Um Modelo de Rede Neurais para a Detecção de Ataques Cibernéticos, Baseado nos Dados da Competição KDD99

Article · December 2019

CITATIONS
0

READS
279

1 author:

Rodrigo Moura Fernandes
Pontifícia Universidade Católica do Rio de Janeiro
1 PUBLICATION **0** CITATIONS

SEE PROFILE

Um Modelo de Rede Neurais para a Detecção de Ataques Cibernéticos, Baseado nos Dados da Competição KDD99.

Dezembro 2019

Rodrigo Moura Fernandes
rmoura@yahoo.com

Introdução

Nos dias de hoje, a preocupação com a segurança digital é maior do que nunca. Nunca tanto se ouviu falar em sistemas com problemas, que foram invadidos, dados que foram roubados e em prejuízos causados por sistemas mal protegidos.

O número de dispositivos conectados a internet cresce significativamente a cada ano e traz, consigo, um número maior de possíveis alvos de ataques. Várias recomendações são dadas para proteger esses sistemas, mas o fato é que é impossível prever que tipo de novos ataques serão usados ou que sistemas estarão susceptíveis a que tipos de ataques, bem como instalar patches regularmente em todos esses dispositivos, considerar o uso de autenticação em duas camadas, implementar o monitoramento de redes e sistemas de detecção de intrusos, entre outros.

Do relatório de 2019 da empresa eSentire, intitulado: "2019 eSentire Threat Intelligence Spotlight":

> *"Cyberattacks against businesses in the U.K. are becoming more frequent, more sophisticated and more successful as the arms race continues between adversaries and targets.*
> *…*
> *"Going forward, threat actors will continue to evolve their techniques and increase attack volume. "*

Portanto, vemos cada vez mais uma maior necessidade de melhores técnicas de detecção de ataques a sistemas corporativos ou individuais expostos a internet.

O que são ataques Cibernéticos

Num mundo conectado, os computadores se comunicam praticamente de maneira ininterrupta uns com os outros. É essa facilidade de comunicação que traz os benefícios que vemos hoje diariamente em nosso cotidiano, com nossos dispositivos conectados. Para essa comunicação, cada dispositivo, ou classe de dispositivos se comunica numa ou mais determinadas línguas, os chamados "protocolos".

Os ataques cibernéticos se caracterizam por uma tentativa, mal ou bem-sucedida, de provocar um comportamento anormal em um dispositivo, visando alguma alteração de comportamento, nesse ou em outro dispositivo, que seja capaz de dar alguma vantagem ao indivíduo que perpetra tais tentativas. Um ataque bem-sucedido pode trazer uma série de benefícios ao perpetrante e uma série de problemas e prejuízos, não só aos perpetrados, mas a uma imensa cadeia de pessoas e empresas ligadas ao dispositivo afetado.

Existem uma série de tipos de ataques cibernéticos diferentes. Cada tipo de ataque tem sua especificidade, grau de dificuldade e de proficiência técnica necessária a realizá-lo com êxito. Cyber criminosos usam uma variedade de métodos, como, por exemplo:

1) **Malware**

   Malwares são programas de computador explicitamente escritos e desenvolvidos para realizar alguma atividade criminosa em algum dispositivo. Os famosos vírus de computador são o tipo de malware mais conhecido e mais comentado. Em geral, os indivíduos se protegem desse tipo de comportamento usando os softwares anti-virus comercializados e usados amplamente.

   Porém existem a mais diversas técnicas de desenvolvimento de software que fazem com que as “assinaturas” que os vírus de computador ou malwares tenham fujam ao padrão conhecido e, portanto, utilizado nos softwares de detecção. Os casos mais comuns são os casos de códigos criados com polimorfismos.

   A infecção com malwares pode ter várias origens. Hoje em dia é muito comum a disseminação desse tipo de conteúdo através da inclusão desses códigos em programas de computador, teoricamente “oficiais”, porém disponíveis gratuitamente na internet em algum site.

   Há de se relatar aqui também que, em muitos casos, um software de uma empresa reconhecida amplamente e com boa reputação pode estar, a revelia de seus clientes tendo comportamentos não condizentes com a sua função inicial. Isso hoje em dia é muito comum e muito utilizado para a coleta indevida dos mais diversos tipos de dados pessoais ou não de seus usuários, com a função de ajudar o desenvolvimento do mesmo, de capitalizar a empresa com a venda desses dados ou fazer propaganda guiada a determinados perfis, por exemplo.

2) **Phishing**

   Os ataques de phishing variam dos mais simples possíveis, até bem sofisticados, categorizados como Spear Fishing. Ataques de fishing (pescaria em inglês) são em geral feitos através de e-mails não solicitados a usuários, tentando captar sua atenção e fazer com que a vítima faça alguma ação que permita que o ataque tenha sucesso. Um caso antigo, mas muito emblemático, foi o caso do vírus “I Love you”. Na primeira década desse século, rodou o mundo um email que tinha o assunto “I Love You” e que, na maioria dos tipos de programas de leitura de email, quando essa mensagem era aberta, se espalhava através de um email para todos os contatos da vítima, e assim por diante.

   Um único ataque desses, bem simples, recorre ao interesse do usuário e sua curiosidade, pontos fracos dos seres humanos, para fazer a sua disseminação. O vírus a época era bem potente e levava as máquinas dos usuários afetados a ficar

"travada"pela alta quantidade de e-mails disparados por segundo. Os sistemas de email empresariais e seus links de internet também era sobrecarregado, impedindo assim que negócios e comunicações legítimas fossem feitas durante a execução e antes da remediação desse evento.

3) **Ransomware**
   Ransomware são programas instalados ou por ataques de fishing ou por outras técnicas de infecção que, após ativados, criptografam todo o computador da vítima e exibem uma mensagem na tela, após as informações da vítima já estarem encriptadas, pedindo um resgate em algum tipo de moeda. Em 2017, dois grandes ataques desse tipo dominaram a mídia: os ataques do WannaCry e NoPetya. Esses ataques tem um alto poder destruidor e ainda assombram as organizações.

4) **Denial of Service**
   Denial of service são ataques em que, o alto volume de dados enviados aos dispositivos atacados faz com que esses fiquem sobrecarregados em processar o tráfego entrante e fiquem, com essa alta carga de processamento, muito lentos ao responder ao tráfego apresentado. Em geral, um criminoso virtual, compromete centenas ou milhares de dispositivos conectados primeiramente e, com esses dispositivos, lança um ataque desse tipo em direção a uma organização. É necessário ter muita banda de internet e um número muito grande de dispositivos para que um ataque desses seja bem-sucedido.
   O sucesso desse tipo de ataque se dá quando o alvo não consegue mais responder ao volume de tráfego enviado a ele e, mesmo o tráfego legítimo, de funcionários ou clientes sofre com isso.
   Para realizar a primeira parte do ataque, ou seja, comprometer uma vasta rede de dispositivos, o criminoso precisa utilizar um vetor de ataque que consiga infectar máquinas remotas sem o devido consentimento ou anuência da parte passiva.

5) **Outros Tipos de Ataque**
   Podemos citar o ping flooding, ping of death, smurf attacks, buffer overflows, heap overflows, stack overflows, format string attacks, direct access attacks, privilegie escalation, SQL Injections, Zero day exploits como outros tipos de ataques que tem um ponto em comum: todos eles dependem exclusivamente da comunicação direta e irrestrita com o alvo. O meio de ataque em comum desses casos é uma comunicação entre dois pontos, que, por algum motivo alheio, descoberto pelo perpetrador, fará com que o alvo se comporte de uma maneira não prevista pelo desenvolvedor do software abusado e com isso, fornecerá um ganho ao atacante.

A parte comum entre todos os cyber ataques é que a vasta eles precisam da camada de rede, ou para infectar ou para serem bem sucedidos de alguma maneira, por isso a investigação da legitimidade de tráfegos de rede é tão importante para a prevenção e remediação desses ataques.

Toda a comunicação entre dois dispositivos quaisquer, conectados entre si através de uma rede de comunicação, seja ela qual for, se dá através de um protocolo de comunicação. É através deles que toda essa comunicação flui e que dispositivos "se falam" e funcionam.

O que são os chamados Protocolos?

Os sistemas de comunicação entre dispositivos trocam, entre si, mensagens bem definidas e que suscitam um tipo de resposta dentro de um conjunto de respostas possíveis para cada situação. A todo esse conjunto de regras, chamamos de protocolos de comunicação.

Fazer com que uma determinada mensagem seja recebida pelo destinatário é apenas uma parte dos problemas que os protocolos de comunicação precisam resolver. Os dados recebidos precisam ser avaliados quanto ao contexto do progresso dessa troca de mensagens em andamento. Por isso, os protocolos de comunicação precisam incluir também, regras de contexto. Esse tipo de regra é expressa na sintaxe desses protocolos. As regras que verificam se os dados trocados são relevantes ao contexto expressam a semântica da comunicação.

Em geral, os protocolos definem regras claras para os formatos de dados, formatos dos endereços de comunicação, mapeamento de endereços (endereços IP para endereços MAC, por exemplo), roteamento, detecção de erros de transmissão, ACKs (notificações de ciência de recebimento de mensagens), perda de informações, direção do fluxo de dados, controle de sequência e controle de fluxo de dados. Um aspecto que deve ser notado é que os protocolos de comunicação estão distribuídos em camadas distintas. Uma comunicação entre dois dispositivos não vai se dar apenas por um determinado protocolo, mas por um conjunto de protocolos, divididos em camadas, que assim são divididas, por tipos de tarefa a serem executados.

Em geral, para protocolos de comunicação, segue-se o modelo OSI, desenvolvido internacionalmente e baseado em redes que funcionavam antes da internet ser inventada. O modelo OSI prevê regras estritas e um sistema de camadas bastante rigoroso. Também nesse modelo, as camadas são divididas da seguinte maneira:

1) **Camada de aplicação:** essa é a camada mais perto de um usuário de um serviço, a mais alta delas, fazendo a interface com o software que está sendo usado por um usuário ou serviço.

2) **Camada de Apresentação:** essa camada estabelece o contexto entre as entidades da camada de aplicação. É uma camada de abstração, que faz o mapeamento entre diferentes sintaxes e semânticas. Essa camada é independente dos dados e os transforma para uma forma que a aplicação a aceite e a saiba tratar.

3) **Camada de Sessão:** a camada de sessão controla o diálogo entre dois dispositivos. Ele inicia, gerencia e termina as conexões entre o dispositivo local e o remoto. Estabelece também pontos de controle, suspensão e reinicialização de cada sessão.

4) **Camada de Transporte:** essa camada é responsável por controlar a confiabilidade na transmissão dos dados, com o controle de fluxo, segmentação e de-segmentação e controle de erro. Essa camada faz a retransmissão de uma mensagem que tenha sido perdida ou que tenha sido detectado um erro em sua

transmissão / recebimento no destinatário. É na camada de transporte, também, que são fornecidos os Acks (notificações de ciência). Uma fácil analogia da camada de transporte é compara-la a um serviço de correio que inspeciona o envelope de uma carta e determina se essa foi ou não entregue intacta, sem se preocupar com o conteúdo dessa.

5) **Camada de Rede:** a camada de rede prove os meios necessários para transferir sequencias dados de tamanho variável, os chamados pacotes de dados entre nós conectados. Nessa camada, cada nó de rede possui um endereço que permite receber e transmitir mensagens. Nessa camada, cada nó só precisa especificar a mensagem e o destinatário, para que as camadas mais abaixo realizem a transferência. Se a mensagem for muito grande, essa camada também ficará responsável por "fatiar" essa mensagem em várias outras de menor tamanho. Essas mensagens fatiadas deverão ser recompostas no destinatário, através do número de sequencia de cada mensagem enviada.

6) **Camada de Link de dados:** essa camada é responsável por iniciar ou terminar a conexão entre dois dispositivos conectados e tenta detectar e corrigir possíveis erros da camada física. Essa camada também define o controle de fluxo entre os nós ligados.

7) **Camada Física:** é a camada que converte bits em sinais elétricos, de rádio ou óticos.

Porque os Protocolos de Comunicação são Importantes?

Existe uma gama imensa de protocolos, públicos ou privados. Cada protocolo estabelece um método de comunicação entre os dispositivos que desejam se comunicar. Eles definem um sistema de regras que se constitui na fundação da comunicação entre dois ou mais dispositivos. Eles também definem a sintaxe, ou seja, a linguagem, a semântica, a sincronização das comunicações e estabelecem regras para corrigir eventuais erros. Portanto, os protocolos regem toda a comunicação síncrona ou assíncrona entre dispositivos conectados a uma rede de telecomunicações. Para entender porque os protocolos de comunicação são tão importantes em relação a cyber ataques, primeiro devemos entender como os ataques acontecem.

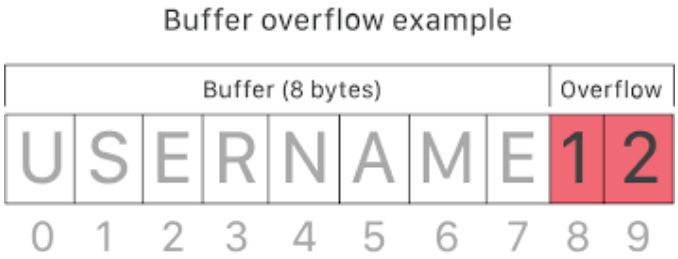
Como os Ataques Cibernéticos Acontecem?

Os vetores de ataque são diversos, mas a grande maioria dos ataques passa através da comunicação entre dois dispositivos conectados.

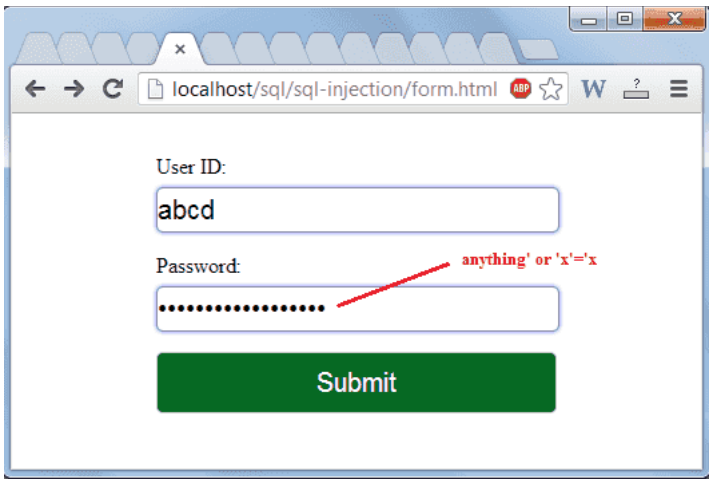
Essa grande parte dos ataques a equipamentos conectados se dá através da exploração de algum tipo de falha ou particularidade de protocolos de comunicação ou da maneira com o que alguns tipos específicos de dispositivos reagem a uma determinada mensagem ou pacote de características pouco convencionais criados por um criminoso.

Um tipo de ataque muito famoso é o “Buffer Overflow”. Esse ataque é, de maneira bem simplificada, simplesmente um endereço de memória que é sobrescrito com outra informação. E dependendo da área que essa nova informação estará armazenada, ela pode corromper a estrutura de memória de um dispositivo e disparar uma ação indesejada ou causar algum outro tipo de evento, que o criminoso explorará.

Do exemplo abaixo, muito simples, ilustrarei uma hipotética situação. Digamos que um sistema espera que todos os nomes de usuários sejam de até 8 bytes. Digamos também, que contiguamente, em sua memória, seja guardado o nível de permissão desse usuário, que varia de 1, convidado, até 12, um super usuário. Se um criminoso conseguir fazer uma mensagem para o sistema informando um nome de usuário de 9 bytes, por exemplo, os dois últimos bytes serão escritos no endereço que está armazenado esse nível de permissão e, assim, concedendo ao usuário, direitos de administrador do sistema. Sistemas mal escritos ou muito complexos abrem as mais diversas portas para ataques desses tipos.



Outro tipo de ataque comum é o do tipo SQL Injection, aonde o criminoso supondo sobre a infraestrutura do alvo, cria uma “mensagem” para o Banco de Dados forçando um tipo de resposta, como ilustra a figura a seguir.



Os Cyber ataques não necessariamente precisam ser executados de um ambiente externo a um alvo. Como em geral as proteções internas de empresas são muito mais fracas do que as proteções de perímetro, em geral os ataques perpetrados de dentro de empresas tem uma chance muito maior de atingir os objetivos do criminoso e de uma maneira mais profunda.

Um caso emblemático também, o STUXNET, revelou que um pen drive contaminado conseguiu instalar um software que se instalou dentro de usinas nucleares Iranianas. Há relatos que outras instalações de outros países também foram afetadas. E o mais incrível desse ataque é que ele ficou quase uma década latente, sem ser descoberto.

Alvos de Alto Potencial de Danos de Ataques
Os ataques a sistemas de infraestrutura podem vir a ser uma ameaça bem maior do que se imagina inicialmente, pois o custo de se atualizar sistemas de infraestrutura é gigantesco e envolve sistemas muito críticos e que afetam uma grande população. Em geral os sistemas de infraestrutura mais visados são:

1) **Sistemas de Controle**
   Ataques a sistemas de controles de fábricas ou de qualquer instalação industrial. Podem controlar desde válvulas ou até o acesso a infraestrutura física.
2) **Energia**
   Criminosos podem atacar sistemas de transmissão de energia que abastecem cidades ou regiões.
3) **Finanças**
   Sistemas financeiros são um alvo escolhido para ataques pela natureza de alta interconexão entre sistemas da mesma espécie e, obviamente, pela motivação de diretos ganhos financeiros.
4) **Telecomunicações**
   Ataques de negação de serviço (DOS) em geral escolhem como alvo empresas de telecomunicação, com o objetivo de diminuir ou silenciar a troca de dados entre pessoas e empresas.
5) **Transportes**
   Ataques bem-sucedidos a infraestrutura de transportes tem um efeito similar ao ataque a infraestrutura de telecomunicações.
6) **Água e Esgoto**
   Sistemas de tratamento e abastecimento de água e tratamento de esgoto também são controlados por dispositivos que podem gerar um risco alto a população se comprometidos.

Por tudo evidenciado nos parágrafos anteriores, pode-se perceber que o controle e o monitoramento do tráfego de informações entre sistemas é chave para a identificação e proteção desses mesmos sistemas de ataques cibernéticos.

Outro ponto crucial para o desenvolvimento dessa ideia é a natureza dos sistemas que usamos hoje em dia. O dinamismo com que esses sistemas precisam ser utilizados, em sua maioria, exige uma **solução que não aumente a latência de resposta dos sistemas**. Por isso, a solução ótima é encontrar um modelo que seja capaz de identificar potenciais ataques com muita rapidez, sem que afete o tráfego e o comportamento normal dos sistemas. Outro ponto relevante é que sistemas de proteção não devem só estar presentes em computadores pessoais ou servidores, com alto poder de processamento, mas em todos os dispositivos conectados, de servidores a telefones, câmeras IP e demais dispositivos de "*Internet of Things*" (ou Internet das coisas, ou IoT).

Esses dispositivos IoT, em geral tem muito pequeno poder de processamento e bateria, e portanto, a solução de sistemas de detecção de intrusos, nesse caso, devem ser muito leves e rápidas para não comprometer o tempo de autonomia de bateria ou processamento dos dispositivos. **Esse é um ponto crucial para a modelagem que será apresentada**.

Como Detectar Cyber Ataques e Remediar Risco

Para a detecção de cyber ataques, pode-se adotar um número vasto de contra medidas nos níveis organizacionais, de procedimentos e técnicos.

No nível organizacional, a medida mais efetiva para o combate as ameaças é treinamento. Sem dúvida, a maioria dos ataques bem-sucedidos foi precedido por uma ação indevida de algum usuário ou empregado da parte afetada. Prover o treinamento adequado e aumentar a exposição do assunto com os colaboradores de uma empresa é uma das maneiras mais eficazes de se deter e mitigar ataques cibernéticos.

No nível de procedimentos, deve fazer um *assessment* de todos os riscos envolvidos na empresa e na sua cadeia de suprimentos, como por exemplo, nos sistemas utilizados por essa e desenvolvidos por terceiros. Talvez essa seja a etapa mais custosa, demorada e, pelo fato de não depender exclusivamente da empresa que faz o *assessment*, ser a etapa mais difícil de ser exaurida por completo.

No nível técnico, há um arsenal de medidas que podem ser tomadas, principalmente em três áreas:

1) **Projeto e Arquitetura de Rede**
   O projeto técnico e a arquitetura dos sistemas e redes de uma empresa devem ser pensados de maneira estratégica e em camadas, propiciando assim a estratificação de seus recursos que ajudam principalmente na tarefa de identificar comportamentos anormais.

2) **Instalação de softwares**
   Vários tipos de softwares ajudam na identificação, bloqueio e monitoramento do tráfego, dispositivos, equipamentos de rede e comportamento de usuários e sistemas. Esses sistemas são geralmente chamados de IDS ou IPS dependendo da sua característica de apenas alarmar o evento, no caso de IDS, ou alarmar o evento e intervir no ataque, no caso de IPS.

   O estabelecimento de níveis padrão de serviço e métricas “basais” ajuda na identificação de comportamentos que fujam a normalidade e, assim, possam ser algum indicativo de risco.

3) **Identificação, acesso e log**
   A correta gestão de Identificação, autorização, acesso e log de todos os sistemas é um pilar fundamental no tripé de boas práticas de segurança empresarial.

Além de se fundamental a identificação de possíveis ataques, a parte de remediação não pode ser deixada de lado, pois é ela que vai garantir a volta a normalidade. Nesse quesito, dois pilares são fundamentais:

1) **Comunicação**
   A comunicação interna e externa é dos fatores mais relevados em ataques cibernéticos e pode se tornar o calcanhar de Aquiles de um evento dessa natureza. A falta de comunicação adequada pode gerar, por si só, prejuízos maiores do que o próprio ataque, na medida em que as pessoas possivelmente afetadas por esse evento não estarão cientes do ocorrido e não poderão assim, tomar medidas preventivas ao ocorrido.
   O risco a empresa que foi atacada então vem na forma de ações judiciais e propaganda negativa fazendo com que, no limite, a empresa afetada possa ir à falência. Fato com precedentes na história.

2) **Seguir o Manual de Procedimentos já estabelecido**
   Em toda área de segurança da informação de empresas deve haver um manual, bem elaborado e de conhecimento dos principais atores do evento, e esse manual deve ser seguido à risca. Parece óbvio, mas as vezes, um ato impulsivo e bem-intencionado tentando uma remediação mais rápida pode trazer mais prejuízos a empresa afetada. O Manual de Procedimentos também serve como uma salvaguarda para a empresa, na medida em que ela segue os melhores passos, auditados e pensados para uma situação especial e específica, fazendo com que o risco de ações judiciais e medidas governamentais sejam diminuídos.

A Importância de Sistemas de Detecção de Intrusão (IDS ou IPS)

A principal tarefa de sistemas de detecção de intrusão é alertar e "preparar" a rede que esse monitora para lidar com possíveis ataques em andamento. As funções de um Sistema de detecção a Intrusão (IDS), incluem:

- Analisar e monitorar atividades de usuários e sistemas;
- Analisar configurações de sistemas e suas vulnerabilidades;
- Examinar a integridade do Sistema e de seus arquivos;
- Habilidade de reconhecer padrões de ataques típicos;
- Analisar padrões de atividade anormais;
- Rastrear violações de politicas de uso dos usuários do Sistema.

O propósito de sistemas de IDS é ajudar a sistemas e redes de computadores a lidar com possíveis ataques a esses, na medida em que os Sistemas de IDS coletam dados de sistemas e redes e os analisam em comparação com padrões já pré-determinados. **No âmbito desse trabalho, não farei distinção entre IDS e IPS pois é a detecção do evento que está sendo analisada.**

Uma pesquisa da empresa israelense Checkpoint estima que 99% das empresas não estejam corretamente protegidas. Portanto, ainda hoje há um vasto horizonte para aplicações de proteção a redes de computadores e dispositivos em geral.

Redes Neurais e IDS

A detecção de uso indevido é o processo de tentar identificar instâncias de ataques à rede, comparando a atividade atual com as ações esperadas de um invasor. As abordagens mais atuais para detecção de uso indevido envolvem o uso de sistemas especializados baseados em regras para identificar padrões de ataques conhecidos. Essas técnicas são menos bem-sucedidas na identificação de ataques que variam de acordo com os padrões esperados. As redes neurais artificiais fornecem o potencial para identificar e classificar as atividades da rede com base em fontes de dados limitadas, incompletas e não lineares.

As redes neurais artificiais oferecem o potencial de resolver vários problemas encontrados pelas outras abordagens atuais da detecção de intrusões. Redes neurais artificiais têm sido propostas como alternativas a análise estatística de IDSs. A análise estatística envolve a comparação estatística dos eventos atuais com um conjunto predeterminado de critérios padrão. A técnica é mais frequentemente empregada na detecção de anomalias de tráfego de redes e na determinação da similaridade de eventos àqueles que são indicativos de um ataque.

Redes neurais foram propostas especificamente para identificar as características típicas de comportamento de um sistema e identificar variações estatisticamente significativas do padrão estabelecido pelos usuários dos sistemas de uma rede corporativa.

Uma Rede Neural artificial consiste matematicamente em um *"somatório de produtos"* visando transformar um conjunto de entradas (parâmetros) em um conjunto de saídas (categorias a serem pesquisadas) esperadas, por meio de um conjunto de nós e tipos de conexões entre eles. Nessas redes existem, portanto, nós de entrada, nós de saída e uma quantidade ótima a ser determinada de nós entre a entrada e a saída, que são as chamadas camadas ocultas. A conexão entre duas unidades deve ter algum peso, usada para determinar quanto uma unidade afetará a outra.

Dois tipos de arquitetura de redes neurais podem ser distinguidos, porém, no âmbito desse trabalho, apenas trabalharemos com o que é chamado de Treinamento supervisionado, aonde na fase de treinamento, a rede aprende a saída desejada para uma determinada entrada ou padrão.

Vantagens dos Sistemas de Detecção de Uso Indevido Baseados em Redes Neurais

1) **Flexibilidade:** uma rede neural seria capaz de analisar os dados da rede, mesmo se os dados estiverem incompletos ou alterados de alguma maneira.

2) **Predição:** uma rede neural, também, pode oferecer um recurso preditivo para a detecção de uso indevido, identificando a probabilidade de um determinado evento ou série de eventos.

3) **Experiência:** uma rede neural ganha experiência para melhorar a capacidade de determinar onde esses eventos provavelmente ocorrerão no processo de ataque.

4) **Capacidade de Aprendizado:** a vantagem mais importante das redes neurais na detecção de comportamento indevido é a capacidade da rede neural de "aprender" as características dos comportamentos e identificar novas ocorrências diferentes de todas as que já foram observadas anteriormente.

KDD99 Dataset

Para esse trabalho, foi escolhido o dataset KDD99, originalmente proposto num desafio da KDD Cup. A KDD Cup é uma competição anual de *Data Mining* e *Knowledge Discovery* organizado pela *ACM Special Interest Group* em *Knowledge Discovery* e *Data Minining* e acontece desde 1997.

O KDD99 é uma modificação do conjunto de dados que se originou de um programa IDS conduzido no Laboratório Lincoln do MIT, foi avaliado primeiro em 1998 e novamente em 1999. Financiado pelo DARPA (*Defense Advanced Research Projects Agency* – agencia militar americana voltada a fazer investimentos em tecnologias de defesa dos Estados Unidos da América), produziu o que é frequentemente referido como o conjunto de dados DARPA98. Posteriormente, esse conjunto de dados foi filtrado para uso no *International Knowledge Discovery and Data*, Competição de Ferramentas de *Data Mining*, resultando no que reconhecemos como os dados do KDD CUP 99, ou KDD99.

Os dados da Competição proposta em 1999 contêm centenas de milhares de citações no Google e servem de base, até hoje, a estudos profundos sobre intrusões de rede e *Data Mining*.

A competição de 1999 propõe a criação de um classificador de tipos de invasões de rede. A tarefa é construir um modelo preditivo capaz de distinguir entre conexões “ruins'', chamadas intrusões ou ataques e conexões normais, ie., “boas''.

Foi fornecido um conjunto padrão de dados a serem auditados, que inclui uma ampla variedade de intrusões simuladas em um ambiente de rede militar. O concurso de detecção de intrusão do KDD de 1999 usa uma versão deste conjunto de dados, obtidos num ambiente para adquirir nove semanas de dados de rede brutos trafegados numa rede local (LAN) simulando uma LAN típica da Força Aérea dos EUA. Eles operavam a LAN como se fosse um verdadeiro ambiente da Força Aérea, mas a enchiam com múltiplos ataques.

Os dados brutos de treinamento eram cerca de quatro gigabytes de dados oriundos de sete semanas de tráfego de rede. Isso resultou em cerca de cinco milhões de registros de conexão, que estão contidos no arquivo analisado.

Uma conexão é definida por uma sequência de pacotes TCP iniciando e terminando em alguns momentos bem definidos, entre os quais os dados transitam para e de um endereço IP de origem para um endereço IP de destino sob algum protocolo bem definido. Cada conexão é rotulada como normal ou como ataque, com exatamente um tipo de ataque específico. Cada registro de conexão consiste em cerca de 100 bytes.

O TCP é um protocolo de comunicação da camada de transporte de rede do Modelo OSI, apresentado anteriormente. Ele tem a função de verificar se os dados são enviados na sequência correta e sem erros via rede. Nesta mesma camada de transporte (camada 4 OSI) trafegam a maioria das aplicações que são usadas amplamente na internet, devido sua versatilidade e robustez. O Protocolo de controle de transmissão provê confiabilidade, entrega na sequência correta e verificação de erros dos pacotes de dados, entre os diferentes nós da rede, para a camada de aplicação.

A seguir, tabelas com as descrições do conjunto de dados usado.

feature name	*description*	*type*
hot	number of ``hot'' indicators	continuous
num_failed_logins	number of failed login attempts	continuous
logged_in	1 if successfully logged in; 0 otherwise	discrete
num_compromised	number of ``compromised'' conditions	continuous
root_shell	1 if root shell is obtained; 0 otherwise	discrete
su_attempted	1 if ``su root'' command attempted; 0 otherwise	discrete
num_root	number of ``root'' accesses	continuous
num_file_creations	number of file creation operations	continuous
num_shells	number of shell prompts	continuous
num_access_files	number of operations on access control files	continuous
num_outbound_cmds	number of outbound commands in an ftp session	continuous
is_hot_login	1 if the login belongs to the ``hot'' list; 0 otherwise	discrete
is_guest_login	1 if the login is a ``guest''login; 0 otherwise	discrete

feature name	*description*	*type*
duration	length (number of seconds) of the connection	continuous
protocol_type	type of the protocol, e.g. tcp, udp, etc.	discrete
service	network service on the destination, e.g., http, telnet, etc.	discrete
src_bytes	number of data bytes from source to destination	continuous
dst_bytes	number of data bytes from destination to source	continuous
flag	normal or error status of the connection	discrete
land	1 if connection is from/to the same host/port; 0 otherwise	discrete
wrong_fragment	number of ``wrong'' fragments	continuous
urgent	number of urgent packets	continuous

feature name	*description*	*type*
count	number of connections to the same host as the current connection in the past two seconds	continuous
	Note: The following features refer to these same-host connections.	
serror_rate	% of connections that have ``SYN'' errors	continuous
rerror_rate	% of connections that have ``REJ'' errors	continuous
same_srv_rate	% of connections to the same service	continuous
diff_srv_rate	% of connections to different services	continuous
srv_count	number of connections to the same service as the current connection in the past two seconds	continuous
	Note: The following features refer to these same-service connections.	
srv_serror_rate	% of connections that have ``SYN'' errors	continuous
srv_rerror_rate	% of connections that have ``REJ'' errors	continuous
srv_diff_host_rate	% of connections to different hosts	continuous

(fonte: https://kdd.ics.uci.edu/databases/kddcup99/task.html)

Testes

Mais de duas centenas de testes foram executados variando-se os parâmetros: camadas “dense”, quantidade de neurônios, dropout, otimizador, learning rate, épocas, batch, função de ativação das camadas dense, função de ativação da camada de saída. Os resultados dos testes foram anotados nas colunas loss e accuracy.
Foram testados de 1 a 7 camadas dense, quando o incremento poderia mostrar algum ganho.
A quantidade de neuronios foi incrementada, começando em 4 e depois, 8, 16, 32, 64, 128, 256, 512 e 1024.
O parâmetro dropout foi testado começando em 0,5, onde se verificou que era muito alto. Durante os testes, foi verificado que o melhor dropout era de 0.2, e assim foi fixado para o restante das iterações.

Vários otimizadores foram testados: Adam, SGD, Adagrad, Adadelta. Houve uma performance próxima entre o SGD e o Adam, mas o Adam ainda saiu-se melhor e foi assim, fixado. Vários parâmetros, como diferentes learning rates, Nesterov e Amsgrad foram testados. O melhor resultado na media, foi o otimizador Adam, com learning rate de 0.001 e amsgrad=false.

No início dos testes comecei com 10 épocas de treinamento, mas foi verificado que na maioria do casos a rede ainda estava aprendendo na ultima época, e por isso, foi aumentado o numero de épocas até 30 e ficado esse valor, para maior equivalência dos resultados obtidos.

O tamanho do batch foi alterado, pois verificou-se que cada caso tem um melhor tamanho de batch associado. Em geral, redes mais complexas tendem a "pedir" um batch menor, (entre 4 a 256 neurônios isso acontece em geral), ao passo que os resultados de melhoria parecem ambíguos com redes de 512 e 1024 neurônios, aonde um aumento ou diminuição do batch pode melhorar o resultado. **Por estratégia, o primeiro teste com uma rede mais complexa era feito com a melhor combinação de parâmetros do teste anterior.**

As funções de ativação foram testadas, inclusive a priori do roll de testes apresentado a seguir, e a função Linear para as camadas densas e a softmax na saída promoveram os melhores resultados, independente das outras configurações da rede.

Conclusão

Para detector os ataques mencionados no desafio KDD99, as redes neurais performam muito melhor do que o algoritmo vencedor do desafio. Conseguimos acurácias acima de 99.5% com redes muito pequenas, de apenas 4 neurônios, em uma só camada.

Redes de 16 neurônios já produziram resultados com 99,88% de acurácia, muito perto do máximo atingido de 99,90% com 1024 neurônios.

Redes com mais de uma camada não foram, nos testes, executados, capazes de produzir resultados que tanto justificassem o seu uso, pela maior complexidade e necessidade de processamento, quanto também em resultado absoluto se compararam as redes de uma camada só e 8 ou 16 neuronios.

Vimos também que, redes de apenas uma camada e 4 neurônios, muito simples, podem ser usadas comercialmente, pois atigiram taxas de sucesso muito altasde acurácia (99.6%) e tem uma velocidade de processamento muito alta, atigindo o objetivo do trabalho que era criar redes neurais capazes de detectar ataques cibernéticos com uma boa acurácia, nesse caso em 99.6% com essa massa de dados fornecida, e um "peso"computacional muito pequeno.

Como sugestão para a evolução futura e outros trabalhos, seria de interesse científico se estudar o comportamento dessa rede de 4 neurônios em outras bases de dados. Vale a pena dizer também, que o caso estudado tinha apenas duas dezenas de ataques propostos e isso não é a realidade. Com isso, é possível e provável que essa topologia de rede seja insuficiente para detectar centenas ou milhares de ataques diferentes, apesar da capacidade de extrapolação das redes neurais.