Deteção de Intrusões em Redes de Computadores com Sistemas Multiagente

Abstract

A segurança de redes de computadores é um dos desafios principais da cibersegurança, dada a crescente sofisticação das ameaças. Este trabalho propõe o uso de um sistema multiagente (SMA) para deteção de intrusões (IDS), que envolve monitorização de tráfego de rede e deteção de atividades suspeitas, como ataques de negação de serviço distribuído (DDoS), *phishing* e *malware*. Os agentes colaboram para identificar padrões anómalos de comportamento e podem ainda integrar técnicas de *machine learning* para aumentar a precisão da deteção de ameaças. A aplicabilidade destes mecanismos é abrangente, tendo potencial para capacitar diferentes dóminios como serviços *Cloud*, *Internet of Things* (IoT), entre outros.

Keywords: Segurança de Redes, Sistemas Multiagente, Deteção de Intrusões, Machine Learning, Cibersegurança

1 Introdução

Com o aumento do tráfego de dados e da interconetividade, também crescem os riscos de ataques cibernéticos. O *Cybersecurity Almanac 2024* [1] aponta que o custo global dos crimes cibernéticos poderá ultrapassar os 10,5 triliões de dólares (US$) até 2025, o que evidencia não apenas a frequência, mas também a magnitude das ameaças atuais. Este crescente aumento de ameaças cibernéticas é representado na Figura 1, que apresenta a estimativa dos custos globais de crimes cibernéticos de 2018 a 2028, destacando a tendência de crescimento exponencial desses prejuízos.

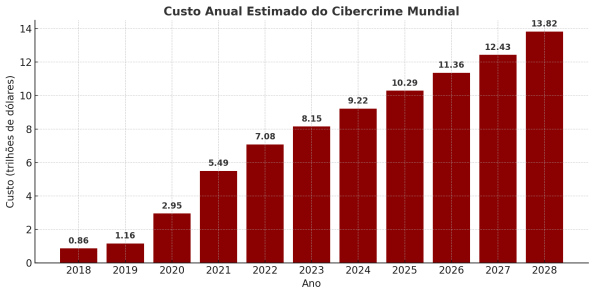

Fig. 1 Estimativa dos custos globais de crimes cibernéticos de 2018 a 2028. Fonte: *Statista* [2].

Além disso, o relatório *Verizon 2024 Data Breach Investigations Report* [3] revela um aumento significativo nos incidentes de segurança, com um novo recorde de 10.626 violações de dados, refletindo a complexidade e a sofisticação dos ataques que desafiam os métodos tradicionais de deteção. Diante esse contexto, torna-se evidente a necessidade de abordagens inovadoras, como a utilização de sistemas multiagente (SMA) para *Intrusion Detection Systems* (IDS).

Ao combinar uma abordagem distribuída com a colaboração entre agentes inteligentes, o sistema não só melhora a escalabilidade, mas também aumenta a precisão na identificação de padrões anómalos em ambientes com tráfego massivo. Desta forma, os desafios de deteção em tempo real podem ser melhor endereçados, proporcionando respostas mais rápidas e eficientes aos ataques cibernéticos. Este trabalho propõe a utilização de SMA para monitorizar e analisar padrões de tráfego, combinando uma abordagem distribuída com possível colaboração entre agentes inteligentes para identificar ameaças de forma eficiente.

O restante deste documento está organizado da seguinte forma: na Secção 2, apresentamos os problemas principais e objetivos desta investigação, destacando os desafios enfrentados na deteção de intrusões em redes. Na Secção 3, discutimos os domínios de investigação relevantes para este estudo, abordando conceitos fundamentais e a metodologia de pesquisa utilizada. A Secção 4 explora o estado da arte, analisando abordagens existentes e as mais recentes contribuições no uso de SMA em IDS. Por fim, na Secção 5, apresentamos as conclusões deste trabalho e delineamos possíveis direções para pesquisas futuras.

2 Problemas e Objetivos

A deteção de intrusões em redes de computadores continua a ser um desafio significativo devido à crescente complexidade dos ataques cibernéticos e ao volume de tráfego gerado em ambientes distribuídos. Os sistemas tradicionais de deteção de intrusões enfrentam diversas limitações, como a dificuldade em processar grandes quantidades de dados em tempo real e a elevada taxa de falsos positivos, que pode comprometer a eficácia da resposta a ameaças. Neste contexto, a abordagem baseada em SMA surge como uma alternativa promissora, permitindo uma monitorização distribuída e colaborativa do tráfego de rede. No entanto, a implementação eficiente desta abordagem requer a superação de vários desafios críticos:

- **Identificação de ataques cibernéticos em tempo real**: A capacidade de identificar ameaças à medida que ocorrem é essencial para mitigar danos e responder rapidamente a incidentes de segurança. No entanto, a análise em tempo real exige um processamento eficiente e distribuído, capaz de lidar com grandes volumes de tráfego sem comprometer o desempenho da rede. [4]
- **Redução de falsos positivos na deteção de ameaças**: Um dos principais problemas enfrentados pelos sistemas de deteção é a elevada taxa de falsos positivos, que pode levar a alertas desnecessários e dificultar a identificação de ataques reais. Métodos mais avançados de correlação de eventos e aprendizagem contínua são necessários para minimizar este problema e melhorar a precisão do sistema.[5]

- **Integração de SMA com *machine learning* para melhoria contínua**: O uso de *machine learning* pode permitir que os agentes se adaptem a novos padrões de ataque e evoluam com o tempo. No entanto, esta integração apresenta desafios relacionados à seleção de modelos apropriados, ao treino eficiente em tempo real e à explicabilidade das decisões tomadas pelos agentes.[6]
- **Escalabilidade na monitorização de redes de grande porte**: A aplicação de SMA para deteção de intrusões em redes de larga escala requer um sistema que possa ser escalado de forma eficiente. Isso inclui a capacidade de coordenar múltiplos agentes sem sobrecarregar a infraestrutura de rede, além de garantir a interoperabilidade com diferentes arquiteturas e protocolos de comunicação.[7]
- **Desafios adicionais de implementação**: A gestão e coordenação de um elevado número de agentes distribuídos impõe exigências quanto à sincronização das ações, quanto a conflitos entre decisões tomadas localmente por diferentes agentes. A comunicação constante entre agentes, necessária para garantir a partilha de informações e a deteção colaborativa, pode introduzir sobrecarga na rede e aumentar a latência, especialmente em ambientes com elevada heterogeneidade de dispositivos. Além disso, assegurar a segurança das comunicações interagente é crítico, uma vez que canais vulneráveis podem ser alvo de ataques que comprometem o próprio sistema de defesa. É necessário, portanto, equilibrar a complexidade dos algoritmos de análise com o consumo reduzido de memória e processamento, garantindo que os agentes mantenham um desempenho aceitável sem degradar os serviços essenciais da rede.[8]

Diante desses desafios, este trabalho tem como objetivo investigar o desenvolvimento de modelos de SMA que permitam a deteção colaborativa e distribuída de ameaças. Para alcançar esse objetivo, serão analisadas técnicas de análise de dados e *machine learning*, para melhorar a precisão e a eficiência da deteção. Além disso, as arquiteturas propostas buscam garantir a escalabilidade e a adaptabilidade do sistema, permitindo a sua aplicação em diferentes cenários de segurança cibernética. A abordagem multiagente será projetada para promover a colaboração entre agentes inteligentes, otimizando a resposta a incidentes e reduzindo o impacto de possíveis ataques nas redes monitoradas.

3 Domínios de Investigação

Compreender o funcionamento dos Sistemas de IDS e dos SMA é essencial para o desenvolvimento de soluções eficazes na área da cibersegurança, uma vez que ambos desempenham papéis fundamentais na identificação e mitigação de ameaças em redes. Desta forma, nesta secção, introduzimos as temáticas pilares na nossa investigação, assim como os métodos de pesquisa.

3.1 Metodologia de Pesquisa

Para sustentar este estudo, foi realizada uma revisão da literatura recorrendo ao Google Scholar como principal ferramenta de pesquisa. A seleção dos artigos teve em consideração a relevância e a atualidade dos trabalhos publicados sobre IDS, segurança de redes e a aplicação de SMA na identificação de ameaças. Privilegiaram-se

estudos indexados em bases de dados reconhecidas, como SpringerLink, IEEE Xplore e ScienceDirect, garantindo a fiabilidade das fontes consultadas.

Entre os materiais analisados, destacam-se os artigos disponíveis em Springer [9] e IEEE Access [10], que apresentam abordagens recentes na área e serviram como referências fundamentais para a definição da estratégia adotada neste trabalho. A análise dos artigos incidiu sobre os desafios identificados e as soluções propostas, permitindo consolidar uma base sólida para uma proposta de um futuro SMA na deteção de intrusões.

3.1.1 Restrições Temporais e Critérios de Seleção

Para garantir a atualidade e relevância das informações, foram estabelecidos critérios na seleção dos artigos, incluindo:

- **Período de Publicação**: Foram considerados apenas artigos publicados a partir de 2020, de forma a assegurar que os estudos analisados reflitam as tendências mais recentes na área de cibersegurança e deteção de intrusões (em casos específicos, foram incluídos artigos de anos anteriores quando considerados altamente relevantes)
- **Qualidade das Fontes**: Apenas artigos de revistas e conferências científicas indexadas em bases de dados reconhecidas (Springer, IEEE Xplore, ScienceDirect) foram incluídos, excluindo trabalhos sem revisão por pares.
- **Relevância Temática**: Foram selecionados artigos que abordassem diretamente IDS, SMA, machine learning para segurança de redes e métodos de deteção de ameaças, excluindo-se trabalhos com foco genérico em inteligência artificial ou segurança sem aplicação prática à deteção de intrusões.
- **Citações e Impacto**: Tentou-se sempre procurar artigos com o maior número de acessos e/ou citações. Garantindo os estudos tenham um impacto significativo na comunidade científica.

3.1.2 Palavras-chave Utilizadas na Pesquisa

A pesquisa foi conduzida utilizando combinações de palavras-chave para abranger diferentes abordagens na área de deteção de intrusões e SMA em segurança de redes. A tabela abaixo apresenta as palavras-chave mais utilizadas e suas combinações:

Categoria	**Palavras-chave**
Network and Cybersecurity	"Network Security", "Secure Network Communication", "Threat Detection", "Cybersecurity Strategies", "AI for Cyber Defense"
Multi-Agent Systems	"Multi-Agent Systems", "MAS for Cybersecurity", "Agent-Based Intrusion Detection"
Intrusion Detection	"Intrusion Detection System (IDS)", "Anomaly-Based IDS", "Signature-Based IDS"

3.2 Sistemas de Deteção de Intrusões

Os IDS são ferramentas essenciais para a segurança de redes, permitindo a identificação de atividades maliciosas que possam comprometer a integridade, confidencialidade ou disponibilidade dos sistemas. Estes sistemas são projetados para monitorizar o tráfego de rede, analisar padrões de comportamento e detetar atividades anómalas que possam indicar a presença de intrusões ou ataques cibernéticos.

Os IDS são amplamente utilizados em diversos contextos, como em *Internet of Things* (IoT) e em *Cloud*. Em IoT, onde a quantidade de dispositivos conectados é muito elevada e muitos têm capacidades limitadas de processamento, os IDS monitorizam o tráfego entre os dispositivos, detetando atividades suspeitas, como tentativas de acesso não autorizado e comportamentos anómalos. Em ambientes de *cloud computing*, os IDS são essenciais para monitorizar o tráfego entre os servidores virtuais, detetando ataques como *exploits* de vulnerabilidades, garantindo a segurança dos dados e a disponibilidade dos serviços.[11]

Atualmente, os IDS dividem-se em dois tipos principais:

- **Baseados em assinaturas (SIDS)[12]:** Estes sistemas dependem de bases de dados de ataques conhecidos, utilizando técnicas de correspondência de padrões (*pattern matching*) para identificar ameaças previamente catalogadas. O funcionamento dos SIDS baseia-se na comparação do tráfego de rede ou das atividades do sistema com uma base de dados de assinaturas de ataques. Quando é detetada uma correspondência, um alarme é acionado. A principal vantagem dos SIDS é a sua eficácia na deteção de ataques cujas assinaturas já são conhecidas e estão presentes na base de dados, resultando numa baixa taxa de falsos positivos para ameaças conhecidas. No entanto, a grande desvantagem dos SIDS é a sua incapacidade de detetar ataques novos ou desconhecidos, uma vez que as assinaturas desses ataques ainda não estão na base de dados. Além disso, estes sistemas exigem uma atualização constante da base de dados para manter a sua eficácia. Alguns dos exemplos destes sistemas são:
  - **Snort:** O Snort [13] é um sistema de deteção de intrusões open source que desempenha um papel crucial na cibersegurança. Este IDS monitoriza o tráfego de rede em tempo real, analisando-o com base em regras predefinidas que descrevem padrões de ataques conhecidos. Quando o tráfego corresponde a uma regra, o Snort gera um alerta, regista o evento e, dependendo da configuração, pode bloquear o tráfego malicioso. O Snort é amplamente utilizado para detetar atividades como *port scans*, *buffer overflows* e atividade de *worms*. Além disso, pode ser usado para registo de pacotes (*packet logging*), depuração de tráfego de rede e investigação em segurança.
  - **Suricata:** O Suricata [14] é outro sistema de deteção de intrusões *open source*, conhecido pela sua alta performance e flexibilidade. Além de detetar ameaças com base em regras predefinidas, o Suricata oferece funcionalidades avançadas, como o registo de pedidos HTTP, a extração de ficheiros de fluxos de rede e a análise de tráfego TLS/SSL.
- **Baseados em anomalias (AIDS)[12]:** Estes sistemas utilizam métodos estatísticos e técnicas de machine learning para identificar comportamentos fora do

normal. Ao contrário dos SIDS, que dependem de assinaturas conhecidas, os AIDS focam-se na deteção de atividades anómalas que possam indicar a presença de uma intrusão. A principal vantagem dos AIDS é a sua capacidade de detetar ameaças novas ou desconhecidas, uma vez que não dependem de assinaturas predefinidas. No entanto, os AIDS têm a desvantagem de poderem gerar um número elevado de falsos positivos, pois as anomalias detetadas podem ser simplesmente novas atividades normais e não intrusões reais.

Os SMA surgem como uma abordagem promissora para a deteção de intrusões, ao permitir uma monitorização descentralizada e uma resposta proativa a ameaças. Estudos recentes [11] demonstram que a integração de agentes inteligentes com técnicas avançadas, como redes neuronais e algoritmos de *clustering*, apresentam melhorias significativas na precisão e eficiência da deteção de intrusões.

3.3 Sistemas Multiagente e Arquitetura Fundamental

Os SMA são compostos por múltiplos agentes autónomos que interagem entre si para atingir objetivos comuns. Cada agente pode possuir diferentes níveis de inteligência, autonomia e capacidade de comunicação, permitindo a descentralização da tomada de decisão. No contexto de deteção de intrusões, um SMA é particularmente vantajoso, pois permite distribuir a análise de tráfego de rede entre diferentes agentes especializados, reduzindo a sobrecarga computacional e aumentando a resiliência do sistema.

3.4 Sistemas Multiagente e a Deteção de Intrusões em Redes

Numa rede, cada agente é uma entidade computacional capaz de percecionar o seu ambiente, tomar decisões e agir de forma autónoma para alcançar objetivos específicos. A colaboração entre agentes permite que o sistema na totalidade seja mais robusto, escalável e adaptável a mudanças no ambiente.

No contexto da deteção de intrusões em redes de computadores, os SMA oferecem uma abordagem promissora para monitorizar e analisar o tráfego de rede de forma distribuída. Ao contrário dos sistemas centralizados tradicionais, que podem tornar-se um ponto único de falha, os SMA distribuem a carga de processamento e tomada de decisão entre múltiplos agentes, aumentando a resiliência e a eficiência do sistema.

3.5 Arquitetura de um SMA para Deteção de Intrusões

A arquitetura de um SMA aplicado à deteção de intrusões pode ser dividida em três componentes principais:

- Agentes de Monitorização:
  1. Responsáveis por recolher dados do tráfego de rede em tempo real.
  2. Cada agente monitoriza um segmento específico da rede, como um *router*, *switch* ou servidor.
  3. Podem ser instalados em *hosts*

4. Utilizam técnicas de análise de pacotes para identificar padrões suspeitos, como portas abertas, tráfego incomum ou tentativas de exploração de vulnerabilidades.

- Agentes de Análise:

1. Processam os dados recolhidos pelos agentes de monitorização.
2. Aplicam algoritmos de *machine learning* e análise estatística para detetar anomalias e comportamentos maliciosos.
3. Podem ser especializados em diferentes tipos de ameaças, como ataques DDoS, *phishing* ou *malware*.

- Agentes de Coordenação:

1. Gerem a comunicação e colaboração entre os agentes de monitorização e análise.
2. Consolidam os resultados das análises e tomam decisões sobre como responder a ameaças identificadas.
3. Podem ativar mecanismos de defesa, como bloquear endereços IP maliciosos ou notificar administradores de rede.

Esta divisão fundamental da arquitetura multiagente é consistente com propostas recentes na área, como a apresentada pelo artigo [6]. A interação entre estes agentes é facilitada por um protocolo de comunicação que permite a partilha de informações e a coordenação de ações. Por exemplo, se um agente de monitorização detetar um padrão suspeito, pode enviar uma notificação aos agentes de análise para uma investigação mais detalhada. Se uma ameaça for confirmada, os agentes de coordenação podem tomar medidas para mitigar o ataque.

3.6 Ética e Privacidade de Dados

A utilização de SMA's para deteção de intrusões levanta preocupações éticas e de privacidade. Estes sistemas processam frequentemente grandes volumes de tráfego de rede, podendo envolver dados sensíveis. Assim, é essencial garantir a conformidade com regulamentos como o GDPR, assegurando a minimização e anonimização dos dados recolhidos. Além disso, é necessário garantir que os agentes atuem de forma transparente e não enviesada, evitando práticas discriminatórias e assegurando a auditabilidade das decisões.[15]

4 Estado da Arte

A revisão do estado da arte aborda a aplicação de Sistemas Multiagente (SMA) na deteção de intrusões, explorando tanto soluções comerciais quanto investigações académicas recentes. São analisadas as principais abordagens, destacando as suas vantagens, desafios e oportunidades de melhoria.

4.1 Aplicações Comerciais

Os SMAs para a deteção de intrusões têm sido amplamente adotados em diversas aplicações comerciais, especialmente em setores que exigem alta segurança e monitorização contínua. A sua capacidade de operar de forma autónoma e cooperativa permite

a identificação e mitigação de ameaças cibernéticas em tempo real, garantindo a integridade e a confiabilidade dos sistemas corporativos. A seguir, destacam-se algumas das principais aplicações comerciais:

- **Bancos e Instituições Financeiras**: A deteção de fraudes e ataques cibernéticos é uma prioridade crítica no setor financeiro, uma vez que transações eletrónicas estão sujeitas a ameaças como *phishing*, *malware*, ataques de engenharia social e acessos não autorizados a contas. Para mitigar esses riscos, SMA têm sido adotados para monitorizar transações em tempo real, identificando padrões suspeitos e prevenindo fraudes antes que causem danos financeiros. Por exemplo, a Akira AI [16] desenvolveu um SMA baseado em inteligência artificial para conformidade financeira, combinando *machine learning* supervisionado e análise comportamental para detetar anomalias em grandes volumes de transações. Esse sistema permite a automação de processos regulatórios, reduzindo em 50% o tempo necessário para auditorias manuais e melhorando significativamente a capacidade de identificar transações potencialmente fraudulentas em tempo real. Além disso, a abordagem multiagente utilizada pela Akira AI possibilita que diferentes entidades do setor financeiro compartilhem informações de forma segura, contribuindo para um modelo colaborativo de segurança que reforça a proteção contra crimes financeiros.
- **Provedores de Serviços de Internet (ISPs)**: Para proteger infraestruturas de rede contra ataques DDoS e outras ameaças, os SMA são integrados para analisar o tráfego e bloquear atividades maliciosas. Diversas abordagens para a deteção de intrusões são utilizadas comercialmente, incluindo técnicas baseadas em *machine learning* e análise comportamental. O artigo [17] destaca a importância de utilizar múltiplas técnicas de deteção, combinando métodos tradicionais com inteligência artificial para aumentar e precisão e reduzir falsos positivos.
- **Empresas de Tecnologia**: Grandes empresas como Google e Microsoft utilizam SMA para monitorizar redes internas e proteger dados sensíveis contra acessos não autorizados. Com o aumento da adoção de infraestruturas *cloud*, a cibersegurança tornou-se um desafio ainda maior, exigindo abordagens mais avançadas para identificar e mitigar ameaças cibernéticas. Os SMA desempenham um papel crucial nessa proteção, permitindo uma deteção proativa de anomalias, resposta autónoma a incidentes e adaptação dinâmica a novas vulnerabilidades. Recentemente, a arquitetura Intelligent Security Service Framework (ISSF) foi proposta como uma solução para operações de segurança nativas em nuvem, combinando aprendizagem profunda e SMA para monitorizar e proteger infraestruturas *cloud* de maneira distribuída e inteligente [18]. O ISSF emprega agentes especializados para analisar continuamente fluxos de tráfego de rede, identificar padrões maliciosos e responder automaticamente a potenciais ataques, minimizando o tempo de exposição das vulnerabilidades. A integração de SMA em infraestruturas *cloud* melhora significativamente a cibersegurança ao reduzir o tempo de deteção de ataques e falsos positivos, além de oferecer uma visão holística e adaptativa da postura de segurança em ambientes altamente dinâmicos. Essa abordagem permite que empresas de tecnologia otimizem as suas políticas de segurança, protejam dados sensíveis e garantam a conformidade com regulamentações internacionais, como o GDPR e a ISO/IEC 27001.

- **Saúde**: Hospitais e clínicas utilizam SMA para proteger sistemas de registos médicos eletrónicos (EHR) e garantir a privacidade dos pacientes. O artigo [19] apresenta uma abordagem baseada em métricas de privacidade multiagente para sistemas de aprendizagem profunda aplicados à saúde. A abordagem combina *deep learning* e agentes inteligentes para monitorizar e controlar o acesso aos dados dos pacientes, garantindo que apenas agentes autorizados possam consultá-los. Além disso, o sistema utiliza métricas de privacidade para avaliar em tempo real a confidencialidade, integridade e disponibilidade das informações, identificando potenciais violações de privacidade. Essa integração permite não apenas um reforço na segurança dos dados médicos, mas também melhora a precisão no diagnóstico e na recomendação de tratamentos, otimizando a assistência aos pacientes de forma segura e eficiente.

Estas aplicações demonstram a versatilidade dos SMA na deteção de intrusões, adaptando-se a diferentes necessidades e ambientes. A capacidade desses sistemas de operar de forma distribuída e colaborativa torna-os particularmente eficazes na identificação e mitigação de ameaças cibernéticas em tempo real, contribuindo para a segurança e a resiliência das infraestruturas críticas em diversos setores.

4.1.1 Dificuldades de Implementação

Apesar da crescente adoção dos SMA em ambientes comerciais, a sua implementação prática apresenta desafios significativos, independentemente do domínio de aplicação. Destacam-se dificuldades relacionadas com privacidade, integração, escalabilidade e segurança.

Em setores como o financeiro e da saúde, garantir a privacidade e integridade das comunicações entre agentes é crucial, exigindo protocolos robustos de encriptação e conformidade com regulamentações rigorosas como o GDPR. Já em ambientes como ISPs e infraestruturas *cloud*, a integração dos SMA com sistemas legados revela-se complexa, dada a necessidade de interoperabilidade entre diferentes arquiteturas e protocolos.

A escalabilidade é outro obstáculo relevante, uma vez que o crescimento do número de agentes aumenta a sobrecarga computacional e o tráfego de comunicação, podendo afetar o desempenho da rede. Soluções como arquiteturas hierárquicas ou *clustering* de agentes são frequentemente propostas, mas trazem maior complexidade de gestão.

Adicionalmente, em contextos como IoT, os recursos limitados dos dispositivos exigem agentes leves, capazes de balancear eficiência de processamento com capacidades analíticas adequadas. Por fim, a comunicação contínua entre agentes pode ser explorada como vetor de ataque, sendo essencial assegurar a autenticidade e integridade das mensagens através de mecanismos de segurança eficazes.

4.2 Investigação Académica

O uso de SMA em Segurança e deteção de ameaças é amplamente investigado devido ao crescimento tecnológico em áreas como os *cloud services* que enfrentam complexas e diversas ameaças diariamente, sendo necessário amplificar as medidas de segurança.

4.2.1 Aplicação em *Clouds*

O estudo [9] propõe um SMA distribuído, capaz de detetar, identificar e prevenir ameaças para infraestruturas de *cloud computing*. Aqui apresentam alguns dos conceitos mais relevantes e atuais para sistemas de deteção de ameaças, começando pelas classes principais, que podem ser baseados em *Host*, *Network* ou Distribuídos. Essencialmente, o que varia entre os dois primeiros tipos são os alvos de instalação, sendo que os *Host Based* monitorizam tráfego de *hosts*, daí precisarem de estar configurados num *host*, enquanto os *Network Based*, tipicamente monitorizam tráfego de rede, ficando instalados em *routers* e *network switches*. Os sistemas distribuídos combinam diversas instancias baseadas nas duas classes apresentadas. Ainda, ficamos a conhecer que os sistemas podem ser ainda categorizados baseado no método adotado para deteção. Desta forma, existem sistemas *signature-based* e *anomaly-based*, como já foi descrito nesta mesma secção.

O estudo [9] ainda propõe uma arquitetura para o sistema. Começando pelos papéis dos agentes, a arquitetura proposta destaca três tipos principais: *Host agent*, que se focam na monitorização local dos *hosts*, recolhendo dados específicos relacionados com as atividades do mesmo e participando na tomada de decisões relativas às anomalias detetadas a nível do *host*; *Network agent*, supervisionam aspetos mais amplos da rede, onde recolhem dados de tráfego e identificam padrões irregulares que possam sugerir intrusões a nível de rede; *Prevention Server*, que ao receber alertas sobre intrusões detetadas pelos agentes, implementa ações para mitigar as ameaças, atualizando as regras da *firewall* e podendo ainda bloquear o tráfego malicioso.

Nesta mesma referência [9], detalha-se ainda um processo de seis etapas para a deteção de intrusões. Na primeira etapa, a **recolha de dados**, tanto os *host agents* como os *network agents* recolhem informação dos seus respetivos ambientes. Na segunda etapa, o **pré-processamento**, os dados recolhidos são limpos e preparados para análise, sendo filtrada a informação irrelevante e o ruído; esta tarefa é realizada pelos mesmos agentes, que organizam a informação para uma análise posterior. Na terceira etapa, o ***clustering***, os dados pré-processados são classificados em *clusters* com base em semelhanças, permitindo a identificação de padrões ou anomalias, mediante a utilização de algoritmos como o *K-means* para reconhecer tendências comuns ou desvios do comportamento normal. Durante o processo de *clustering*, o objetivo é determinar o nível de prioridade dos dados. Na quarta etapa, a **identificação de dados semelhantes**, os agentes examinam os dados agrupados à procura de pontos assemelhados a padrões já conhecidos, quer sejam de comportamento normal ou malicioso, facilitando assim a deteção de potenciais intrusões. Na quinta etapa, a **geração de regras de associação**, os agentes criam normas que definem as relações entre os pontos de dados, detalhando o que constitui atividade normal e o que é considerado anómalo, contando ambos os tipos de agentes com o seu contributo na formação destas regras. Finalmente, na sexta etapa, a **tomada de decisão**, os agentes avaliam a informação e decidem se uma atividade é maliciosa com base nas regras geradas; se for detetada atividade suspeita, os agentes alertam o *prevention server*, o qual implementa medidas para mitigar as ameaças.

A arquitetura do SMA subjacente do estudo [9] não é diretamente mencionada pelos autores. Da nossa análise, acreditamos que o sistema pode ser visto como uma

arquitetura híbrida. Primeiramente, os agentes do tipo *Host* ou *Network* são exemplos da presença de uma arquitetura reativa, onde, de forma simplificada, reagem a perceções que obtém sobre a observação do seu alvo, seja ficheiros no caso dos *hosts*, ou pacotes no caso de uma rede. Estes agentes ainda podem ser equipados com *machine learning*, aproximando a uma *Subsumption architecture*. No caso do *prevention server*, se considerarmos como um agente, encontramos uma arquitetura deliberativa, onde o agente atua apenas sobre uma *firewall*, que pode ser representada simbolicamente.

4.2.2 Aplicação em Internet of Things (IoT)

O estudo [20] apresenta uma solução multiagente desenhada para a segurança de dispositivos IoT. Para este efeito, há considerações importantes sobre a limitação de recursos dos dispositivos disponíveis, como também a colaboração entre diferentes agentes, dispostos em diferentes dispositivos. Na arquitetura, há distinção entre dois tipos de camadas. Primeiramente, a *edge layer*, contém diversos agentes que se encontram próximos dos dispositivos IoT. A sua função principal é monitorizar, transformar e analisar dados do seu ambiente, contando ainda com a possibilidade de colaboração com outros *edge agents* para tarefas que não conseguem concretizar sozinhos. Outra característica destes agentes são os modelos *lightweight* que suportam para a tarefa de deteção, usando o mínimo de recursos possíveis. A camada seguinte é a *cloud layer* que serve como complemento da camada anterior ao fornecer poder computacional e capacidade de análise superior. É esperado que um *cloud agent* consiga ter um conhecimento da rede mais amplo que o outro tipo de agentes, e também é esperado que consiga gerir, através de atualizações, os demais modelos utilizados pelos *edge agents*. A comunicação entre agentes da *cloud layer* serve para aumentar o poder de análise desta camada, onde podem experimentar com *datasets* diversificados.

A arquitetura da referência [20] tem como pilar a comunicação entre camadas. A colaboração entre agentes de camadas distintas permite aos *cloud agents* fornecerem conhecimento valioso e facilita o aprimoramento de modelos com base em operações observadas em toda a rede, enquanto os *edge agents* concentram se em dados imediatos e localizados. Desta forma, alcançasse uma inteligência distribuída, onde o sistema pode melhorar a sua escalabilidade, adaptabilidade a ameaças emergentes e tempos de resposta durante ataques. Podemos afirmar que o sistema adota o conceito de aprendizagem federada, onde os *edge agents* atualizam os seus modelos com base na análise local e os *cloud agents* agregam estes dados para aprimorar e distribuir as atualizações necessárias, mantendo um controlo rigoroso das versões e promovendo a aprendizagem adaptativa. Para proteger a privacidade e a segurança, os dados sensíveis são tratados localmente e a troca de informação com a cloud é cuidadosamente controlada e, quando necessário, anonimizada.

Relativamente ao desenho da arquitetura do sistema do estudo [20], com foco no SMA, encontramos uma situação bastante semelhante ao modelo anterior. Os *edge agents* correspondem a um desenho reativo, visto que percecionam os ambientes e agem em função do conhecimento assimilado através dos seus sensores. Neste caso concreto, dependendo de cada dispositivo, a informação levantada pode ser bastante dispare. Os *cloud agents*, assemelham-se a uma arquitetura *Beliefs, Desires and Intention* (BDI), onde, colaborativamente, integram e analisam os dados provenientes dos *edge*

agents para construir uma visão global do ambiente. Dessa forma, formulam hipóteses e estratégias de resposta que são, então, distribuídas de maneira a otimizar a defesa contra possíveis ameaças, garantindo uma atuação adaptativa em tempo real.

4.2.3 Aplicação em Sistemas Físicos

A referência [10] mostra como uma arquitetura multiagente pode ser aplicada à cibersegurança em relés de distância, combinando técnicas de *deep learning* para detetar de forma eficaz ataques cibernéticos. Cada agente é responsável por monitorizar as medições locais de tensão e corrente elétrica, enquanto interage com os seus vizinhos através de uma estrutura onde os relés correspondem aos nós e as ligações de comunicação às arestas. Esta abordagem permite uma troca de dados localizada, reduzindo os custos de conexão e aumentando a eficiência do sistema. Além disso, a colaboração entre os agentes, que recolhem dados próprios e partilham informações com os vizinhos, possibilita uma deteção mais robusta de anomalias que possam indicar ataques cibernéticos. Para reforçar esta capacidade, o método utiliza um mecanismo de controlo de consenso, que permite aos agentes estimar os valores reais de tensão e corrente em toda a rede, assegurando um entendimento comum do estado do sistema, mesmo na presença de dados potencialmente comprometidos.

O estudo [10] consta ainda com a integração de *deep learning*, onde cada agente está equipado com uma rede neural (DNN) treinada para analisar os dados recolhidos para identificar padrões anómalos. O processo de treino das DNNs envolve a simulação de diversos cenários operacionais da rede elétrica, abrangendo condições normais e diferentes tipos de falhas, sejam estas simétricas ou assimétricas. Esta abordagem permite que os agentes aprendam a distinguir entre o comportamento típico do sistema e as atividades anormais associadas a ataques cibernéticos. Uma vez implementadas, as DNNs avaliam os dados em tempo real, classificando-os como normais, indicativos de falhas ou de possíveis ataques cibernéticos, através da comparação com os padrões previamente aprendidos. Desta forma, cada agente avalia o seu desempenho de forma independente, contribuindo para uma deteção descentralizada e reforçando a resiliência do sistema global.

A arquitetura do SMA em [10] alinha-se com uma abordagem híbrida, semelhante aos casos relatados anteriormente. Neste contexto, cada agente monitoriza localmente medições elétricas, um comportamento típico reativo, e comunica com vizinhos para uma análise mais aprofundada do estado da rede, um comportamento deliberativo. Além disso, a utilização de DNNs indica um nível de processamento avançado de padrões, típico de agentes com aprendizagem. A presença de um mecanismo de consenso reforça ainda mais a natureza colaborativa e distribuída do sistema

4.3 Análise Crítica

A abordagem proposta por [9] evidencia uma estratégia robusta para a deteção de ameaças em ambientes de *cloud computing*. A divisão do processo em etapas claramente definidas, desde a recolha e pré-processamento dos dados até à tomada de decisão, facilita a identificação de anomalias e a resposta rápida a incidentes. Contudo, embora a independência permita respostas rápidas e localizadas, a ausência de

uma comunicação estruturada pode levar a uma visão fragmentada do ambiente, dificultando a correlação de informações e a identificação de padrões mais amplos de ameaças. Além disso, a capacidade de atualização constante das regras de segurança é um processo bastante difícil e complexo, possivelmente exigente computacionalmente. Assim, outras investigações poderão explorar mecanismos de comunicação entre peças do sistema, assim como métodos adicionais de aprendizagem e adaptação em tempo real.

A solução apresentada por [20] evidencia uma tendência positiva para a segurança de dispositivos IoT, especialmente ao integrar uma *edge layer* com modelos *lightweight* e a colaboração com a *cloud layer* para análises mais complexas. Essa divisão permite uma resposta rápida a eventos locais, beneficiando, simultaneamente, do poder de processamento centralizado para tarefas de *deep learning*. Entretanto, a eficácia global do sistema depende criticamente da qualidade da comunicação entre as camadas, que pode introduzir desafios como latência e riscos de exposição dos dados.

De forma semelhante, a aplicação de *deep learning* na arquitetura descrita em [10] demonstra o potencial dos agentes para identificar anomalias de forma autónoma e em tempo real. Contudo, a adaptação contínua dos modelos às condições reais da rede e a necessidade de um treino abrangente para cobrir cenários imprevistos permanecem como possíveis obstáculos do sistema.

4.3.1 Vantagens e Desvantagens

De seguida, apresentam-se as principais vantagens e desvantagens associadas à utilização de Sistemas Multiagente (SMA) na deteção de intrusões, com base nas aplicações e estudos discutidos anteriormente.

Vantagens

- **Estratégia robusta para deteção de ameaças:** A abordagem proposta por [9] oferece uma metodologia clara e estruturada para a deteção de ameaças em ambientes de *cloud computing*.
- **Adaptabilidade e Escalabilidade:** Os SMA são adaptáveis a diversas situações e facilmente escaláveis, o que os torna adequados para diferentes cenários e necessidades.
- **Resposta em Tempo Real:** A capacidade de os agentes reagirem de forma autónoma a ameaças detetadas permite uma resposta imediata, reduzindo o tempo de exposição a vulnerabilidades.
- **Redução de Falsos Positivos:** Quando combinados com técnicas de *Machine Learning*, os SMA podem reduzir significativamente a ocorrência de falsos positivos, aumentando a precisão do sistema.

Desvantagens

- **Complexidade de Implementação:** A implementação de SMA pode ser complexa e exigir recursos computacionais significativos, o que pode representar um desafio para organizações com capacidades técnicas ou financeiras limitadas.
- **Dependência da Qualidade dos Dados para Monitorização:** A eficácia dos SMA depende da qualidade dos dados recolhidos e processados. Dados incompletos

ou ruidosos podem levar a falsos positivos ou negativos, comprometendo a fiabilidade do sistema.
- **Latência e Comunicação:** A comunicação entre agentes, especialmente em sistemas distribuídos, pode introduzir latência, o que pode afetar a eficiência da resposta a ameaças.
- **Riscos de Segurança na Comunicação:** A comunicação entre agentes pode ser um ponto vulnerável, expondo o sistema a riscos de interceptação ou manipulação de dados, o que requer medidas adicionais de segurança.

Resumidamente, os SMA oferecem uma abordagem promissora para a deteção de intrusões, com vantagens significativas em termos de eficiência de deteção e resposta em tempo real. No entanto, como em qualquer tecnologia, também apresentam as suas limitações. Questões como a latência nas comunicações e os desafios de segurança associados à troca de informações são desvantagens que precisam de ser cuidadosamente consideradas para maximizar o potencial destes sistemas.

5 Conclusão e Trabalho Futuro

A segurança de redes de computadores enfrenta desafios cada vez mais complexos devido à sofisticação e ao volume crescente de ameaças cibernéticas. A revisão do estado da arte neste trabalho, evidenciou que os sistemas tradicionais de deteção de intrusões podem ser baseados em assinaturas ou baseados em anomalias. Os primeiros dependem de bases de dados de ataques conhecidos, o que os torna eficazes contra ameaças inéditas, enquanto os segundos utilizam métodos estatísticos e de *machine learning*, mas ainda enfrentam desafios na redução de falsos positivos e adaptação a padrões dinâmicos.

Neste contexto, este trabalho propôs a utilização de SMA como uma solução inovadora para a monitorização de tráfego de rede e deteção de atividades suspeitas, como ataques de negação de serviço distribuído (DDoS), *phishing* e *malware*. A abordagem multiagente, combinada com técnicas de *machine learning* (seja redes neuronais ou algoritmos de *clustering*), demonstrou ser eficaz na identificação de padrões anómalos e na resposta rápida a intrusões, especialmente em redes de grande escala.

A arquitetura proposta, que integra agentes de monitorização, análise e coordenação, mostrou-se escalável e resiliente, permitindo uma deteção distribuída e colaborativa de ameaças. A capacidade dos agentes de aprenderem e adaptarem-se a novos padrões de ataque, aliada à redução de falsos positivos, reforça a eficácia do sistema. Além disso, a aplicação de SMA em diversos setores, como financeiro, saúde e tecnologia, evidenciou a versatilidade e o potencial desta abordagem para proteger infraestruturas críticas.

A investigação académica e as aplicações comerciais destacadas neste trabalho confirmam que os SMA são uma solução promissora para os desafios atuais da cibersegurança. No entanto, a implementação eficiente destes sistemas requer a superação de desafios como a integração de *machine learning* em tempo real, a escalabilidade em redes de grande porte e a garantia de uma comunicação eficaz entre agentes.

Futuros trabalhos poderão explorar três direções principais para o aprimoramento dos SMA:

- **Integração com IA explicável (XAI)**: Um dos desafios da aplicação de *machine learning* em cibersegurança é a interpretabilidade dos modelos. Atualmente, muitas abordagens baseadas em redes neurais atuam como "caixas pretas", tornando difícil entender como uma ameaça foi identificada. A introdução de XAI (Explainable Artificial Intelligence) permitirá que os agentes justifiquem as suas decisões, aumentando a confiabilidade e possibilitando auditorias mais eficazes sobre a tomada de decisão.[21]
- **Uso de aprendizagem federada**: SMA operam em redes distribuídas, tornando desafiador o treino centralizado de modelos de deteção de ameaças. A aprendizagem federada surge como uma solução promissora, permitindo que os agentes treinem modelos localmente e compartilhem apenas parâmetros relevantes, garantindo maior privacidade e eficiência computacional. Estudos futuros podem explorar como diferentes agentes podem colaborar de forma descentralizada para melhorar a precisão da deteção de intrusões sem comprometer a confidencialidade dos dados.
- **Incorporação de *blockchain***: Para fortalecer a integridade das decisões dos agentes e prevenir ataques contra os próprios SMA, o uso de *blockchain* pode ser investigado como uma forma de assegurar a imutabilidade dos eventos registados. Uma abordagem baseada em *blockchain* pode permitir que cada interação ou alerta gerado por um agente seja validado e armazenado de forma segura, evitando falsificações e garantindo transparência no funcionamento do sistema.[22]
- **Redução de latência na tomada de decisão dos agentes**: Embora SMA sejam altamente eficientes na análise descentralizada de ameaças, ainda enfrentam desafios relacionados à latência na comunicação entre agentes, especialmente em infraestruturas de grande porte. Futuros estudos podem investigar estratégias de comunicação mais otimizadas, como redes de comunicação hierárquicas entre agentes ou o uso de redes neurais neuromórficas, para acelerar a resposta a incidentes.[23]
- **Avaliação do impacto da combinação de técnicas híbridas**: Uma abordagem futura interessante seria a combinação de diferentes técnicas de *machine learning* e heurísticas baseadas em comportamento, criando SMAs híbridos que aproveitem o melhor de diferentes metodologias. A exploração de modelos neuro-simbólicos, que combinam aprendizagem estatística e regras explícitas, pode aumentar a precisão e confiabilidade das decisões dos agentes.[24]

Em suma, os SMAs representam um avanço significativo na proteção de redes de computadores, oferecendo uma abordagem distribuída, adaptável e eficiente para enfrentar as ameaças cibernéticas do mundo moderno. A sua adoção em larga escala poderá fortalecer a segurança de infraestruturas críticas, garantindo a integridade e a confidencialidade dos dados em ambientes cada vez mais interconectados.

Referências

[1] Cybersecurity Venture, Url = https://cybersecurityventures.com/cybersecurity-almanac-2024/, Urldate = 2025-03-01

[2] Cybercrime Expected To Skyrocket in Coming Years. https://www.statista.com/chart/28878/expected-cost-of-cybercrime-until-2027/ Accessed 2025-03-03

[3] Verizon: 2024 Data Breach Investigations Report. Accessed: 2025-03-01 (2024). https://verizon.com/dbir

[4] Sethi, K., Venu Madhav, Y., Kumar, R., Bera, P.: Attention based multi-agent intrusion detection systems using reinforcement learning. Journal of Information Security and Applications **61**, 18 (2021) https://doi.org/10.1016/j.jisa.2021.102923

[5] Gupta, N., Srivastava, K., Sharma, A.: Reducing false positive in intrusion detection system: A survey. International Journal of Computer Science and Information Technologies, 1600–1603 (2016)

[6] Tellache, A., Mokhtari, A., Korba, A.A., Ghamri-Doudane, Y.: Multi-agent reinforcement learning-based network intrusion detection system. arXiv preprint arXiv:2407.05766 (2024) https://doi.org/10.1109/NOMS59830.2024.10575541

[7] Bacha, A., Ktata, F., Louati, F.: Improving intrusion detection systems with multi-agent deep reinforcement learning: Enhanced centralized and decentralized approaches. Proceedings of the 20th International Conference on Security and Cryptography (SECRYPT 2023), 772–777 (2023) https://doi.org/10.5220/0012124600003555

[8] Owoputi, R., Ray, S.: Security of multi-agent cyber-physical systems: A survey. IEEE Access **10**, 121465–121479 (2022) https://doi.org/10.1109/ACCESS.2022.3223362

[9] Javadpour, A., Pinto, P., Ja'fari, F., Zhang, W.: Dmaidps: a distributed multi-agent intrusion detection and prevention system for cloud iot environments. Cluster Computing **26**, 367–384 (2023) https://doi.org/10.1007/s10586-022-03621-3

[10] Rajaee, M., Mazlumi, K.: Multi-agent distributed deep learning algorithm to detect cyber-attacks in distance relays. IEEE Access **11**, 10842–10849 (2023) https://doi.org/10.1109/ACCESS.2023.3239684

[11] Soltani, M., Khajavi, K., Jafari Siavoshani, M., Hossein Jahangir, A.: A multi-agent adaptive deep learning framework for online intrusion detection. Cybersecurity **6**(1), 1–16 (2023) https://doi.org/10.1186/s42400-023-00199-0

[12] Khraisat, A., Gondal, I., Vamplew, P., Kamruzzaman, J.: Survey of intrusion detection systems: techniques, datasets and challenges. Cybersecurity **2**(1), 20 (2019) https://doi.org/10.1186/s42400-019-0038-7

[13] Snort. https://www.snort.org/faq/what-is-a-signature Accessed 2025-03-01

[14] Suricata. https://suricata.io/features/ Accessed 2025-03-01

[15] Bougueroua, N., Mazouzi, S., Belaoued, M., Seddari, N., Derhab, A., Bouras, A.:

A survey on multi-agent based collaborative intrusion detection systems. Journal of Artificial Intelligence and Soft Computing Research **11**(2), 111–142 (2021) https://doi.org/10.2478/jaiscr-2021-0008

[16] Multi-Agent System for Flawless Financial Compliance. https://www.akira.ai/blog/multi-agent-system-for-financial-compliance Accessed 2025-03-01

[17] Ozkan-Okay, M., Samet, R., Aslan, , Gupta, D.: A comprehensive systematic literature review on intrusion detection systems. IEEE Access **9**, 34 (2021) https://doi.org/10.1109/ACCESS.2021.3129336

[18] Ozkan-Okay, M., Samet, R., Aslan, , Gupta, D.: ISSF: The Intelligent Security Service Framework for Cloud-Native Operation. Accessed: 2025-03-26. https://arxiv.org/abs/2403.01507

[19] Dhasarathan, C., Shanmugam, M., Kumar, M., Tripathi, D., Khapre, S., Shankar, A.: A nomadic multi-agent based privacy metrics for e-health care: a deep learning approach. Multimedia Tools and Applications **83**, 7249–7272 (2024) https://doi.org/10.1007/s11042-023-15363-4

[20] Funchal, G., Pedrosa, T., De La Prieta, F., Leitao, P.: Edge multi-agent intrusion detection system architecture for iot devices with cloud continuum. In: 2024 IEEE 7th International Conference on Industrial Cyber-Physical Systems (ICPS), pp. 1–6 (2024). https://doi.org/10.1109/ICPS59941.2024.10639952 . IEEE

[21] Pan, Z., Mishra, P.: Explainable ai for cybersecurity. Springer Nature Link, 53 (2023) https://doi.org/10.1007/978-3-031-46479-9

[22] Sharma, S., Pandey, A., Sharma, V., Mishra, S., Alkhayyat, A.: Federated learning and blockchain: A cross-domain convergence, 1121–1127 (2023) https://doi.org/10.1109/ICTACS59847.2023.10390227

[23] Jiao, W., Zhao, H., Feng, P., Chen, Q.: A blockchain federated learning scheme based on personalized differential privacy and reputation mechanisms, 630–635 (2023) https://doi.org/10.1109/ISPDS58840.2023.10235627

[24] Capuano, N., Fenza, G., Loia, V., Stanzione, C.: Explainable artificial intelligence in cybersecurity: A survey, 630–635 (2022) https://doi.org/10.1109/ACCESS.2022.3204171