CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
CAMPUS TIMÓTEO

Audrey Mistris

MODELAGEM DE AMEAÇA PERSISTENTE AVANÇADA POR RANSOMWARE USANDO TEORIA DE JOGOS

Timóteo - MG

2022

Audrey Mistris

MODELAGEM DE AMEAÇA PERSISTENTE AVANÇADA POR RANSOMWARE USANDO TEORIA DE JOGOS

Monografia apresentada ao Curso de Engenharia de Computação do Centro Federal de Educação Tecnológica de Minas Gerais - Campus Timóteo, para obtenção do grau de Bacharel em Engenharia de Computação.

Orientador: Lucas Pantuza Amorim

Timóteo - MG

2022

Audrey Mistris

Modelagem de Ameaça Persistente Avançada por *Ransomware* usando Teoria de Jogos

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia de Computação do Centro Federal de Educação Tecnológica de Minas Gerais, campus Timóteo, como requisito parcial para obtenção do título de Engenheiro de Computação.

Trabalho aprovado. Timóteo, 19 de dezembro de 2022.

_____

Prof. Dr. LUCAS PANTUZA AMORIM
Orientador

_____

Prof. Me. ADILSON MENDES RICARDO
Professor Convidado

_____

Prof. Me. MARCELO DE SOUSA BALBINO
Professor Convidado

Timóteo
2022

Emitido em 19/12/2022

FOLHA DE ROSTO (PLATAFORMA BRASIL) Nº 2/2022 - DCCTM (11.63.05)

(Nº do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 21/12/2022 10:14)
ADILSON MENDES RICARDO
PROFESSOR ENS BASICO TECN TECNOLOGICO
CECOMTM (11.51.22)
Matrícula: ###493#8

(Assinado digitalmente em 20/12/2022 17:57)
LUCAS PANTUZA AMORIM
PROFESSOR ENS BASICO TECN TECNOLOGICO
DCCTM (11.63.05)
Matrícula: ###974#1

(Assinado digitalmente em 22/12/2022 09:46)
MARCELO DE SOUSA BALBINO
PROFESSOR ENS BASICO TECN TECNOLOGICO
DCCTM (11.63.05)
Matrícula: ###093#2

Visualize o documento original em https://sig.cefetmg.br/documentos/ informando seu número: **2**, ano: **2022**, tipo: **FOLHA DE ROSTO (PLATAFORMA BRASIL)**, data de emissão: **20/12/2022** e o código de verificação: **dd0df037e8**

Às memórias de Paul Mistris.
Dedico esta monografia ao meu pai
que apesar de não estar mais entre nós
sempre esteve do meu lado.

“Le monde est trop calme sans toi à proximité.”
— Lemony Snicket

Agradecimentos

Agradeço aos meus pais, Paul e Wania, que sempre estiveram me apoiando ao longo de toda minha trajetória, possibilitando que eu chegasse até aqui. Obrigada pela compreensão, pela paciência, por todo o esforço investido na minha educação e por acreditar no meu potencial, como sempre.

Ao meu orientador, Lucas Pantuza Amorim, por aceitar conduzir meu trabalho, pelas valiosas contribuições que fizeram toda a diferença e por me empurrar até o final do meu projeto de pesquisa.

A todos os meus professores do curso de Engenharia de Computação do CEFET-MG, Campus Timóteo, pelo comprometimento e a excelência da qualidade do ensino oferecido.

Também agradeço aos meus amigos do curso de graduação, que participaram dos diversos desafios que surgiram, descobertas e aprendizados, pelas trocas de ideias, ajuda mútua e apoio moral.

E por fim, às mil e uma noites em claro, durante as quais pesquisei arduamente e escrevi tanto. Sem elas não teria tido a mesma inspiração para produzir este trabalho. Já dizia alguém sábio: a noite traz conselho.

Com muita gratidão,
“Até mais, e obrigada pelos peixes!”

“If we wait until we’re ready,
we’ll be waiting for the rest of our lives.”
— Lemony Snicket, The Ersatz Elevator

“Se esperarmos até estarmos prontos,
esperaremos pelo resto de nossas vidas.”
— Lemony Snicket, O Elevador Ersatz

Resumo

Nesta pesquisa, foram estudados *Advanced Persistent Threats* (APT) por *ransomware*, ataques cibernéticos seletivos e direcionados que obtêm acesso não-autorizado a sistemas e informações para tomar dados confidenciais e causar danos a indústrias e organizações governamentais. Esses ataques são difíceis de detectar e podem causar danos significativos, como perda e exposição de dados, interrupção das operações e danos à imagem institucional. Ao realizar modificações internas em sistemas, o invasor mantém o acesso indevido pelo maior tempo possível na infraestrutura. Consequentemente, ao afetar os computadores dos alvos, o acesso aos arquivos e dados do sistema é inativado, sendo exigido um pagamento de resgate para a devolução das funcionalidades e dados. Esses ataques foram aprimorados com o crescimento da Internet, apresentando mais características como ameças de vazamento ou venda de dados da vítima do ataque. Aplicou-se um modelo de Teoria de Jogos para analisar cenários de ataque e defesa. Assim, foi possível determinar as condições sob as quais o defensor tem vantagem em neutralizar o ataque com sucesso, para estimar a lucratividade dos ataques de APT por *ransomware*. Desta forma, procurou-se compreender as variáveis críticas de decisão envolvidas por determinado *ransomware* de venda de dados. Comparando as variações da reputação de ataque e da lucratividade dos diferentes *ransomwares*, os resultados da pesquisa demonstraram que os ataques de APT por *ransomware* com venda e ameaça de dados são mais lucrativos do que o *ransomware* tradicional e, de forma geral, o *ransomware* com venda de dados se mantém mais lucrativo que o *ransomware* com ameaça de dados. Deduz-se que o motivo desta ocorrência vem da possibilidade do atacante obter um pagamento do resgate ou vender os dados da vítima. O estudo de caso também sugere que as incertezas introduzidas por esse modelo de lucros podem ter um impacto na reputação do atacante e na disposição de pagar da vítima, inferindo que a venda de dados nem sempre pode aumentar a lucratividade do ataque. Também observou-se que a maximização da reputação é importante para os *ransomwares* tradicionais e com ameaça de dados, mas não para o *ransomware* com venda de dados, uma vez que o invasor consegue manipular sua variável de reputação para maximizar a recompensa. Os resultados da análise concluem que a lucratividade do ataque pode ser afetada pela incerteza e disposição da vítima de pagar. Por fim, a melhor estratégia para as vítimas do ataque é não pagar pelo resgate dos dados, independentemente da reputação da campanha de ataque APT por *ransomware*, visto que o pagamento do resgate pode encorajar mais ataques.

Palavras-chave: Ameaça Persistente Avançada (APT), Estratégia, Lucratividade, *Ransomware*, Teoria de Jogos.

Abstract

In this research, Advanced Persistent Threats (APT) by ransomware were studied — selective and targeted cyberattacks that gain unauthorized access to systems and information to seize confidential data and cause damage to industries and government organizations. These attacks are difficult to detect and can cause significant damage, including loss and exposure of data, disruption of operations, and damage to institutional reputation. By making internal modifications to systems, the intruder maintains unauthorized access for as long as possible. Consequently, when the target computers are affected, access to the files and data of the system is disabled, requiring a ransom payment to return functionality and data. These attacks have been enhanced with the growth of the Internet, presenting more characteristics, such as threats to leak or sell the victim's data. A Game Theory model was applied to analyze attack and defense scenarios. It was then possible to determine the conditions under which the defender has an advantage in successfully neutralizing the attack, to estimate the profitability of APT ransomware attacks. Critical decision variables in data-selling ransomware were explored, comparing variations in attack reputation and profitability among different ransomware types. Research results demonstrated that APT ransomware attacks with data-selling and data threat are more profitable than traditional ransomware, with data-selling ransomware consistently proving more lucrative. It is deduced that this occurrence is due to the attacker's ability to obtain ransom payment or sell the victim's data. The case study also suggests that uncertainties introduced by this profit model can impact the attacker's reputation and the victim's willingness to pay, inferring that selling data may not always increase attack profitability. It was also observed that maximizing reputation is important for traditional and data-threat ransomware, but not for data-selling ransomware, as the attacker can manipulate reputation variables to maximize reward. The results of the analysis conclude the attack's profitability may be affected by the victim's uncertainty and willingness to pay. Ultimately, the best strategy for ransomware victims is not to pay the ransom, regardless of the reputation of the APT ransomware campaign, as ransom payments may encourage further attacks.

Keywords: Advanced Persistent Threat (APT), Game Theory, Profitability, Ransomware, Strategy.

Lista de ilustrações

Lista de abreviaturas e siglas

APT	*Advanced Persistent Threat* – Ameaça Persistente Avançada
IEC	*International Electrotechnical Commission* – Comissão Eletrotécnica Internacional
ISO	*International Organization for Standardization* – Organização Internacional para Padronização
NIST	*National Institute of Standards and Technology* – Instituto Nacional de Padrões e Tecnologia
SGSI	Sistema Gerenciamento de Segurança da Informação

Lista de símbolos

P	Probabilidade
π	Vetor de recompensa esperada pelo atacante
N	Jogo definido
a	Atacante
v	Vítima
V	Conjunto de dados da vítima
R	Valor de resgate dos dados
p	Escolha da vítima de pagar o resgate
r	Escolha do atacante de devolver os dados
s	Escolha do atacante de vender os dados
i	Indivíduo específico
C_r	Custo de devolução de dados
C_d	Custo de transação dos dados
D_i	Valor de mercado dos dados roubados de um indivíduo
A_i	Lucro do atacante
u	Recompensa da vítima
$V_{r,i}$	Valor dos dados ainda bloqueados para a vítima i
$L_{d,i}$	Perda da vítima i se os dados roubados forem vendidos pelo invasor
Π_b	Lucro total recebido pelo atacante se $p = 0$ e $r = 0$
Π_t	Lucro do atacante se $p = 1$
β_r	Probabilidade do atacante devolver os dados
β_d	Probabilidade de manter os dados roubados confidenciais
u_B	Recompensa esperada da vítima considerando β_r e β_d
π_B	Recompensa esperada do atacante considerando β_r e β_d
R_t	Valor de resgate no jogo de reputação ótima
R_u	Valor de resgate no jogo competitivo de reputação falha

Sumário

1 Introdução

Com o aumento da conectividade à *Internet*, o número de vulnerabilidades aumentou ao longo dos últimos anos, assim como a quantidade de ataques (FONTES, 2017). Corporações e indivíduos em todo o mundo se conectam e se organizam no meio cibernético (TSAKANYAN, 2017) e, à medida que o uso da *Internet* se expande, essa infraestrutura se tornará cada vez mais integrada à vida cotidiana (NATIONALSECURITYSTRATEGY, 2011).

Dito isso, considerando o crescimento da dependência da tecnologia, esta precisa proporcionar três princípios básicos para garantir a segurança da informação: confidencialidade, integridade e disponibilidade. Além disso, defende-se que, para que uma informação seja considerada segura, o sistema que a administra ainda deve respeitar os seguintes critérios: autenticidade, não-repúdio, privacidade e auditoria (TIPTON; KRAUSE, 2006).

A análise dos ataques cibernéticos nos últimos anos mostra que as ameaças à infraestrutura virtual não estão apenas aumentando em volume, mas também se tornando mais sofisticadas (CHUNG; KAMHOUA; KWIAT, 2016). Acredita-se que ataques maliciosos que raptam dados de forma extorsiva (*ransomwares*) são muito lucrativos, segundo um estudo (SIMOIU et al., 2019) que diz que nos últimos tempos, ataques de *ransomware* impactaram organizações governamentais, serviços de saúde, instituições de educação e corporações de negócios. Foi estimado em 2019 o custo desses ataques em aproximadamente US$7,5 bilhões. De acordo com os relatórios anuais de segurança cibernética da empresa CyberEdge Group (2022), o *ransomware* está listado como uma das três primeiras ameaças cibernéticas de 2017 a 2020, além de que 63% das vítimas de *ransomware* pagaram resgates no ano de 2021, incentivando o aumento desses ataques.

Ataques de computador são uma ameaça generalizada, e considera-se que esses evoluem à medida que os sistemas mirados evoluem. Esse tipo de ameaça pode ser causado por vários motivos, como erros de programação ou configuração de sistema. Assim, é sensato afirmar que qualquer sistema de computador é vulnerável e portanto, deve-se tentar detectar os ataques que ele sofre, possibilitando a interrupção e análise de modo a eliminar a fonte (SINGER; FRIEDMAN, 2014).

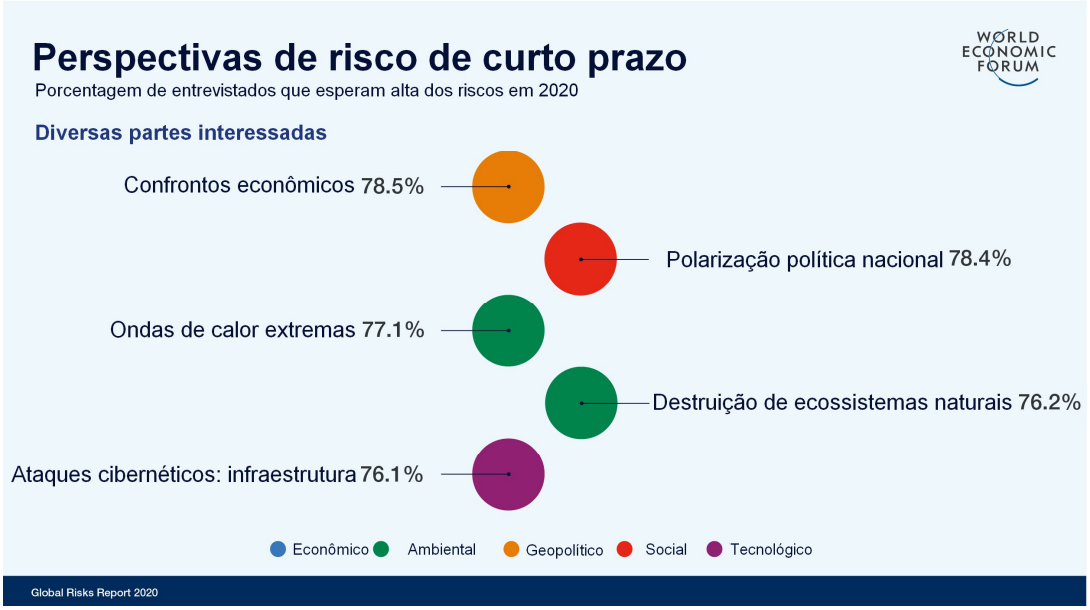
Vários anos de pesquisa em segurança da informação proporcionaram uma quantidade de mecanismos preventivos de segurança tais como criptografia, monitoramento e segurança de rede, controles de acesso, mapeamento de ativos, entre outros. Esses procedimentos reduzem a probabilidade de comprometimento de plataformas computacionais (LIU; TANAKA; KANTA, 2007). No entanto, a despeito desses mecanismos, ainda ocorrem intrusões em sistemas. Segundo o laboratório DFNDR (2018), o Brasil sofre mais de 120 milhões de ataques cibernéticos somente no ano de 2018. De acordo com *The Global Risks Report 2020* (WORLDECONOMICFORUM, 2020), ataques cibernéticos e ameaças de violação de dados estão entre os dez principais riscos globais durante os próximos dez anos, conforme ilustrado na Figura 1.

Figura 1 – Perspectiva de risco de longo prazo, de um ponto de vista multissetorial.

Fonte: The Global Risks Report 2020, WORLDECONOMICFORUM.

Os ciberataques em infraestruturas críticas foram classificados como o quinto maior risco em 2020 pelo pela *World Economic Forum* (2020), enquanto a probabilidade de detecção das entidades organizadas de crimes cibernéticos é estimada em 0,05% nos Estados Unidos. Já a curto prazo, houve alta de riscos desses ataques em 76,1% no ano de 2020. A Figura 2 ilustra tal hipótese.

Figura 2 – Perspectiva de risco de curto prazo durante o ano de 2020.



Fonte: The Global Risks Report 2020, WORLDECONOMICFORUM.

"A atual falta de governança de tecnologia global e a presença de pontos cegos de segurança cibernética aumentam o risco de um ciberespaço fragmentado e de regulamentações de tecnologia concorrentes." (WORLDECONOMICFORUM, 2020).

Em concordância com a revista *World Economic Forum* (2020), é desejável otimizar a aplicação dos controles de segurança sem comprometer a rede ou interromper operações.

1.1 Problema

A segurança dos sistemas computacionais é um tema preocupante e significativo. Conforme o crescimento do uso da tecnologia digital, as ameaças evoluem dinamicamente (GORE; PADILLA; DIALLO, 2017).

O *ransomware* é uma categoria de *malware* que consiste em criptografar arquivos de um sistema e o atacante exige um pagamento para fornecer os meios de descriptografar esses arquivos. De acordo com a revista de saúde Becker's Hospital Review (2016), o primeiro ataque de *ransomware* conhecido ocorreu no ano de 1989 e teve como alvo o setor de saúde. Vale ressaltar que a revista também apontou que o setor de saúde continua entre os principais alvos de ataques por *ransomware*. Com o aprimoramento das táticas de *ransomware*, apareceram ataques mais sofisticados, como os ataques CryptoLocker em 2013 (JARVIS, 2013) e WannaCry de 2017 (MOHURLE; PATIL, 2017). Estima-se que a campanha de *ransomware* de 2017 infectou em torno de 200.000 computadores em diversos países, atingindo em sua maioria hospitais. Segundo a empresa Cyence (CBSNEWS, 2017), as perdas econômicas em consequência ao ataque cibernético passaram de US$4 bilhões naquele ano.

Um *ransomware* pode ser do tipo APT ou básico, dependendo de sua natureza e nível de complexidade. O *ransomware* APT é normalmente desenvolvido para lançar ataques altamente sofisticados realizados em uma série de etapas secretas (BAKSI; UPADHYAYA, 2018). Embora o ganho monetário seja geralmente o objetivo principal de tais ataques, eles podem ter outras finalidades ocultas e/ou disfarçadas. Por outro lado, em um ataque de *ransomware* padrão, o invasor criptografa os dados e exige um resgate. Se o resgate for pago, o invasor liberará os recursos criptografados; caso contrário, a vítima perderá os recursos indefinidamente. Esses ataques geralmente têm um único objetivo: tornar os recursos da vítima inacessíveis até que o resgate seja pago.

Segundo a revista Suno (MOUTINHO, 2021), um levantamento feito pela ISH Tecnologia apontou cerca de 13 mil empresas brasileiras são vitimas de ataques *hackers* por mês, a maioria do tipo *ransomware*. As estimativas do setor de segurança e da aplicação da lei sobre a quantidade de dinheiro extorquida com sucesso, além dos danos subsequentes, causados por ataques de *ransomware* variam muito, porém são consideráveis. Os autores Young e Yung (2017) desaprovam a falta de uma reação efetiva focada em ataques de *ransomware*, apesar do perigo ser conhecido há mais de 20 anos.

Com o aumento de ataques direcionados, específicos e lucrativos, isto é, Ameaças Persistentes Avançadas (APT) por *ransomware*, o estudo do Ransomware Task Force (2021)

destaca que o *ransomware* é semelhante a um assalto ou rapto.

O problema levantado por esta pesquisa é a falta de estratégia e preparo diante do lançamento de ciclos de ataques (campanhas) APT, especificamente por ataque de *ransomware*. Os mais recentes avanços de ataques de *ransomware* vêm com recursos adicionais como, por exemplo, estratégias de cancelamento de campanha, ou um plano de contingência de ataque ao ser descoberto antes do lançamento do ataque APT (BAKSI; UPADHYAYA, 2020), que os qualificam como Ameaças Persistentes Avançadas. A interação entre o atacante e a vítima de um ataque de *ransomware* possibilita modelar cenários de Teoria de Jogos.

Após sofrer um ataque, a vítima tem as seguintes opções: não fazer nada para lidar com a ameaça ou escolher uma estratégia apropriada para a reintegração dos arquivos roubados. Responder a um ataque de *ransomware* avançado envolve investimento em tecnologias de *backup* e recuperação de dados, ou pagar o resgate exigido no caso de um ataque bem-sucedido.

Em termos da decisão no pagamento de resgate, a IBM (2016) entrevistou aproximadamente 600 empresas nos Estados Unidos sobre ataques de *ransomware*. Quase metade das organizações relatou na pesquisa a ocorrência desses ataques e cerca de 70% dessas indicaram o pagamento de resgate para tentar recuperar os dados. Cerca de 50% pagaram mais de US$10.000 e 20% relataram mais de US$40.000. Outra pesquisa, da Kaspersky Labs (2021a), mostra que 56% das vítimas de *ransomware* pagaram o resgate para restaurar o acesso aos seus dados. Por outro lado, de acordo com uma pesquisa sobre práticas de resposta a *ransomware*, em torno de 96% das empresas afetadas decidiram não pagaram o resgate.

1.2 Justificativa

Ataques por APT são direcionados, isto é, têm um propósito previamente definido (*targeted attacks*). Ao contrário de um ataque comum em que o atacante procura um número limitado de vulnerabilidades em muitos alvos de um sistema, uma APT tem seus alvos escolhidos com antecedência e leva um tempo para analisar cada um, de modo a encontrar vulnerabilidades possíveis de explorar e invadir o sistema (CISCO, 2021). Esse tipo de invasão é tipicamente usado com o objetivo de roubar informações de entidades ricas e poderosas, como governos ou multinacionais. São normalmente motivadas por espionagem corporativa e sabotagem (KASPERSKY, 2021b), possibilitadas pela capacidade dos invasores de modificar os ataques existentes para evitar a detecção por mecanismos de defesa reativos e desenvolver novos ataques usando brechas anteriormente desconhecidas pelos desenvolvedores até o momento em que *hackers* as descobrem (vulnerabilidades *zero-day*) (SECUREWORKS, 2016).

Os ataques de *ransomware* oriundos de APT causam estragos em empresas e organizações governamentais não apenas financeiramente, mas também em comprometimento de imagem, confiança (em resguardar dados pessoais e informações sensíveis) e produtividade. Os ataques de *ransomware* podem fazer com que serviços, organizações e infraestruturas

suspendam suas atividades (MILOSEVIC; SKLAVOS; KOUTSIKOU, 2016).

Recentes relatórios da Check Point Research (2022) demonstram que empresas vítimas de ataques de *ransomware* acabam gastando cerca de sete vezes mais do que o valor pago nos resgates; as perdas a longo prazo são muito mais significativas devido a reparos posteriores, incluindo ações de resposta e restauração de sistemas, trabalhos de rotina, honorários advocatícios, custos de monitoramento, entre outros. Os relatórios apontaram uma grande evolução dessas ações criminosas nos últimos anos, sendo hoje o tipo de ataque mais prejudicial que as organizações têm enfrentado.

Além disso, um estudo da Veeam Software (2022) diz que 72% das empresas analisadas tiveram ataques parciais ou completos em seus repositórios de *backup*, e que 76% dessas organizações que pagaram o resgate. Aproximadamente 1/3 delas não conseguiu recuperar seus dados. Segundo o *Chief Technology Officer* (CTO) da Veeam Software ou diretor de tecnologia, Danny Allan (2022):

> "Pagar a criminosos para restaurar dados não é uma estratégia de proteção de dados. Não há garantia da recuperação de dados, os riscos de danos à reputação e perda de confiança do cliente são altos e, o mais importante, alimenta uma profecia autorrealizável de que a atividade criminosa compensa."

A Teoria dos Jogos foi introduzida em diversas ocasiões em segurança de rede para descrever as interações entre atacante e defensor e as maneiras como eles podem afetar mutualmente (SANKARDAS et al., 2010). Como esse vetor de ataque age com base em um comportamento sistematizado por criminosos cibernéticos, pode-se considerar uma abordagem sob a premissa da Teoria dos Jogos. Além disso, a Teoria de Jogos permite sugerir várias ações prováveis e prever seus resultados relacionados para controlar ameaças futuras em sistemas de segurança. Apresenta-se um jogo atacante-defensor não cooperativo de soma zero. Deste modo, projeta-se um jogo estratégico de segurança entre um atacante e uma vítima para investigar as interações dinâmicas de jogadores racionais com interesses concorrentes.

As ferramentas automatizadas de análise de ataque (SHEYNER; WING, 2003), são impraticáveis no contexto de *ransomware*, onde o invasor pode empregar táticas de engenharia social e lançar o ataque em etapas. Entretanto, esse tipo de ataque pode ser efetivamente modelado usando a Teoria de Jogos, que registra as interações entre o atacante e o defensor.

1.3 Objetivo

Procura-se modelar a interação entre parte atacante e parte vítima do ataque de APT por ransomware aplicando Teoria de jogos, através da análise de diferentes tipos de *ransomwares*.

1.3.1 Objetivos específicos

Mais intrinsecamente, esta pesquisa tem por finalidades:

- Caracterizar o jogo;
- Estabelecer os tipos de *ransomware*;
- Comparar cada modelo separadamente;
- Estabelecer um modelo de tomada de decisão estratégica com recompensa máxima;
- Determinar a importância do grau de reputação de *ransomware*.

1.4 Organização do Texto

O Capítulo 2 aborda algumas informações básicas e realiza uma análise de *ransomware* baseada na Teoria de Jogos. O Capítulo 3 apresenta trabalhos relacionados neste campo de pesquisa. No Capítulo 4, demonstra-se a metodologia utilizada, além dos parâmetros de tomada de decisão informados para cada situação proposta. No Capítulo 5, observa-se o andamento dos modelos propostos e verifica-se a influência, ou efeitos, dos parâmetros escolhidos para esta pesquisa. Em seguida, comparam-se as partidas dos diferentes jogos apresentados na análise dos modelos, baseados nas estratégias adotadas, averíguam-se os resultados. Finalmente, no Capítulo 6, propõe-se uma solução estratégica para preparação e mitigação de *ransomware*. Encerra-se o trabalho ao abordar ponderações da pesquisa para lidar com ataques de *ransomware* mais sofisticados.

2 Fundamentação Teórica

Este capítulo apresenta conceitos que foram necessários para desenvolver a pesquisa e podem auxiliar o entendimento da pesquisa.

2.1 Segurança da Informação

A Segurança da Informação, segundo Stoneburner, Hayden e Feringa (2004), protege as informações e os sistemas de informação contra acessos, divulgações, interrupções, modificações, utilizações indevidas de dados e até a possível destruição desses, de modo a fornecer integridade, confidencialidade e disponibilidade.

2.1.1 ISO/IEC 27001

A ISO 27001 é uma norma internacional do Sistema de Gestão de Segurança da Informação (SGSI). Sua estrutura foi projetada para auxiliar as empresas a gerenciar suas operações de segurança de maneira centralizada, uniforme e econômica conforme as melhores práticas globais, avaliada e certificada de modo independente (BSIBRASIL, 2021). O padrão possibilita a proteção dos dados sensíveis e confidenciais de forma mais eficiente ao minimizar os acessos indevidos ou ilegais (ISO, 2013).

2.1.2 Vulnerabilidade e Ameaça Cibernética

Segundo o NIST (2021), uma vulnerabilidade é uma fraqueza em um sistema de informações, procedimentos de segurança do sistema, controles internos ou implementação, que pode ser explorada por agentes externos ou internos, ou acionada por uma fonte de ameaça.

As vulnerabilidades diferem das ameaças cibernéticas porque não são introduzidas em um sistema, elas estão lá desde o início (SECURITYSCORECARD, 2021). Por outro lado, as ameaças cibernéticas são introduzidas como resultado de um evento externo ao sistema ou processo, tal como uma ação indevida de um funcionário ou de um cibercriminoso.

2.2 Ameaças Persistentes Avançadas

Advanced Persistent Threats (APT) permitem ataques seletivos que obtêm acesso não-autorizado a sistemas e informações para filtrar dados confidenciais, ou causar danos a uma empresa, indústria ou organização governamental. Atualmente, muitas dessas ameaças permanecem não-detectadas e, mesmo detectadas, realizam modificações internas no sistema ou aplicação para atingir seu objetivo (FIREEYE, 2016) e manter o acesso indevido.

A definição de uma APT pode ser resumida pelos componentes da sua terminologia:

- **(A)** *Advanced*: o atacante possui conhecimentos e ferramentas de intrusão à sua disposição, tais como tecnologia e técnicas de coleta de informações. Ele está familiarizado com métodos de direcionamento para atingir, comprometer e manter o acesso ao alvo, além de poder desenvolver *exploits* personalizados e específicos para o ataque.
- **(P)** *Persistent*: o atacante tem um propósito específico a cumprir ao invés de buscar informações de forma oportunista. O direcionamento do ataque é realizado por meio de monitoramento e interação contínuos para atingir os objetivos definidos. No entanto, não implica em vários ataques constantes na esperança que alguns dêem certo; um dos objetivos do invasor é manter o acesso ao alvo a longo prazo, por isso a estratégia *"low-and-slow"* costuma ser mais bem-sucedida.
- **(T)** *Threat*: o atacante é qualificado, motivado, organizado e financiado. O planejamento prévio e custeamento do ataque são características da APT, uma vez que o ataque está direcionado para afetar especificamente um alvo.

A Figura 3 ilustra o conceito de APT, destacando sua característica de persistência por um ciclo de vida em *loop* infinito. Uma vez definido o alvo, reúnem-se ferramentas e recursos de modo a efetuar o ataque devidamente e manter seu ponto de apoio.

Figura 3 – Descrição da abordagem em estágios do ciclo de vida de uma APT, que se repete após concluída.

Fonte: Advanced Persistent Threats: Learn the ABCs of APTs - Part A, SECUREWORKS.

A Kaspersky (2021b) ressalta cinco principais estágios de uma APT:

1. **Obter acesso ao sistema:** encontrar uma vulnerabilidade ou brecha na estrutura do sistema;
2. **Estabelecer um ponto de apoio:** por uma rede de *backdoors*, por exemplo;
3. **Aprofundar o acesso:** encontrar outras vulnerabilidades no intuito de aproveitar delas;
4. **Mover-se lateralmente:** com o ganho de acesso com direitos de administradores, o atacante pode se mover no sistema ou rede à vontade;
5. ***Look, Learn, and Remain*, ou Olhe, Aprenda e Permaneça:** uma vez no sistema, o entendimento do funcionamento dele é mais rápido e permite ao atacante para se retirar quando o objetivo for atingido, ou deixar um *backdoor* para futuros acessos.

Segundo a Secureworks (2016), os motivos que incentivam ataques APT variam muito e dependem do propósito do ataque. Isso significa que a categoria do ataque, o esforço despendido nele e sua importância, estão ligados ao tipo de alvo escolhido. Essas razões vão de coletas de informações, obtenção de ganhos financeiros ou uma vantagem competitiva, até instalar um ponto de apoio para acessos e controles posteriores. A Figura 4 apresenta uma classificação de ataques APT considerando alvo, motivo, sofisticação, importância e o objetivo do ataque por APT.

Figura 4 – Relação entre Sofisticação *versus* Prevalecimento, classificando por motivo e alvo do ataque APT.

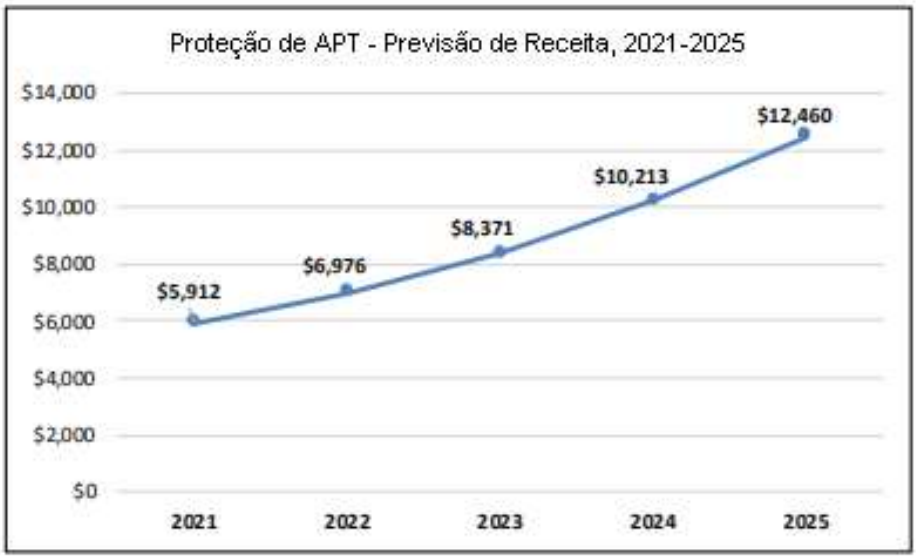
Fonte: Advanced Persistent Threats: Learn the ABCs of APTs - Part A, SECUREWORKS.

2.2.1 Investimento em defesa contra APT

Soluções de detecção de APT podem ser implantadas em vários formatos, incluindo *softwares* de detecção, ferramentas físicas ou virtuais, em nuvem ou modelos híbridos. A virtualização e implementação de soluções híbridas estão sendo mais utilizadas pelos fornecedores de segurança em APT. Algumas das principais empresas deste mercado são Bitdefender, Cisco, Kaspersky, McAfee, Symantec, VMware e FireEye (THERADICATIGROUP, 2021).

Constatou-se que a receita mundial das soluções de proteção de APT cresceu para mais de US$5,9 bilhões em 2021, e espera-se que aumente para mais de US$12,4 bilhões até 2025. A Figura 5 mostra esse crescimento.

Figura 5 – Predição do crescimento da receita mundial em proteção de APT.



Fonte: Advanced Persistent Threat (APT) Protection - Market – Quadrant 2021, THERADICATIGROUP.

2.3 *Exploits*

Um exploit é qualquer tipo de ataque que usa falhas de *software*, *hardware*, redes, sistemas operacionais ou outros recursos do sistema. Os *exploits* são geralmente códigos ou programas para obter o controle de computadores e/ou roubar arquivos (LATTO, 2020).

Exploits não são *malwares* em si, mas métodos para espalhar *malware*. Um computador não pode ser infectado por um *exploit*, no entanto, um *exploit* permite a entrada de *malwares*, tal qual uma porta (ZAMORA, 2017).

2.4 *Zero-day* / 0-day, ou dia zero

Um ataque *zero-day* consiste em um ataque virtual que acontece no mesmo dia em que uma vulnerabilidade do sistema é descoberta. É efetuada antes que o responsável pelo sistema providencie um *patch* ou a correção da falha (ZETTER, 2014).

Depois que uma correção é desenvolvida, a probabilidade do *exploit* ter sucesso diminui à medida que mais usuários aplicam a correção. Para *exploits* de dia zero, a menos que a vulnerabilidade seja corrigida inadvertidamente, como por uma atualização não relacionada que aconteça para corrigir a vulnerabilidade, a probabilidade de que um usuário tenha aplicado um *patch* fornecido pelo fornecedor que corrige o problema é baixa. Então o *exploit* permaneceria acessível, tornando-se um agravante para a ameaça apresentada por ataques *zero-day* (KASPERSKY, 2020).

2.4.1 *Trojan Horse*, ou Cavalo de Troia

Um Cavalo de Troia é um programa malicioso disfarçado de aplicação confiável. Esse malware consiste em códigos maliciosos injetados em *softwares* legítimos, para tentar obter acesso ao sistema dos usuários com seu software (FORTINET, 2022). Uma vez baixado, o código malicioso executa atividades projetadas pelo invasor, tais como obter acesso *backdoor* a sistemas corporativos, espionar a atividade online dos usuários ou roubar dados confidenciais. O *Trojan* apresenta um risco alto na segurança da rede, por poder ignorar todas as políticas de segurança de rede; um invasor pode obter acesso a uma máquina usando credenciais de rede armazenadas, comprometendo toda a rede (OWASP, 2022).

2.5 Vetor de ataque

Um vetor de ataque é um método ou cenário específico que pode ser explorado para invadir um sistema ou uma aplicação, comprometendo assim sua segurança. Um vetor de ataque pode ser explorado manual, automaticamente ou por meio de uma combinação de atividade manual e automática. Os ataques cibernéticos são lançados usando um vetor de ataque. Existem dois tipos principais de vetores de ataques:

1. Um ataque passivo ocorre quando um invasor monitora um sistema em busca de vulnerabilidades de modo a obter ou coletar informações sobre seu alvo. Os ataques passivos podem ser difíceis de detectar porque não envolvem alterar dados ou recursos do sistema. Além disso, ao invés de causar danos aos sistemas de uma organização, o invasor ameaça a confidencialidade de seus dados;

2. Um ataque ativo interrompe ou causa danos aos recursos do sistema de uma organização, afetando assim suas operações regulares. Os invasores podem lançar ataques contra vulnerabilidades do sistema, visando senhas fracas de usuários ou por meio de ataques de *malware* e *phishing*.

Um vetor de ataque comum em campanhas APT são os *ransomwares* (Seção 2.5.1).

2.5.1 *Ransomware*

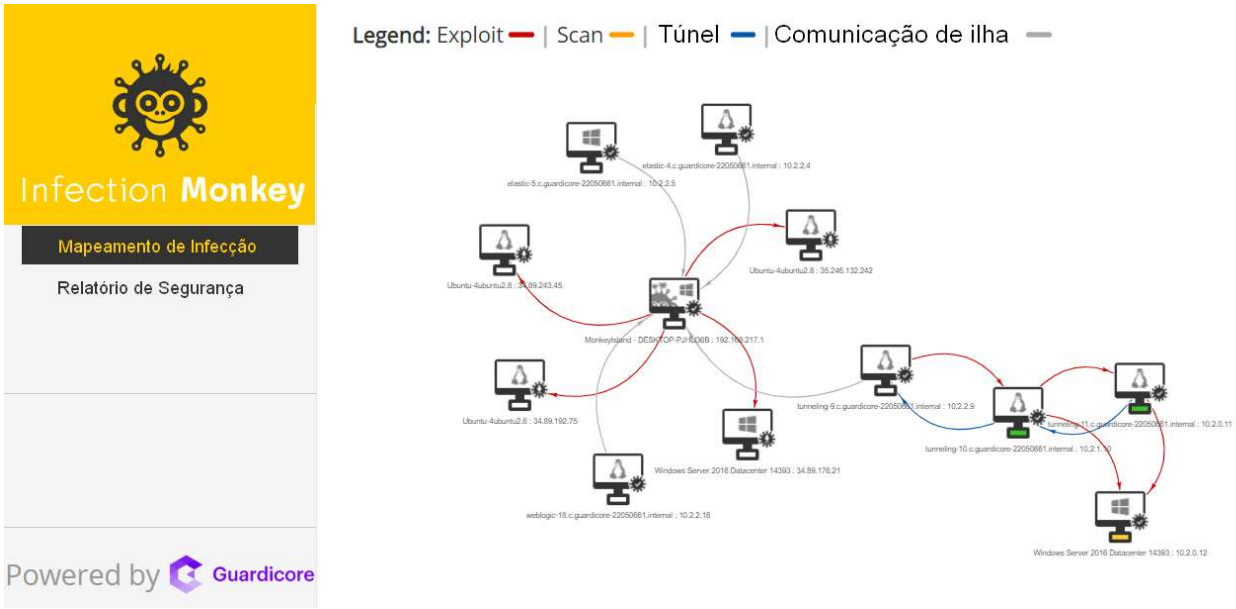
O *ransomware* é uma categoria de *malware* que criptografa arquivos em um sistema e exige pagamento pela capacidade de descriptografar esses arquivos. As vítimas recebem

um certo tempo, geralmente 72 horas, para pagar o resgate, que pode variar de US$100 a US$1.000 para pessoas físicas e muito maior para empresas e organizações.

Segundo a Mandiant (2022), o *ransomware* e a extorsão de dados estão entre as ameaças mais ativas e significativas que as organizações de todos os setores enfrentam atualmente. O impacto de uma implantação bem-sucedida de APT por *ransomware* inclui desafios técnicos e não técnicos, que podem prejudicar serviços e negócios. Além disso, os atacantes estão desenvolvendo técnicas avançadas que agora exigem estratégias abrangentes em mitigação de riscos.

A Figura 6 demonstra uma rede de computadores com seus respectivos sistemas operacionais e IP, onde o progresso de um ataque por *ransomware* é simulado.

Figura 6 – Mapeamento de infecção de máquina por *ransomware* em uma rede.



Fonte: Network Breach: Ransomware Simulation Infection Map, GUARDICOREINFECTIONMONKEY.

As principais consequências de um ataque por *ransomware* envolvem os seguintes itens, entre outros:

- Perda de Dados;
- Prejuízo Financeiro;
- Vazamento de Dados;
- Dano à Imagem.

No ataque de *ransomware* padrão, o invasor criptografa os recursos em risco e exige um resgate. Se o resgate for pago, o invasor liberará os recursos criptografados; caso contrário,

a vítima perderá os recursos indefinidamente. Esses ataques geralmente têm um único objetivo: tornar os recursos da vítima inacessíveis até que o resgate seja pago. Um *ransomware* pode ser categorizado como APT ou padrão, dependendo de sua natureza e nível de sofisticação, conforme a Figura 4 (BAKSI; UPADHYAYA, 2018). Embora o ganho monetário seja geralmente o objetivo principal de tais ataques, eles podem ter outros objetivos ocultos e/ou disfarçados.

A seguir, dois exemplos de *ransomwares* famosos foram escolhidos para exemplificar e basear este trabalho.

2.5.2 CryptoLocker

O ataque do *ransomware* CryptoLocker foi um ataque cibernético que ocorreu em setembro de 2013 e novamente em maio de 2014. O ataque utilizou um *Trojan* (Seção 2.4.1) que tinha como alvo computadores que executavam o Microsoft Windows. O ZDNET tentou avaliar os ganhos do ataque em dezembro de 2013; esses se elevaram a aproximadamente 41.928 Bitcoins (BTC), cerca de US$27 milhões na época. No Brasil, 304 máquinas foram infectadas, isto é, 4,8% do total do ataque (JARVIS, 2013).

No lançamento da sua campanha APT, o CryptoLocker oculta sua presença das vítimas até que tenha invadido com sucesso um servidor de comando e controle e criptografa os arquivos localizados nas unidades conectadas. Antes dessas ações, o *ransomware* garante sua permanência em execução nos sistemas infectados e persiste nas reinicializações das máquinas. Quando executado pela primeira vez, o *ransomware* cria uma cópia de si em `%AppData%` ou `%LocalAppData%` e exclui o arquivo executável original.

No primeiro ataque, o valor do resgate foi fixado em US$300 ou 2 BTC, apesar de variar levemente dependendo da moeda do país em que ocorreu a infecção. No final de 2013, o mercado de Bitcoin experimentou uma grande volatilidade e aumentou drasticamente de preço; a inflação do preço do Bitcoin nos últimos meses de 2013 levou os responsáveis pelo ataque a reduzir o resgate para 1 BTC, mais tarde 0,5 BTC e depois 0,3 BTC. Uma pesquisa da Dell SecureWorks (2013) relatou que os atacantes do CryptoLocker cumpriram o acordo e instruíram as vítimas que optaram por pagar o resgate a descriptografar os arquivos e retirar o *ransomware*.

O CryptoLocker, por ser caracterizado originalmente como um *ransomware* “barato” em relação ao valor do resgate, os invasores terem regulado esse valor em momentos de inflação e por devolverem os dados após o pagamento, faz com que o *ransomware* seja atribuído uma notoriedade positiva no cenário dos *ransomwares*.

2.5.3 WannaCry

O WannaCry é um *ransomware* de criptografia usado por cibercriminosos para extorquir dinheiro. Foi um ataque cibernético mundial em maio de 2017, visando computadores que executavam o sistema operacional Microsoft Windows, criptografando dados e exigindo pagamentos de resgate em Bitcoin. Estima-se que a campanha do *ransomware* infetou mais de

200.000 computadores em 150 países (MALWAREBYTES, 2021).

Os invasores exigiram primeiro US$300 em Bitcoins pela devolução, porém aumentaram o valor do resgate mais tarde para US$600 em Bitcoins. Se as vítimas não pagassem o resgate dentro de três dias, os responsáveis pelo ataque do WannaCry ameaçaram excluir os arquivos permanentemente.

Devido ao aumento do resgate imposto pelos responsáveis pelo WannaCry, além da ameaça à integridade e disponibilidade dos dados, este se tornou um exemplo de *ransomware* de notabilidade desfavorável, apesar de sua importância no avanço dos *ransomwares*.

2.6 Teoria de Jogos

O estudo do uso de modelos matemáticos para avaliar sistemas interativos é conhecido como Teoria de Jogos, definido como a análise das interações de tomadores de decisão ou atores racionais independentes (ROUSE, 2018).

À medida que as tecnologias se tornam capazes de armazenar um conjunto de regras e extrapolar resultados cognitivos, a Teoria de Jogos pode determinar como elas funcionam juntas ou competem em um determinado sistema (EXPERIENCE, 2021).

Na segurança cibernética, a Teoria de Jogos é usada para observar a natureza de incidentes cibernéticos em que defensores, invasores e usuários da rede interagem entre si e produzem um resultado. A Teoria de Jogos torna-se útil ao modelar o comportamento e as estratégias de cada jogador e captura as interações dos jogadores adversários (THREATPOST, 2020).

Como mencionado anteriormente, os participantes envolvidos no jogo de *ransomware* são o atacante e a vítima. Considera-se o jogador definido por $N = \{a, v\}$, onde a representa o Atacante e v representa a Vítima.

2.6.1 O atacante

Consideram-se diversos fatores que influenciam a tomada de decisão do atacante, entre os quais:

1. o valor ou importância dos arquivos criptografados, até o momento desconhecido;
2. a dificuldade de recuperar os arquivos criptografados sem pagamento pelo resgate, seja por *backup* ou decriptamento dos arquivos;
3. a probabilidade de receber a chave correta do atacante para descriptografar os arquivos com sucesso.

Enquanto os dois primeiros itens dependem da vítima, o item 3 é uma característica do atacante, em que decorrem três novas situações:

1. O atacante visa conseguir o pagamento da vítima. Após o resgate, ele entrega as chaves à vítima para recuperar os dados;
2. O atacante visa lucrar com crimes cibernéticos, e possivelmente não entregará as chaves, mesmo após o pagamento pelo resgate;
3. O atacante planeja lucrar com crimes cibernéticos e, além de não devolver o acesso aos dados encriptados, tem a intenção de vender essas informações.

Dados esses três itens, é possível destacar três tipos diferentes de ataques por *ransomware*. Considera-se primeiro o ataque de *ransomware* clássico, que pede apenas o pagamento do resgate e devolve os arquivos após receber o pagamento. Considera-se em seguida o ataque de *ransomware* com ameaça de dados, que pode vazar os arquivos com ou sem pagamento do resgate. Considera-se, por fim, o ataque de *ransomware* com venda de dados, que venderá os dados caso não haja pagamento do resgate.

2.6.2 A vítima

Retomando os dois primeiros fatores influenciadores na decisão do atacante na Seção 2.6.1, assume-se como desconhecido o valor dos arquivos da vítima. Assim, qualquer arquivo pode ser atribuído com qualquer valor monetário. Como arquivos têm diferentes valores monetários, considera-se o valor monetário dos arquivos de uma vítima como $[0, \infty)$, onde 0 significa que os arquivos não são importantes para o proprietário, e ∞ significa que os arquivos são indispensáveis e não substituíveis.

Além disso, uma vez que os arquivos são bloqueados e o pagamento do resgate é exigido, a vítima do *ransomware* clássico reavalia o valor monetário dos arquivos. O proprietário deve avaliar seus arquivos de forma consistente, pois serve de informações para realizar o modelo do jogo. Lembrando que o atacante não tem conhecimento da avaliação exata dos arquivos de dados da vítima. Esta informação restrita à vítima se refere ao conjunto de dados da vítima, e denota-se o conjunto de dados da vítima por $V = [0, \infty)$.

2.6.3 Modelo e Representação do Jogo

O jogo é uma matriz em que são representados os jogadores, estratégias e recompensas. Na interação de forças opostas, sejam indivíduos ou equipes, cada lado está interessado em vencer ou evitar perder.

Por exemplo, podemos considerar a matriz do jogo de pedra, papel e tesoura, conforme ilustrado na Figura 7 a seguir, em que dois jogadores estão usando as mesmas estratégias, ambos com chances iguais de ganhar. A recompensa, ou *payoff*, é dada por:

- 1, uma vitória;
- 0, um empate;
- -1, uma derrota.

Figura 7 – Matriz de Jogo de Estratégia de Pedra, Papel, Tesoura.

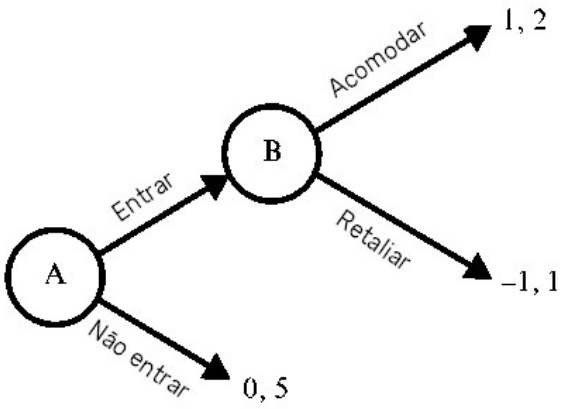
		J2		
		Pedra	Papel	Tesoura
J1	Pedra	0 0	1 -1	-1 1
	Papel	-1 1	0 0	1 -1
	Tesoura	1 -1	-1 1	0 0

Fonte: Fighting Cyber Attacks with Game Theory, THREATPOST.

Os jogadores agem simultaneamente e sem conhecimento da ação dos outros jogadores. Se os jogadores têm informações sobre as escolhas dos outros, o jogo passa a ser apresentado em forma extensiva, explicado adiante.

Um jogo na forma extensiva considera a ordem das ações importante. Os jogos são apresentados como diagramas de árvores, conforme na Figura 8. Lê-se da esquerda para a direita, ou seja, da raiz aos galhos, e cada vértice representa um ponto de tomada de decisão para um jogador. As recompensas são especificadas na parte inferior da árvore. Neste exemplo, apresenta-se um cenário de *Entry Game*, ou Jogo de Entrada, que consiste em um participante potencial escolher se deseja entrar em um mercado controlado por um monopólio. Se o participante entrar, o monopolista pode iniciar uma guerra de preços ou compartilhar. Neste caso ilustrado, Jogador A deve decidir primeiro se entrará ou não na partida. Se Jogador A entrar, Jogador B verá o movimento do Jogador 1 e então terá a opção de escolher entre acomodar ou retaliar a ação do Jogador A.

Figura 8 – Representação de Forma Extensiva de um *Entry Game*, ou Jogo de Entrada.



Fonte: *The Entry Game* Game, MCCAIN.

2.6.4 Categorização do Jogo

A Teoria dos Jogos emprega um conjunto de modelos matemáticos e ferramentas analíticas projetadas para descrever e analisar os fenômenos observados na interação entre dois ou mais tomadores de decisão. A descrição formal da interação entre os jogadores é denotada como um jogo de conflito e cooperação entre eles.

A Teoria dos Jogos parte do princípio de que os jogadores são considerados racionais e inteligente, ou seja, tomam decisões consistentes em busca de seu próprio propósito e tentam maximizar seus resultados, e levam em consideração decisões de outros jogadores. A solução de um jogo é a descrição das estratégias que cada jogador tem que seguir para alcançar o melhor resultado possível.

No jogo, os participantes escolhem entre suas táticas disponíveis para maximizar o valor de seu pagamento pessoal (jogos competitivos) ou o retorno de ambas as coalizões (jogos cooperativos). O resultado é julgado em uma escala de utilidade única para cada participante. Durante um turno de jogo, cada lado empregará uma estratégia baseada nas informações à sua disposição.

Os jogos são divididos em diversas categorias, baseado no tipo de interação:

Cooperativos e não cooperativos: Em jogos cooperativos, todos os jogadores tentam maximizar a recompensa geral. Já em jogos não cooperativos, cada jogador se preocupa apenas com seu próprio ganho e custo. Em caso de não cooperação, os jogos são subdivididos em jogos estáticos e dinâmicos;

- Em jogos estáticos, todos os jogadores tomam suas decisões simultaneamente sem conhecer as estratégias de outros jogadores. Cada jogada representa uma estratégia a partir da qual deve-se escolher o melhor lance para maximizar sua recompensa, ou *payoff* (lucro-custo). Este reflete a desejabilidade de um possível resultado de um jogador, expressando sua utilidade, e se refere ao ganho líquido de um jogador quando ele escolhe jogar uma estratégia;
- Em jogos dinâmicos, um jogador pode alterar sua jogada durante o jogo. O jogo é aplicado em etapas onde o jogador tem que escolher sua jogada. A estratégia nesses jogos é definida como a combinação de movimentos sequenciais escolhidos pelo jogador para maximizar sua recompensa total. Cada etapa de um jogo dinâmico pode ser considerado um jogo estático, levando assim a uma estrutura de jogos estáticos sequenciais;

Jogos de soma-zero e diferente de zero: Nos jogos de soma-zero, o benefício total para todos os jogadores, para cada combinação de estratégias, sempre soma zero. Um jogador só lucra com base no prejuízo de outro e o vencedor recebe exatamente a soma das perdas de seus oponentes. Já nos jogos de soma diferente de zero, o ganho de um dos jogadores não necessariamente corresponde à perda dos outros, portanto algumas saídas têm resultados combinados maiores ou menores que zero;

Jogos de informações perfeitas e imperfeitas: Um jogo imperfeito acontece quando os jogadores escolhem suas estratégias simultaneamente, sem conhecer as escolhas dos outros jogadores. Em jogos perfeitos, cada jogador sabe exatamente as estratégias que outros jogadores seguiram antes da sua vez. Portanto, apenas jogos onde os jogadores jogam em sequência podem ser considerados jogos perfeitos;

Jogos de informações completas e incompletas: Os de informações completas indicam que os jogadores conhecem as estratégias disponíveis e as recompensas dos outros jogadores, mas não necessariamente conhecem suas estratégias. Já em jogos de informações incompletas, os jogadores não podem ter acesso às estratégias e recompensas dos demais durante o jogo;

Estratégias puras e mistas: As estratégias puras referem-se às ações determinísticas tomadas por um jogador no jogo para todas as situações possíveis criadas pelos adversários que se pode enfrentar. Em estratégias mistas, o movimento de um jogador não é baseado em uma ação determinística, mas envolve uma combinação probabilística das estratégias puras disponíveis.

2.7 Teorema de Bayes

O teorema de Bayes descreve a probabilidade de um evento, baseado em um pré-conhecimento que pode estar relacionado a esse evento. O teorema mostra como alterar as probabilidades tendo em vista novas evidências.

2.7.1 Definição Formal

A regra de Bayes é um corolário da lei da probabilidade total, expressa matematicamente como na equação 2.1, onde A e B são eventos e $P(B) \neq 0$:

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} \tag{2.1}$$

O teorema de Bayes também pode ser escrito da seguinte maneira:

$$P(A \mid B)\, P(B) = P(A \cap B) = P(B \cap A) = P(B \mid A)\, P(A) = P(A \mid B)\, P(B), \tag{2.2}$$

onde:

$P(A)$ e $P(B)$ são as probabilidades *a priori* de A e B;

$P(A|B)$ é a probabilidade *a posteriori* de A condicionada a B;

$P(B|A)$ é a probabilidade *a posteriori* de B condicionada a A.

2.7.2 Interpretação Bayesiana

A interpretação Bayesiana diz que a probabilidade mede o "grau de crença". O teorema de Bayes aplica o grau de crença em uma posição antes e após se considerar as evidências.

Se, por exemplo, acreditar–se a princípio com 50% de certeza de que uma moeda tem duas vezes maior probabilidade de cair em cara do que coroa e se a moeda for lançada várias vezes seguidas, o grau de crença pode aumentar, diminuir ou permanecer igual, dependendo dos resultados observados. Para a proposição A e a evidência B, tem-se que:

- $P(A)$, probabilidade anterior de A, é o grau de crença inicial em A;
- $P(A|B)$, probabilidade posterior de A condicionada a B, é o grau de crença após afirmar que B é verdadeiro;
- O quociente $\frac{P(B \mid A)}{P(B)}$ representa o suporte que a evidência B fornece para a proposição A.

Nesta pesquisa adotou-se o grau de crença como "reputação" do *ransomware* e formaram-se dois cenários:

1. O *ransomware* com má reputação;
2. O *ransomware* com ótima reputação.

3 Trabalhos Correlatos

Vários estudos foram realizados nos últimos anos sobre a Teoria dos Jogos em pagamentos de *ransomware*. Spyridopoulos (2013) descobriu um equilíbrio de estratégias combinando possíveis custos de mitigação com o custo de um ataque bem-sucedido. Este estudo analisou o equilíbrio de diferentes técnicas de defesa usando modelos inspirados em propagação de epidemia para a disseminação de *malware* em rede. O jogo é construído em um modelo unificado de proliferação, incluindo fatores tais como taxas de infecção, imunização e desinfecção que orientam a estratégia dos jogadores. As recompensas foram então determinadas para um conjunto de estratégias de acordo com os parâmetros controlados pelo invasor e defensor (taxas de infecção, correção e remoção). Esse trabalho definiu uma estratégia ideal para o defensor em resposta a um ataque por *ransomware*.

O trabalho de Caporusso (2018) utilizou um modelo de negociação mostrado como um jogo extensivo. Este artigo trata do *ransomware* em circunstâncias em que a renegociação do resgate é viável, um fenômeno bastante difundido visto com certos grupos de ataque por *ransomware*, segundo Monroe (2021). Caporusso, ao fazer o reconhecimento da dinâmica pós-ataque entre a vítima humana e o atacante humano, observa que existem fatores humanos significativos fora da negociação do resgate a serem considerados no processo de tomada de decisão.

Cartwright (2019) examina em seu trabalho se o resgate deveria ou não ser pago. Seu estudo aborda o problema de *ransomware* como um sequestro seguido de resgate. Cartwright se baseou em dois modelos de sequestro: os modelos de Selten (2013) e Lapan e Sandler (1988). Enquanto o primeiro propõe um valor ótimo de resgate a ser escolhido pelo atacante, otimizando suas chances de receber sua recompensa, o segundo contribui com medidas para evitar o sequestro dos dados em primeiro lugar.

4 Procedimentos Metodológicos

Segundo as classificações de pesquisa apresentadas por Wazlawick (2014), o trabalho realizado se classifica como pesquisa de análise e modelagem, no ponto de vista dos procedimentos técnicos. Seus objetivos são descritivos, com abordagem de análise qualitativa, uma vez que busca relacionar conceitos e a identificação dos resultados não é numérica e exata, e sim valorativa.

Neste presente trabalho foi feito o estudo do jogo baseado no modelo de Teoria de Jogos, sob um conjunto de diferentes fatores que podem influenciar a tomada de decisão tanto do atacante, quanto do defensor. Após analisar o comportamento esperado de ambos lados, estimou-se com base no custo-benefício do vetor de ataque a decisão de pagar ou não em um ataque por *ransomware* bem-sucedido.

Propõe-se que, ao sempre devolver os arquivos, a reputação do *ransomware* é aprimorada e a "confiança" da vítima é aumentada na convicção de que, ao pagar o resgate, recuperará seus dados roubados. Logo, sugere-se que o peso da reputação afete o modelo de confiança e pode influenciar positiva ou negativamente a decisão de pagar, ou não, o resgate.

Criou-se um jogo sequencial para dois jogadores em que o atacante é um deles, enquanto a vítima, ou defensor, é o outro. Foram determinadas as estratégias para o atacante e o defensor para diversas condições de jogos, além de analisar se os valores dos parâmetros de decisão afetam as escolhas de ambos os jogadores.

Uma vez que o ataque ocorreu, partindo do princípio que foi um ataque de *ransomware* bem-sucedido, a vítima tem duas opções: a primeira é pagar o resgate e esperar que o invasor libere a chave de descriptografia, enquanto a segunda opção é se recusar a pagar o resgate e assumir o prejuízo.

Essas decisões podem ser tomadas pela vítima dependendo das circunstâncias. Nesta seção, são examinadas duas condições que ajudariam a vítima do ataque a decidir sobre o pagamento de resgate e descriptografia dos recursos criptografados mantidos para resgate. De acordo com a hipótese inicial, a disposição da vítima em pagar o resgate é determinada principalmente por dois fatores: o valor dos recursos recuperados sob o cerco e a reputação do atacante.

A principal contribuição deste trabalho é a apresentação de parâmetros para quantificar a importância do valor dos recursos sob ataque, bem como a introdução de um novo parâmetro para melhor compreensão da reputação do atacante.

A metologia utilizada para realizar a pesquisa foi organizada da seguinte forma:

1. Classificação do jogo;
2. Representação dos jogadores;

3. Cálculo da matriz de recompensa;
4. Construção da base do jogo;
5. Aplicação da inferência de Bayes pelo grau de crença.

Na segurança de rede, o estudo do *ransomware* na vida real se encaixa na categoria de jogos não cooperativos, uma vez que não há cooperação entre o atacante e o defensor. O atacante e o defensor são descritos como duas partes em um jogo de *ransomware*: o primeiro começa o jogo infectando com sucesso um *host*. Isso resulta instantaneamente em um pedido de resgate à vítima, ou defensor, com a promessa de descriptografar ou liberar os dados mediante pagamento. As vítimas então escolherão se pagam ou não a taxa, apesar do perigo de perder seus dados. Em algum momento, os invasores liberarão ou destruirão os arquivos.

A partir da premissa da Teoria de Jogos, o atacante é um jogador racional, logo só liberará a chave de descriptografia se um pagamento de resgate for recebido. Se não receber o resgate, não liberará a chave de descriptografia.

No entanto, em alguns casos, o invasor pode optar por fazer o oposto. O atacante escolhe não ser racional dessa maneira, ou seja, atuar fora da sequência lógica de opções estabelecidas no modelo. As razões para a falta de racionalidade do invasor podem ser inúmeras, mas isso está além do escopo deste trabalho. Como o atacante pode ser racional ou irracional, sua reputação pode desempenhar um papel importante para que o defensor, ou vítima nesta representação, tome uma decisão quando estiver sob ataque.

Quando o atacante age racionalmente, sua reputação mediante a todos os ataques relatados e relacionados a ele melhora. Se agirem irracionalmente, incorre em uma penalidade e sua reputação decai, resultando em uma menor disposição por parte do defensor em pagar o resgate.

4.0.1 Características do Jogo

Segundo as classificações tratadas na Seção 2.6.4, pode-se considerar que uma situação de *ransomware* é competitiva, pois tanto o invasor quanto a vítima estão preocupados com seus resultados individuais e como suas escolhas podem afetar a distribuição de compensação dentro de sua própria coalizão. No caso de jogos recorrentes e informações semi-imperfeitas, os invasores têm a opção de executar uma técnica não cooperativa para reforçar a credibilidade de sua ameaça (ou seja, destruir os dados se o resgate não for pago) e influenciar as vítimas. Nesse sentido, os invasores podem usar um método cooperativo (liberando os arquivos se o resgate for pago) para desenvolver confiança.

O jogo é de informações incompletas, uma vez que os jogadores não sabem sobre as estratégias dos outros. Por exemplo, a vítima, ao decidir se pagará ou não o resgate, não tem certeza do que o invasor fará após receber o dinheiro; assim como ao criptografar os arquivos da vítima, o atacante não sabe o quanto a vítima valoriza seus arquivos, ou se essa possui *backup* dos mesmos.

4.1 Modelo do Jogo

Nesta seção, é proposto um jogo Bayesiano não cooperativo com informações incompletas. As partidas apresentadas neste trabalho não são de soma zero, pois o ganho do Atacante não vem da perda da Vítima. O modelo consiste nas partes a seguir:

1. o conjunto de jogadores envolvidos no jogo;
2. a categorização dos jogadores;
3. o conjunto de ações possíveis dos jogadores.

4.2 Matriz de recompensa

O jogo de *ransomware* é sequencial, composto por vários turnos envolvendo as ações do invasor e a vítima. A linha do tempo do jogo é estabelecida da seguinte maneira:

- 1º turno: O invasor lança um ataque de *ransomware* bem-sucedido em uma vítima, marcando o início do jogo. A vítima perde o acesso aos arquivos e o atacante exige um pagamento de resgate R;
- 2º turno: após receber a informação R, a vítima decide se paga ou não o resgate. Esta etapa é o ponto de tomada de decisão da vítima sobre o pagamento do resgate;
- 3º turno: ao obter a decisão da vítima sobre o pagamento do resgate, o atacante escolhe se quer devolver os arquivos à vítima. Esta etapa é o ponto de tomada de decisão sobre a devolução de dados;
- 4º turno: O atacando determina se venderá os dados roubado ou não. Esta etapa é o ponto de tomada de decisão sobre a venda de dados.

Seja p a escolha da vítima de pagar o resgate no 2º turno.

$$p = \begin{cases} 0, \textit{Não pagar o resgate,} \\ 1, \textit{Pagar o resgate.} \end{cases} \tag{4.1}$$

Seja r a escolha do atacante de devolver os dados no 3º turno.

$$r = \begin{cases} 0, \textit{Não devolver os dados,} \\ 1, \textit{Devolver os dados.} \end{cases} \tag{4.2}$$

Seja s a escolha do atacante de vender os dados no 4º turno.

$$s = \begin{cases} 0, \textit{Não vender os dados,} \\ 1, \textit{Vender os dados.} \end{cases} \tag{4.3}$$

Seja A_i o lucro do atacante se este vender os dados, em que $C_d > 0$ representa o custo de transação e $D_i \geq 0$ é o valor de mercado dos dados roubados da vítima i.

$$A_i = \begin{cases} D_i - C_d, & \textit{se } D_i \geq C_d, \\ 0, & \textit{se } D_i < C_d. \end{cases} \tag{4.4}$$

Seja uma vítima i, a recompensa π esperada pelo atacante será definida como:

$$\pi = pR - rC_r + sA_i \tag{4.5}$$

onde $C_r > 0$ é o custo de devolução de dados para a vítima.

A recompensa da vítima i é então:

$$u_i = -pR - (1 - r)V_{r,i} - sL_{d,i} \tag{4.6}$$

onde $V_{r,i} \geq 0$ é o valor dos dados ainda bloqueados para a vítima, e $L_{d,i} \geq 0$ é a perda da vítima se os dados roubados forem vendidos pelo invasor.

A Figura 9 demonstra a matriz de recompensa em um cenário de *ransomware* "clássico" em que p e r são as variáveis binárias de decisão no jogo. Este cenário possui, portanto, quatro resultados possíveis.

Figura 9 – Matriz de Recompensa do jogo com *ransomware* "clássico".

Resultado		Atacante (π)	Vítima (u)
$p = 0$	$r = 0$	0	$-V_{r,i}$
$p = 0$	$r = 1$	$-C_r$	0
$p = 1$	$r = 0$	R	$-R - V_{r,i}$
$p = 1$	$r = 1$	$R - C_r$	$-R$

Fonte: Pela autora.

Nota-se que não foi considerado o custo dos dados da vítima, uma vez que o modelo do jogo considera apenas fatores que influenciam o ataque. Esse custo, no ponto de vista da vítima, consta na manutenção de seus dados, fora da situação de um ataque por *ransomware*.

Com a possibilidade de venda de dados, surge um novo fator que influencia a matriz de recompensa de ambos jogadores. O ataque pode variar em sua lucratividade para o atacante, na decisão de pagamento do resgate, na manutenção da confidencialidade dos arquivos roubados, entre outros.

Para isso, compara-se o *ransomware* com venda de dados com o ransomware "clássico", demonstrando a distinção entre os resultados e recompensas de jogos.

Desta forma, s é uma variável de estratégia binária envolvida no jogo do *ransomware* com venda de dados, logo com oito resultados possíveis, conforme à Figura 10.

Figura 10 – Matriz de Recompensa do jogo com *ransomware* com venda de dados.

Resultado			Atacante (π)	Vítima (u)
$p = 0$	$r = 0$	$s = 0$	0	$-Vr,i$
$p = 0$	$r = 0$	$s = 1$	$Di - Cd$	$-Vr,i - Ld,i$
$p = 0$	$r = 1$	$s = 0$	$-Cr$	0
$p = 0$	$r = 1$	$s = 1$	$Di - Cd - Cr$	$-Ld,i$
$p = 1$	$r = 0$	$s = 0$	R	$-R - Vr,i$
$p = 1$	$r = 0$	$s = 1$	$R + Di - Cd$	$- R - Vr,i - Ld,i$
$p = 1$	$r = 1$	$s = 0$	$R - Cr$	$-R$
$p = 1$	$r = 1$	$s = 1$	$R + Di - Cd - Cr$	$-R - Ld,i$

Fonte: Pela autora.

Sabendo que a vítima do ataque está em desvantagem no jogo *ransomware*, vale ressaltar que os objetivos tanto do atacante quanto da vítima são maximizar suas recompensas, que dependem dos resultados do jogo.

Com base nas Figuras 9 e 10, o melhor resultado para a vítima é uma recompensa zero. Essa situação ocorre se o atacante devolver os dados gratuitamente e não vender os arquivos roubados. Em todas as outras situações, a vítima recebe uma recompensa negativa.

5 Desenvolvimento e Resultados

Com o impacto de campanhas APT, principalmente utilizando vetores de ataque por *ransomware*, faz-se necessário a análise de ações envolvidas pelos atores de modo a construir e definir uma estratégia, e reduzir os danos causados.

Nesta seção, serão apresentados os antecedentes e suposições para especificar os ataques de *ransomware* que serão analisados. Será então analisado um jogo baseado em modelos teóricos em três tipos de reputação variável do atacante. Foi comparado o lucro do *ransomware* de venda de dados com o do *ransomware* clássico em cada modelo.

Pode-se usar o modelo de jogo Bayesiano. Os jogadores nesse modelo podem ser de vários tipos, e pelo menos um jogador tem informações incompletas sobre o tipo de outros jogadores no jogo. Também assume-se que a Vítima tem informações limitadas sobre o tipo de Atacante.

Vale ressaltar que, enquanto muitas vítimas que optam por não pagar o resgate acabam perdendo seus arquivos, pesquisas indicam que as vítimas que pagam nem sempre conseguem recuperar seus arquivos. Inclusive, de acordo com estudos recentes, aproximadamente 62% das vítimas que pagaram o resgate recuperaram seus arquivos (CYBEREDGE-GROUP, 2020).

5.1 Análise do jogo sem grau de crença

A análise do modelo de base trata um jogo sem aplicação do grau de crença, em que a tomada de decisão do atacante nas etapas 3 e 4 (Seção 4.2) são independentes. Considera-se que o atacante somente vende os dados se o valor de mercado dos dados for superior ao custo de transação, isto é:

$$s = \begin{cases} 1, & \textit{se } D_i \geq C_d, \\ 0, & \textit{se } D_i < C_d. \end{cases} \tag{5.1}$$

Quando o valor de mercado dos dados roubados excede o custo de transação, o invasor sempre os vende. O atacante recebe um ganho de $A_i = D_i - C_d$ das vítimas cujos dados valem mais do que o custo de transação. Por outro lado, o atacante recebe uma recompensa $A_i = 0$ das vítimas cujos valores de dados são menores que o custo de transação.

A Figura 9 mostra que, independentemente do pagamento do resgate, não devolver os dados às vítimas é sempre a estratégia prevalecente do atacante. Por consequência, se o fator de reputação não for considerado, o atacante não tem motivos para devolver os dados.

O resultado do jogo de um *ransomware* clássico é $p = 0$, $r = 0$, onde o lucro do atacante é 0 e a recompensa da vítima i é $-V_{r,i}$. Se a reputação do invasor não for levada

em consideração, o *ransomware* com venda de dados é mais lucrativo do que o *ransomware* clássico. Os dois são equivalentes apenas se não for possível vender os dados.

O lucro total recebido pelo atacante é então:

$$\Pi_b = \sum_{n=1}^{N} A_i \tag{5.2}$$

Esse resultado indica que, mesmo que o pagamento do resgate seja zero, ou seja, a vítima não pagou o resgate ($p = 0$), o atacante ainda pode lucrar caso o valor de mercado dos dados roubados seja maior que o custo de venda dos dados.

Esta é uma grande vantagem dos ataques do *ransomware* com venda de dados sobre o *ransomware* clássico. Vale destacar que medidas defensivas como *backup* de dados e a decisão de não pagar o resgate, que funcionariam para mitigar ataques de *ransomware* clássico, podem não funcionar no *ransomware* com venda de dados.

Sem pagamento de resgate, o lucro sobre o *ransomware* clássico é zero, porém o lucro do *ransomware* com venda de dados pode ser positivo. No entanto, isso não implica necessariamente que o *ransomware* com venda de dados seja sempre mais lucrativo do que o *ransomware* clássico.

5.1.1 Utilidade do grau de crença

Envolvendo o grau de crença, verificou-se que a vítima tem mais disposição a pagar se tiver certeza que o atacante devolverá os arquivos após o pagamento, isto é, com uma boa reputação. A vítima se beneficia do pagamento do resgate se o valor do mesmo for inferior ao valor dos dados bloqueados, isto é, se $R \leq V_r$. Já se $R \geq C_r$, ou seja, o resgate é maior que o custo de devolução dos dados, o atacante se beneficia ao devolver os dados e não vendê-los, apesar de que, uma condição não impede a outra.

Com a possibilidade de que o valor dos arquivos da vítima exceda o custo de devolução pelo atacante, existe uma faixa de resgate $R \in [C_r, V_r]$ que pode ser mutuamente benéfica. Entretanto, esta situação necessita cooperação entre ambas as partes, além de que a vítima deve confiar no atacante. Esse cenário implica um jogo do tipo cooperativo, em que os jogadores tentam maximizar a recompensa geral, tornando o *ransomware* se tornar um modelo de negócios viável para ambas partes, contradizendo o conceito do modelo da Teoria de Jogos.

5.2 Análise do jogo cooperativo com grau de crença crescente

Se o resultado de um jogo influencia as decisões da vítima, pode-se dizer que a reputação é importante no desenvolvimento do jogo. É plausível dizer que quanto pior a reputação, isto é, se o atacante não devolver os dados após o pagamento dos dados, maior o receio da vítima em pagar o resgate.

Já no modelo de jogo com reputação ótima, isto é, supondo que o atacante sempre devolva os dados e não os venda após o pagamento do resgate, têm-se as seguintes condições:

- $r = 1$ e $s = 0$ se o resgate for pago;
- $r = 0$ e $s = 1$ se o resgate não for pago e $D_i \geq C_d$.

Isto significa que o atacante teria a reputação de seguir o "acordo" com a vítima e, em reação à decisão desta, a estratégia do atacante será de devolver o acesso aos dados e mantê-los confidenciais caso o resgate seja pago, ou não devolver e vender caso o resgate não seja pago.

Logo, tem-se a vítima i que define:

$$p = \begin{cases} 0, & \textit{se } R > V_{r,i} + L_{d,i}, \\ 1, & \textit{se } R \leq V_{r,i} + L_{d,i}. \end{cases} \tag{5.3}$$

Se a vítima tiver certeza que receberá os arquivos de volta, sua disposição em pagar o resgate exigido se dá a partir de $V_{r,i} + L_{d,i}$. Portanto, ao pagar o resgate, evita-se a perda dos dados.

Com $p = 1$, o lucro do atacante Pi_t é:

$$\Pi_t = n(R - C_r) + \sum_{i=n+1}^{N} A_i \tag{5.4}$$

Verifica-se que no jogo com reputação ótima, a disposição das vítimas em pagar o resgate é limitada por $V_{r,i}$ e o lucro do atacante é $n(R - C_r)$. Lembrando que no modelo base sem fator reputação, é possível que a disposição das vítimas em pagar o resgate seja zerada, pode-se dizer que construir uma reputação positiva pode aumentar a disposição das vítimas em pagar o resgate.

Para o *ransomware* com venda de dados, quando a vítima decide não pagar o resgate, o lucro do invasor aumenta para o *ransomware* de venda de dados, enquanto o *ransomware* clássico não tem lucro. Baseado nesta lógica, se o invasor tiver uma reputação ótima, o *ransomware* de venda de dados é mais lucrativo do que o *ransomware* clássico.

5.3 Análise do jogo competitivo com grau de crença decrescente

Nesta seção, caracteriza-se um jogo competitivo, onde os jogadores tentam maximizar suas próprias recompensas e não cooperam entre si. Pode-se analisar a influência da reputação negativa no ataque de *ransomware* com venda de dados.

Partindo do princípio que a disposição da vítima em pagar o resgate é determinada pela reputação do atacante, baseado na análise da seção anterior, estima-se que ocorre uma avaliação do histórico do atacante em relação à decisão de devolver o acesso aos arquivos e mantê-los confidenciais; quanto maior esse "hábito", maior o grau de crença.

Seja $\beta_r \in [0, 1]$ a probabilidade do atacante devolver os dados e $\beta_d \in [0, 1]$ a probabilidade de manter os dados roubados confidenciais.

A recompensa u_B da vítima considerando essas variáveis é:

$$u_B = -pR - V_r(1 - p\beta_r) - L_d(1 - p\beta_d) \tag{5.5}$$

A vítima, se não pagar o resgate, recebe uma recompensa de $-V_r - L_d$, logo $p = 0$, $\beta_r = 0$ e $\beta_d = 0$, uma vez que as probabilidades do atacante devolver e não vender os dados são zeradas. Por outro lado, ao pagar o resgate, a vítima recebe $-R - V_r(1 - \beta_r) - L_d(1 - \beta_d)$, logo $p = 1$. Neste último caso, ainda não se garante a decisão do atacante, pois depende da sua reputação.

Como se trata de um jogo competitivo, espera-se que a vítima escolherá pagar o resgate se a recompensa esperada for mais alta, ou seja, $p = 1$ se $-R - V_r(1 - \beta_r) - L_d(1 - \beta_d) \geq -V_r - L_d$.

Nesta linha de raciocínio, pode-se supor que no jogo competitivo, a vítima escolherá pagar o resgate se $R \leq \beta_r V_r + \beta_d L_d$, ou seja, se o valor do resgate for menor ou igual à soma do valor dos dados da vítima e da perda se forem vendidos. Logo, atendendo esta condição, a vítima teria mais disposição a pagar o resgate.

No jogo sem reputação, tem-se $\beta_r = \beta_d = 0$ e no jogo de reputação ótima tem-se $\beta_r = \beta_d = 1$. O lucro π_B do atacante é então:

$$\pi_B = pR - p\beta_r C_r + A_i(1 - p\beta_d) \tag{5.6}$$

Se a vítima de um *ransomware* com venda de dados não pagar o resgate, o atacante receberá um lucro de A_i e $R - \beta_r C_r + A_i(1 - \beta_d)$ se o resgate for pago. Já se o ganho de resgate esperado não for maior do que o lucro da venda dos dados, o atacante opta por vendê-los, logo tem-se que $\beta_d = 0$ se $\beta_d L_d \leq A_i$, e $\beta_d = 1$, no caso contrário.

Se por um lado, no modelo de jogo de base, é melhor que o atacante não retorne os dados, visto que maximiza sua recompensa, no modelo do jogo cooperativo e reputação ótima, o atacante deve sempre devolver os arquivos após receber o pagamento do resgate. Quando o jogo é competitivo e a reputação do atacante é falha, verifica-se que pode não ser ideal a decisão de sempre devolver os arquivos com pagamento de resgate, nem nunca devolvê-los. Esta análise será demonstrada no Capítulo 5.

Baseado nesta última análise, não está claro qual estratégia é a mais lucrativa. O atacante tem a possibilidade de duas recompensas:

1. Sempre devolver os aquivos tem como consequência a construção de uma reputação ótima, e aumenta a disposição da vítima a pagar;
2. Nunca devolver os arquivos é uma vantagem exclusiva do *ransomware* com venda de dados, uma vez que além do valor do resgate existe o lucro sobre a venda de dados.

Em seguida, pode-se dizer que no jogo competitivo, o atacante deve devolver arquivos com pagamento de resgate se $\beta_r V_r \geq C_r$, ou seja, quando o custo para devolver os arquivos for menor ou igual ao lucro na devolução dos mesmos.

Seja R_t o valor de resgate no jogo de reputação ótima, R_u no jogo competitivo de reputação falha, o lucro Π é:

$$\Pi_u - \Pi_t = n[C_r(1-\beta_r) - (R_t - R_u)] + \sum_{i=1}^{n} A_i(1-\beta_d) \tag{5.7}$$

Com base nesta equação, o jogo competitivo com reputação falha, o atacante pode receber o lucro do resgate e não devolver os dados, representado por $C_r(1-\beta_r)$, além da venda dos mesmos, representado por $A_i(1-\beta_d)$.

Desta forma, a venda de dados não apenas diminui o grau de crença no desenvolvimento do jogo competitivo, mas também diminui a disposição da vítima a pagar o resgate.

5.4 Resultados

Considera-se o lucro de *ransomwares* em que n é o número de vítimas dispostas a pagar o resgate. Nota-se que o valor de n pode variar em função da campanha APT lançada, como o tipo de alvo e outros parâmetros de campanha que não foram considerados nesta pesquisa. Vale ressaltar que o lucro do *ransomware* com ameaça de dados é igual ao do *ransomware* clássico, pois ambos ganham apenas o valor do resgate.

5.4.1 Lucro esperado de *ransomwares*

O modelo exposto trata do cenário no qual os dados bloqueados são usados de modo a forçar o pagamento do resgate ao invés de serem vendidos. A Figura 11 mostra como o lucro esperado do *ransomware* varia com a probabilidade de haver vazamento de dados, observando valores diferentes de probabilidade de devolução dos arquivos.

Figura 11 – Comparação de lucros entre o *ransomware* clássico e com venda de dados, variando o grau de crença do *ransomware*

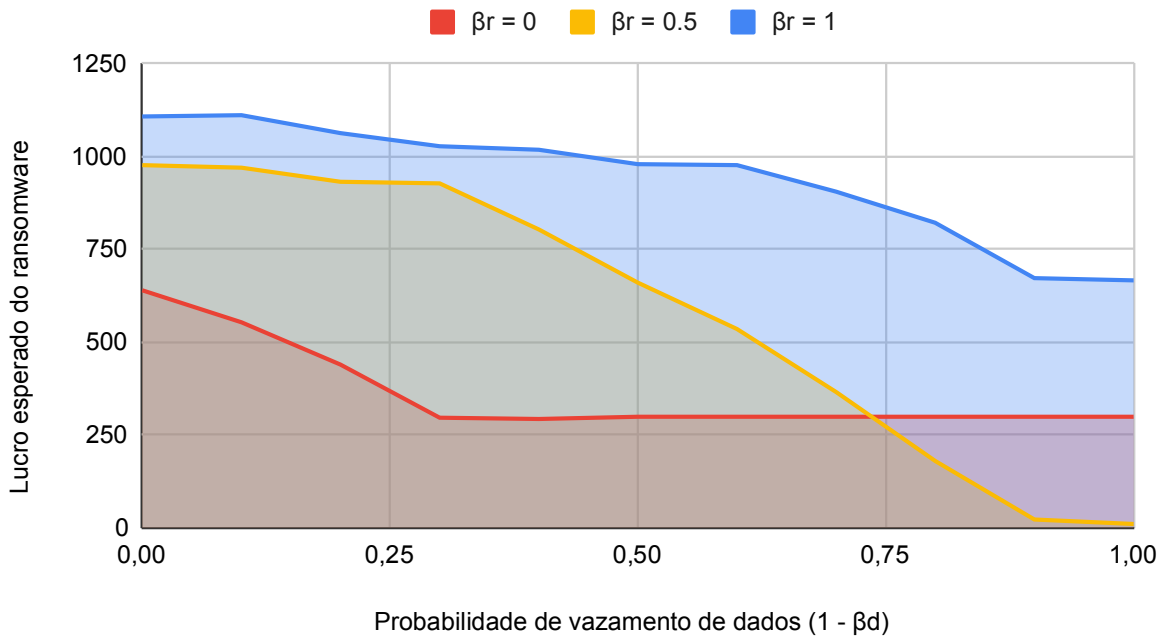
grau de crença	ransomware clássico	ransomware com venda de dados
$\beta r = \beta d = 0$	0	ΣAi
$\beta r = \beta d = 1$	$n(R - Cr)$	$n(R - Cr) + \Sigma Ai$
$0 < \beta r < \beta d < 1$	$n(R - \beta rCr)$	$n(R - \beta rCr) + \Sigma(1 - \beta d)Ai + \Sigma Ai$

Fonte: Pela autora.

Como foi verificado no capítulo anterior, independentemente da decisão do atacante em devolver os arquivos ou não, o lucro do *ransomware* com venda de dados é maximizado em $1 - \beta_d = 0$, ou seja, mantendo a confidencialidade dos dados. Isto significa que os dados são sempre mantidos em sigilo, e diminui à medida que β_d diminui. Esse resultado corresponde à hipótese levantada no Capítulo 4.

Ademais, se o atacante sempre devolver os dados para a vítima quando o resgate é pago, logo tem-se $r = 1$, a progressão ultrapassa consistentemente em $r = 0$ e $r = 0,5$, em que, respectivamente, os lucros esperados caem para zero quando a probabilidade de vazamento de dados está maximizada, tal como na Figura 12. No entanto, o lucro esperado em $\beta_r = 1$, isto é, a probabilidade de devolução dos dados é máxima, torna-se constante, porém nunca chega a zero, mesmo com taxa de vazamento de dados muito alta.

Figura 12 – Rentabilidade do *ransomware* com ameaça de dados versus Probabilidade de vazamento de dados, após o pagamento do resgate.



Fonte: Pela autora.

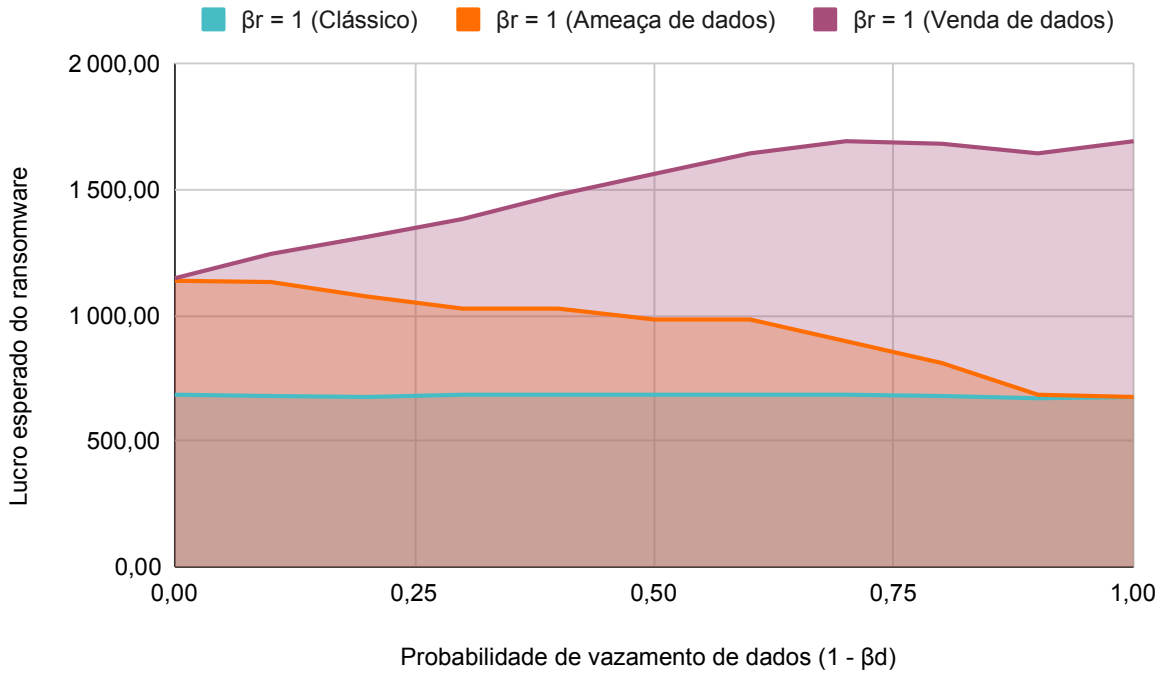
Como o ganho do atacante vem apenas do pagamento do resgate, a estratégia ideal deste é maximizar a disposição da vítima através do grau de crença, com simultaneamente a maximização da probabilidade de vazamento de dados. Assim dizendo, a ameaça efetiva de vazamento dos dados depende da credibilidade da ameaça. O atacante consegue maximizar seu lucro ao manter um padrão de reputação ótima, isto é, sem vazamento de dados, não apenas para aumentar a disposição de pagar da vítima, mas também para construir um histórico para ataques futuros. Em outras palavras, o invasor do *ransomware* com ameaça de dados não recebe recompensa entre o lucro sobre o pagamento do resgate e da venda de dados, uma vez que ele tem a opção de vazar os dados ou manter o sigilo.

Nota-se que embora o jogo de *ransomware* com ameaça de dados seja menos incerto do que o jogo de *ransomware* com venda de dados, ele não é tão lucrativo.

A Figura 13 compara os lucros do *ransomware* clássico, com ameaça de dados e com venda de dados em diferentes taxas de vazamento de dados, propondo a probabilidade de devolução dos arquivos em $\beta_r = 1$, ou seja, com reputação ótima. O lucro do *ransomware*

clássico se mantém constante, conforme o gráfico, apesar da reputação ótima do atacante. Além disso, embora os outros tipos de *ransomwares* estejam mais lucrativos do que o clássico, observa-se que o *ransomware* com venda de dados permanece mais rentável do que o *ransomware* com ameaça de dados.

Figura 13 – *Ransomware* clássico versus *ransomware* com ameaça de dados versus *ransomware* com venda de dados, com a devolução de dados garantida.



Fonte: Pela autora.

Pelos gráficos anteriores, constata-se que independentemente da probabilidade de devolução de dados pelo atacante, o *ransomware* com venda de dados é mais lucrativo do que o *ransomware* com ameaça de dados, e este último é mais rentável do que o *ransomware* clássico.

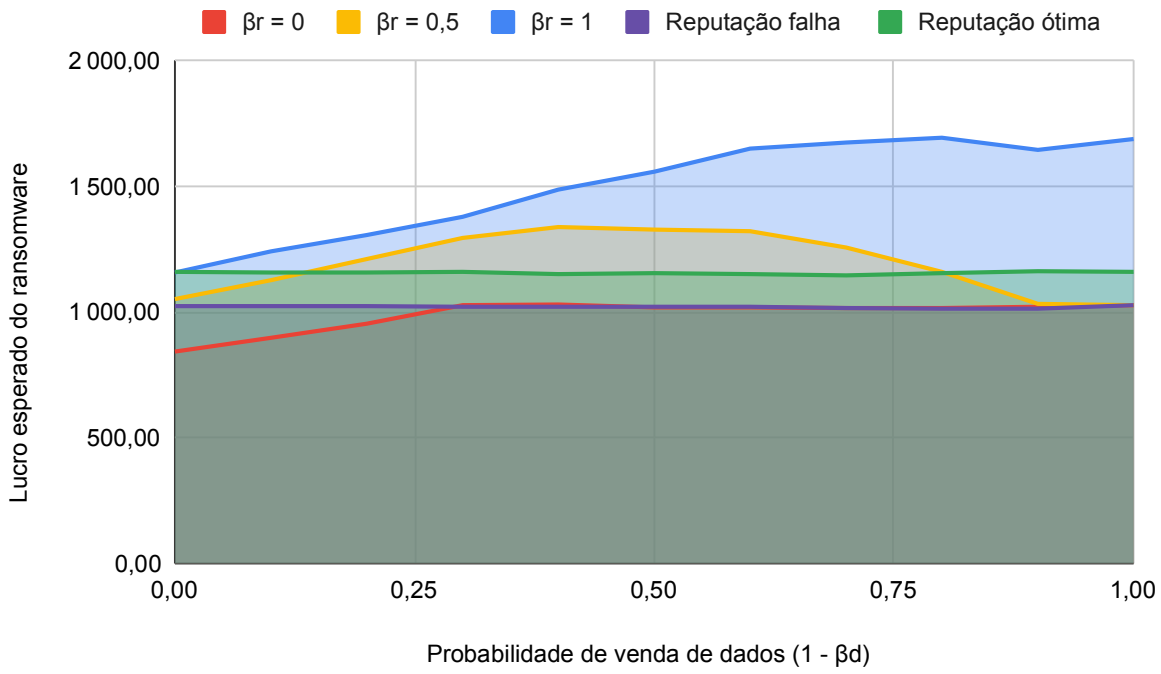
5.4.2 Lucro esperado com reputação falha e reputação ótima

Nos parâmetros mencionados no capítulo anterior, os valores dos dados V_r e da perda sofrida pela vítima L_d foram gerados aleatoriamente. Observa-se que o lucro do *ransomware* com venda de dados com reputação falha $(\beta_r = \beta_d = 0)$, obtido com a venda dos dados, é superior ao lucro produzido pelo *ransomware* clássico.

Já com reputação ótima $(\beta_r = \beta_d = 1)$, a vítima decide pagar o resgate se $R \leq V_r$ para *ransomware* clássico, isto é, quando o valor do resgate é inferior ao valor dos dados. Para o *ransomware* com venda de dados, a vítima escolhe pagar o resgate se o resgate for inferior à soma do valor dos dados e da perda da vítima se os arquivos forem vendidos, ou seja, se $R \leq V_r + L_d$.

A Figura 14 demonstra que uma baixa probabilidade de devolução de arquivos e uma alta probabilidade de venda de dados reduzem a disposição da vítima em pagar o resgate. À medida que a probabilidade de vender dados aumenta, a chance de pagamento diminui, mas o lucro sobre a venda dos dados aumenta. As mudanças relacionadas ao lucro esperado do *ransomware* decorrem da relação entre o ganho na venda de dados e o lucro do *ransomware*.

Figura 14 – Rentabilidade do *ransomware* com venda de dados versus Probabilidade de venda de dados, variando expectativa de devolução de arquivos.



Fonte: Pela autora.

Os resultados da modelagem mostram que o atacante que vende os dados recebe um ganho adicional sobre o clássico, em caso de não pagamento do resgate. Em vista disso, o *ransomware* com venda de dados é mais rentável do que o clássico, tanto em caso reputação falha, quanto em caso de reputação ótima.

Com reputação falha, a disposição da vítima a pagar está limitada a $\beta_r V_r + \beta_d L_d$, ou seja, a soma do lucro sobre o valor dos dados e da sua perda em caso de venda deles. Dada a avaliação das vítimas dos arquivos bloqueados e dos dados roubados, a escolha do atacante entre devolver (β_r) e vender dados $(1 - \beta_d)$ determina a progressão do grau de crença, logo a disposição das vítimas em futuros ataques.

O atacante recebe sua recompensa ao definir β_r e β_d, que resultará entre o lucro sobre o pagamento do resgate e o sobre a venda de dados. Desta forma, ao estabelecer uma maior probabilidade de devolver os dados arquivos e mantê-los confidenciais, o atacante poderá ganhar com o aumento dos pagamentos de resgate, todavia perderá com o ganho proporcionado pela venda de dados.

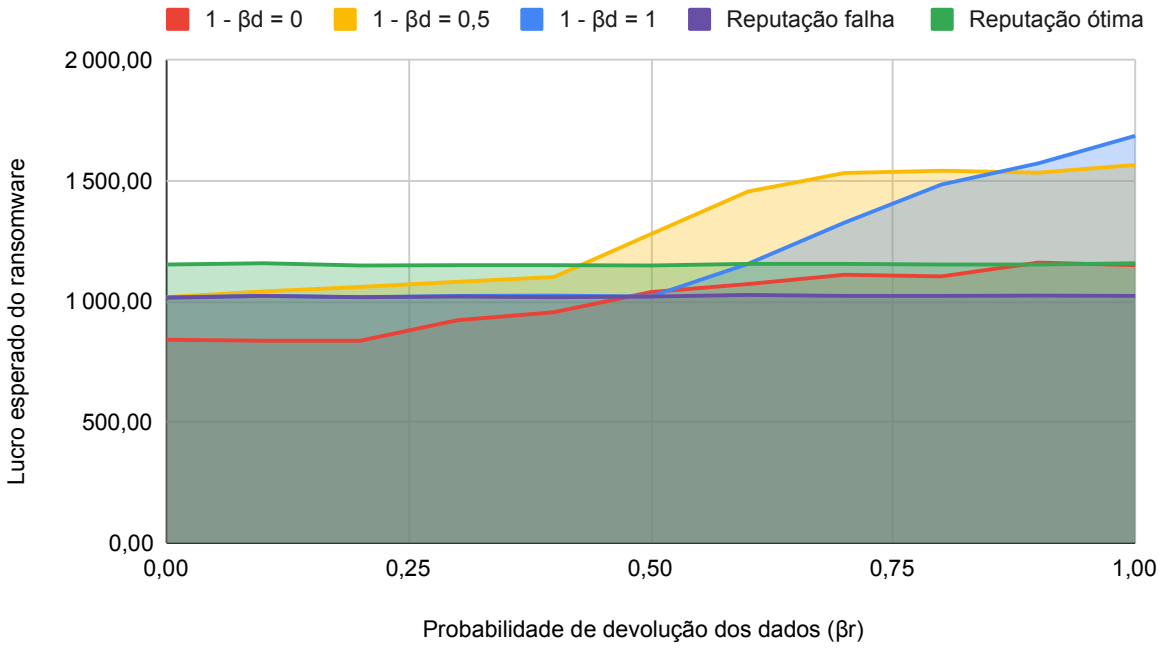
Verifica-se também que vender dados a uma taxa de 0,5 supera as situações de nenhuma venda ou de venda garantida, tanto com reputação ótima quanto falha. Por consequente, analisa-se que a melhor estratégia de *ransomware* é uma mescla de devolução de arquivos e venda de dados.

5.4.3 Influência da probabilidade de devolução de dados sobre o grau de crença

Por fim, examinou-se como a probabilidade de devolução de dados afeta o lucro do *ransomware*, variando o fator de probabilidade de venda de dados. Conforme à Figura 15, o lucro do *ransomware* com venda de dados se altera à medida que a probabilidade de devolver os arquivos varia, dada uma determinada probabilidade de venda de dados $(1 - \beta_d)$.

Se, de modo geral, a lucratividade do *ransomware* com venda de dados aumenta à medida que a probabilidade de devolução dos arquivos aumenta; pode-se dizer que o aumento do grau de crença resulta em mais disposição das vítimas em pagar o resgate.

Figura 15 – Rentabilidade do *ransomware* com venda de dados versus Probabilidade de devolução dos dados.



Fonte: Pela autora.

Esses resultados são confirmados pela recompensa do atacante ao definir β_r e β_d. Aumentar a probabilidade da devolução dos arquivos aumenta a disposição da vítima em pagar o resgate, resultando em mais lucros sobre resgates, porém uma perda no lucro sobre os dados.

Vale realçar que a probabilidade de vender os dados da vítima que não pagou o resgate é de 1, contudo a probabilidade de vender os da vítima que pagou é definida por $(1 - \beta_d)$; isto significa que o lucro dos dados diminui à medida que mais vítimas pagam o resgate. Em-

bora tenha oscilações, o *ransomware* com venda de dados é mais lucrativo quando o invasor aumenta a probabilidade de devolução de arquivos, mantendo a probabilidade de venda de dados.

Com base nos resultados desta pesquisa, conclui-se que tanto nos jogos de reputação falha, quanto nos modelos de reputação ótima, o *ransomware* com venda de dados é sempre mais lucrativo do que o *ransomware* clássico. Mesmo assim, é uma estratégia ideal para ataques de *ransomware* clássico construir uma reputação ótima, sempre devolvendo dados, a fim de maximizar sua recompensa.

Contudo, estruturar uma reputação ótima nem sempre é lucrativo para o *ransomware* com venda de dados, pois o atacante deve escolher entre ganhar com o resgate e ganhar com a venda de dados. As Figuras 14 e 15 mostram que o lucro esperado do *ransomware* no caso da reputação falha não é determinístico, o que implica que a estratégia ideal do atacante é uma estratégia mesclada com combinações equilibradas das variáveis β_r e β_d, segundo a avaliação da vítima em relação aos dados roubados.

6 Conclusão

Neste trabalho, foi desenvolvido um modelo de Teoria de Jogos focado em aspectos da interação entre atacantes cibernéticos de *ransomware* e organizações visadas, com foco nos aspectos financeiros do negócio. Além de calcular a decisão estratégica de pagar o resgate ou não, aplicou-se o fator de reputação, relacionado e também influenciado pelo comportamento tanto do atacante quanto da vítima. Verificou-se que, dependendo do tipo de ataque por *ransomware*, quanto melhor a reputação de um invasor, maior é a disposição da vítima a pagar o resgate.

Utilizou-se a Teoria de Jogos para analisar um cenário de defesa contra ataques envolvendo um *ransomware* clássico e uma vítima. Adicionou-se um parâmetro para auxiliar a mesma a tomar uma decisão informada quando estiver sob ataque. O parâmetro de reputação é a relação entre o valor dos recursos recuperados após o pagamento do resgate e o valor total dos bens da vítima. Isso permite que a vítima avalie o valor dos recursos bloqueados. Desse modo, quanto maior a disposição do da vítima em pagar o resgate para desbloquear os dados, maior o valor da reputação.

Foram analisados três tipos de *ransomware*, sendo com um deles possível de dar lucro financeiro na venda dos dados roubados. Os modelos teóricos de jogo desenvolvidos examinaram as possíveis táticas do atacante e da vítima, aplicadas em diversos cenários, considerando o peso de reputação da parte atacante.

Pode-se dizer que neste cenário, o grau de crença das vítimas é moldado pelas ações passadas do atacante. Nesse caso, o atacante sempre devolverá os arquivos, devido a dois fatores distintos:

- Devolver os arquivos leva à melhoria da reputação do atacante, até a reputação ótima;
- O atacante pode se beneficiar dessa boa reputação a longo prazo.

O grau de crença torna-se, portanto, um incentivo para devolução de arquivos. O método de construção de reputação do invasor, isto é, sempre liberando os dados e mantendo-os confidenciais quando o resgate é pago, diminui a incerteza do *ransomware* de ameaça de dados, tornando-o menos lucrativo do que o *ransomware* de venda de dados. Na maioria das análises, as modelagens indicam que o *ransomware* de venda de dados é mais lucrativo financeiramente do que o *ransomware* clássico.

Baseado nas análises anteriores, esse modelo não pode ser sustentado a longo prazo, o que implica que o *ransomware* com venda de dados não tem futuro viável. A curto prazo, algumas vítimas pagam o resgate na “esperança” de recuperar o acesso aos seus arquivos. Porém, quanto mais se souber que os arquivos não são devolvidos e sim vendidos, menos confiança haverá. Isto significa um grau de crença baixo, logo uma reputação falha.

Contudo, perceber as possíveis recompensas financeiras dependem muito da comercialização dos dados roubados, bem como o medo do vazamento de dados afeta a disposição das vítimas de pagar o resgate. Nesse sentido, o *ransomware* de venda de dados representa um perigo maior tanto para o invasor quanto para a vítima. Ter uma boa reputação beneficia ambas partes, mas ter uma reputação perfeita nem sempre é lucrativo para o atacante com *ransomware* de venda de dados. A análise mostra que esse invasor pode se beneficiar ao desbloquear, vender dados seletivamente e manipular as vítimas de seus ataques.

É provável que a viabilidade a longo prazo do *ransomware* dependa de sua reputação.

6.1 Considerações finais

Esta pesquisa fornece uma análise da lucratividade e estratégia dos ataques APT por *ransomware*, entretanto possui algumas limitações. O modelo de Teoria de Jogos usado na pesquisa é simplificado e pode não refletir com precisão a realidade. Além disso, a pesquisa se baseia em dados estatísticos e pode não representar os ataques de ransomware atuais.

6.2 Trabalhos futuros

O *ransomware* continua sendo um problema significativo no mundo, e a análise realizada demonstra haver efetivamente um incentivo infinito para usar o *ransomware*. Como o custo é baixo e as recompensas potenciais são altas, os atores motivados financeiramente são incentivados a seguir essa linha de ataque. Além disso, é provável que as vítimas de ataques bem-sucedidos paguem por vários motivos, incluindo a capacidade de interpretar o resgate como uma despesa comercial.

6.2.1 Aumento do valor de resgate

O *ransomware* clássico pode incluir alguns recursos extras. Uma característica importante é o prazo de pagamento antecipado do resgate. Após esse prazo inicial, o resgate exigido é frequentemente dobrado. Significa que se o defensor desejar pagar o resgate, mas demorar a reunir o valor exigido, deverá pagar o dobro após o prazo inicial.

6.2.2 Possibilidade de negociação

Outra característica do jogo que não foi considerada pode ser a presença de uma fase de negociação entre o atacante e o defensor. Se o valor do resgate e/ou o valor dos recursos mudarem após o processo de entendimento, o defensor pode atualizar o valor dos parâmetros no jogo.

Referências

BAKSI, R. P.; UPADHYAYA, S. J. A comprehensive model for elucidating advanced persistent threats (apt). In: SECURITY, P. of the International Conference on; (SAM), M. (Ed.). The Steering Committee of The World Congress in Computer Science, 2018. Disponível em: <https://www.proquest.com/openview/ced5b969cf8baa0a46ad092a71ef4562/1?pq-origsite=gscholar&cbl=1976342>. Acesso em: 19 jul. 2022. Citado nas páginas 14 e 24.

BAKSI, R. P.; UPADHYAYA, S. J. Decepticon: a theoretical framework to counter advanced persistent threats. In: NATURE, S. (Ed.). Information Systems Frontiers, 2020. Disponível em: <https://doi.org/10.1007/s10796-020-10087-4>. Acesso em: 18 jul. 2022. Citado na página 15.

BECKER'SHEALTHIT. First known ransomware attack in 1989 also targeted healthcare. In: BECKER'S HEALTH IT. 2016. Disponível em: <https://www.beckershospitalreview.com/healthcare-information-technology/first-known-ransomware-attack-in-1989-also-targeted-healthcare.html>. Acesso em: 9 jul. 2022. Citado na página 14.

BSIBRASIL. Iso/iec 27001 - segurança da informação. In: THE BRITISH STANDARDS INSTITUTION. BSI Brasil, 2021. Disponível em: <https://www.bsigroup.com/pt-BR/ISO-IEC-27001-Seguranca-da-Informacao/>. Acesso em: 6 jul. 2021. Citado na página 18.

CBSNEWS. Wannacry ransomware attack losses could reach $4 billion. In: BERR, J. (Ed.). Money Watch, 2017. Disponível em: <https://www.cbsnews.com/news/wannacry-ransomware-attacks-wannacry-virus-losses/>. Acesso em: 21 jun. 2022. Citado na página 14.

CHECKPOINT. Behind the curtains of the ransomware economy. In: CHECK POINT RESEARCH. 2022. Disponível em: <https://research.checkpoint.com/2022/behind-the-curtains-of-the-ransomware-economy-the-victims-and-the-cybercriminals>. Acesso em: 4 mai. 2022. Citado na página 16.

CHUNG, K.; KAMHOUA, C. A.; KWIAT, K. A. Game theory with learning for cybersecurity monitoring. In: IEEE. [S.l.]: 2016 IEEE 17th International Symposium on High Assurance Systems Engineering (HASE), 2016. p. 1–8. Citado na página 12.

CISCO. What is an advanced persistent threat (apt)? In: CISCO. Cisco, 2021. Disponível em: <https://www.cisco.com/c/en/us/products/security/advanced-persistent-threat.html>. Acesso em: 15 mar. 2021. Citado na página 15.

CYBEREDGEGROUP. 2017cyberthreat defense report. In: . LLC, 2017. Disponível em: <https://cyber-edge.com/wp-content/uploads/2021/02/CyberEdge-2020-CDR-Report-v1.0.pdf>. Acesso em: 11 jul. 2022. Citado na página 12.

CYBEREDGEGROUP. 2020 cyberthreat defense report. In: . LLC, 2020. Disponível em: <https://cyber-edge.com/wp-content/uploads/2021/02/CyberEdge-2020-CDR-Report-v1.0.pdf>. Acesso em: 11 jul. 2022. Citado nas páginas 12 e 37.

CYBEREDGEGROUP. 2021 cyberthreat defense report. In: . LLC, 2021. Disponível em: <https://cyber-edge.com/wp-content/uploads/2021/04/CyberEdge-2021-CDR-Report-v1.1-1.pdf>. Acesso em: 11 jul. 2022. Citado na página 12.

CYBEREDGEGROUP. 2022 cyberthreat defense report. In: . LLC, 2022. Disponível em: <https://cyber-edge.com/cdr/>. Acesso em: 11 jul. 2022. Citado na página 12.

DFNDR. Relatório da segurança digital no brasil,terceiro trimestre - 2018. In: DFNDR. dfndr lab, 2018. Disponível em: <https://www.psafe.com/dfndr-lab/pt-br/relatorio-da-seguranca-digital>. Acesso em: 17 mar. 2021. Citado na página 12.

EXPERIENCE, A. Game theory explained. In: AMERICAN EXPERIENCE PUBLIC BROADCASTING SERVICE. Public Broadcasting Service, 2021. Disponível em: <https://www.pbs.org/wgbh/americanexperience/features/nash-game/>. Acesso em: 15 jul. 2022. Citado na página 25.

FIREEYE. Advanced persistent threat groups (apt groups) | fireeye. In: FIREEYE. FireEye, 2016. Disponível em: <https://www.fireeye.com/current-threats/apt-groups.html>. Acesso em: 8 abr. 2021. Citado na página 18.

FONTES, E. L. G. Segurança da informação. In: . [S.l.]: Saraiva Educação SA, 2017. Citado na página 12.

FORCE, R. T. The ransomware task force report combating ransomware: A comprehensive framework for action. In: . Institute for Security+Technology, 2021. Disponível em: <https://securityandtechnology.org/wp-content/uploads/2021/09/IST-Ransomware-Task-Force-Report.pdf>. Acesso em: 21 jun. 2022. Citado na página 14.

FORTINET. Trojan horse virus. In: . Fortinet, 2022. Disponível em: <https://www.fortinet.com/resources/cyberglossary/trojan-horse-virus>. Acesso em: 10 out. 2022. Citado na página 22.

GORE, R.; PADILLA, J.; DIALLO, S. Markov chain modeling of cyber threats. In: . [S.l.]: The Journal of Defense Modeling and Simulation, 2017. v. 14, n. 3, p. 233–244. Citado na página 14.

GUARDICOREINFECTIONMONKEY. Infection monkey documentation hub. In: GUARDICORE INFECTION MONKEY. Guardicore Infection Monkey, 2021. Disponível em: <https://www.guardicore.com/infectionmonkey/docs>. Acesso em: 3 out. 2021. Citado na página 23.

ISO. Iso/iec 27001 — information security management. In: . [s.n.], 2013. Disponível em: <https://www.iso.org/isoiec-27001-information-security.html>. Acesso em: 20 mar. 2021. Citado na página 18.

JARVIS, K. Cryptolocker ransomware. In: . DELL SecureWorks, 2013. Disponível em: <https://www.secureworks.com/research/cryptolocker-ransomware>. Acesso em: 18 jul. 2022. Citado nas páginas 14 e 24.

KASPERSKY. O que é um ataque de dia zero? – definição e explicação. In: KASPERSKY. Kaspersky, 2020. Disponível em: <https://www.kaspersky.com.br/resource-center/definitions/zero-day-exploit>. Acesso em: 14 jul. 2022. Citado na página 22.

KASPERSKY. Consumer appetite versus action: the state of data privacy amid growing digital dependency. In: KASPERSKY. Kaspersky Consumer IT Security Risks Report 2021, 2021. Disponível em: <https://media.kasperskydaily.com/wp-content/uploads/sites/92/2021/03/16090300/consumer-appetite-versus-action-report.pdf>. Acesso em: 15 jun. 2022. Citado na página 15.

KASPERSKY. What is an advanced persistent threat (apt)? In: KASPERSKY. Kaspersky, 2021. Disponível em: <https://www.kaspersky.com/resource-center/definitions/advanced-persistent-threats>. Acesso em: 12 out. 2021. Citado nas páginas 15 e 19.

LATTO, N. Exploits: What you need to know. In: AVAST. 2020. Disponível em: <https://www.avast.com/c-exploits>. Acesso em: 25 set. 2021. Citado na página 21.

LIU, W.; TANAKA, H.; KANTA, M. Empirical-analysis methodology for information-security investment and its application to reliable survey of japanese firms. In: . [S.l.]: IPSJ Digital Courier, 2007. p. 585–599. Citado na página 12.

MALWAREBYTES. Wannacry. In: MALWAREBYTES. Malwarebytes, 2021. Disponível em: <https://www.malwarebytes.com/wannacry>. Acesso em: 15 jul. 2022. Citado na página 25.

MANDIANT. Ransomware. In: . Mandiant, 2022. Disponível em: <https://www.mandiant.com/node/2246>. Acesso em: 16 jul. 2022. Citado na página 23.

MCCAIN, R. A. Game theory and public policy. In: . [s.n.], 2010. p. 10–11. Disponível em: <https://books.google.com.br/books?id=d_gBXgutHkcC&lpg=PA3&ots=9CjxmZ3hnV&dq=game%20theory%20and%20public%20policy&lr&hl=pt-BR&pg=PA11#v=onepage&q=game%20theory%20and%20public%20policy&f=false>. Citado na página 27.

MILOSEVIC, J.; SKLAVOS, N.; KOUTSIKOU, K. Malware in iot software and hardware. In: WORKSHOP ON TRUSTWORTHY MANUFACTURING AND UTILIZATION OF SECURE DEVICES. 2016. Disponível em: <https://www.researchgate.net/publication/317011595_Malware_in_IoT_Software_and_Hardware>. Acesso em: 16 jun. 2022. Citado na página 16.

MOHURLE, S.; PATIL, M. A brief study of wannacry threat: Ransomware attack 2017. In: . [s.n.], 2017. v. 8, n. 5, p. 1938–1940. Disponível em: <https://sbgsmedia.in/2018/05/10/2261f190e292ad93d6887198d7050dec.pdf>. Acesso em: 15 jun. 2022. Citado na página 14.

MOUTINHO, L. Ransomware: pagar resgate não é a melhor opção em ataques hackers. mas como evitá-los? In: GRUPO SUNO. 2021. Disponível em: <https://www.suno.com.br/noticias/ransomware-ataque-hacker-lojas-renner-lren3-jbs-jbss3/>. Acesso em: 15 jun. 2022. Citado na página 14.

NATIONALSECURITYSTRATEGY. Strategy for operating in cyberspace. In: DEPARTMENT OF DEFENSE, UNITED STATES OF AMERICA. [S.l.], 2011. p. 1–19. Citado na página 12.

NIST. Glossary | csrc. In: NIST. 2021. Disponível em: <https://csrc.nist.gov/glossary>. Acesso em: 15 set. 2021. Citado na página 18.

OWASP. Trojan horse. In: . The OWASP® Foundation, 2022. Disponível em: <https://owasp.org/www-community/attacks/Trojan_Horse>. Acesso em: 12 out. 2022. Citado na página 22.

ROUSE, M. Game theory. In: TECHNOPEDIA. TechDictionary, 2018. Disponível em: <https://www.techopedia.com/definition/32765/game-theory#:~:text=Game%20theory%20is%20the%20study,rational%20decision%2Dmakers%20or%20actors>. Acesso em: 15 jul. 2022. Citado na página 25.

SANKARDAS, R. et al. A survey of game theory as applied to network security. In: IEEEXPLORE. [S.l.]: Hawaii International Conference on System Sciences (HICSS), 2010. p. 1–10. Acesso em: 15 jun. 2022. Citado na página 16.

SECUREWORKS. Advanced persistent threats - learn the abcs of apt: Part a. In: SECUREWORKS. Secureworks, 2016. Disponível em: <https://www.secureworks.com/blog/advanced-persistent-threats-apt-a>. Acesso em: 17 mar. 2021. Citado nas páginas 15, 19 e 20.

SECURITYSCORECARD. What is a cybersecurity vulnerability? In: SECURITY SCORECARD. 2021. Disponível em: <https://securityscorecard.com/blog/what-is-a-cybersecurity-vulnerability>. Acesso em: 12 out. 2021. Citado na página 18.

SHEYNER, O.; WING, J. Tools for generating and analyzing attack graphs. In: INTERNATIONAL SYMPOSIUM ON FORMAL METHODS FOR COMPONENTS AND OBJECTS. 2003. p. 344–371. Disponível em: <https://www.researchgate.net/publication/221047658_Tools_for_Generating_and_Analyzing_Attack_Graphs>. Acesso em: 15 jul. 2022. Citado na página 16.

SIMOIU, C. et al. "i was told to buy a software or lose my computer. i ignored it": A study of ransomware. In: *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. Santa Clara, CA: USENIX Association, 2019. p. 155–174. ISBN 978-1-939133-05-2. Disponível em: <https://www.usenix.org/conference/soups2019/presentation/simoiu>. Citado na página 12.

SINGER, P. W.; FRIEDMAN, A. Cybersecurity: What everyone needs to know. In: . [S.l.: s.n.], 2014. Citado na página 12.

SJOUWERMAN, S. Ibm study: 70% of businesses attacked pay ransomware. In: KNOWBE4 (Ed.). KnowBe4 Security, 2016. Disponível em: <https://blog.knowbe4.com/ibm-study-70-percent-of-businesses-attacked-pay-ransomware>. Acesso em: 18 jul. 2022. Citado na página 15.

STONEBURNER, G.; HAYDEN, C.; FERINGA, A. Engineering principles for information technology security. In: NIST. 2004. Disponível em: <https://www.govinfo.gov/content/pkg/GOVPUB-C13-5bce946ec882edc69dec9b9adc98c2b4/pdf/GOVPUB-C13-5bce946ec882edc69dec9b9adc98c2b4.pdf>. Citado na página 18.

THERADICATIGROUP. Advanced persistent threat (apt) protection - market quadrant 2021. In: THE RADICATI GROUP, INC. The Radicati Group, Inc, 2021. p. 5–6. Disponível em: <https://docs.broadcom.com/doc/radicati-apt-market-quadrant-2021>. Acesso em: 11 jan. 2022. Citado na página 21.

THREATPOST. Fighting cyber attacks with game theory. In: . ThreatPost, 2020. Disponível em: <https://threatpost.com/trapx-fighting-cyber-attacks-with-game-theory/156545/>. Acesso em: 16 jul. 2022. Citado nas páginas 25 e 27.

TIPTON, H. F.; KRAUSE, M. Information security management handbook. In: . [S.l.]: CRC press, 2006. v. 3. Citado na página 12.

TSAKANYAN, V. T. The role of cybersecurity in world politics. In: . [S.l.]: Vestnik RUDN. International Relations, 2017. v. 17, n. 2, p. 339–348. Citado na página 12.

VEEAM. Cybersecurity research: 76% of organizations admit to paying ransomware criminals, with one-third still unable to recover data. In: VEEAM. 2022. Disponível em: <https://www.veeam.com/news/cybersecurity-research/76-of-organizations-admit-to-paying-ransomware-criminals-still-unable-to-recover-data>. Acesso em: 10 jul. 2022. Citado na página 16.

WAZLAWICK, R. *Metodologia de pesquisa para a ciência da computação*. Rio de Janeiro, Brasil: Campus, 2014. v. 2. Citado na página 32.

WORLDECONOMICFORUM. The global risks report 2020. In: WORLD ECONOMIC FORUM. World Economic Forum, 2020. Disponível em: <https://www.weforum.org/reports/the-global-risks-report-2020>. Acesso em: 17 mar. 2021. Citado nas páginas 12, 13 e 14.

YOUNG, A. L.; YUNG, M. Cryptovirology: The birth, neglect, and explosion of ransomware. In: . Communications of the ACM, 2017. v. 60, n. 7, p. 24–26. Disponível em: <https://cacm.acm.org/magazines/2017/7/218875-cryptovirology/abstract>. Acesso em: 15 jun. 2022. Citado na página 14.

ZAMORA, W. What are exploits? (and why you should care). In: MALWAREBYTES. 2017. Disponível em: <https://www.malwarebytes.com/blog/news/2017/03/what-are-exploits-and-why-you-should-care>. Acesso em: 25 set. 2021. Citado na página 21.

ZETTER, K. Hacker lexicon: What is a zero day? In: WIRED. Wired, 2014. Disponível em: <https://www.wired.com/2014/11/what-is-a-zero-day/>. Acesso em: 14 jul. 2022. Citado na página 21.