****

TEORIA DOS JOGOS

Aplicada na segurança informática

Semedo, Héricles Emanuel. **Universidade Lusófona Humanidades e Tecnologias (a21 801 188) – LEIRT –2021**

****

Abstract

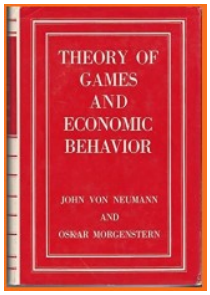
Uma das definições da palavra segurança é: "Conjunto das acções e dos recursos utilizados para proteger algo ou alguém."["segurança", in Dicionário Priberam da Língua Portuguesa] O facto da segurança ser um bem indispensável nos sistemas informáticos, desenvolvem-se 'acções e recursos' para garantir a segurança de tais sistemas. Este artigo tem como príncipal objetivo apresentar alguns casos em que usa-se a Teoria dos Jogos para modelar e analisar a segurança em sistemas computacionais.

Palavras-chave: Teoria dos Jogos, dilema dos Prisioneiros, IDS, Jogo estático, Jogo dinâmico, Botnets.

1. Teoria dos Jogos – Conceito

Teoria dos Jogos é um ramo da Matemática aplicada que estuda situações estratégicas, com o objetivo modelar matematicamente e de forma lógica o comportamento racional das entidades que decidem e executam ações, ou seja, jogadores.

Embora vários matemáticos terem contribuído para o surgimento do conceito de Teoria dos Jogos, aplicando a teoria das probabilidades em dilemas e situações estratégicas, a Teoria dos Jogos, apareceu pela primeira vez, como um ramo da matemática (framework), nos anos 40, com a publicação do livro The Theory of games and economic behavior – *John V. Neumann & Oskar Morgenstern.*



Apesar do ramo ser vulgarmente designado por Teoria dos Jogos, são na verdade várias teorias e modelos que permitem estudar tais situações.

2. Jogos

O comum conceito da palavra jogo implica a existência de mais do que um interveniente (jogador) e a tentativa de cada uma das partes de obter o melhor resultado possível que depende das suas decisões e comportamento.

A teoria dos jogos descreve de forma matemática três elementos fundamentais que definem todos os jogos:

1. Os jogadores;
2. Lista de estratégias possíveis para cada jogador;
3. Resultados que correspondem a cada estratégia ou combinação de estratégias.

A Teoria dos Jogos também define vários tipos de jogos de acordo com os ganhos potenciais para cada interveniente.

Outros teóricos defendem que uma interação pode ser considerado um jogo só se cumprirem os seguintes princípios:

1. Há pelo menos dois jogadores;
2. Há interação entre os jogadores;
3. Há recompensa ou objetivo;
4. Os jogadores agem racionalmente;
5. Os jogadores agem segundo interesses próprios.

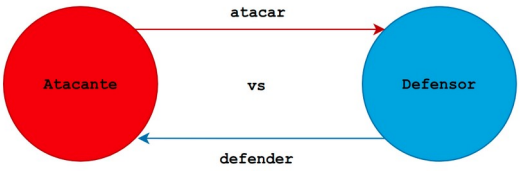
De acordo com estes princípios, pode-se considerar as ameaças de segurança informática como um jogo, visto que:

1. Há múltiplos jogadores

Em um ambiente cibernético, é possível definir pelo menos dois tipos de "jogadores" quando se trata de segurança, pois, de um lado há sistemas e/ou pessoas com perfil de atacante e sistemas e/ou pessoas que adotam o perfil de defensores.

2. Há interação entre os jogadores

Pode-se considerar que os jogadores estão em constante interação, pois como a própria designação dos perfis indicam, O "atacante" visa **atacar** os sistemas informáticos, enquanto que o "defensor" quer sempre **defender** os sistemas dos atacantes.

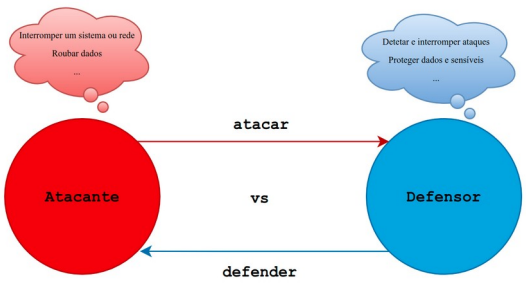


3. Há recompensa ou objetivo final

O ato de atacar ou defender dos jogadores têm sempre um ou mais objectivos.

Um dos objetivos principais do atacante pode ser simplesmente interromper um sistema informático ou rede de computadores, ou ainda, roubar dados importantes que circulam numa rede ou que são armazenados numa base de dados.

Já os defensores têm objetivos opostos aos dos atacantes, que pode ser detetar e interromper um ataque e assim garantir a segurança e integridade de um sistema computacional ou de dados/informações sensíveis.



4. Os jogadores agem racionalmente;

5. Os jogadores agem segundo interesses próprios.

Pode-se verificar ainda o cumprimento dos dois últimos princípios, pois todos os intervenientes agem racionalmente para atingir o objetivo ou recompensa.

3. Dilema dos prisioneiros

É quase imperativo referir o famoso **dilema dos prisioneiros** ao abordar o tema da Teoria dos Jogos.

O dilema dos prisioneiros foi desenvolvido na década de 50, pelo matemático *Albert W. Tucker*, que apesar de não ser o primeiro a formular a estrutura do dilema, foi o primeiro a aplicá-lo no tema penal e passou a ser o modelo clássico.

Pode ser definido pelo seguinte enunciado:

"Dois suspeitos (A e B) foram presos pela polícia. Não tendo provas suficientes para os condenar, a polícia separa os prisioneiros e oferece a ambos o mesmo acordo: se um dos prisioneiros confessar ou testemunhar contra o outro e o outro prisioneiro permanecer em silêncio, o que confessou sai livre enquanto o cúmplice silencioso cumpre 9 anos de sentença. Se ambos ficarem em silêncio, a polícia só pode condená-los a uma pena de 1 ano a cada um. Se ambos traírem o comparsa, cada um leva 3 anos de cadeia. Cada um dos prisioneiros faz a sua decisão sem saber qual é a decisão que o outro vai tomar, e nenhum tem certeza da decisão do outro. A questão que o dilema propõe é: o que vai acontecer? Como o prisioneiro vai reagir?"

As hipóteses apresentadas aos prisioneiros da ilustração podem ser resumidas da seguinte forma:

1. Se ambos confessarem, ambos cumprem 3 anos de prisão.

2. Se ninguém confessar, ambos cumprem 1 ano de prisão

3. Se apenas um suspeito confessar, aquele que confessou não cumpre a pena e aquele que não confessou cumpre 9 anos de prisão.

Com as hipóteses definidas, o enunciado pode ser representado pela tabela indicada a seguir, que é designada por tabela de ganhos (payoffs matrix):

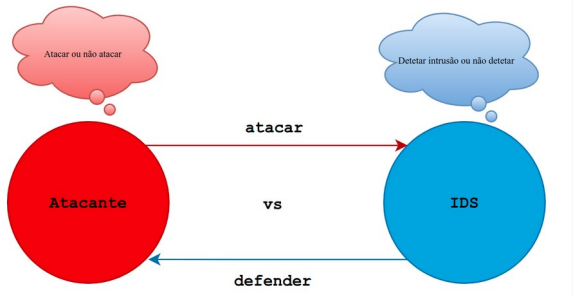
Preso A / Preso B	Confessar	Não confessar
Confessar	(3,3)	(0,9)
Não confessar	(9,0)	(1,1)

A matriz dos ganhos permite analisar mais rapidamente os ganhos ou perdas possíveis que resultam das decisões tomadas pelos prisioneiros (jogadores).

4. Dilema dos prisioneiros em IDS

O mesmo modelo de jogo usado no dilema dos prisioneiros é usado na análise e validação de sistemas de detecção ou prevenção de intrusão, normalmente designados por IDSs (*Intrusion Detection System*).

Usando a mesma analogia apresentada no segundo ponto do artigo, em interações envolvendo os IDSs, de um lado temos o atacante que deve decidir uma das duas jogadas: Atacar e/ou não atacar. E o seu oponente é o próprio sistema anti-intrusão cujas jogadas são: Detetar intruso e/ou não detectar.



Representando numa matriz de ganho, as jogadas do atacante e do sistema de detecção, obtém-se a tabela indicada abaixo:

Atacante / IDS	Atacar	Não atacar
Detetar	(0,-1)	(-2,0)
Não detetar	(-4,4)	(1,0)

A tabela de ganhos, embora tenha valores apenas representativos, pode ser interpretada identificando as seguintes combinações de jogadas:

1. O atacante ataca e IDS detecta a intrusão (0, −1)

Para o lado do IDS, pode-se reparar que o resultado está representado pelo valor 0, visto que ao detectar o ataque, o sistema em si não ganha nada nem perde nada. Enquanto que o resultado para o lado do atacante é representado pelo valor negativo de menos 1, visto que quando o atacante é detectado perde sempre alguma coisa, que na vida real pode ser apenas os recursos gastos para efetuar o ataque ou mesmo penalizações jurídicas.

2. O atacante não ataca e IDS detecta a intrusão (−2, 0)

Esta hipótese representa os casos em que uma IDS, lança um falso alarme. Esta é a razão pela qual o resultado da ação do IDS é representado pelo valor menos 2, já que os alarmes falsos normalmente são custosos para os IDSs já que é necessário consertar as falhas que causaram o falso alarme. Já o resultado para o lado do atacante é nulo, pois não ganha nem perde nada, visto que nem ao menos atacou o sistema.

3. O atacante ataca e IDS não deteta a intrusão (−2, 0)

Esta é sem dúvida a pior hipótese de combinação de jogadas para o lado do IDS e deste modo, o resultado do facto de não ter detectado um ataque efetuado é representado pelo valor menos 4, já que além ser necessário investigar e consertar as falhas que causaram a falha, a intrusão permitida pode interromper ou corromper sistemas informáticos críticos e expor dados e informações sensíveis. Enquanto que o atacante neste caso pode ser considerado como vencedor do jogo, já que cumpriu seu principal objetivo que é atacar um sistema sem ser pego. Por esta razão, o resultado do atacante é representado pelo valor 4.

4. O atacante não ataca e IDS não deteta a intrusão (−2, 0)

Esta é a melhor hipótese para ambos os jogadores, e como indicam os resultados, o atacante quando não ataca um sistema não corre o risco de ser identificado e sofrer qualquer perda e o estado ideal de um IDS é que não seja efetuado nenhum ataque.

Como referido anteriormente, os valores numéricos do modelo de jogo, são apenas representativos, já que os ganhos e perdas em um jogo semelhante no mundo real é principalmente de ordem qualitativa e não quantitativa, não permitindo assim refleti-los de forma exata.

Na avaliação e validação de IDSs o objetivo de modelar a interação atacante – IDS com o modelo apresentado, é mais do que apenas conhecer os ganhos e perdas de cada jogada, mas sim manipulá-los.

Um sistema ou os engenheiros que desenvolvem os sistemas de deteção de intrusão, não têm poder de decidir quando é que um atacante ataca ou não ataca. Conjugando o dilema dos prisioneiros com a teoria das probabilidades diminui essa impossibilidade dos sistemas.

A tabela indicada indica como seria uma matriz de ganho aplicada com a probabilidade de cada hipótese ou combinação de jogadas:

IDS \ Atacante	Atacar	Não atacar
Detetar	(0,-1) 25%	(-2,0) 25%
Não detetar	(-4,4) 25%	(1,0) 25%

Esta tabela, ao ser interpretada, permite concluir que a probabilidade de cada combinação é de 25%, dando assim a cada jogador uma probabilidade de 50% de ganho ou perda.

Esta combinação de probabilidade-hipótese é semelhante ao dos famosos jogos de sorte e azar ou de simplesmente "jogar uma moeda ao ar". É sem dúvida os piores valores que se deseja obter ou garantir em jogos deste tipo, principalmente se envolve sistemas reais críticos.

Mas lembrando que a teoria das probabilidades postula que a probabilidade de um acontecimento é a razão entre casos favoráveis e casos possíveis, é possível manipular as probabilidades no lado dos IDS por aprimorá-los e fazer com que sejam mais robustos a ataques o que claramente é possível conseguir realizando testes de avaliação.

Após melhorar as probabilidades das hipóteses, uma possível combinação hipótese-probabilidade que seria razoável para uma interação atacante – IDS seria de 95% para o caso de um IDS detectar um ataque efetuado ou de não haver nenhum ataque e 5% de probabilidade de um atacante ser bem sucedido no seu ataque ou de um IDS lançar um falso alarme.

Este modelo pode ser modelado na forma como apresenta a seguinte matriz de ganhos:

IDS \ Atacante	Atacar	Não atacar
Detetar	(0,-1) 95%	(-2,0) 5%
Não detetar	(-4,4) 5%	(1,0) 95%

Além de manipular as probabilidades de cada acontecimento, é possível piorar a perda de um atacante no caso de ser efetuado um ataque e o IDS detectar o ataque. Esta ação possibilita que o modelo acima seja melhorado para o modelo representado na tabela seguinte:

IDS \ Atacante	Atacar	Não atacar
Detetar	(0,-8) 95%	(-2,0) 5%
Não detetar	(-4,4) 5%	(1,0) 95%

Além de ter um IDS eficiente, um dos objetivos deve ser garantir que o valor absoluto que representa a perda de um atacante que é detectado seja maior que o valor do ganho que ele teria se fosse bem sucedido no seu ataque.

Como referido anteriormente os sistemas ou as pessoas que monitoram os sistemas de defesa, não têm o poder de saber se um atacante vai atacar ou não atacar um sistema. Mas como se pode observar, manipulando as probabilidades e as perdas do fracasso dos atacantes, há o aumento do custo de um ataque. E como a teoria dos jogos também indica, os jogadores tomam sempre decisões de forma racional. Deste modo um atacante ao perceber que tem muito mais a perder se não for bem sucedido em um ataque e que a probabilidade de fracasso é maior do que o de sucesso, de forma indireta obriga-se ao atacante a não efetuar o ataque.

Embora este modelo seja um pouco otimista, o objetivo de um bom sistema é sempre a possibilidade ideal e impossível, de ser 100% possível ter sucesso e 0% de fracasso, o que seria um sistema perfeito.

Assim sendo, este modelo de Teoria dos Jogos, permite demonstrar um dos princípios básicos da segurança: ***"Nunca é possível obter 100% de segurança"***. O resultado que se pode atingir é o máximo de segurança nos sistemas informáticos e redes de computadores.

5. Tipos de Jogos

Considerando alguns critérios, tais como os resultados (ganhos ou perdas) de um jogo, ou natureza das estratégias dos jogadores, a Teoria dos Jogos, sub classifica os jogos em vários tipos.

De acordo com o comportamento ou perfil de um jogador, um jogo pode ser classificado por estático ou dinâmico.

1. Jogo estático

Um jogo pode ser classificado como estático, se os jogadores envolvidos decidem quais estratégia ou jogadas utilizar, sem ter nenhum conhecimento das ações tomadas pelos outros jogadores envolvidos.

Um exemplo de jogo estático, é o famoso e já referido, dilema dos prisioneiros. O próprio enunciado do dilema deixa claro que os prisioneiros são separados e cada um tem que decidir sem saber qual será a decisão do outro prisioneiro.

2. Jogo dinâmico

Em um jogo considerado dinâmico, normalmente os jogadores repetem o mesmo processo inúmeras vezes em sequência.

Também, em jogos dinâmicos, os jogadores usam estratégias ou jogadas cooperativas, que permitem que dois ou mais jogadores estejam em um consenso no processo de decisão. A ideia predominante em um modelo de jogo dinâmico é: "Se todos cooperarem todos vencerão."

Em jogos dinâmicos aparece o conceito de grupo, que é um conjunto de jogadores cooperantes e com objetivos comuns.

6. Jogos dinâmicos & Botnets

O modelo de jogo dinâmico pode ser verificado em Botnets maliciosos, que, como o próprio nome indica são redes de bots utilizadas para executar ataques informáticos tais como a negação de serviços ou para roubar dados, enviar mensagens de spam.

Redes de Botnets maliciosos (jogadores) trabalham juntos, ou cooperam entre si, com o mesmo objetivo que é de lançar ataques informáticos. O mesmo modelo pode ser observado botnets individuais, em que o jogador é cada *bot* da rede.

Uma análise baseada na Teoria dos Jogos, permitiu concluir que uma ideia ou estratégia cooperativa entre dois botnets, sempre resulta na sobrevivência de ambos os botnets, enquanto uma estratégia competitiva elimina sempre um dos botnets.

Esta conclusão permitiu obter um novo método de combater ataques maliciosos envolvendo botnets. Este método consiste em introduzir nas redes de botnets, outros botnets que não sejam cooperativos com os botnets maliciosos. A competição causada entre tais botnets, não só dificulta o sucesso de um botnet malicioso, mas também pode causar a eliminação do botnet malicioso, como indica o estudo feito com a Teoria dos Jogos.

Este passou a ser um dos critérios para classificar e distinguir os botnets legais dos botnets ilegais, que são os que visam efetuar ataques maliciosos aos sistemas informáticos.

7. Mais aplicações

A Teoria dos Jogos também tem sido aplicada na gestão de algoritmos do protocolo Client Puzzle (CPP), que é um dos protocolos de comunicação que é usado para combater ataques de negação de serviços (DOS).

Esta aplicação, assim como outras, envolvem modelos mais complexos, pelo facto de terem mais componentes matemáticas associadas. Embora sejam mais difíceis de demonstrar, contribuem na resolução de problemas de segurança informática e aprimoramento dos sistemas de prevenção.

8. Vantagem e limitação

Certo matemático afirma que a teoria dos jogos "*ajuda a entender teoricamente os processos de decisão entre jogadores, a partir da compreensão da lógica da situação em que estão envolvidos*".

Permite entender ***teoricamente*** visto que na Teoria dos Jogos, utilizam-se abstrações que excluem alguns fatores particulares e acidentais que podem afetar o resultado do processo em estudo ou jogo. Permite assim, uma boa aproximação e não um resultado concreto das decisões e estratégias utilizadas, sendo esta a principal limitação deste ramo matemático.

Mas é digno de observação que a Teoria dos Jogos é uma poderosa ferramenta que ajuda a perceber, modelar e avaliar interações estratégicas (jogos), principalmente as que envolvem a segurança informática.

9. Referências

[1] *Vicky Papadopoulou and Andreas Gregoriades.* **Network Security Validation Using Game Theory**.

[2] *Attacks Oleksii Ignatenko, Oleksander Sinetskiy.* **Game Theoretic Modeling of Network Security.**

[3] *Quanyan Zhu and Stefan Rass.* **Game Theory Meets Network Security.**

[4] *Mohammad Hossein Manshaei, Quanyan Zhu, Tansu Alpcan and Tamer Bąsar, Jean-Pierre Hubaux.* **Game Theory Meets Network Security and Privacy**

[5] *Antonis Michalas, Nikos Komninos and Neeli R. Prasad.* **Cryptographic Puzzles and Game Theory against DoS and DDoS attacks in Networks**.

[6] *Kong-wei Lye and Jeannette M. Wing.* **Game strategies in network security.**