Detecção de intrusos: Uma abordagem usando redes neurais

RICARDO BERNARDO DOS SANTOS
WALMIR MATOS CAMINHAS
LUCIANO DE ERRICO

UFMG - Universidade Federal de Minas Gerais
CPDEE - Centro de Pesquisa e Desenvolvimento em Engenharia Elétrica
CEP 31.270-901 - Belo Horizonte (MG)
{rbsantos,caminhas,l.errico}@cpdee.ufmg.br

Resumo. Para promover a segurança necessária em sistemas computacionais são usados mecanismos de controle de acessos, geralmente senhas. Entretanto, se estas senhas forem comprometidas um acesso não autorizado pode ser conseguido por um invasor, podendo causar danos. Este trabalho fornece uma introdução à área de detecção de intrusos e propõe a aplicação de um modelo de rede neural como ferramenta no auxílio a esta detecção. O modelo de rede neural utilizado é do tipo MLP *(multilayer perceptron)* e é proposto para um sistema *offline* de detecção por anomalia. Esta estrutura obteve resultados significativos, reportando acertos de 95.5% para detecção de comportamento intruso baseando-se em dados gerados por um simulador de ambiente para 15 usuários.

Palavras Chaves: Segurança de redes, Detecção de Intrusos, Redes Neurais.

1 Introdução

Atualmente, com o crescimento acelerado das redes de computadores [16, 13], a integridade e a privacidade das informações que trafegam nestas redes tem gerado uma constante preocupação com a sua segurança. A tecnologia mais utilizada para promover segurança em redes, principalmente aquelas conectadas à uma estrutura externa (Internet), é o firewall [12]. Contudo, é perceptível que um sistema que implemente uma proteção muito severa penaliza os usuários, restringindo a liberdade e a flexibilidade de utilização dos recursos protegidos. Visando desenvolver sistemas e modelos que, ao mesmo tempo que promovam uma proteção contra intrusos, não interfiram de forma muito profunda na flexibilidade e utilização do ambiente, é que surge a área de pesquisa chamada Detecção de Intrusão [1].

Deste então, vários mecanismos de segurança em redes de computadores estão sendo projetados para impedir o acesso não autorizado a sistemas e a dados restritos. Contudo nem sempre é possível prevenir o acesso não autorizado. O que pode ser feito é uma tentativa de detecção de intrusão e uma possível reparação das falhas do sistema e dos danos ocorridos após a invasão, impedindo que um ataque semelhante ocorra futuramente.

A dificuldade na prevenção do acesso não autorizado está na grande incerteza sobre a segurança existente nos atuais sistemas de computadores e o verdadeiro conhecimento possuído por quem utiliza os mesmos. Vários sistemas, por serem projetados em tempo limitado, com financiamentos restritos ou até mesmo por uma equipe de técnicos não bem qualificados, muitas vezes não são testados da forma adequada antes de serem colocados disponíveis no mercado. Ocorre assim uma indesejável situação quando estes são expostos a intrusos inescrupulosos e sem limites.

É esperado que os sistemas possam prover confiabilidade, integridade e garantia contra falhas de serviços e tentativas de invasão. Porém, nenhum sistema pode ser considerado totalmente seguro, pois nem tudo é sabido sobre ele antes que o mesmo seja colocado para o uso de usuários finais. Há muitas formas de promover um sistema seguro, no entanto há muitas maneiras, também, de burlarmos a segurança fornecida por eles.

2 Intrusos

O principal problema relativo a um sistema sob ameça está formulado no princípio que o atacante pode ser ou não um usuário autorizado no sistema. Se este atacante for um usuário autorizado a usar os recursos disponíveis pelo ambiente ele é um intruso interno. Caso ele não seja autorizado, ele é um intruso externo. Estas duas grandes categorias de intrusos foram definidas em [1].

Uma representação de ameaça em um sistema pode ser vista na figura 1, onde os recursos protegidos são vistos como anéis de controle e anéis de usuários.

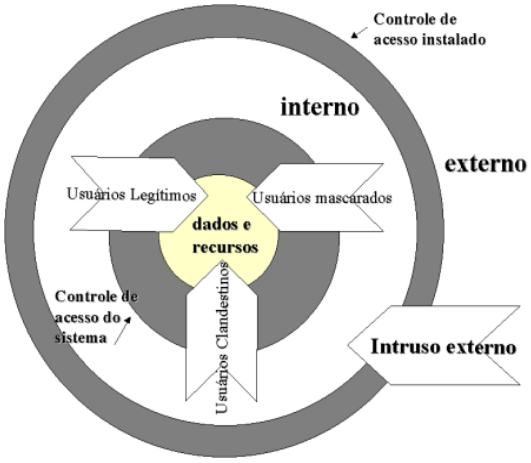

Figura 1: *representação de perigo do sistema*

Os intrusos internos são usuários que abusam de seus previlégios para tirarem algum proveito. Eles podem ser vistos como mascarados (aqueles que se fingem como usuários legítimos do sistema) e clandestinos (aqueles que tem o poder de direcionar os dados de controle para si próprios).

Os intrusos mascarados são caracterizados como intrusos internos, contudo eles podem ser vistos, também, como invasores externos que tiveram sucesso em uma penetração no sistema. A principal característica que diferencia o mascarado do usuário legítimo do sistema é o seu perfil de uso. Talvez esta peculiaridades seja a forma mais aconselhada de detectarmos este tipo de usuário.

Os intrusos clandestinos são o tipo mais difícil de intruso a ser detectado. Eles são usuários legítimos do sistema e aproveitam das condições privilegiadas que possuem para conseguir vantagens não autorizadas.

Esta hipótese é confirmada considerando que estes intrusos podem manipular os dados auditados que estão sendo armazenados e desta forma direcioná-los para outro caminho.

Informações mais detalhadas sobre tipos de usuários de um sistema pode ser obtida em [1].

3 Detecção de Intrusos

3.1 Introdução

Para que seja possível encontrar uma solução para o problema de invasão em sistemas de computadores, algumas técnicas foram desenvolvidas. Estas são divididas em duas grandes categorias: detecção de intrusos por anomalia [5, 4, 8, 14] e detecção de intrusos por abuso [2, 10, 8].

Ambas as categorias visam à descoberta de intrusos em um sistema. Na primeira, vários perfis de comportamento dos usuários do ambiente são determinados com o uso de métricas pré-estabelecidas. Então, quando uma ação culminar em uma variação extrema deste perfil, esta ação é considerada anômala. Já na segunda, as ações dos usuários são comparadas a várias ações anômalas e intrusivas conhecidas como abusivas pelo ambiente. Assim, caso ocorra muita semelhança entre as ações do usuário e as abusivas, uma situação de abuso será constatada.

3.2 Detecção de Intrusos por Anomalia

Comportamento anômalo é uma atividade intrusiva que foge aos padrões de algum tipo de comportamento considerado normal. Geralmente dentro de um sistema computacional cada usuário possui um padrão de uso, o qual é considerado o seu perfil específico de uso do sistema. Quando qualquer ação ocorre dentro da conta de um usuá-rio e esta ação não é um subconjunto de ações normais, ela é considerada anormal e dentro deste conjunto de a-ções anormais existe um outro subconjunto que pode ser avaliado como atividades intrusiva. Estas atividades intrusivas são provenientes de uma variação muito brusca no perfil conhecido do usuário da conta e podem ter sido executadas por algum invasor ou até mesmo pelo próprio dono da conta, proveniente de algum avanço ou regressão em seus conhecimentos.

Há uma dificuldade muito grande em se determinar quem, realmente, esta causando uma atividade anômala, e também se a atividade normal é realmente proveniente do usuário da conta. Isto porque o invasor nem sempre está envolvido em mudanças bruscas de padrão na conta invadida, tornando desta forma suas ações consideravelmente normais e não avaliadas como intrusivas.

3.3 Detecção de Intrusos por Abuso

Neste tipo de detecção é considerado que existem situações de intrusão que podem ser armazenadas e quando qualquer invasor executar um sequência de atividades semelhantes a alguma atividade intrusiva já conhecida pelo sistema, o mesmo detecta aquela ação como sendo intrusiva.

Este sistema possui o inconveniente de que nem sempre é possível ou viável armazenar todos os padrões de intrusão existentes. Além disso os padrões depois de armazenados não serão, na maioria das vezes, atualizados automaticamente com o aparecimento de novos padrões. Contudo este tipo de abordagem para detecção de

intrusão está sendo muito utilizado, pois não acarreta um custo computacional elevado e consequentemente não interfere muito no desempenho do sistema.

4 Princípios de Sistemas de Detecção de Intrusos (SDIs)

Os sistemas de detecção de intrusos podem ser classificados a partir de três premissas: baseado na fonte de dados, baseado nos modelos de detecção de intrusão e baseado na forma de aplicação.

- baseado na fonte de dados os sistemas podem ser: baseados em *host* e *multihost*, isto é, análises de *logs* provenientes do uso do sistema com o decorrer do tempo; baseados em rede(*network based*), que usam, análises do fluxo de dados que trafegam através da estrutura de rede que está sendo monitorada; ou ambos.
- baseado nos modelos de detecção de intrusão, os sistemas podem ser: baseados em detecção por anomalia; baseados em deteção por abuso; ou híbridos.
- baseados na forma de aplicação os sistemas podem ser: *online* (tempo real) [14], que procuram detectar a invasão no momento que ela está ocorrendo; *offline* [7], que buscam detectar a invasão após ela já ter ocorrido, usando os dados armazenados durante um período de tempo de uso do sistema.

4.1 Características de um bom SDI

Como mostrado em [11], um sistema de detecção de intrusos deve ser capaz de:

- trabalhar continuamente sem supervisão humana e ser capaz também de executar em segundo plano (*background*), de forma transparente para os usuários e não ser visualizado como uma caixa preta (*black box*);
- ser tolerante a falhas, sobrevivendo a quedas do sistema e não tendo a base de conhecimentos corrompida;
- ser resistente a subversões, monitorando a si próprio e evitando auto-ataques;
- gerar o mínimo de *overhead* ao sistema, não acarretando quedas bruscas no desempenho;
- observar variações de comportamentos intrusos;
- adaptar-se com facilidade a mudanças de padrões e mecanismo de defesas inerentes do ambiente computacional, bem como a mudanças provenientes de evoluções do sistema avaliado;
- ser difícil de ser enganado.

5 O uso de Redes Neurais em Sistemas de Detecção de Intrusos

5.1 Introdução as redes neurais

Um modelo de rede neural é identificado pela sua topologia e pelo seu método de aprendizado. Como as redes neurais possuem inspiração biológica, elas assemelham-se, a modelos neurais do cérebro humano. O modelo artificial pioneiro de neur ônio biológico foi proposto por McCulloch e Pitts em 1943 [9]. Após esta proposição varias atualizações e varias idéias foram surgindo para adaptar as redes neurais a problemas de cunho real.

A aprendizagem em redes neurais é caracterizada pela capacidade que as redes possuem de modificar o seu comportamento em resposta a eventos ou situações que ocorrem no ambiente externo e que fornecem um conjunto de entradas, o qual pode ser associado a um conjunto de saídas desejadas ou não. Através de um algoritmo de treinamento, este conjunto de entrada acarreta um ajuste dos pesos da rede, produzindo um conjunto de resposta adequado que concorda com os padrão de entrada ou com os padrões armazenados pela rede. Após a execução consistente e correta deste aprendizado a rede torna-se capaz de compor similaridades e generalizar situações que ainda não foram aprendidas. Durante o treinamento da rede é muito importante a monitoração de quanto tempo ela deve ficar treinando, pois um treinamento muito prolongado pode levá-la a um estado de *overlearning*. Nesta situação, a rede perde a capacidade de generalização pois tenta decorar os padrões de entrada.

O comportamento de uma rede neural, após treinada, é determinado pelos pesos existentes entre as conexões de seus neurônios e as funções de ativação usadas para o treinamento da rede. Estas funções são consideradas os limiares de ativação da rede. Toda rede neural possui uma topologia, e esta topologia está diretamente ligada ao problema que se deseja resolver, à complexidade deste problema e a outras abordagens.

Maiores detalhes sobre redes neurais podem ser encontrados em [6].

5.2 Modelos de SDIs baseados em Redes Neurais

As Redes Neurais são conhecidas pela sua alta capacidade de adaptação, aprendizado e generalização. Os modelos baseados em redes neurais visam explorar estas caracteristicas e, através de treinamento, gerar uma estrutura que seja capaz de classificar padrões de intrusão ou normalidade.

Estes modelos podem ser usados para definir tanto sistemas de detecção por anomalia quanto sistemas de

detecção por abusos, bem como *host-based* e *network-based*. Para que seja possível desenvolver um sistema destes é necessário definir qual será a topologia da rede, seu algoritmo de treinamento, as variáveis quantitativas e qualitativas que representem o modelo e também os dados que iram compor o treinamento da rede. Além disso deve-se definir como serão feitas as verificações e adaptabilidades dos dados após a rede ter sido treinada.

Sistemas de detecção de intrusos por anomalia usando redes neurais podem ser vistos em: Debar [3], onde é proposto um sistema *online* que aprende a predizer qual será o próximo comando a ser usado por um usuário do sistema; Ryan [7], onde é proposto um sistema *offline* que usa métricas referentes a frequência de utilização de comandos para identificar o legítimo perfil de um usuário e em Tan [15], onde são usadas várias métricas para compor um vetor de dados a ser aplicado à rede neural, objetivando assim, a detecção de padrões inesperados em sessões de uso destas métricas.

Um sistema de detecção de intrusos por abuso usando redes neurais pode ser visto em Cansian [2], onde um módulo usando uma rede neural do tipo MLP com 126 entradas e 1 saída é usado para testar assinaturas de intrusão previamente treinadas e verificar se o sistema está ou não sob ataque.

Outro ponto importante na utilização de modelos baseados em redes neurais para detecção por anomalia é a forma como será tratada a entrada dos dados. Em [7] uma discussão é feita a respeito da determinação da janela w a ser aplicada a uma rede de detecção de comportamento intruso. Esta janela indica quantos dados (comandos) serão capturados para serem aplicados à entrada da rede neural. A rede neural é recorrente, isto é, parte da saída é retornada para a entrada do próximo passo. Assim, ela está sujeita a esquecer os padrões mais antigos de comportamento com o passar do tempo. Desta forma se o tamanho da janela for muito pequeno, ocorrerão muitos falsos-positivos; e se o tamanho da janela for muito grande, a rede não generalizará bem novos perfis, isto é, poderão ocorrer mais falsos-negativos.

5.2.1 Vantagens e desvantagens dos modelos baseados em Redes Neurais

- Vantagens:
  - são adaptáveis e generalistas, conseguindo possivelmente detectar novos ataques com os quais ainda não teve contato;
  - não necessita de especialistas para codificarem novas regras de ataques e/ou normalidade;
  - as redes neurais não dependem que os dados obedeçam certas distribuições de probabilidade (ex.: gaussiana), ao contrário de métodos estatísticos que as vezes utilizam destas hipó-teses para conseguir bons resultados.
  - após o treinamento não promove grande perda de desempenho na verificação dos dados.
- Desvantagens:
  - alguns algoritmos de treinamento de redes neurais não garantem convergência e outros são computacionamente exaustivos;
  - possuem dificuldades de escalabilidade, isto é, com o aumento da quantidade de usuários talvez seja necessário agrupá-los ao invés de usar grupos individuais; e nem sempre é garantido que o aumento das classes mantenha bons resultados.
  - promovem uma abstração muito grande nos resultados, não facilitando os processos de extra-ção de regras.

6 A Proposta deste trabalho

A proposta deste trabalho é mostrar como uma rede neural é utilizada como parte de um sistema *offline* de detec-ção de intrusão por anomalia. A rede utilizada é do tipo MLP *(multilayer perceptron)* e a filosofia é utilizar o sistema operacional Linux e algums comandos mais usados por seus usuários, para analisar o perfil de comportamento dos mesmos.

A opção de utilização do Linux é baseada nas condi-ções de trabalho que ele fornece, dentre elas: ser principalmente um clone do UNIX, sistema operacional onde ocorrem a maioria dos ataques, nos dias de hoje; ser um sistema operacional de rede que funciona em plataformas PCs convencionas; estar ganhando um grande mercado como servidor de serviços para Internet(principal fonte dos ataques).

6.1 A topologia da rede neural utilizada

A rede neural projetada possui 81 nodos de entrada (os quais representam as métricas monitoradas), trinta nodos intermediários e quinze nodos de saídas (os quais representam o número de usuários do sistema que está sendo monitorado). O conjunto de entradas usado para testes e validação do modelo é composto de 405 vetores divididos em: 80% para treinamento e 20% para validação. Um conjunto de 68 vetores anômalos é usado para testes da capacidade de distinção de comportamento intruso. A topologia da rede neural utilizada é mostrada na figura 2.

6.2 As métricas

As métricas usadas como entrada do sistema são dependentes de quais características são consideradas importantes para representarem o perfil de um usuário. A defini-ção destas características é essencial para o sucesso da monitoração do ambiente e a seleção delas deve ser bem cuidadosa.

Neste trabalho foram escolhidas 81 métricas para avaliação do comportamento de um usuário. Este valor está vinculado à quantidade de comandos que estão sendo auditados e foram considerados mais importantes no ambiente que está sendo monitorado. Este dado é variável e empiricamente escolhido dadas as necessidades de moni-toração do ambiente.

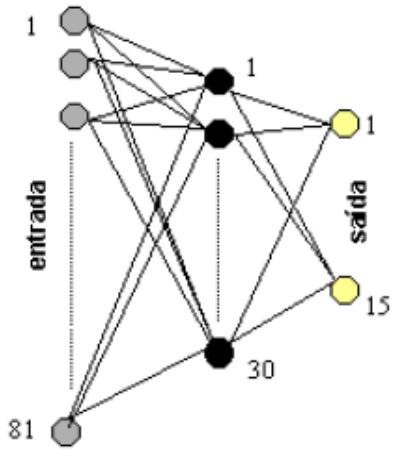

Figura 2: *Topologia da rede neural utilizada*

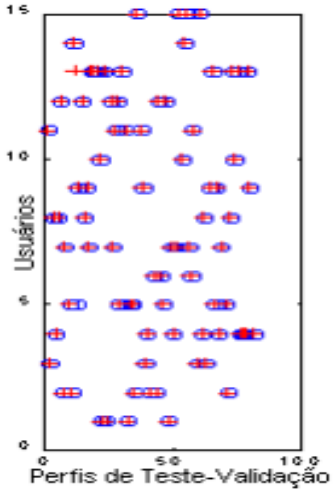

Figura 3: *Resultado observado na validação*

6.3 Algoritmo de treinamento da rede

O algoritmo usado para o treinamento da rede foi uma variação otimizada do algoritmo de *backpropagation* convencional (adicionando *momentum*). A escolha do algoritmo backpropagation com variação está vinculada a quantidade de dados que o problema possui, ao esforço computacional exigido, e ao tempo esperado de convergên-cia da rede.

O *momentum* é usado no algoritmo de *backpropagation* para prevenir a rede de convergir em um mínimo local caso as alterações nos pesos dependam da média do gradiente do erro quadrático em uma pequena região, ao invés do gradiente em um ponto [9].

7 Resultados observados

Os resultados observados na rede durante a validação e a aplicação de comportamento intrusivo podem ser vistos na análise abaixo e nas figuras 3 e 4.

- Taxa de erro observada nos dados de treinamento: 0%;
- Taxa de erro observada nos dados de teste: 3%;
- Taxa de erro observada na detecção de comportamento intrusivo: 4.5%;

Estas figuras mostram, respectivamente, o índice de acertos na validação e testes e na aplicação de perfis de intrusão. Os pontos circulares representam o resultado desejado e os ponto em cruz representam os resultados obtidos. Na figura 4 é perceptível que raras duplas de pontos (círculo - cruz) concordaram, reportando desta forma que o perfil aplicado era intrusivo.

Diante dos testes executados, o ponto mais importante a ser analisado é a ocorrência de falsos positivos e a ocorrência de falsos negativos. A taxa de erro no conjunto de teste significa a taxa de falsos positivos no modelo de detecção de intrusão, isto é, vetores de dados que deveriam apontar perfis corretos do

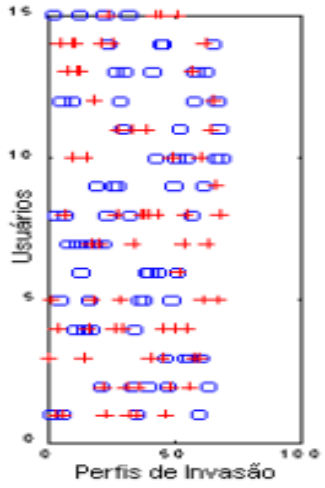

Figura 4: *Resultado observado na intrusão*

sistema e foram apontados como anormais. Desta forma, foram considerados intrusivos, erroneamente. Já a taxa de erro na detecção de comportamento intruso significa a taxa de falsos negativos e compreende as atividades intrusivas que foram consideradas como normais pelo sistema e não foram detectadas.

8 Conclusões e direcionamentos da área de detecção de intrusos

A área de detecção de intrusos ainda é um campo novo de pesquisa. Contudo, com o crescimento dos ambientes computacionais modernos, ela está tornando-se de extrema importância para o funcionamento de um sistema de computador. A combinação de fatos como o crescimento da Internet e

o grande número de transações comerciais que estão acontendo através dela, incentivam ainda mais os trabalhos na área.

Neste estudo de caso, como proposto inicialmente, foi discutido como as redes neurais podem ser aplicadas à modelos de detecção de intrusos. O treinamento é um ponto muito importante para a obtenção de bons resultados. Foi observado que, quando a rede era treinada por um período muito grande de tempo, os resultados nos dados de testes eram prejudicados e algumas vezes a rede perdia sua capacidade de generalização. O conjunto de dados usados no treinamento também deve refletir bem o perfil dos usuários do sistema, contudo nem sempre é possível obter tais dados e a distinção dos perfis pode acabar prejudicada. Finalmente é possível perceber que o aprendizado é uma excelente ferramenta para identificação e classificação de usuários e que o modelo de rede neural constitui uma prática efetiva para detecção de intrusos por anomalia. Contudo muitos cuidados devem ser tomados na coleta dos dados e no treinamento da rede.

A tendência das atuais pesquisas é convergir para um modelo que seja híbrido (detecção por anomalia e abuso) e utilize algum modelo de inteligência computacional aplicado. No entanto vários outros mecanismos estão sendo estudados para suprir deficiências observadas em modelos isolados.

Referências

[1] J.P Anderson. Conputer security threat monitoring and surveillance. Technical report, James P. Anderson Co., Fort Washington, Pensylvania, April 1980.

[2] A. Cansian. *Modelo Adaptativo para Detecção de Comportamento Suspeito em Redes de Computadores.* PhD thesis, IBILCE-UNESP, May 1997.

[3] Becker-M. Siboni D. Debar, H. A neural network component for an intrusion detection system. *In: Proceedings of the 1992 IEEE Computer Society Symposium on Research in Computer Security and Privacy*, pages 240–250, 1992.

[4] Dorothy E. Denning. An intrusion-detection model. *In: IEEE Transactions on Software Engineering*, 13(2):222–232, February 1987.

[5] G. E. Liepins H. S. Vaccaro. Detection of anomalous computer session activity. *In: IEEE Computer Society Symposium on Security and Privacy*, pages 280–289, May 1989.

[6] S. Haykin. *Neural Networks - A Comprehensive Fondation.* Prentice-Hall, 1994.

[7] R. Miikkulainen J. Ryan, M-J. Lin. Intrusion detection with neural networks. *In: Advances in Neural Information Processing Systems 10, Cambridge, MA*, 1998.

[8] T. F. Lunt. Detecting intruders in computer systems. *In: Conference on Auditing and Computer Technology*, 1993.

[9] W.S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, (5):115–133, 1943.

[10] D. B. Parker. P. Neumann. A summary of computer misure techniques. *In: Proceedings 12th National Computer Security Conference*, pages 280–289, May 1989.

[11] Katherine Price. *COAST - Intrusion Detection Pages*. http://www.cs.purdue.edu/coast/intrusion-detection/, setembro, 1999.

[12] M. J. Ranum. A internet firewall. *In: Proceedings of World Conference on Systems Management and Security*, 1992.

[13] W. Stallings. *Data Computer and Comunications*. Prentice Hall, 5 ed.,1997.

[14] F. Gilham. T. F. Lunt, A. Tamaru. A real time intrusion detection expert system(ides). Technical report, Project 6784,SRI International,Washington, February 1992.

[15] K. Tan. The aplication of neural networks to unix computer security. Technical report, Computer Science Department of University of Melbourne, Australia, 1995.

[16] A.S. Tanenbaum. *Computer Networks*. Prentice Hall, 3 ed.,1996.