma diferente avaliação do que os modelos de produção 120 indicam. Por exemplo, quando um falso negativo para um sinal malicioso foi indicado pelo modelo de produção 120, o analista pode avaliar que o sinal é de fato malicioso e indicar a ação que deve ser tomada. Em outro exemplo, quando um falso positivo para um sinal benigno foi indicado, o analista pode avaliar que o sinal é de fato benigno e indicar que nenhuma ação deve ser tomada. Em um exemplo adicional, quando um positivo verdadeiro para uma ação maliciosa foi indicado, o analista pode indicar que nenhuma ação ou diferente ação do que a recomendada pelo sistema de segurança 100 deve ser tomada. Correções de analistas são por meio disto utilizadas para adicionalmente treinar e aperfeiçoar os modelos.

[0017] Os resultados de detecção 125 são também alimentados, em vários aspectos, para um banco de dados que armazena uma janela rolante 130 dos sinais de segurança observados 115 nos d dias passados (onde d é configurável pelo usuário analista ou outro administrador de rede, por exemplo, como dois, dez, quinze, etc. dias), e um banco de dados que armazena sinais históricos 135 para sinais de segurança 115 que devem ser utilizados para treinamento independentemente se estes foram vistos nos d dias passados. Os sinais históricos 135 são curados pelo usuário analista para incluir sinais de segurança 115 associados com ataques externos conhecidos. Em aspectos adicionais, um usuário analista cura os sinais históricos 135 para incluir sinais benignos que podem parecer suspeitos ou de outro modo retornar falsos positivos para intrusão rede para assegurar que os modelos preditivos sejam treinados para apropriadamente responder a sinais que foram historicamente provados difíceis de apropria-

damente identificar.

[0018] O atacante automatizado 140 utiliza padrões conhecidos de ataques e explora para testar a segurança do serviço online 110 e prover resultados conhecidos para utilização em conjunto com os resultados de detecção 125 produzidos pelos modelos de produção 120. Quando os resultados de detecção 125 para um sinal de segurança 115 que é o resultado de um ataque de um atacante automatizado 140 não especifica que o ataque foi malicioso, o sinal de segurança 115 será tratado como malicioso, já que o atacante automatizado 140 indica que este era malicioso. Em vários aspectos, o atacante automatizado 140 é um componente opcional do sistema de segurança 100 ou do serviço online 110.

[0019] Os sinais de segurança 115 (incluindo aqueles na janela rolante 130 e os sinais históricos 135, quando disponíveis) são alimentados para um divisor de sinal 145 juntamente com os resultados de detecção 125 dos modelos de produção 120 (e correções do usuário analista) que indicam se um sinal de segurança 115 foi determinador ser benigno ou malicioso. Similarmente, em aspectos nos quais um atacante automatizado 140 é desenvolvido, a identidade benigna / maliciosa dos sinais de segurança 115 gerada de suas ações sobre o serviço online 110 é provida para o divisor de sinal 145. O divisor de sinal 145 está configurado para dividir os sinais de segurança 115 em sinais benignos, providos para balanceador de sinais benignos 150, e sinais maliciosos, providos para um balanceador de sinais de ataque 155.

[0020] O balanceador de sinais benignos 150 e o balanceador de sinais de ataque 155 desenvolvem o conjunto de sinais de segurança 115 utilizados para popular o conjunto de dados utilizado pelo autocarregador de dados de treinamento 160 para prover sinais benignos e maliciosos pelos quais treinar os modelos para detectar façanhas atua-

lizadas do serviço online 110. O autocarregador de dados de treinamento 160 remove os sinais benignos recebidos de dispositivos comprometidos no serviço online 110, deixando para trás somente os sinais maliciosos dos dispositivos comprometidos. Os sinais benignos de dispositivos limpos são juntados cruzados com os sinais maliciosos de dispositivos comprometidos, resultando em BxM exemplos de ataque, onde B representa o número de exemplos benignos e M o número de exemplos maliciosos. Isto produz um conjunto de dados expandido que sobrepõe os exemplos de ataques por sobre exemplos benignos como se os ataques acontecessem sobre dispositivos limpos.

[0021] Como os dispositivos limpos têm diferentes variações de sinais benignos, e os dispositivos comprometidos têm diferentes variações de sinais de ataque, a ligação cruzada dos dois conjuntos de dados cria um grande número de cenários com uma grande quantidade de variação. No entanto, se os cenários forem escolhidos randomicamente, tal como por um atacante automatizado 140, um número desigual de cada tipo de ataque pode estar presente no conjunto de treinamento, o que poderia distorcer o treinamento dos modelos (resultando em alguns ataques sendo melhor preditos do que outros). Os exemplos de ataque são portanto, balanceados contra os cenários de ataque para assegurar que existe um número substancialmente igual (por exemplo, ± 5%) de cada ataque exemplar no conjunto de treinamento. Em vários aspectos, tipos de ataque sub-representados (isto é, tipos de ataque de uma quantidade abaixo de um número balanceado) têm sinais maliciosos existente copiados para aumentar seu número relativo e/ou tipos de ataque sobrerrepresentados (ou seja, tipos de ataque de uma quantidade acima de um número balanceado) têm sinais maliciosos existentes apagados ou substituídos / sobrepostos por exemplos de tipos de ataque sub-representados para atingir um conjunto exemplar de ataques balanceados.

[0022] Similarmente aos sinais maliciosos, os sinais benignos são balanceados uns em relação aos outros com relação ao tipo de dispositivo ou desempenho para o qual os sinais foram recebidos, de modo que que um dado tipo de dispositivo ou desempenho não estava sobre representado no conjunto de dados de treinamento (resultando em alguns ataques sendo melhor preditos sobre dados tipos de dispositivos / desempenhos do que outros). Os exemplos benignos são portanto, balanceados contra o tipo de dispositivo disponível para assegurar que existe um número substancialmente igual (por exemplo, ± 5%) de cada tipo de dispositivo, que provê exemplos benignos. Em vários aspectos, os tipos de dispositivos sub-representados (isto é, tipos de dispositivos de uma quantidade abaixo de um número balanceado) têm sinais benignos existentes copiados para aumentar o seu número relativo e/ou tipos de dispositivos sobrerrepresentados (isto é, tipos de dispositivos de uma quantidade acima de um número balanceado) têm os sinais benignos existentes apagados ou substituídos / sobrepostos por exemplos benignos de tipos de dispositivos sub-representados para atingir um conjunto exemplar benigno balanceado.

[0023] A Figura 1B é um sistema de treinamento e seleção de modelo exemplar 105 para utilização com o sistema de segurança exemplar 100 da Figura 1A com o qual a presente descrição pode ser praticada. O conjunto de dados balanceado de sinais benignos e maliciosos do autocarregador de dados de treinamento 160 é fornecido a um divisor de treinamento / teste 165 para tanto treinar quanto avaliar vários modelos pelos quais proteger o serviço online 110. O conjunto de dados é dividido em k subconjuntos, onde k -1 dos subconjuntos disponíveis (por exemplo, dois terços) são usados para treinar os modelos, e um subconjunto do conjunto de dados (por exemplo, um terço) é reservado para avaliar os modelos. Em vários aspectos, diferentes frações são previstas para dividir o conjunto de dados em subconjuntos

de treinamento e avaliação, os quais são providos para um treinador de modelo 170 e um avaliador de modelo 175, respectivamente.

[0024] O treinador de modelo 170 está configurado para treinar uma pluralidade de modelos de desenvolvimento 180 através de uma ou mais técnicas de aprendizado de máquina através do subconjunto de treinamento dos dados balanceados. As técnicas de aprendizado de máquina treinam modelos para precisamente fazer predições sobre dados alimentados nos modelos (por exemplo, se os sinais de segurança 115 são benignos ou maliciosos; se um substantivo é uma pessoa, local ou coisa; como será o tempo amanhã). Durante uma fase de aprendizado, os modelos são desenvolvidos em relação a um conjunto de dados de treinamento de entradas conhecidas (por exemplo, amostra A, amostra B, amostra C) para otimizar os modelos para corretamente predizerem o resultado de uma dada entrada. Geralmente, a fase de aprendizado pode ser supervisionada, semissupervisionada, ou não supervisionada; indicando um nível de diminuição ao qual os resultados "corretos" são providos em correspondência às entradas de treinamento. Em uma fase de aprendizado supervisionada, todos os resultados são providos para o modelo e o modelo é direcionado para desenvolver uma regra geral ou algoritmo que mapeia a entrada para o resultado. Em contraste, em uma fase de aprendizado não supervisionada, um resultado desejado não é provido para as entradas, de modo que o modelo pode desenvolver suas próprias regras para descobrir as relações dentro do conjunto de dados de treinamento. Em uma fase de aprendizado semissupervisionada, um conjunto de treinamento incompletamente identificado é provido, com alguns dos resultados conhecidos e alguns desconhecidos para o conjunto de dados de treinamento.

[0025] Os modelos podem ser executados em relação a um conjunto de dados de treinamento por diversas épocas, nas quais o con-

junto de dados de treinamento é repetidamente alimentado no modelo para refinar seus resultados. Por exemplo, em uma fase de aprendizado supervisionada, um modelo é desenvolvido para predizer um resultado para um dado conjunto de entradas, e é avaliado sobre diversas épocas para mais confiavelmente prover um resultado que é especificado como correspondendo à dada entrada pelo maior número de entradas para o conjunto de dados de treinamento. Em outro exemplo, para uma fase de aprendizado não supervisionada, um modelo é desenvolvido para agrupar o conjunto de dados em n grupos, e é avaliado sobre diversas épocas em quão consistentemente este coloca uma dada entrada em um dado grupo e quão confiavelmente este produz os n grupamentos desejados através de cada época.

[0026] Em vários aspectos, uma validação cruzada é aplicada no topo de cada fase de treinamento, onde uma porção do conjunto de dados de treinamento é utilizada como um conjunto de dados de avaliação. Por exemplo, o conjunto de dados de treinamento pode ser dividido em k segmentos, onde segmentos (k-1) são utilizados em épocas de treinamento, e o segmento restante é usado para determinar quão bem os modelos treinados funcionaram. Neste modo, cada modelo é treinado em relação a todas as combinações disponíveis de parâmetros de entrada, de modo que cada modelo é treinado k vezes, e os melhores parâmetros modelo são selecionados com base em seus desempenhos médios através das épocas.

[0027] Uma vez uma época é executada, os modelos são avaliados e os valores de suas variáveis são ajustados para tentar refinar melhor o modelo. Em vários aspectos, as avaliações são tendenciosas contra falsos negativos, tendenciosas contra os falsos positivos, e igualmente tendenciosas com relação à precisão geral do modelo. Os valores podem ser ajustados em diversos modos dependendo da técnica de aprendizado de máquina utilizada. Por exemplo, em um algo-

ritmo genético ou evolucionário, os valores para os modelos que têm mais sucesso em predizer os resultados desejados são usados para desenvolver valores para os modelos utilizarem durante a época subsequente, que podem incluir variação / mutação randômica para prover pontos de dados adicionais. Alguém versado na técnica estará familiarizado com diversos outros algoritmos de aprendizado de máquina que podem ser aplicados com a presente descrição, incluindo regressão linear, florestas randômicas, aprendizado de árvores de decisão, redes neurais, etc.

[0028] O modelo desenvolve uma regra ou algoritmo sobre diversas épocas variando os valores de uma ou mais variáveis que afetam as entradas para mapear mais de perto um resultado desejado, mas como o conjunto de dados de treinamento pode ser variado, e é de preferência muito grande, uma acurácia e precisão perfeitas podem não ser alcançáveis. Um número de épocas que compõem uma fase de aprendizado, portanto, pode ser determinado como um dado número de tentativas ou um orçamento de tempo / computação fixo, ou pode ser terminado antes que o número / orçamento seja alcançado quando a precisão de um dado modelo é alta o suficientemente ou baixa o suficiente ou um patamar de precisão foi alcançado. Por exemplo, se a fase de treinamento for projetada para executar n épocas e produzir um modelo com pelo menos 95% de precisão, e se tal modelo for produzido antes da enésima época, a fase de aprendizado pode terminar mais cedo e utilizar o modelo produzido satisfazendo o limite de precisão do objetivo final. Similarmente, se um dado modelo for impreciso o bastante para satisfazer um limite de chance randômico (por exemplo, o modelo é somente 55% preciso em determinar resultados verdadeiro / falsos para dadas entradas), a fase de aprendizado para este modelo pode ser terminada mais cedo, apesar que outros modelos na fase de aprendizado podem continuar treinando. Simi-

larmente, quando um dado modelo continua a prover uma precisão similar ou vacila em seus resultados através de múltiplas épocas - tendo atingido uma plataforma de desempenho - a fase de aprendizado para o dado modelo pode terminar antes que o orçamento de número / computação de época seja atingido.

[0029] Uma vez que a fase de aprendizado está completa, os modelos são finalizados. Os modelos que são finalizados são avaliados em relação a critérios de teste. Em um primeiro exemplo, um conjunto de dados de teste que inclui resultados conhecidos para suas entradas é alimentado nos modelos finalizados para determinar uma precisão dos modelos em manipular dados sobre os quais estes não foram treinados. Em um segundo exemplo, uma taxa de falso positivo, taxa de falso negativo, pode ser utilizada para avaliar os modelos após a finalização. Em um terceiro exemplo, uma delineação entre agrupamentos é utilizada para selecionar um modelo que produz os limites mais claros para seus agrupamentos de dados. Em outros exemplos, métricas adicionais dos modelos são avaliadas, tal como áreas sob precisão e curvas de recuperação.

[0030] Os modelos de desenvolvimento 180 (e portanto os modelos de produção 120) são modelos preditivos que são inicialmente desenvolvidos por um configurador de características de modelo 185 com base em seleções feitas por um usuário administrativo. O usuário administrativo seleciona uma ou mais características de um sinal de segurança 115 que devem ser ouvidos nos dispositivos do serviço online 110 e como estas características devem ser analisadas para significar se um dado sinal de segurança 115 é malicioso ou benigno. Em vários aspectos, as características estão providas em arquivos de texto estruturado (por exemplo, utilizando Extensible Markup Language (XML) ou caracteres de JavaScript Object Notation (JSON)) que o usuário administrativo é capaz de selecionar e definir uma caracteristi-

ca ajustada para um novo modelo de desenvolvimento 180. Com base na configuração de características, as características são dinamicamente extraídas como um vetor de característica do dado conjunto de sinais de segurança para um dispositivo. Diferentes características podem ser extraídas para diferentes modelos com base em sua respectiva configuração de características. Os arquivos de texto estruturados, com isto, permitem o usuário administrativo adicionar ou modificar características e como estas são examinadas para um modelo sem precisar adicionar ou modificar um código para uma base de códigos; o arquivo de texto estruturado invoca segmentos de código de uma base de códigos que pode ser expandida ou modificada por um desenvolvedor para fornecer novos tipos de característica para o usuário administrativo selecionar. Por exemplo, um usuário administrativo pode selecionar como um tipo de exame de característica para utilização com um dado parâmetro ou campo de dados dos sinais de segurança 115: uma contagem de valores distintos em um conjunto de dados (Count), um valor máximo em um conjunto de dados (Max), uma contagem do valor que ocorre mais frequentemente em uma lista (MaxCount), uma soma máxima de valores em uma lista que não excede um limite (MaxSum), etc. Exemplos de campos de dados / parâmetros para observar nos sinais de segurança incluem, mas não estão limitados a: tipos de sinal (por exemplo, exfiltração de dados, tentativas de logon, solicitação de acesso para dados arquivos), portas utilizadas, bytes utilizados em um processo / comunicação, bytes transferidos para / de um dado endereço de Protocolo de Internet (IP) e conjunto de portas, um identificador de usuário, se um dado endereço de IP ou ação está em uma lista negra ou lista branca, etc.

[0031] O avaliador de modelo 175 está configurado para avaliar os modelos de desenvolvimento 180 para determinar quais modelos devem ser utilizados como modelos de produção 120 no sistema de se-

gurança 100. Em vários aspectos, os modelos de produção 120 são reincorporados nos modelos de desenvolvimento 180 para avaliação, ou os limites de precisão dos modelos de produção 120 são utilizados para determinar se substituir um dado modelo de produção 120 por uma modelo de desenvolvimento 180. Em outros aspectos, os modelos de desenvolvimento 180 são comparados com os modelos de produção 120 com relação a outras métricas, tal como, por exemplo, precisão, áreas sob precisão e curvas de recuperação, etc., nas quais os melhores modelos são selecionados como modelos promovidos 190 para utilização como modelos de produção 120. Modelos podem ser continuamente promovidos de modelos de desenvolvimento 180 para modelos de produção 120 (e rebaixados de modelos de produção 120 para modelos de desenvolvimento 180) conforme o avaliador de modelo 175 determina que sua eficiência em apropriadamente identificar sinais maliciosos como maliciosos e sinais benignos como benignos. Em vários aspectos, os superiores n modelos de desenvolvimento mais precisos 180 ou todos os modelos de desenvolvimento 180 que excedem um limite de precisão são promovidos como modelos promovidos 190 para modelos de produção 120. Em outros aspectos, um usuário administrativo pode manualmente promover um modelo de desenvolvimento 180 para um modelo de produção 120, tal como, por exemplo, quando nenhum outro modelo monitora uma dada característica dos sinais de segurança 115.

[0032] O sistema de segurança 100, o sistema de treinamento e seleção de modelo 105, e os seus respectivos elementos componentes são ilustrativos de uma multiplicidade de sistemas de computação que incluem, sem limitação, sistemas de computador desktop, sistemas de computação com fio e sem fio, sistemas de computação móveis (por exemplo, telefones móveis, netbooks, tablet ou computadores do tipo slate, computadores notebook, e computadores laptop), dispo-

sitivos portáteis, sistemas de multiprocessador, eletrônica de consumidor baseada em microprocessador ou programável, minicomputadores, impressoras, e computadores mainframe. O hardware destes sistemas de computação está discutido em maiores detalhes com relação à Figura 4.

[0033] Apesar dos elementos componentes do sistema de segurança 100 e do sistema de treinamento e seleção de modelo 105 serem mostrados remotamente uns dos outros para propósitos ilustrativos, deve ser notado que diversas configurações de um ou mais destes dispositivos hospedados localmente em outro dispositivo ilustrado são possíveis, e cada dispositivo ilustrado pode representar múltiplas instâncias daquele dispositivo. Vários servidores e intermediários familiares daqueles versados na técnica podem ficar entre os elementos componentes ilustrados nas Figuras 1A e 1B para rotear as comunicações entre estes sistemas, os quais não estão ilustrados de modo a não distrair dos novos aspectos da presente descrição.

[0034] A Figura 2 é um fluxograma que mostra estágios gerais envolvidos em um método exemplar 200 para desenvolver um conjunto de dados de treinamento pelo qual treinar de modelos preditivos para utilização em proteger um serviço online 110. O método 200 começa com OPERAÇÃO 210, onde os sinais de segurança 115 são reunidos. Em vários aspectos, os sinais de segurança 115 podem ser recebidos em tempo real (ou quase em tempo real, levando em conta retardos de processamento e transmissão) e podem ser recebidos e colocados em cache em um banco de dados para revisão periódica, tal como, por exemplo, em um processo de lote para revisar eventos de segurança a cada m minutos. Os sinais de segurança 115 incluem eventos e parâmetros escutados de várias ações que acontecem em máquinas no serviço online 110.

[0035] Os eventos e parâmetros escutados são utilizados na

OPERAÇÃO 220 para identificar se um dado sinal de segurança 115 corresponde a uma ação que é maliciosa ou benigna. Em vários aspectos, os sinais de segurança reunidos 115 são alimentados para modelos preditivos projetados para utilização com serviços online ao vivo 110 (isto é, os modelos de produção 120) para determinar se cada sinal de segurança 115 é malicioso ou benigno. Estas determinações são apresentadas para usuários analistas, os quais podem atuar sobre as determinações para proteger o serviço online 110 contra indivíduos maliciosos, mas podem também revogar a determinação feita pelos modelos preditivos; indicando que a determinação é um falso positivo ou um falso negativo. Similarmente, em aspectos onde um atacante automatizado 140 é utilizado para simular um ataque sobre o serviço online 110, o atacante automobilizado 140 provê uma notificação que identifica os sinais de segurança 115 produzidos em resposta ao ataque como maliciosos de modo que estes sinais de segurança 115 são tratados como maliciosos, independentemente do resultado de detecção dos modelos preditivos.

[0036] Na OPERAÇÃO 230 uma janela rolante 130 é ajustada para definir um quadro de tempo do tempo corrente no qual analisar os sinais de segurança 115 relevantes para as façanhas e ataques mais correntes sendo executados no serviço online 110. A janela rolante 130 define um conjunto de sinais de segurança 115 que caem dentro de um período de tempo designado do tempo corrente; os sinais de segurança 115 reunidos dentro dos últimos d dias. Uma janela de múltiplos dias é utilizada para treinar e predizer ataques lentos, os quais são executados sobre múltiplos dias para evitar a detecção por sistemas de segurança convencionais. Conforme os sinais de segurança 115 são reunidos, os sinais de segurança mais recentes são adicionados ao conjunto de sinais de segurança 115 para a janela rolante 130 e os sinais de segurança 115 que foram reunidos antes do período de

tempo designado para a janela rolante 130 são continuamente removidos do conjunto de sinais de segurança 115.

[0037] Em alguns aspectos, sinais históricos 135 são opcionalmente recebidos na OPERAÇÃO 240. Os sinais históricos 135 são curados por um usuário analista de sinais de segurança 115 previamente observados para incluir sinais de segurança 115 historicamente significativos que representam certos tipos de ataque ou casos de utilização benigna que são designados para propósitos de treinamento independentemente se um ataque ou caso de utilização similar foi visto dentro do período de tempo da janela rolante 130. Em um exemplo, uma façanha historicamente perigosa pode ter sinais de segurança 115 relativos à sua detecção adicionados aos sinais históricos 135 para constantemente permanecer em guarda contra aquela façanha. Em outro exemplo, um desenvolvedor pode descobrir uma façanha de dia zero e não saber se indivíduos maliciosos já a estão utilizando, e prover um sinal de segurança exemplar 115 simulando as ações da façanha de dia zero para utilização como um sinal histórico 130 para preventivamente guardar contra a façanha, mesmo se esta nunca for vista. Em ainda um exemplo adicional, um sinal de segurança 115 que frequentemente resulta em falsos positivos pode ser adicionado aos sinais históricos 135 para assegurar que os modelos preditivos sejam treinados contra este sinal de segurança específico 115. Os sinais históricos 135, se disponíveis, são adicionados ao conjunto de sinais de segurança 115 reunidos dentro da janela rolante 130.

[0038] Prosseguindo para a OPERAÇÃO 250, o método 200 balanceia os sinais maliciosos e benignos reunidos que caem dentro da janela rolante 130 e quaisquer sinais históricos 135 adicionados ao conjunto na OPERAÇÃO 240 opcional. Quando balanceando os sinais maliciosos, o tipo de ataque de cada sinal é determinado de modo que as quantidades relativas dos sinais que representam cada tipo de ata-

que são trazidas em equilíbrio (isto é, equalizadas), de modo que nenhum dado tipo de ataque é sobrerrepresentado ou sub-representado na população de sinais maliciosos. Quando balanceando os sinais benignos, os sinais benignos recebidos de dispositivos que produziram os sinais maliciosos dentro da janela rolante 130 são descartados, e as quantidades relativas dos sinais benignos recebidos de cada tipo de dispositivo no serviço online 110 são trazidos em equilíbrio, de modo que nenhum dado tipo de dispositivo é sobrerrepresentado ou sub-representado na população de sinais benignos.

[0039] Além disso, como o conjunto de sinais maliciosos é esperado ser menor em número do que o conjunto de sinais benignos, uma porção do conjunto de sinais benignos pode ser selecionada na OPERAÇÃO 260 para junção cruzada com os sinais maliciosos para produzir um novo, maior conjunto de sinais maliciosos de modo que os dois conjuntos conterão uma razão de sinais malignos para benignos desejada. Em vários aspectos, uma vez que o conjunto de sinais maliciosos e o conjunto de sinais benignos são trazidos em uma razão desejada (por exemplo, equilíbrio), os dois conjuntos são utilizados juntos como um conjunto de treinamento.

[0040] Na OPERAÇÃO 270 o conjunto de treinamento de vários cenários de ataque compostos dos sinais maliciosos e benignos balanceados que ocorrem na janela rolante 130 (e quaisquer sinais históricos 135) é tornado disponível para treinar modelos preditivos. Por exemplo, os modelos de produção 120 utilizados para analisar os sinais de segurança 115 são continuamente retreinados e/ou substituídos por diferentes modelos preditivos conforme os conteúdos da janela rolante 130 são atualizados ao longo do tempo para melhor avaliar os ataques e façanhas ativamente sendo utilizados contra o serviço online 110. O método 200 portanto pode concluir após a OPERAÇÃO 270 ou retornar para a OPERAÇÃO 210 para continuar a reunir sinais

de segurança 115 para periodicamente ou constantemente suprir um conjunto de dados de treinamento com base em uma janela rolante 130.

[0041] A Figura 3 é um fluxograma que mostra estágios gerais envolvidos em um método exemplar 300 para treinar e selecionar modelos preditivos para utilização em proteger um serviço online 110. O método 300 começa com a OPERAÇÃO 310, onde um conjunto de dados de treinamento de sinais maliciosos e benignos balanceados, tal como aquele desenvolvido de acordo com o método 200, é recebido. Em vários aspectos, o método 300 é invocado em uma base periódica (por exemplo, cada h horas), em resposta a uma atualização da janela rolante 130 (e portando do conjunto de dados de treinamento), ou um comando de usuário.

[0042] Prosseguindo para a OPERAÇÃO 320, o conjunto de dados de treinamento é dividido em um subconjunto de avaliação e um subconjunto de aprendizado. Em vários aspectos, o tamanho do subconjunto de avaliação em relação ao conjunto de dados de treinamento pode variar, mas é geralmente menor em tamanho do que o subconjunto de aprendizado. Por exemplo, o subconjunto de avaliação pode ser um terço do conjunto de treinamento inicial, e o subconjunto de aprendizado seria portanto, os dois terços restantes do conjunto de treinamento inicial. Alguém versado na técnica apreciará que outras frações do conjunto de dados de treinamento podem ser divididas para utilização como um subconjunto de avaliação.

[0043] Na OPERAÇÃO 330 as características de configuração são recebidas para produzir modelos de desenvolvimento 180 como modelos preditivos potenciais para utilização na produção (isto é, como modelos de produção 120) para proteger um serviço online 110. Um usuário administrativo, tal como um analista de segurança, escolhe um ou mais parâmetros listados para serviço online 110 através dos sinais de

segurança 115 e um tipo de característica pelo qual examinar estes parâmetros. Os sinais de segurança 115 incluem, mas não estão limitados a: registros de eventos, rastreamentos de rede, relatórios de erros, relatórios de ouvinte de eventos especiais, detecção atômica, e suas combinações, e os parâmetros para as características selecionadas podem incluir quaisquer dos elementos incluídos nos sinais de segurança 115.

[0044] Por exemplo, quando os sinais de segurança 115 incluem rastreamentos de rede, um parâmetro de um par de endereços de remetente / receptor pode ser selecionado e avaliado de acordo com um tipo de característica de "contagem" de modo que o número de vezes que o par é visto dentro do conjunto de treinamento incrementa uma pontuação / valor para avaliar aquela característica. Em outro exemplo, quando os sinais de segurança 115 incluem traços de rede, um parâmetro de um número de bytes transmitidos entre um par de remetente / receptor é provido como um valor / pontuação para avaliar por esta característica. Em um outro exemplo, um parâmetro de um balanço de transmissões entre um par de remetente / receptor para indicar uma relativa razão de upload / download está provido como um valor / pontuação para avaliar por esta característica. Alguém versado na técnica reconhecerá os acima como exemplos não limitantes; outros parâmetros e outros tipos de característica destes parâmetros pelos quais estes podem ser avaliados pelos modelos preditivos são previstos para utilização com o presente pedido.

[0045] Prosseguindo para a OPERAÇÃO 340 modelos de desenvolvimento 180 são criados com base na configuração de características recebida e são refinados de acordo com o subconjunto de aprendizado com um ou mais algoritmos de aprendizado de máquina. Cada modelo preditivo é criado para aceitar um vetor de característica específico (que especifica as características selecionadas pelo usuário ad-

ministrativo) onde cada característica que compõe o vetor de característica está associada a um coeficiente. Cada vetor de característica é dinamicamente extraído dos sinais de segurança 115 com base na configuração de características. Os valores dos coeficientes são ajustados sobre diversas épocas de um algoritmo de aprendizado da máquina de modo que quando um dado modelo de desenvolvimento 180 receba uma entrada de um vetor de característica, as interações entre os vários valores de característica podem ser ajustadas para confiavelmente produzir um resultado de malicioso ou benigno para coincidir com os resultados designados no subconjunto de aprendizado.

[0046] Prosseguindo para a OPERAÇÃO 350, os modelos preditivos refinados com relação ao conjunto de dados de treinamento da OPERAÇÃO 340 são avaliados contra o subconjunto de avaliação dividido do conjunto de dados de treinamento da OPERAÇÃO 320. O subconjunto de avaliação inclui entradas (os sinais de segurança 115 coletados do serviço online 110) com resultados conhecidos de se o sinal é malicioso ou benigno. Além disso, os pares de entrada / resultado do subconjunto de avaliação não foram utilizados para diretamente treinar os modelos de desenvolvimento 180, e assim proveem um teste para quanto a se os modelos de desenrolamento 180 proveem uma regra funcional geral para determinar se um sinal desconhecido é malicioso ou benigno.

[0047] Um limite de promoção é aplicado aos modelos de desenvolvimento 180 para determinar se promover um dado modelo de desenvolvimento 180 para um modelo de produção 120. Os limites de promoção especificam quão precisamente um modelo de desenvolvimento 180 precisa estar predizendo se os sinais são maliciosos ou benignos com base em um vetor de característica extraído dos sinais de segurança 115. Em alguns aspectos, o limite de promoção é ajustado como uma constante, tal como, por exemplo, pelo menos n% de

precisão, uma dada área sob uma curva de precisão e recuperação sobre os dados de teste, etc. Em outros aspectos, o limite de promoção é ajustado pela precisão de um modelo de produção corrente 120 para um dado vetor de característica ou tipo de ataque de modo que para um modelo de desenvolvimento 180 substituir um modelo de produção 120 no sistema de segurança 100, o modelo de desenvolvimento 180 deve ser mais preciso do que o modelo de produção corrente 120.

[0048] Na OPERAÇÃO 360 os modelos de desenvolvimento 180 e os modelos de produção reavaliados 120 que executam melhor de acordo com o subconjunto de avaliação e o limite de promoção são promovidos para utilização pelo sistema de segurança 100 para proteger o serviço online 110. Os modelos de produção 120 que não mais satisfazem o limite de promoção ou foram substituídos por modelos de desenvolvimento 180 podem ser apagados ou rebaixados para modelos de desenvolvimento 180 para treinamento e correção adicionais, e para posterior reavaliação. O método 300 pode então concluir.

[0049] Apesar de implementações terem sido descritas no contexto geral de módulos de programa que executam em conjunto com um programa de aplicação que executa em um sistema de operação em um computador, aqueles versados na técnica reconhecerão que aspectos podem também ser implementados em combinação com outros módulos do programa. Geralmente, os módulos de programa incluem rotinas, programas, componentes, estruturas de dados, e outros tipos de estruturas que executam tarefas específicas ou implementam tipos de dados abstratos específicos.

[0050] Os aspectos e funcionalidades aqui descritos podem operar através de uma multiplicidade de sistemas de computação, incluindo, sem limitação, sistema de computador desktop, sistema de computação com fio e sem fio, sistemas de computação móveis (por exemplo,

telefones móveis, netbooks, tablets ou computadores tipo slate, computadores notebooks e computadores laptop), dispositivos portáteis, sistemas de multiprocessador, eletrônica de consumidor baseada em microprocessadores ou programáveis, minicomputadores e computadores mainframe.

[0051] Além disso, de acordo com um aspecto, os aspectos e funcionalidades aqui descritos operam sobre sistemas distribuídos (por exemplo, sistema de computação baseados em nuvem), onde a funcionalidade de aplicação, memória, armazenamento e recuperação de dados e várias funções de processamento são operadas remotamente umas das outras sobre uma rede de computação distribuída, tal como a Internet ou uma intranet. De acordo com um aspecto, as interfaces de usuários e informações de vários tipos são exibidas através de display de dispositivo de computação integrados ou através de unidades de display remotas associadas com um ou mais dispositivos de computação. Por exemplo, as interfaces de usuário e informações de vários tipos são exibidas e interagidas sobre uma superfície de parede sobre a qual interfaces de usuário e informações de vários tipos são projetadas. Uma interação com a multiplicidade de sistema de computação com os quais as implementações são praticadas inclui, entrada de digitação, entrada de tela de toque, entrada de voz ou outro áudio, entrada de gesto onde um dispositivo de computação associado está equipado com uma funcionalidade de detecção (por exemplo, câmera) para capturar e interpretar gestos de usuário para controlar a funcionalidade do dispositivo de computação, e similares.

[0052] A Figura 4 e a descrição associada proveem uma discussão de uma variedade de ambientes de operação nos quais exemplos são praticados. No entanto, os dispositivos e sistemas ilustrados e discutidos com relação à Figura 4 são para propósitos de exemplo e ilustração e não são limitantes de um vasto número de configurações de

dispositivos de computação que são utilizadas para a prática de aspectos, aqui descritos.

[0053] A Figura 4 é um diagrama de blocos que ilustra os componentes físicos (isto é, hardware) de um dispositivo de computação 400 com o qual exemplos da presente descrição podem ser praticados. Em uma configuração básica, o dispositivo de computação 400 inclui pelo menos uma unidade de processamento 402 e uma memória de sistema 404. De acordo com um aspecto, dependendo da configuração e tipo de dispositivo de computação 400, a memória de sistema 404 é um dispositivo de armazenamento de memória que compreende, mas não está limitado a, armazenamento volátil (por exemplo, memória de acesso randômico), armazenamento não volátil (por exemplo, memória somente de leitura), memória instantânea, ou qualquer combinação de tais memórias. De acordo com um aspecto, a memória de sistema 404 inclui um sistema de operação 405 e um ou mais módulos de programa 406 adequados para executar as aplicações de software 450. De acordo com um aspecto, a memória de sistema 404 inclui o sistema de segurança 100, o sistema de treinamento e seleção de modelo 105, e quaisquer modelos utilizados ou produzidos por meio disto. O sistema de operação 405, por exemplo, é adequado para controlar a operação do dispositivo de computação 400. Mais ainda, aspectos são praticados em conjunto com uma biblioteca gráfica, outros sistemas de operação, ou qualquer outro programa de aplicação, e não estão limitados a nenhuma aplicação ou sistema específico. Esta configuração básica está ilustrada na Figura 4 por aqueles componentes dentro de uma linha tracejada 408. De acordo com um aspecto, o dispositivo de computação 400 tem características ou funcionalidade adicionais. Por exemplo, de acordo com um aspecto, o dispositivo de computação 400 inclui dispositivos de armazenamento de dados adicionais (removíveis e/ou não removíveis) tais como, por exemplo, discos magnéticos, dis-

cos óticos, ou fita. Tal armazenamento adicional está ilustrado na Figura 4 por um dispositivo de armazenamento removível 409 e um dispositivo de armazenamento não removível 410.

[0054] Como acima declarado, de acordo com um aspecto, um número de módulos de programa e arquivos de dados estão armazenados na memória de sistema 404. Enquanto executando na unidade de processamento 402, os módulos de programa 406 (por exemplo, sistema de segurança 100, sistema de treinamento e seleção de modelo 105) executam processos que incluem, mas não limitados a, um ou mais dos estágios dos métodos 200 e 300 ilustrados nas Figuras 2 e 3, respectivamente. De acordo com um aspecto, outros módulos do programa são utilizados de acordo com exemplos e incluem aplicações tais como correio eletrônico e aplicações de contatos, aplicações de processamento de palavra, aplicações de planilha, aplicações de banco de dados, aplicações de apresentação de slides, programas de aplicação de desenho ou auxiliado por computador, etc.

[0055] De acordo com um aspecto, o dispositivo de computação 400 tem um ou mais dispositivo(s) de entrada tal como um teclado, um mouse, uma caneta, um dispositivo de entrada de som, um dispositivo de entrada de toque, etc. O(s) dispositivos de saída 414 tais como um display, alto-falantes, uma impressora, etc. estão também incluídos de acordo com um aspecto. Os dispositivos acima mencionados são exemplos e outros podem ser utilizados. De acordo com um aspecto, o dispositivo de computação 400 inclui uma ou mais conexões de comunicação 416 que permitem comunicações com outros dispositivos de computação 418. Exemplos de conexões de comunicação adequadas 416 incluem, mas não estão limitados a, transmissor de frequência de rádio (RF), receptor, e/ou circuito de transceptor; barramento serial universal (USB), portas paralelas, e/ou seriais.

[0056] O termo meio legível por computador, como aqui utilizado,

inclui um meio de armazenamento de computador. O meio de armazenamento de computador inclui um meio volátil e não volátil, removível e não removível implementado em qualquer método ou tecnologia para armazenamento de informações, tal como instruções legíveis por computador, estruturas de dados ou módulos de programa. A memória de sistema 404, o dispositivo de armazenamento removível 409, e o dispositivo de armazenamento não removível 410 são todos exemplos de meio de armazenamento de computador (isto é, armazenamento de memória). De acordo com um aspecto, o meio de armazenamento de computador inclui RAM, ROM, Memória somente de leitura programável eletricamente apagável (EEPROM), memória instantânea ou outra tecnologia de memória, CD-ROM, discos versáteis digitais (DVD) ou outro armazenamento ótico, cassetes magnéticos, fita magnética, armazenamento de disco magnético ou outros dispositivos de armazenamento magnético, ou qualquer outro artigo de manufatura o qual pode ser utilizado para armazenar informações e o qual pode ser acessado pelo dispositivo de computação 400. De acordo com um aspecto, qualquer tal meio de armazenamento de computador faz parte do dispositivo de computação 400. O meio de armazenamento de computador não inclui uma onda portadora ou outro sinal de dados propagado.

[0057] De acordo com um aspecto, os meios de comunicação são incorporados por instruções legíveis por computador, estruturas de dados, módulos de programa, ou outros dados em um sinal de dados modulado, tal como uma onda portadora ou outro mecanismo de transporte, e incluem qualquer meio de fornecimento de informações. De acordo com um aspecto, o termo "sinal de dados modulado" descreve um sinal que tem uma ou mais características ajustadas ou mudadas de tal modo a codificar informações no sinal. Por meio de exemplo, e não limitação, os meios de comunicação incluem um meio

com fio tal como uma rede com fio ou conexão com fio direta, e um meio sem fio tal como um meio acústico, frequência de rádio (RF), infravermelho e outros meios sem fio.

[0058] As implementações, por exemplo, estão acima descritas com referência a diagramas de bloco e/ou ilustrações operacionais de métodos, sistemas e produto de programa de computador de acordo com aspectos. As funções / atos notados nos blocos podem ocorrer fora da ordem como mostrada em qualquer fluxograma. Por exemplo, dois blocos mostrados em sucessão podem de fato ser executados substancialmente concorrentemente ou os blocos podem algumas vezes ser executados na ordem inversa, dependendo da funcionalidade / atos envolvidos.

[0059] A descrição e ilustração de um ou mais exemplos providos nesta aplicação não pretendem limitar ou restringir o escopo como reivindicado em nenhum modo. Os aspectos, exemplos, e detalhes providos neste pedido são considerados suficientes para transmitir possessão e permitir outros fazer e utilizar o melhor modo. As implementações não devem ser consideradas como sendo limitadas a qualquer aspecto, exemplo, ou detalhe provido neste pedido. Independentemente de se mostradas e descritas em combinação ou separadamente, as várias características (tanto estruturais quanto metodológicas) pretendem ser seletivamente incluídas ou omitidas para produzir um exemplo com um conjunto específico de características. Tendo sido provido com a descrição e ilustração do presente pedido, alguém versado na técnica pode imaginar variações, modificações e exemplos alternativos que caem dentro do espírito dos aspectos mais amplos do conceito inventivo geral incorporado neste pedido que não se afastam do escopo mais amplo.

REIVINDICAÇÕES

1. Método para proteger um serviço online através de um modelo de aprendizado contínuo, caracterizado pelo fato de compreender:

reunir um conjunto de sinais de segurança do serviço online, em que o conjunto de sinais de segurança são reunidos em uma janela rolante de tempo;

identificar se cada sinal de segurança do conjunto de sinais de segurança é malicioso ou benigno;

balancear os sinais maliciosos do conjunto de sinais de segurança com sinais benignos do conjunto de sinais de segurança para produzir um conjunto de dados de treinamento balanceado; e

produzir um modelo preditivo com base no conjunto de dados de treinamento balanceado, em que o modelo preditivo está configurado para identificar se um sinal de segurança é malicioso ou benigno.

2. Método de acordo com a reivindicação 1, caracterizado pelo fato de que identificar se cada sinal de segurança do conjunto de sinais de segurança é malicioso ou benigno ainda compreende:

examinar cada sinal de segurança em um modelo de produção, em que o modelo de produção é produzido por um treinador de modelo de acordo com o conjunto de dados de treinamento balanceado e configurado para produzir um resultado de detecção de se um dado sinal de segurança é malicioso ou benigno;

transmitir o resultado de detecção para um usuário analista; e

em resposta a receber uma ação do usuário analista em relação ao resultado de detecção, atualizar o resultado de detecção para indicar se o dado sinal de segurança é malicioso ou benigno.

3. Método de acordo com a reivindicação 2, caracterizado

pelo fato de que um atacante automatizado simula um ataque sobre o serviço online, e em que identificar se cada sinal de segurança do conjunto de sinais de segurança é malicioso ou benigno ainda compreende:

receber uma notificação do atacante automatizado que identifica os sinais de segurança produzidos em resposta ao ataque; e

tratar os sinais de segurança produzidos em resposta ao ataque como maliciosos independentemente do resultado de detecção.

4. Método de acordo com a reivindicação 2, caracterizado pelo fato de que identificar se cada sinal de segurança do conjunto de sinais de segurança é malicioso ou benigno ainda compreende:

extrair características do dado sinal de segurança;

determinar se as características extraídas do dado sinal de segurança satisfazem um conjunto de características projetado por um usuário administrativo como definindo um tipo de ataque;

em resposta a determinar que as características extraídas satisfazem o conjunto de características, designar o dado sinal de segurança como malicioso; e

em resposta a determinar que as características extraídas não satisfazem o conjunto de características, designar o dado sinal de segurança como benigno.

5. Método de acordo com a reivindicação 4, caracterizado pelo fato de que balancear os sinais maliciosos com os sinais benignos ainda compreende:

identificar o tipo de ataque de cada um dos sinais maliciosos;

balancear um número relativo de tipos de ataque para um conjunto de tipos de ataque observados para os sinais maliciosos por pelo menos um de:

aumentar uma quantidade relativa de tipos de ataque sub-representados no conjunto de tipos de ataque; e

diminuir uma quantidade relativa de tipos de ataque sobrerrepresentados no conjunto de tipos de ataque.

6. Método de acordo com a reivindicação 4, caracterizado pelo fato de que os conjuntos de características são identificados em um documento estruturado submetido pelo usuário administrativo, o documento estruturado especificando tipos de características e campos de dados para observar nos sinais de segurança, e as características dos conjuntos de características são dinamicamente extraídas dos sinais de segurança com base no documento estruturado sem precisar modificar o código.

7. Método de acordo com a reivindicação 1, caracterizado pelo fato de que sinais de dados históricos estão incluídos no conjunto de sinais de segurança.

8. Método de acordo com a reivindicação 1, caracterizado pelo fato de balancear os sinais maliciosos com os sinais benignos ainda compreende:

identificar um dispositivo no serviço online do qual pelo menos um sinal malicioso foi reunido dentro da janela rolante; e

remover os sinais benignos associados com o dispositivo do conjunto de sinais de segurança.

9. Sistema para proteger um serviço online através de um modelo de aprendizado contínuo, caracterizado pelo fato de compreender:

um processador; e

um dispositivo de armazenamento de memória, que inclui instruções que quando executada pelo processador são operáveis para:

receber sinais de segurança de dispositivos dentro do

serviço online;

extrair vetores de característica de cada um dos sinais de segurança, em que que um dado vetor de característica provê valores numéricos que representam um status de um dado dispositivo do qual um dado sinal de segurança é recebido;

produzir resultados de detecção para cada um dos vetores de característica através de modelos preditivos associados, em que um dado resultado de detecção identifica se o dado sinal de segurança é indicativo de atividade maliciosa ou benigna sobre o dado dispositivo;

definir uma janela rolante, em que a janela rolante inclui uma pluralidade de sinais de segurança e resultados de detecção associados que foram recebidos dentro de um quadro de tempo de um tempo corrente;

produzir um conjunto de dados de treinamento balanceado para a janela rolante, em que para produzir o conjunto de dados de treinamento balanceado o sistema está ainda configurado para:

identificar um tipo de ataque de cada um dos sinais de segurança na janela rolante identificado como sendo indicativo de atividade maliciosa;

aumentar uma quantidade de sinais de segurança identificados com tipos de ataque sub-representados na janela rolante em relação a sinais de segurança identificados com tipos de ataque sobrerrepresentados; e

juntar cruzados os sinais de segurança identificados como sendo indicativos de atividade maliciosa com os sinais de segurança identificados como sendo indicativos de atividade benigna para produzir cenários de ataque para a janela rolante; e

atualizar, de acordo com um algoritmo de aprendizado de máquina, os modelos preditivos associados com base no

conjunto de dados de treinamento balanceado.

10. Sistema de acordo com a reivindicação 9, caracterizado pelo fato de que atualizar os modelos preditivos associados inclui substituir os modelos de produção, utilizados para produzir os resultados de detecção, por modelos de desenvolvimento, desenvolvidos do conjunto de dados de treinamento balanceado de acordo com o algoritmo de aprendizado de máquina, em resposta ao algoritmo de aprendizado de máquina indicar que os modelos de desenvolvimento mais precisamente identificam se os sinais de segurança são indicativos de atividade maliciosa ou benigna sobre os dispositivos de acordo com o conjunto de dados de treinamento balanceado.

11. Sistema de acordo com a reivindicação 9, caracterizado pelo fato de que sinais históricos estão incluídos na janela rolante, em que os sinais históricos incluem sinais de segurança reunidos fora do quadro de tempo.

12. Sistema de acordo com a reivindicação 9, caracterizado pelo fato de que os sinais de segurança recebidos dos dispositivos dentro do serviço online incluem sinais de segurança gerados em resposta a um atacante automatizado executando as atividades maliciosas conhecidas sobre o serviço online, e em que os resultados de detecção produzidos pelos sinais de segurança gerados em resposta ao atacante automatizado executando as atividades maliciosas conhecidas são ajustados para indicar que o dado sinal de segurança é indicativo de atividade maliciosa com base em uma notificação do atacante automatizado.

13. Sistema de acordo com a reivindicação 9, onde para produzir o conjunto de dados de treinamento balanceado, o sistema está ainda caracterizado pelo fato de estar configurado para:

remover os sinais de segurança identificados como sendo indicativos de atividade benigna da janela rolante em resposta a identi-

ficar que um dispositivo específico do qual os sinais de segurança identificados como sendo indicativos de atividade benigna foram recebidos está associado com um ou mais sinais de segurança identificados como sendo indicativos de atividade maliciosa na janela rolante.

14. Sistema de acordo com a reivindicação 9, onde para produzir o conjunto de dados de treinamento balanceado, o sistema está ainda caracterizado pelo fato de estar configurado para:

identificar um tipo de dispositivo do qual cada um dos sinais de segurança na janela rolante identificados com sendo indicativos de atividade benigna foi recebido; e

aumentar uma quantidade de sinais de segurança identificados com tipos de dispositivo sub-representados na janela rolante em relação a sinais de segurança identificados com tipos de dispositivo sobrerrepresentados.

15. Dispositivo de armazenamento legível por computador que inclui instruções executáveis por processador para proteger um serviço online através de um modelo de aprendizado contínuo, caracterizado pelo fato de compreender:

reunir um conjunto de sinais de segurança do serviço online, em que o conjunto de sinais de segurança são reunidos em uma janela rolante de tempo;

examinar cada sinal de segurança do conjunto de sinais de segurança através de modelos preditivos para identificar se cada sinal de segurança é malicioso ou benigno, em que os modelos preditivos estão configurados para produzir um resultado de detecção de se um dado sinal de segurança é malicioso ou benigno com base em um vetor de característica definido por um usuário administrativo;

associar os sinais de segurança com os resultados de detecção para identificar os sinais de segurança como sinais maliciosos ou sinais benignos;

balancear os sinais maliciosos com os sinais benignos para produzir um conjunto de dados de treinamento balanceado, que inclui:

identificar um tipo de ataque de cada um dos sinais maliciosos;

identificar um tipo de dispositivo do qual cada um dos sinais benignos foi reunido;

equalizar números relativos de sinais maliciosos na janela rolante com base em tipos de ataque identificados para produzir um conjunto de exemplos de ataque;

equalizar números relativos de sinais benignos na janela rolante com base em tipos de dispositivo identificados para produzir um conjunto de exemplos benignos identificados; e

juntar cruzados o conjunto de exemplos de ataque com pelo menos uma porção do conjunto de exemplos benignos para balancear um número de exemplos de ataque no conjunto de exemplos de ataque relativo a um número de exemplos benignos no conjunto de exemplos benignos; e

refinar os modelos preditivos com base no conjunto de dados de treinamento balanceado e um algoritmo de aprendizado de máquina.

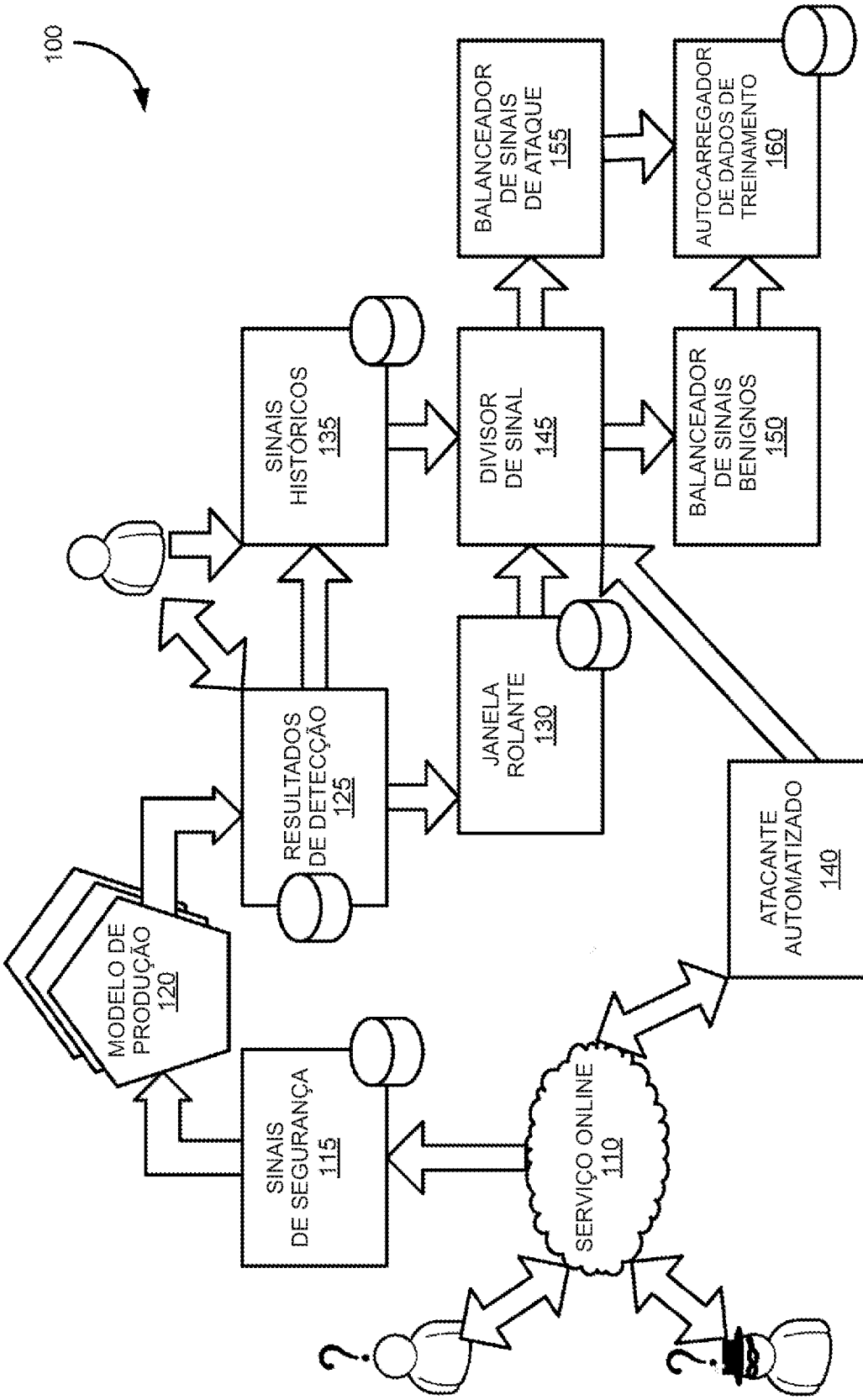
100
SERVIÇO ONLINE
110
SINAIS
DE SEGURANÇA
115
MODELO DE
PRODUÇÃO
120
RESULTADOS
DE DETECÇÃO
125
SINAIS
HISTÓRICOS
135
JANELA
ROLANTE
130
DIVISOR
DE SINAL
145
BALANCEADOR
DE SINAIS
DE ATAQUE
155
BALANCEADOR
DE SINAIS
BENIGNOS
150
AUTOCARREGADOR
DE DADOS DE
TREINAMENTO
160
ATACANTE
AUTOMATIZADO
140

FIG. 1A

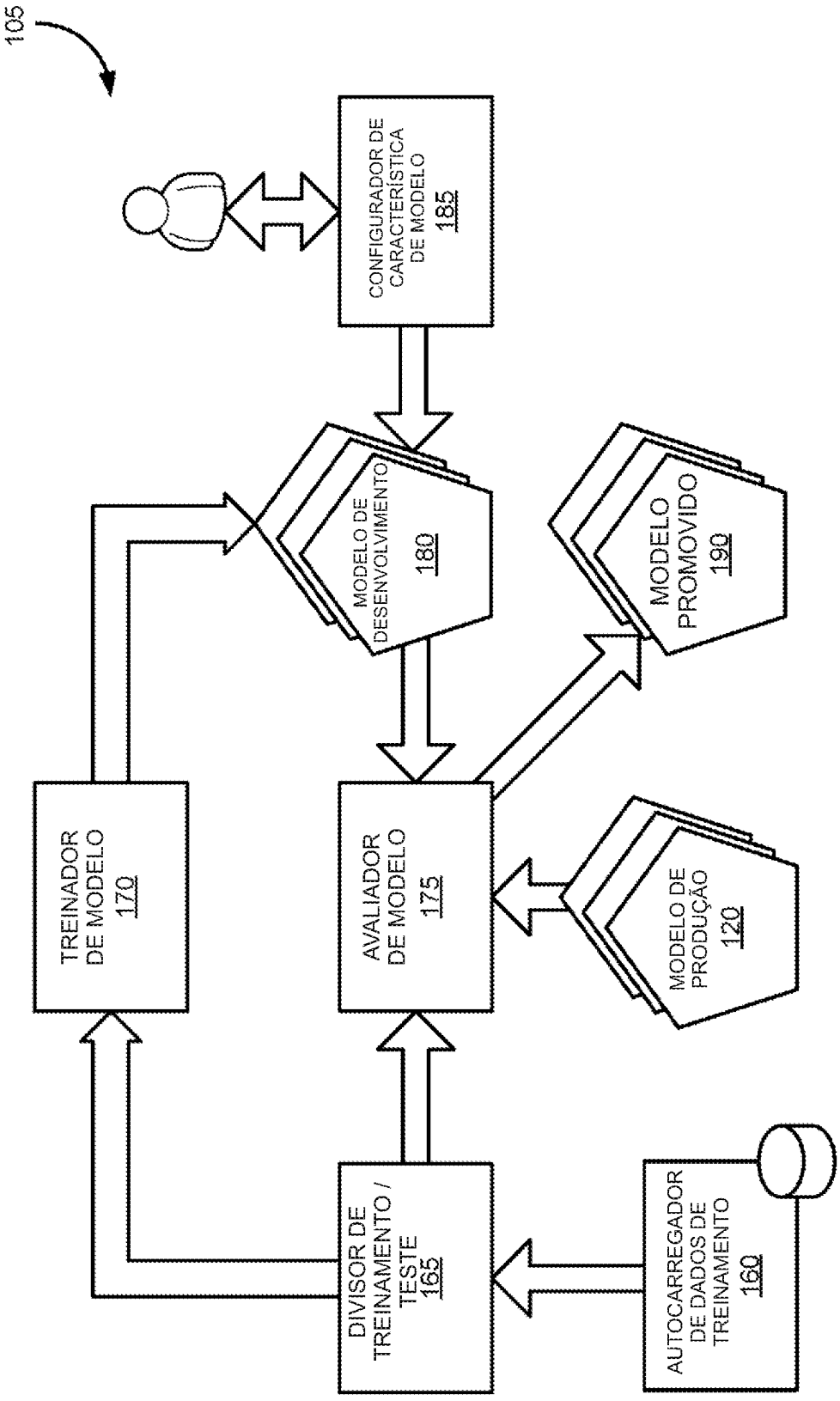
105
CONFIGURADOR DE CARACTERÍSTICA DE MODELO 185
MODELO DE DESENVOLVIMENTO 180
MODELO PROMOVIDO 190
TREINADOR DE MODELO 170
AVALIADOR DE MODELO 175
MODELO DE PRODUÇÃO 120
DIVISOR DE TREINAMENTO / TESTE 165
AUTOCARREGADOR DE DADOS DE TREINAMENTO 160

FIG. 1B

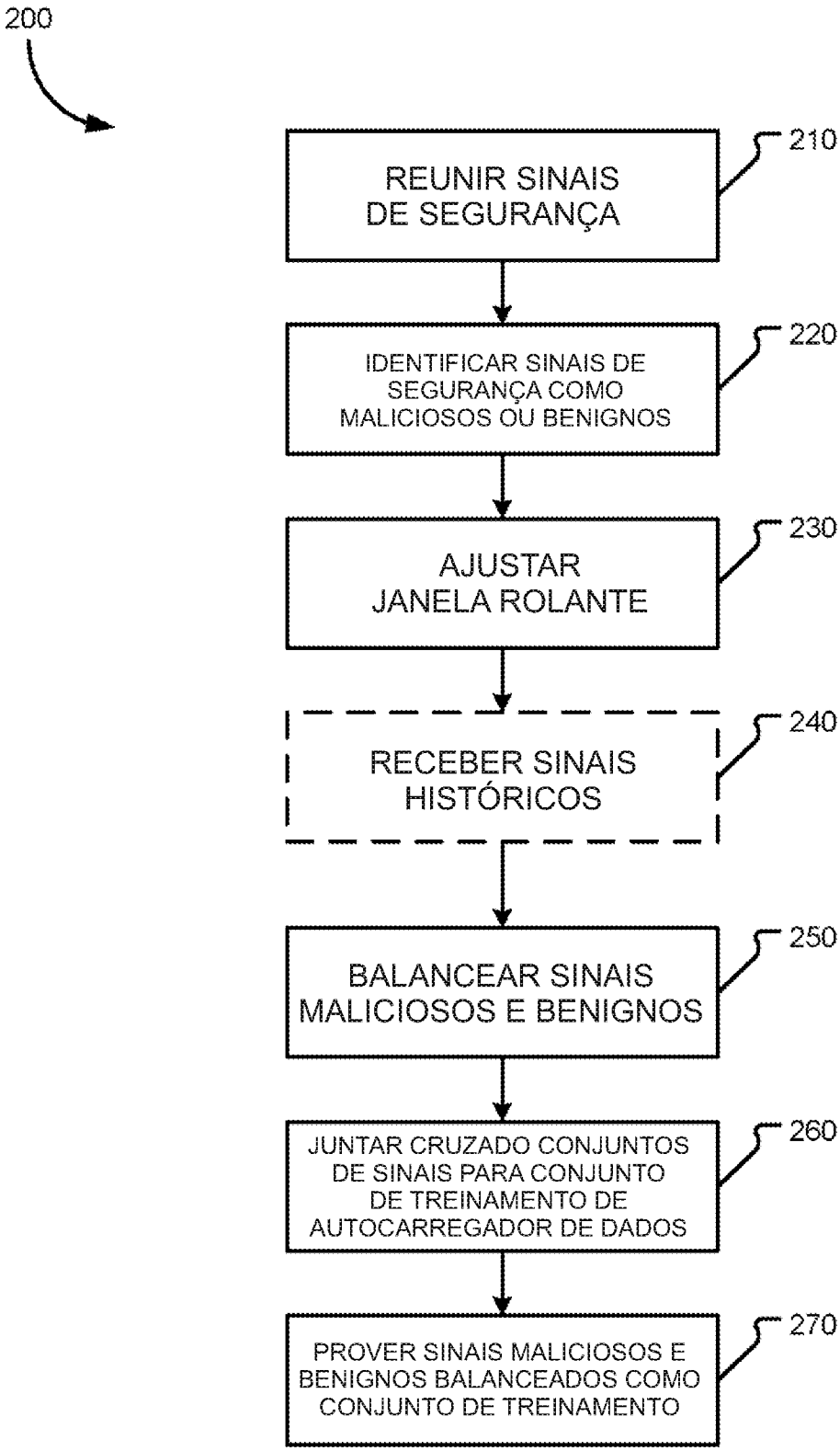
200
REUNIR SINAIS DE SEGURANÇA
210
IDENTIFICAR SINAIS DE SEGURANÇA COMO MALICIOSOS OU BENIGNOS
220
AJUSTAR JANELA ROLANTE
230
RECEBER SINAIS HISTÓRICOS
240
BALANCEAR SINAIS MALICIOSOS E BENIGNOS
250
JUNTAR CRUZADO CONJUNTOS DE SINAIS PARA CONJUNTO DE TREINAMENTO DE AUTOCARREGADOR DE DADOS
260
PROVER SINAIS MALICIOSOS E BENIGNOS BALANCEADOS COMO CONJUNTO DE TREINAMENTO
270

FIG. 2

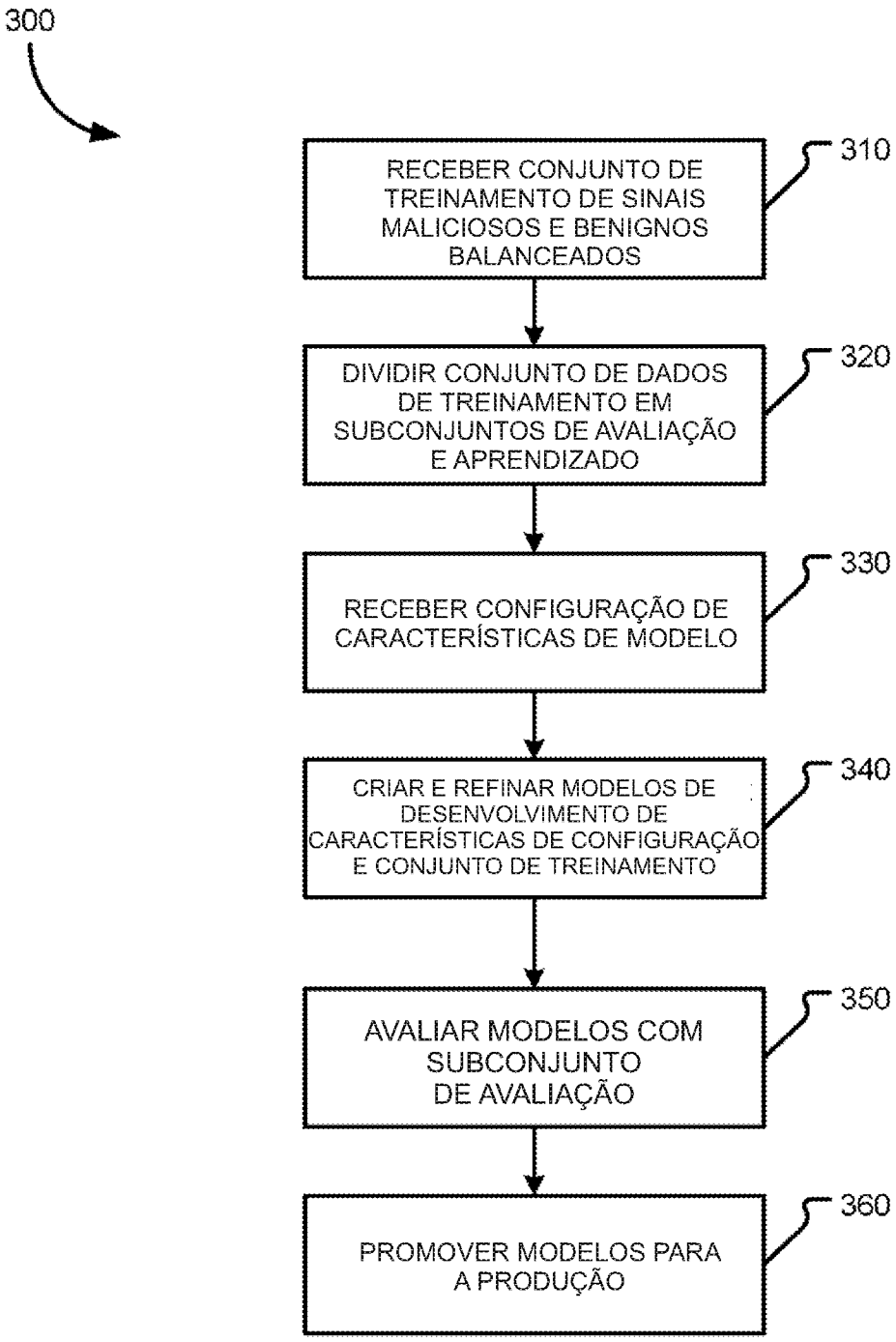
300
RECEBER CONJUNTO DE TREINAMENTO DE SINAIS MALICIOSOS E BENIGNOS BALANCEADOS
310
DIVIDIR CONJUNTO DE DADOS DE TREINAMENTO EM SUBCONJUNTOS DE AVALIAÇÃO E APRENDIZADO
320
RECEBER CONFIGURAÇÃO DE CARACTERÍSTICAS DE MODELO
330
CRIAR E REFINAR MODELOS DE DESENVOLVIMENTO DE CARACTERÍSTICAS DE CONFIGURAÇÃO E CONJUNTO DE TREINAMENTO
340
AVALIAR MODELOS COM SUBCONJUNTO DE AVALIAÇÃO
350
PROMOVER MODELOS PARA A PRODUÇÃO
360

FIG. 3

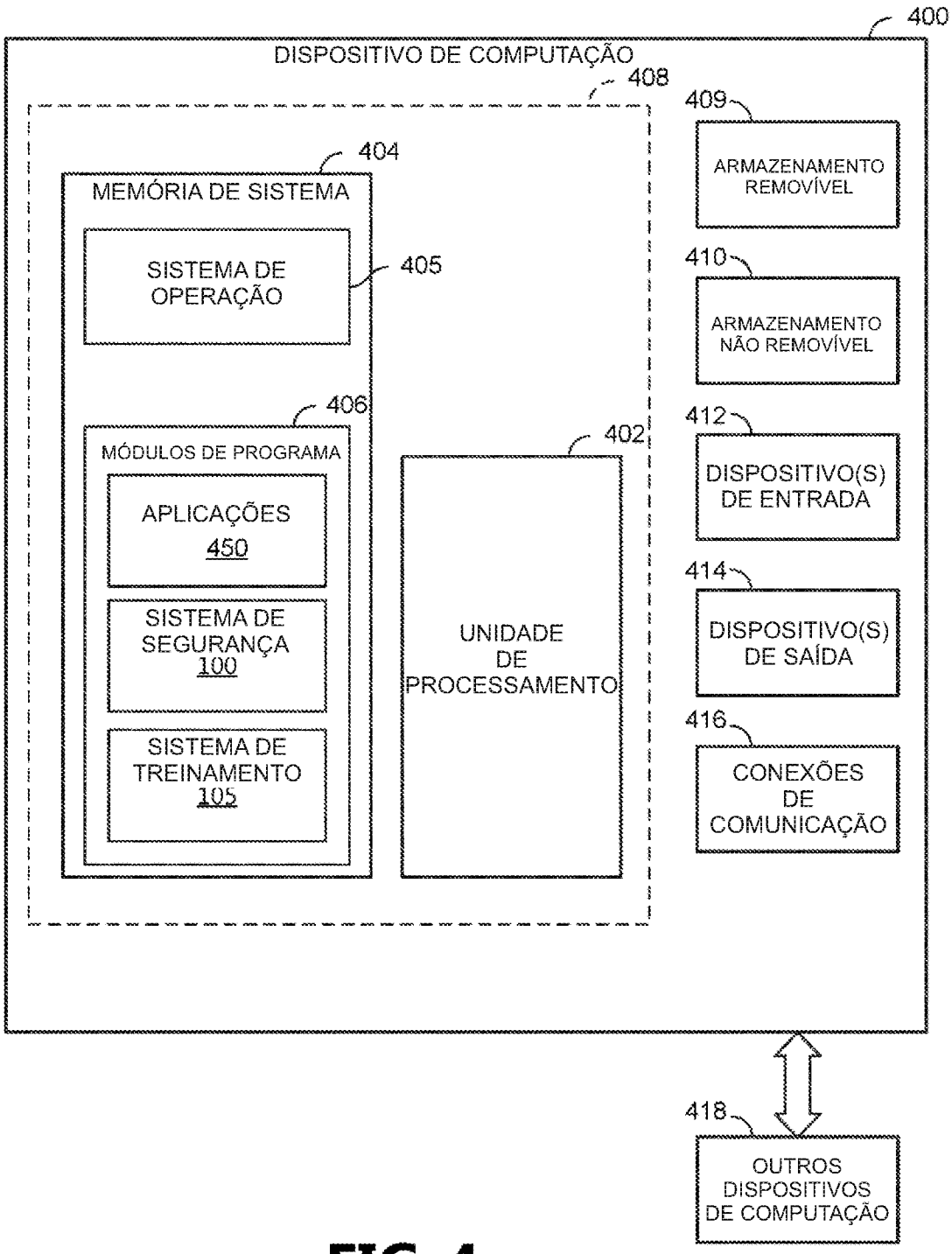
400
DISPOSITIVO DE COMPUTAÇÃO
408
409
ARMAZENAMENTO REMOVÍVEL
404
MEMÓRIA DE SISTEMA
SISTEMA DE OPERAÇÃO
405
410
ARMAZENAMENTO NÃO REMOVÍVEL
406
MÓDULOS DE PROGRAMA
APLICAÇÕES 450
SISTEMA DE SEGURANÇA 100
SISTEMA DE TREINAMENTO 105
402
UNIDADE DE PROCESSAMENTO
412
DISPOSITIVO(S) DE ENTRADA
414
DISPOSITIVO(S) DE SAÍDA
416
CONEXÕES DE COMUNICAÇÃO
418
OUTROS DISPOSITIVOS DE COMPUTAÇÃO

FIG. 4

RESUMO

Patente de Invenção: **"APRENDIZADO CONTÍNUO PARA DETECÇÃO DE INTRUSÃO"**.

A presente invenção refere-se a balancear os sinais observados utilizados para treinar modelos de detecção de intrusão de rede que permite uma alocação mais precisa de recursos de computação para defender a rede de indivíduos maliciosos. Os modelos são treinados contra dados ao vivo definidos dentro de uma janela rolante e dados históricos para detectar características definidas pelo usuário nos dados. Ataques automatizados asseguram que vários tipos de ataques estão sempre presentes na janela de treinamento de rolante. O conjunto de modelos é constantemente treinado para determinar qual modelo colocar em produção para alertar os analistas de intrusões, e/ou para automaticamente desenvolver contramedidas. Os modelos são continuamente atualizados conforme as características são redefinidas e conforme os dados na janela rolante mudam, e o conteúdo da janela rolante é balanceado para prover dados suficientes de cada tipo observado pelo qual treinar os modelos. Quando balanceando o conjunto de dados, sinais de baixa população são sobrepostos sobre sinais de alta população para balancear os seus números relativos.