ANDERSON APARECIDO ALVES DA SILVA

Uma Arquitetura para Agrupamento de Controles de Segurança em Ambientes de
Tecnologia da Informação Baseada em Barganhas Cooperativas Irrestritas

São Paulo
2017

ANDERSON APARECIDO ALVES DA SILVA

Uma Arquitetura para Agrupamento de Controles de Segurança em Ambientes de
Tecnologia da Informação Baseada em Barganhas Cooperativas Irrestritas

Tese apresentada à Escola Politécnica
da Universidade de São Paulo para
obtenção do título de Doutor em
Ciências

São Paulo
2017

ANDERSON APARECIDO ALVES DA SILVA

Uma Arquitetura para Agrupamento de Controles de Segurança em Ambientes de
Tecnologia da Informação Baseada em Barganhas Cooperativas Irrestritas

Tese apresentada à Escola Politécnica
da Universidade de São Paulo para
obtenção do título de Doutor em
Ciências

Área de Concentração:
Sistemas Eletrônicos

Orientador:
Prof. Dr. Sergio Takeo Kofuji

São Paulo
2017

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, _____ de _____ de _____

Assinatura do autor: _____

Assinatura do orientador: _____

Catálogo-na-publicação

Alves da Silva, Anderson Aparecido

Uma Arquitetura para Agrupamento de Controles de Segurança em Ambientes de Tecnologia da Informação Baseada em Barganhas Cooperativas Irrestritas / A. A. Alves da Silva -- versão corr. -- São Paulo, 2017.
192 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo.
Departamento de Engenharia de Sistemas Eletrônicos.

1.Teoria dos Jogos 2.Barganhas cooperativas irrestritas 3.Mobile Ad hoc NETwork (MANET) 4.Controles de segurança 5.Detecção e previsão de ataques I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Sistemas Eletrônicos II.t.

DEDICATÓRIA

Para

Olirio e Alice ...

Para

Carla ...

Para

Julia, Lucas e Vinícius.

AGRADECIMENTOS

Em primeiríssimo lugar ao Senhor Deus Todo Poderoso, meu Deus, que pela graça do Senhor Jesus me deu a oportunidade, a saúde e a perseverança para cumprir esta tarefa. Na medida em que rendo ao Senhor meu Deus toda a honra, Ele não tem deixado de me cobrir com todas as bênçãos.

Agradeço aos meus pais Olirio Alves da Silva e Alice Alves da Silva pela criação e pelos bons ensinamentos que me passaram durante a vida (graças a Deus ainda passam!).

Agradeço a minha esposa Carla Cristina da Silva e aos meus filhos Julia Alves da Silva, Lucas Alves da Silva e Pedro Vinicius Alves da Silva pela força, incentivo e principalmente pela paciência que tiveram comigo durante todo este percurso. Afinal eles já não aguentavam mais tanto estudo e trabalho sem descanso e diversão.

Sou muito grato a minha sogra Marlene José Maria por toda a ajuda que nos prestou durante meu período de estudos (e em outros períodos também!).

Quero deixar agradecimentos também a todos da Donatelli, em especial aos diretores Roberto Donatelli Vani e Antonio Carlos Donatelli Barbosa, por toda a confiança que depositaram em mim ao longo dos anos. Agradeço também ao meu parceiro Luciano Casado de Oliveira que sempre se desdobrou para cumprir suas tarefas, proporcionando tempo para que eu me dedicasse à pesquisa.

Meus agradecimentos também se estendem aos acadêmicos. Agradeço ao meu orientador Prof. Dr. Sergio Takeo Kofuji e aos meus co-orientadores informais Prof. Dr. Adilson Eduardo Guelfi e Prof. Dr. Fen Zhou por todas as correções e observações relevantes sobre o trabalho. A toda turma do Laboratório de Sistemas Integráveis (LSI) que está comigo no mesmo barco. À inestimável ajuda do meu grande amigo, parceiro e co-orientado Norisvaldo Ferraz Junior, o Noris, pela formatação da tese.

E claro, ao meu saudoso amigo e parceiro de tantas publicações que se foi tão cedo, já no fim do doutorado: Elvis Pontes o *“bode do asfalto”*.

Muitíssimo obrigado a todos.

RESUMO

Controles de segurança, também chamados de mecanismos de proteção, voltados para previsão e detecção de eventos indesejados são cada vez mais empregados em ambientes de Tecnologia da Informação (TI). O pouco entendimento sobre as características dos eventos indesejados que agem nos sistemas e a baixa compatibilidade existente entre os diversos mecanismos de proteção são problemas que se destacam neste tipo de cenário. Diferentes configurações dificultam a combinação dos resultados destes mecanismos e raramente dois ou mais controles de segurança se complementam. Por esse motivo, o agrupamento entre mecanismos de detecção e de previsão não é trivialmente resolvido. Neste trabalho é proposta uma arquitetura, denominada de Arquitetura Estratégica de Agrupamento - Strategic Grouping Architecture (SGA) - para agrupamento de controles de segurança voltados para detecção e/ou previsão, que tem como base a busca de um equilíbrio entre as configurações e os resultados individuais de cada mecanismo de proteção envolvido. Para alcançar este equilíbrio a arquitetura proposta divide a análise dos eventos (legítimos e maliciosos) que passam pelos controles de segurança em dois níveis de abstração: {1} o técnico, onde são coletadas as configurações e os resultados dos controles de segurança; e {2} o estratégico, onde os dados obtidos no nível técnico são analisados por meio de barganhas cooperativas irrestritas – Unrestricted Cooperative Bargains (UCB), conceito proveniente da Teoria dos Jogos, que busca a otimização e equilíbrio entre resultados. Justamente por ser realizada em um nível de abstração diferente, a análise gerada pelo SGA identifica a influência que cada configuração exerce nos resultados agrupados. Para explorar a capacidade da arquitetura proposta, dois experimentos, bastante diferentes, que envolvem a ação de eventos indesejados em ambientes de TI são conduzidos. Os resultados obtidos mostram a viabilidade do agrupamento de controles de segurança de detecção e previsão e a possibilidade do uso do SGA em outros ambientes, que não estejam necessariamente ligados à segurança de TI. Baseada na literatura científica a validação do SGA consiste de uma transformação prévia na categoria dos jogos estratégicos usados – cooperativos para não-cooperativos - e na busca de situações como o Equilíbrio de Nash (EN) e o ótimo de Pareto, que indicam os melhores resultados de um jogo.

Palavras-chave: Teoria dos Jogos. Barganhas cooperativas irrestritas. Mobile Ad hoc NETwork (MANET). Controles de segurança. Detecção e previsão de ataques.

ABSTRACT

Security controls, also called protection mechanisms, focused on forecasting and detection of unwanted events are increasingly employed in Information Technology (IT) environments. The little understanding about the characteristics of unwanted events which act on the systems and the low rate of compatibility among several protection mechanisms are both problems that arise in that scenario. Different settings make difficult combining the results of these mechanisms and two or more controls rarely complement each other. Due to that, grouping mechanisms of detection and forecasting is not a trivial matter. In this work a framework called Strategic Grouping Architecture (SGA) is proposed to grouping security controls focused on detection and/or forecasting. SGA is based on the search for equilibrium between the settings and the individual results of each protection mechanism involved. In order to reach this equilibrium the proposed framework divide the analysis of events (legitimates and malicious) which go through the security controls in two abstract levels: {1} the technical level, where the settings and the results of security controls are collected; and {2} the strategic level, where the data obtained in the technical level are analyzed through Unrestricted Cooperative Bargains (UCB), concept from Game Theory that seeks to optimize and balance the results. Precisely because it is performed on a different level of abstraction, the analysis generated by the SGA identifies the influence that each setting has on the clustered results. In order to exploit the capability of the proposed architecture, two experiments, quite different, involving the action of unwanted events in IT environments, are conducted. The obtained findings show the feasibility of grouping detection and forecasting security controls and the possibility of using the SGA in other environments that are not necessarily related to IT security. Based on scientific literature SGA validation consists of a previous transformation in the category of strategy games used - cooperative to non-cooperative - and the search for situations such as the Nash Equilibrium (NE) and the Pareto optimal, indicating the best results a game.

Key-words: Game Theory. Unrestricted cooperative bargains. Mobile Ad hoc NETwork (MANET). Security controls. Detection and forecasting attacks.

LISTA DE FIGURAS

Figura 1 - Exemplo de equilíbrio perfeito em subjogos	39
Figura 2 - Exemplo de barganha com paciência	40
Figura 3 - TVC e o conceito de compartilhamento de valores.....	42
Figura 4 - TVC e o conceito de valor compartilhado em uma barganha indeterminada.....	45
Figura 5 - Níveis do SGA.....	58
Figura 6 - Módulos do nível técnico.....	59
Figura 7 - Detalhes sobre os módulos do nível estratégico.....	62
Figura 8 - Topologia da MANET usada no Experimento 1	69
Figura 9 - Representação de M e $M2$, usados para determinar a posição de um nó	77
Figura 10 - Resultados do primeiro experimento - Amplitude do desvio padrão - GM(1,1)	87
Figura 11 - Resultados do primeiro experimento - Média do desvio padrão - GM(1,1)	89
Figura 12 - Resultados do primeiro experimento - Amplitude do desvio padrão - Tendência polinomial	91
Figura 13 - Resultados do primeiro experimento - Média do desvio padrão - Tendência polinomial	93
Figura 14 - Comparativo entre a variabilidade de séries previstas e falsas com 1% de aleatoriedade para o método GM(1,1)	96
Figura 15 - Comparativo entre a variabilidade de séries previstas e falsas com 1% de aleatoriedade para o método de tendência polinomial	96
Figura 16 - Distribuição de energia ao longo de 1500 s para diferentes tamanhos e intervalos de envio de pacotes	101
Figura 17 - Topologia de rede com dois controles de segurança: <i>Filtro 1 e 2</i>	108
Figura 18 - Comparação entre VP real e VP baseado no valor mínimo de <i>ranking</i> ($S_1=400$)	137
Figura 19 - Comparação entre VP real e VP baseado no valor mínimo de <i>ranking</i> ($S_2=201$)	138
Figura 20 - Comparação entre VN real e VN baseado no valor mínimo ($S_1=400$) ..	139

Figura 21 - Comparação entre VP real e VP baseado no valor mínimo de ranking ($S_2=201$)	139
Figura 22 - Valores mínimos e máximos de limite de cada amostra ($S_1=400$)	140
Figura 23 - Valores mínimos e máximos de limite de cada amostra ($S_2=201$)	141

LISTA DE TABELAS

Tabela 1 - Forma básica de uma matriz de recompensas	31
Tabela 2 - Comparativo entre o SGA e trabalhos relacionados	56
Tabela 3 - Exemplo de formação de jogadores	63
Tabela 4 - Comparativo entre o Primeiro experimento e trabalhos relacionados	71
Tabela 5 - Parâmetros gerais do Primeiro Experimento	78
Tabela 6 - Nível técnico: configurações do Primeiro Experimento	84
Tabela 7 - Nível estratégico: configurações do Primeiro Experimento	85
Tabela 8 - Valores das variáveis independentes e das variantes das variáveis resposta.....	86
Tabela 9 - Percentual de aumento no desvio padrão	97
Tabela 10 - Relevância de parte das variáveis independentes com base em um modelo de regressão linear.	99
Tabela 11 - Vetor de energia com os quartis obtidos em simulação de 1500 s	102
Tabela 12 - Cinco registros com maior CM da matriz UCB para o método de tendência polinomial.....	103
Tabela 13 - Cinco registros com maior CM da matriz <i>UCB</i> para o método GM(1,1).	103
Tabela 14 - Descrição da classificação dos campos detectados pelos filtros 1 e 2	113
Tabela 15 - Duas linhas da tabela com contabilização dos registros da amostra A003 com $S_1=400$	124
Tabela 16 - Nível técnico: configurações do Segundo Experimento	126
Tabela 17 - Nível estratégico: configurações do Segundo Experimento.	127
Tabela 18 – Média dos <i>p-value</i> dos modelos lineares das 77 amostras.	128
Tabela 19 - Contabilização dos <i>rankings</i> das amostras a_{002} e a_{003} com nível de segurança $s_1=400$	130
Tabela 20 - Contabilização dos <i>rankings</i> das amostras a_{002} e a_{003} com nível de segurança $s_2=201$	131
Tabela 21 - Valores de WTP e custo após o processo de padronização - Amostras a_{002} e a_{003} com nível de segurança $s_1=400$	133
Tabela 22 - Matriz UCB em ordem decrescente de CM, calculados para a amostra a_{002} com nível de segurança $s_1=400$	134

Tabela 23 - Faixas ideais para algumas amostras do nível de segurança $s_1=400$..	136
Tabela 24 - Percentuais finais reais e baseados nos valores mínimos ideais.....	141
Tabela 25 - Os cinco registros com CM mais alto (valores reproduzidos da tabela 30 (APÊNDICE E)	145
Tabela 26 - Cálculo de UCB do Exemplo 1	172
Tabela 27 - Cálculo de UCB do Exemplo 2.....	173
Tabela 28 – Fonte do programa de regressão linear (R-free).....	175
Tabela 29 – Regressão linear para ajuste dos dados - resultados de relevância das variáveis independentes do Primeiro Experimento.....	176
Tabela 30 – Cinco registros com maior CM das matrizes UCB para o método de previsão tendência polinomial.	177
Tabela 31 – Cinco registros com maior CM das matrizes UCB para o método de previsão GM(1,1).....	178
Tabela 32 – Regressão linear para ajuste dos dados - resultados de relevância das variáveis independentes do Primeiro Experimento.....	179
Tabela 33 – Faixas ideais e limites de <i>ranking</i> para as amostras do Segundo Experimento.	184

LISTA DE QUADROS

Quadro 1 - Parte do processo de dominância iterada com EN do Primeiro Experimento.....	147
Quadro 2 - Parte do processo de dominância iterada e faixas com EN do Segundo Experimento.....	150
Quadro 3 – Processo de dominância iterada para a amostra A_{002} – SEM ITERAÇÕES.....	187
Quadro 4 – Processo de dominância iterada para a amostra A_{002} – 1ª ITERAÇÃO.....	188
Quadro 5 – Processo de dominância iterada para a amostra A_{002} – 2ª ITERAÇÃO.....	189
Quadro 6 – Processo de dominância iterada para a amostra A_{002} – 3ª ITERAÇÃO.....	190
Quadro 7 – Processo de dominância iterada para a amostra A_{002} – 4ª ITERAÇÃO.....	191
Quadro 8 – Processo de dominância iterada para a amostra A_{002} – RESULTADO FINAL.....	192

LISTA DE ABREVIATURAS E SIGLAS

AGO	Accumulated Generation
CBR	Constant BitRate
CM	Contribuição Marginal
DDoS	Distributed Denial of Service
EN	Equilíbrio de Nash
FN	Falsos Negativos
FP	Falsos Positivos
FUT	Função Utilidade Transferível
GB	Gigabyte
GHz	Giga-Hertz
GM	Grey forecasting Model
IAGO	Inverse Accumulated Generation
IVC	Individual Value Created
KPCA	Kernel Principal Component Analysis
MANET	Mobile Ad-hoc NETwork
RAM	Random Access Memory
RSSF	Redes de Sensores Sem Fio
RSSI	Received Signal Strength Indicator
SBN	Solução de Barganha de Nash
SDI/SPI	Sistemas de Detecção/Prevenção de Intrusões
SGA	Strategic Grouping Architecture
SQL	Standard Query Language
SSH	Secure SHell
SVM	Support Vector Machines
TB	Terabyte
TI	Tecnologia da Informação
TVC	Total Value Created
UCB	Unrestricted Cooperative Bargain
UDP	User Datagram Protocol
VANET	Vehicular Ad-hoc Networks
VN	Verdadeiro Negativo

VNN	Verdadeiro Negativo Novo
VP	Verdadeiro Positivo
WAF	Web Application Firewall

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Objetivo	21
1.2	Metodologia	21
1.3	Contribuições Originais	22
1.4	Organização da Tese	23
2	REFERENCIAL TEÓRICO	25
2.1	Eventos Indesejados em Ambientes de TI	25
2.2	Teoria dos Jogos	28
2.2.1	Introdução à Teoria dos Jogos	28
2.2.2	Jogos Não Cooperativos	31
2.2.3	Jogos Cooperativos	34
2.2.4	Situações de Barganha	35
2.2.5	Unrestricted Cooperative Bargains (UCB)	42
3	TRABALHOS RELACIONADOS	49
4	PROPOSTA: STRATEGIC GROUPING ARCHITECTURE (SGA).....	57
4.1	A Arquitetura.....	57
4.1.1	Nível Técnico	58
4.1.2	Nível Estratégico	62
4.2	Discussão Sobre a Arquitetura Proposta.....	66
5	RESULTADOS E VALIDAÇÃO DO TRABALHO.....	67
5.1	Primeiro Experimento	67
5.1.1	Conceitos Fundamentais do Primeiro Experimento	72
5.1.2	Formalização do Primeiro Experimento	75
5.1.3	Resultados do Primeiro Experimento.....	86
5.1.4	Discussões sobre o Primeiro Experimento.....	105
5.2	Segundo Experimento	107
5.2.1	Formalização do Segundo Experimento	110
5.2.2	Resultados do Segundo Experimento.....	128
5.2.3	Discussões sobre o Segundo Experimento.....	142

5.3 Validação.....	144
5.3.1 Validação do Primeiro Experimento	145
5.3.2 Validação do Segundo Experimento	149
6 CONCLUSÕES	152
6.1 Resumo e Conclusão	152
6.2 Limitações e Trabalhos Futuros	155
6.3 Publicações Relacionadas	158
REFERÊNCIAS	161
APÊNDICE A – Unrestricted Cooperative Bargain (UCB): Exemplo 1	171
APÊNDICE B – Unrestricted Cooperative Bargain (UCB): Exemplo 2	173
APÊNDICE C – programa de regressão linear (R-free)	175
APÊNDICE D – Regressão linear para relevância das variáveis independentes do Segundo Experimento	176
APÊNDICE E – Matrizes UCB.....	177
APÊNDICE F – Regressão linear para relevância das variáveis independentes do Segundo Experimento	179
APÊNDICE G – Resultados da aplicação do SGA no para o Segundo Experimento	184
APÊNDICE H – Demonstração de dominância iterada na amostra A_{002} do Segundo Experimento.....	187

1 INTRODUÇÃO

Em geral, redes e sistemas computacionais estão sujeitos a diversos eventos indesejados¹ como ameaças cibernéticas, ataques ou mudanças nos fluxos de comunicação. As atividades de detecção, recuperação, e principalmente, prevenção de tráfego malicioso podem ser prejudicadas, uma vez que é muito complexo adquirir o conhecimento, sobre o *modus operandi* e as características de todos estes eventos (MELL; KENT; NUSBAUM, 2005; SCARFONE; MELL, 2007).

Boa parte das características dos eventos indesejados – por exemplo, bibliotecas de *software*, campos de protocolos, recursos do sistema operacional em uso, endereços de memória, portas lógicas, entre outras - está relacionada com informações recolhidas pela ocorrência destes próprios eventos nos sistemas em operação.

O número e a diversidade de eventos indesejados dificultam o entendimento das ameaças cibernéticas e o desenvolvimento de estratégias de defesa (SCARFONE; SOUPPAYA, 2012; JANSEN; GRANCE, 2011; KENT; SOUPPAYA, 2006; O'REILLY, 2012). Formas para mitigar essas ameaças incluem análise (AXELSSON, 2000; ZHANG; LEE; HUANG, 2003; HODGE; AUSTIN, 2004; GARCIA-TEODOROA et al., 2009), classificação e agrupamento de informações de eventos com características comuns (STINE et al., 2008; ZHANG; LEE, 2000; GIACINTO; ROLI; DIDACI, 2003).

Uma alternativa bastante utilizada contra a ação de eventos indesejados são os controles de segurança, também chamados de mecanismos de proteção². Tratam-se de ferramentas largamente utilizadas para detectar ou prevenir ameaças cibernéticas (GARCIA-TEODOROA et al., 2009; ZHANG; LEE; HUANG, 2003; HOFMEYR; FORREST; SOMAYAJI, 1998).

¹ Eventos indesejados ou maliciosos é um termo genérico para designar todos os tipos de ameaças cibernéticas e tráfego de rede indesejado.

² Neste trabalho mecanismo de proteção é um sinônimo para controle de segurança. Ambos os termos referem-se a qualquer método ou técnica utilizada para prevenir (mecanismo de prevenção) ou para detectar (mecanismo de detecção). O termo mecanismo de previsão também é usado e é relacionado à métodos e técnicas de previsão voltadas para segurança.

Apesar de não ser livre de erros, o alcance de um controle de segurança é bastante amplo e os resultados gerados podem ser úteis até de formas não convencionais. Por exemplo, mecanismos de previsão não precisam ser usados exclusivamente para prevenir a ação de ameaças, mas em um contexto estratégico podem também ser utilizados como ferramentas de decisão, antecipando tendências (DENG, 1989; KENT; SOUPPAYA, 2006). Percebe-se, portanto, que o gerenciamento de controles de segurança é um processo bastante específico. Em relação à análise estratégica³, esse processo pode ser visto a partir de duas perspectivas: {1} não-cooperativa, quando lida com eventos indesejados e o conflito entre atacante *versus* vítima (LIU; COMANICIU; MAN, 2006; CHEN; LENEUTRE, 2009); e {2} cooperativa, quando se busca a união dos mecanismos de proteção em prol de um mesmo objetivo (ZHU et al., 2009).

Há uma grande variedade de controles de segurança (HODGE; AUSTIN, 2004; AXELSSON, 2000) que podem ter diferentes configurações, parâmetros, padrões e resultados. Essas diferenças podem se tornar um obstáculo para a análise conjunta dessas ferramentas (ZHU et al., 2009; LIU; COMANICIU; MAN, 2006; CHEN; LENEUTRE, 2009).

Como a heterogeneidade de controles de segurança pode representar um problema na geração de uma segurança mais efetiva, existem trabalhos que apresentam modelos e metodologias para agrupar mecanismos de detecção: Srinivas; Andrew e Ajith (2005) implementam redes neurais para classificar ameaças cibernéticas; Beauquier e Hu (2007) comparam diferentes métodos de agrupamento usando modelos de dissimilaridade como distância estatística; Chebrolu; Abraham e Thomas (2005) fazem redução de parâmetros com correlação; e Giacinto, Roli e Didaci (2003) usam reconhecimento de padrões para agrupar múltiplos classificadores. Por outro lado, trabalhos como Palit e Popovic (2000), Stuber; Gemmar e Greving (2000), Riedel e Gabrys (2005), e Armstrong (2001) tem como foco métodos para agrupar mecanismos de previsão.

A literatura referenciada mostra que o processo de agrupamento tem algumas características muito específicas, bem diferentes de outros procedimentos de

³ Os termos análise e abordagem estratégica não devem ser entendidos como termos pertencentes à governança da Tecnologia de Informação (TI). Dentro do contexto deste trabalho, esses termos se referem aos critérios de escolha dos melhores resultados, uma das principais bases da Teoria dos Jogos que funciona como pilar das hipóteses desta tese.

segurança. Como um exemplo, ao contrário de ambientes que possuem apenas um mecanismo de proteção, há casos nos quais o processo de agrupamento de dois ou mais controles de segurança exige a verificação do viés dos dados (LIPOVETSKY; CONKLIN, 2001). Na sequência são levantadas pelo menos quatro questões que representam pontos abertos relacionados ao agrupamento de controles de segurança e que precisam ser mais bem estudadas:

1. Quando um modelo usado para combinar controles de segurança já está estabelecido é muito difícil agregar novos resultados⁴ de outro mecanismo de proteção. Como consequência, dados recolhidos de um controle de segurança podem não servir para complementar outro;
2. Análise estratégica é uma abordagem raramente escolhida quando o assunto diz respeito a agrupamento de mecanismos de detecção, sendo mais comum o uso no agrupamento de mecanismos de previsão. Isso ocorre porque estimação ou previsão geralmente depende de dados confiáveis. Essa é uma característica muito importante para a análise estratégica. Mesmo quando dados precisos e confiáveis são coletados, os atuais mecanismos de detecção usados em Tecnologia da Informação (TI) muitas vezes não podem compartilhar toda a informação sobre um evento malicioso, seja porque os formatos são incompatíveis, seja porque parte da informação é desconhecida. Na verdade, apesar de mais difíceis de implementar (RIEDEL; GABRYS, 2005), o agrupamento de controles de previsão costuma melhorar o desempenho das previsões individuais e ter um melhor custo benefício (TIMMERMAN, 2005). Além disso, há modelos que podem ser usados para medir o desempenho de previsões agrupadas (CLEMENTS; SMITH, 2002; GONZÁLEZ-RIVERA; YOLDAS, 2012; CHRISTOFFERSEN, 1998; BIDARKOTA, 1998; PREEZ; WITT, 2003; KO; PARK, 2013; WANG; BESSLER, 2004);
3. Com base nos primeiros dois itens, o agrupamento de mecanismos de detecção com mecanismos de previsão não é algo trivial. Parece haver falta de estudos que agrupem ou mesmo que realizem algum tipo de balanceamento entre os resultados obtidos de mecanismos de detecção e

⁴ Resultados de controles de segurança podem incluir, entre outros exemplos, conteúdo de *logs*, alertas ou valores estatísticos extraídos de séries históricas previstas.

mecanismos de previsão. Basicamente os resultados de mecanismos de detecção são coletados apenas para alimentar séries históricas usadas pelos mecanismos de previsão ao invés de prover algum tipo de ganho real nos resultados;

4. Um processo de agrupamento costuma carregar efeitos colaterais, como consequência dos resultados gerados pelos controles de segurança. Por exemplo, aumentar o nível de segurança de um sistema manipulando um mecanismo de proteção pode levar ao incremento no tráfego da rede ou no consumo de energia.

Sendo assim, percebe-se que o processo de agrupamento de controles de segurança, principalmente se estes forem de detecção e previsão ainda apresenta pontos em aberto que precisam ser mais bem trabalhados.

A principal hipótese apresentada nesta tese sugere que uma forma eficiente de agrupar controles de segurança – mesmo se estes forem mecanismos de detecção e previsão - é realizar uma abordagem estratégica, na forma de um balanceamento que envolva as configurações e os resultados dos controles envolvidos, pois esse procedimento tende a minimizar os efeitos colaterais. Ou seja, este trabalho sugere que encontrar os resultados mais equilibrados entre diferentes controles de segurança é uma forma válida de agrupamento. É importante frisar que a proposta deste trabalho é totalmente baseada em estratégias cooperativas, que buscam coordenar esforços para agrupar os controles de segurança. De forma geral, ao contrário da proposta desta tese, quando se trata de segurança de TI, a literatura referenciada se concentra em jogos não-cooperativos⁵ que avaliam o conflito entre atacante e vítima (AGAH et al., 2004; LIU; COMANICIU; MAN, 2006; CHEN; LENEUTRE, 2009; KUMAR; KUMAR, 2012).

A análise estratégica leva em conta as interações entre as variáveis de um sistema ou rede (CHEBROLU; ABRAHAM; THOMAS, 2005) e funciona em um nível acima do nível técnico – onde os dados são coletados pelos controles de segurança (BIERMAN; FERNANDEZ, 2011). Esta abordagem pode potencializar o desenvolvimento de metodologias preventivas, por exemplo, que impeçam

⁵ Estudos baseados em jogos não-cooperativos são mais comuns no tratamento de ameaças cibernéticas. Há poucos trabalhos focados em redes e sistemas computacionais baseados em jogos cooperativos. Os trabalhos de Khan e Ahmad (2009) e Khan e Ahmad (2009) podem ser citados como exemplo, contudo, não são pesquisas relacionadas com a área de segurança.

configurações desvantajosas durante o processo de agrupamento dos controles de segurança.

Em termos práticos, uma opção interessante para uma abordagem estratégica é o uso de um método denominado de Barganha Cooperativa Irrestrita (Unrestricted Cooperative Bargain - UCB). Essencialmente, uma situação de UCB entre dois jogadores é um subconjunto da Teoria dos Jogos Cooperativos e implica em colaboração mútua para benefício de ambos (NASH JR, 1950).

Assim, a análise estratégica baseada em UCB é uma forma interessante de equilibrar configurações e resultados, e lidar com os já mencionados pontos em aberto presentes nos processos de agrupamento de controles de segurança. Outra vantagem é que algumas características exigidas em UCB, como racionalidade, igualdade nas condições de barganha, total conhecimento dos resultados, gerenciamento centralizado e monitoramento constante estão presentes no gerenciamento de controles de segurança (CHEN; LENEUTRE, 2009; KHAN; AHMAD, 2009; STUART JR, 2001; NASH JR, 1950).

Outro ponto que conta a favor da adoção da UCB em processos de agrupamento é que se pode lidar com os resultados finais exibidos pelos mecanismos de proteção envolvidos. Ao contrário de uma análise individual das características de cada controle de segurança durante o processo de agrupamento, uma análise de resultados finais tende a trazer mais flexibilidade para o processo já que menos variáveis estarão envolvidas.

Em linhas gerais e de acordo com as referências previamente citadas, {1} a possibilidade de se trabalhar em um nível de abstração superior ao técnico (nível estratégico) que se adapta a diversas situações e permite o agrupamento de mecanismos de detecção e previsão tendo como base a maximização dos benefícios; {2} a vantagem advinda do fato de características frequentemente presentes em ambientes de gerenciamento de controles de segurança, como a racionalidade e o conhecimento dos resultados serem pré requisitos importantes em situações que envolvam barganhas cooperativas; e {3} a diminuição de eventuais efeitos colaterais do processo de agrupamento devido ao balanceamento entre as configurações e os resultados finais dos controles de segurança; servem para

justificar o uso da análise estratégica baseada em UCB nos processos de agrupamento de mecanismos de previsão e detecção..

1.1 Objetivo

O objetivo deste trabalho é propor uma arquitetura denominada de Arquitetura de Agrupamento Estratégico (Strategic Grouping Architecture - SGA) baseada em UCB para agrupar controles de segurança de detecção e/ou previsão que atuem contra eventos indesejados em ambientes de rede e TI.

1.2 Metodologia

Para agrupar controles de segurança este trabalho adapta conceitos como UCB, provenientes da Teoria dos Jogos e estabelecidos por Nash Jr (1950, 1953). Largamente usados em economia, esses conceitos são estendidos nos trabalhos de Brandenburger e Stuart Jr (1996, 2007), Stuart Jr (2001), Brandenburger e Dekel (1987), Brandenburger e Nalebuff (1995) e Brandenburger (2007) que definem formas de combinar estrategicamente resultados e estabelecer equilíbrio entre jogadores cooperativos.

Como forma de ilustrar a diversidade de situações em que o SGA pode ser utilizado, dois experimentos distintos que envolvem a ação de eventos indesejados são construídos como prova de conceito.

No Primeiro Experimento a principal preocupação é determinar se o posicionamento informado por nós móveis de uma Mobile Ad-hoc NETWORKS (MANET) é verdadeiro. Para isso, dois métodos de previsão são agrupados a um mecanismo de detecção para identificação de ação maliciosa na rede. Devido à grande diversidade de configurações e resultados envolvidos no processo de agrupamento, o SGA é implementado para que sejam encontrados valores que representem o balanceamento ideal entre a variabilidade das previsões de

posicionamento – um forte indicativo de atividade maliciosa – e o consumo de energia dos nós.

No Segundo Experimento um ambiente real de comércio eletrônico que conta com dois mecanismos de detecção tem o tráfego de rede analisado e previsto por meio de um *ranking* baseado em pesos. O SGA é aplicado nos percentuais deste *ranking* para que sejam verificadas quais são as configurações mais adequadas de trabalho, tendo como critérios principalmente os índices de Verdadeiro Positivo (VP) e Verdadeiro Negativo (VN) gerados no experimento.

Como forma de validação os resultados obtidos com o SGA de ambos os experimentos são comparados com as otimizações – ótimo de Pareto e Equilíbrio de Nash (EN) - sugeridas nos trabalhos de Wu et al. (2010), Nadkar et al. (2012), Ma; Rao e Yau (2011), Khan e Ahmad (2009) e Agah et al. (2004). Para isso, os dados usados no SGA passam por uma transformação prévia sugerida por Nudelman et al. (2004) para que obtenham um caráter não cooperativo.

1.3 Contribuições Originais

Este trabalho apresenta algumas contribuições originais, listadas a seguir:

- A possibilidade de agrupamento de controles de segurança voltados para previsão e detecção de eventos maliciosos em ambientes de TI;
- A separação da arquitetura proposta em dois níveis de abstração que ampliam a diversidade de ambientes onde o SGA pode ser aplicado; nível técnico e nível estratégico⁶;
- A aplicação de uma abordagem estratégica baseada em UCB na análise das configurações e dos resultados gerados pelos controles de segurança.

Em ambientes de TI, sobretudo relacionados à segurança, um gerenciamento muito engessado não é adequado - as configurações adotadas necessitam de certo grau de flexibilidade. Dessa forma, além das citadas contribuições originais, há pelo menos duas diferenças entre o processo de preparação e análise dos dados

⁶ No nível estratégico os dados coletados no nível técnico são analisados por meio de UCB.

realizados nesta proposta e os trabalhos que trazem os principais conceitos sobre UCB (STUART JR, 2001; BRANDENBURGER; STUART JR, 1996; BRANDENBURGER; STUART JR, 2007; BRANDENBURGER; DEKEL, 1987; BRANDENBURGER; NALEBUFF, 1995; BRANDENBURGER, 2007):

- A geração de diversos valores de barganha que permitem a um administrador escolher os resultados e configurações mais adequados para o ambiente onde os controles de segurança estão sendo utilizados;
- A forma de agrupamento dos controles de segurança baseada no relacionamento entre variáveis independentes⁷ e variáveis resposta⁸.

1.4 Organização da Tese

Este trabalho é organizado da seguinte forma:

- Capítulo 2 REFERENCIAL TEÓRICO: descreve conceitos relacionados a medidas usadas contra eventos indesejados, alguns conceitos estatísticos, a Teoria dos Jogos e a UCB;
- Capítulo 3 TRABALHOS RELACIONADOS: este capítulo tem por objetivo descrever o estado da arte com os principais trabalhos relacionados aos eixos teóricos desta tese, realizando comparações que evidenciem as principais contribuições e os diferenciais obtidos;
- Capítulo 4 PROPOSTA: STRATEGIC GROUPING ARCHITECTURE (SGA): neste capítulo, a proposta de arquitetura (SGA) é apresentada evidenciando seu funcionamento, módulos funcionais, resultados e formas de análise;
- Capítulo 5 RESULTADOS E VALIDAÇÃO DO TRABALHO: neste capítulo pretende-se expor os cenários e os resultados de dois experimentos

⁷ As variáveis independentes (também conhecidas como variáveis preditivas, regressivas ou covariáveis) são aquelas que são observáveis e, na maioria dos casos, podem ser configuradas de acordo com a conveniência (DRAPER; SMITH, 1998). Neste trabalho as variáveis independentes podem ser entendidas como as configurações usadas nos controles de segurança. Por exemplo, se resultados de uma ferramenta de tráfego for coletada, as variáveis independentes podem ser os campos do protocolo ou a quantidade de *bytes* transmitidos.

⁸ Variáveis resposta (também conhecidas como variáveis dependentes) são aquelas cujos valores são influenciados pelas independentes (DRAPER; SMITH, 1998). Por exemplo, consumo de energia dos nós em uma rede sem fio pode ser explicado pela quantidade de *bytes* enviada.

realizados. Discussões sobre cada um dos experimentos são conduzidas, assim como conclusões individuais sobre os resultados obtidos. Ao fim, como forma de validação, otimizações realizadas em outros trabalhos – ótimo de Pareto e EN – são reproduzidas e comparadas aos resultados do SGA;

- Capítulo 6 CONCLUSÕES: mostra um resumo do trabalho, com as contribuições e considerações finais sobre os principais resultados. Também apresenta as principais limitações e propostas para pesquisas futuras.

2 REFERENCIAL TEÓRICO

Neste capítulo são descritos os principais conceitos utilizados neste trabalho. A Seção 2.1 define conceitos relacionados à segurança em ambientes de TI, como eventos maliciosos e controles de segurança. Na Seção 2.2 são descritos diversos aspectos sobre a Teoria dos Jogos relacionados aos processos de UCB.

2.1 Eventos Indesejados em Ambientes de TI

Administradores de sistemas computacionais são constantemente confrontados com problemas que exigem decisões precisas e rápidas. Para isso, o administrador deve ter nas mãos a maior quantidade de informações possíveis sobre os eventos que ocorrem em um sistema. Eventos são ações que acontecem e afetam um sistema computacional de forma benéfica ou maléfica, podendo ser ou não registrados pelo sistema computacional na forma de *logs* (KENT; SOUPPAYA, 2006).

Existe uma crença que os eventos indesejados, também chamados de maliciosos, que agem em sistemas sejam relacionados especificamente a ações de intrusão. Contudo, de acordo com Mell; Kent e Nusbaum (2005) há diversas outras categorias nas quais o termo pode ser aplicado. Exemplos incluem os *malwares* ou os ataques contra a disponibilidade de sistemas.

De forma geral, quando se tratam de sistemas cibernéticos, eventos indesejados se referem a ações que afetam de forma prejudicial os principais serviços de segurança dos dados: confidencialidade, integridade, autenticação, irretratabilidade, controle de acesso e disponibilidade. O conceito de evento indesejado pode ser facilmente estendido, por exemplo, ao fluxo de informações. A justificativa é que o tráfego de rede também pode ser encarado como um evento, pois afeta as aplicações e os dispositivos de comunicação (*switches*, roteadores, *modems*, etc.) presentes em uma rede (SILVA; GUELF, 2010).

De acordo com Feitosa, Souto e Sadok (2012), o tráfego de rede malicioso pode ser definido como qualquer fluxo de informações inesperado e/ou não requisitado que possui como único propósito degradar a disponibilidade do sistema por meio do desperdício de recursos técnicos (processamento, armazenamento, etc.) e tempo, potencialmente ocasionando prejuízos financeiros.

Desta forma, enquadram-se como exemplos de eventos maliciosos ligados ao fluxo de rede:

- Ataques cibernéticos de negação de serviço: ações remotas ou locais impetradas contra sistemas ou redes computacionais, com o objetivo de comprometer a disponibilidade de uma rede;
- Tráfego malicioso: fluxo gerado com o objetivo de diminuir a disponibilidade de um determinado canal de comunicação. Também se enquadra nesta categoria o falso tráfego de rede gerado para confundir ou prejudicar um determinado elemento de rede;
- Programas maliciosos como vírus, *trojans* e *malwares*: cujos variados objetivos passam pela obtenção de informações sobre a rede ou *hosts*, disponibilização de *backdoors* e execução de *scripts* maliciosos (MELL; KENT; NUSBAUM, 2005);
- Intrusão: ataque cibernético cujos principais objetivos são a obtenção acesso e/ou o uso não autorizado de sistemas ou privilégios.
- Ações de roubo de identidade, *phishing*, *spoofing*, geração de *spam* e *spyware*, entre outros (MELL; KENT; NUSBAUM, 2005).

Na verdade, essa forte ligação com o tráfego de rede, ao invés de limitar, expande a necessidade de monitoramento dos eventos indesejados para outras áreas ou mecanismos utilizados em TI, como por exemplo, redes de sensores e criptografia.

Há diversos controles de segurança que podem ser empregados na contenção dos eventos indesejados. Tendo em vista que os mecanismos de prevenção agem por antecipação, procurando evitar a concretização de determinadas ameaças, a viabilidade econômica destes costuma ser melhor que a viabilidade oferecida pelos mecanismos de detecção (SCARFONE; MELL, 2007).

Um controle de segurança voltado para prevenção tem a função de monitorar e analisar eventos que ocorrem em um sistema computacional ou rede, de forma que seja possível perceber sinais, alertas ou comportamentos que podem levar a incidentes, evitando ou minimizando a ação destes antes que ocorram (SCARFONE; MELL, 2007). Há duas categorias listadas como serviços de prevenção: {1} métodos de previsão: métodos e técnicas usados para previsão de anomalias e intrusões em sistemas computacionais e redes (LAI-CHENG, 2007; HASLUM; ABRAHAM; KNAPSKOG, 2008; ISHIDA et al., 2005); e {2} análise de anomalias: métodos e técnicas usadas para determinar comportamentos considerados padronizados e analisar os desvios de conduta prejudiciais que ocorrem tendo como referência estes comportamentos (LAZAREVIC et al., 2002; MANIKOPOULOS; PAPAVALASSILIOU, 2002; PATCHA; PARK, 2007).

Entretanto, mecanismos de prevenção são bem difíceis de serem implantados, pois dependem de informações precisas sobre o ambiente monitorado (SCARFONE; MELL, 2007).

Os controles usados na detecção de eventos indesejados têm a disposição uma ampla quantidade de ferramentas como os Sistemas de Detecção/Prevenção de Intrusões (SDI/SPI), *firewall*, sistemas de análise de tráfego e programas de monitoramento como *anti-spywares*, entre outros.

As implementações de SDI/SPI costumam monitorar eventos de três formas: {1} comparando o tráfego de rede com assinaturas de ataques conhecidos; {2} comparando o tráfego com desvios do comportamento normal (anomalia); e {3} de forma híbrida, usando detecção por assinatura em conjunto com detecção por anomalia ou estatística. Contudo, estas ferramentas também estão sujeitas à falhas e imprecisões como Falsos Negativos (FN) e Falsos Positivos (FP) (ZHANG; LEE, 2000; AXELSSON, 2000; BASS, 2000; SILVA; GUELF, 2010).

Outro problema que afeta os sistemas que contam com mecanismos de detecção em seu perímetro de defesa diz respeito ao fato que eventuais ações maliciosas detectadas já se encontram em andamento e podem provocar algum dano caso a reação não seja rápida e precisa.

2.2 Teoria dos Jogos

Nesta seção são descritas as ligações entre os principais aspectos da Teoria dos Jogos com os conceitos de construção e negociação dos processos de UCB.

2.2.1 Introdução à Teoria dos Jogos

As primeiras ideias sobre a Teoria dos Jogos começaram a serem traçadas no século XVIII, mas apenas entre os anos 1920-1944 houve um impulso maior através dos trabalhos realizados por Emile Borel e principalmente por Von Neumann e Morgenstern em 1944. Por volta de 1949, John F. Nash Jr. contribuiu muito com o desenvolvimento da teoria ao estabelecer conceitos de equilíbrio entre decisões estratégicas que permitiram a evolução das pesquisas e a aplicação prática da Teoria dos Jogos em diversos segmentos (OSBORNE, 2004).

A Teoria dos Jogos busca entender situações nas quais indivíduos ou grupos de indivíduos (também chamados de agentes ou jogadores) interagem por meio de decisões estratégicas. Formalmente, uma estratégia é um plano de ações que especifica, para um determinado jogador, que ação tomar em momentos de decisão (DIXIT, 2004). Estratégias são classificadas como:

- Estratégias puras: correspondem a estratégias determinísticas escolhidas por um jogador. Ocorrem, por exemplo, quando a escolha representa um determinado valor que durante a análise é literalmente comparado ao valor de outro jogador (BIERMAN; FERNANDEZ, 2011; OSBORNE, 2004);
- Estratégias mistas: são estratégias representadas pela aleatoriedade aplicada sobre as estratégias puras (valores das decisões literais) dos agentes. Uma estratégia mista representa uma probabilidade sobre as possíveis decisões tomadas pelos agentes envolvidos no jogo (FIANI, 2006). De acordo com Osborne (2004), estratégias mistas ocorrem quando uma distribuição de probabilidades é escolhida a partir das escolhas disponíveis para um jogador.

Um jogo estratégico pode ser definido como um modelo formal, normalmente simples, composto por descrições e análises, jogado por agentes cujas ações são parte de um comportamento estratégico de caráter racional que afeta os demais

jogadores (FIANI, 2006). Um comportamento estratégico acontece quando um jogador leva em consideração o fato de que existem interações entre outros jogadores e tem ciência que sua decisão tem consequências sobre os outros (FIANI, 2006; DIXIT, 2004).

A racionalidade é a qualidade que leva os agentes a empregar os meios mais adequados para atingir os objetivos estabelecidos. O termo racional, num primeiro momento, pode levar a constatação de que, um jogo estratégico deve necessariamente ser jogado por jogadores que mantenham um comportamento decisório dentro dos padrões esperados ou normais. Contudo, a racionalidade na tomada de decisões está fortemente ligada às recompensas oferecidas e ao aprendizado adquirido durante o jogo. A racionalidade sofre forte influência da quantidade de informações disponíveis e do cenário onde o jogo é aplicado (FIANI, 2006). Dixit (2004) acrescenta ainda que a racionalidade de um agente depende do conhecimento de dois fatores essenciais: {1} quais são os interesses do jogador; e {2} quais ações são as mais viáveis para atingir esses interesses.

Nada impede, portanto, que máquinas ou algoritmos sejam agentes de jogos estratégicos. Por exemplo, um mecanismo de previsão de ataques baseado em um modelo estatístico pode ser complementado com informações sobre o tipo de alertas ou registros gerados por um SDI/SPI e tomar decisões consideradas racionais. Dentro do mesmo exemplo, um administrador de sistemas pode tomar decisões racionais a partir das informações geradas pelo mecanismo de previsão. Ou seja, dentro do escopo que este trabalho abrange, o entendimento sobre a racionalidade depende muito da quantidade de informações conhecidas sobre a dinâmica do sistema.

Assim como em outras ciências, o estudo dos jogos se dá por meio da criação de modelos que representam a realidade e capturam a interação ocorrida entre os participantes. Formalmente, a representação de um jogo (OSBORNE, 2004; SARTINI et al., 2004) envolve um conjunto finito de jogadores $G=\{g_1, g_2, \dots\}$; um conjunto com todas as decisões (estratégias) disponíveis aos jogadores $S=\{S_1, S_2, \dots\}$; um subconjunto $S_i=\{S_{i1}, S_{i2}, \dots\}$, também chamado de vetor ou perfil de estratégias, com as estratégias puras de um determinado jogador g_i ; uma função de

incentivo $v=\{v_1, v_2, \dots\}$ que associa um valor de recompensa $u=\{u_1, u_2, \dots\}$ a cada decisão S_{ij} de um determinado jogador g_i .

As estratégias puras de um jogador i formam um produto cartesiano:

$$S = \prod_{i=1}^n S_i = S_1 \times S_2 \times \dots \times S_n \quad (1)$$

Há diversas classificações usadas para jogos estratégicos. Em relação ao conhecimento das estratégias dos agentes participantes, jogos podem ser: {1} simultâneos, quando os jogadores decidem sem conhecer as decisões dos outros jogadores e sem preocupações com consequências futuras destas escolhas; e {2} sequenciais, quando os jogadores sabem o que foi decidido em etapas anteriores e os fatos acontecem em uma determinada ordem (FIANI, 2006; BIERMAN; FERNANDEZ, 2011).

Se o momento decisório for levado em conta, jogos podem ser classificados como: {1} estáticos, quando as decisões dos jogadores são tomadas simultaneamente; e {2} dinâmicos, quando os jogadores tomam decisões em momentos diferentes (FIANI, 2006; BIERMAN; FERNANDEZ, 2011).

Em relação ao conhecimento das recompensas dos demais jogadores existem: {1} jogos de informação completa: aqueles nos quais todas as recompensas dos jogadores são conhecidas; e {2} jogos de informação incompleta ou Bayesianos: aqueles em que as recompensas não são conhecidas por todos os jogadores. Como exemplo deste último, pode ser citado um leilão com lances fechados (FIANI, 2006; BIERMAN; FERNANDEZ, 2011).

Quando a recompensa de um jogador não implica necessariamente na perda total para o outro jogador temos um jogo de soma não zerada. O inverso é conhecido como jogo de soma zero.

Uma das principais classificações usadas é relacionada à interação entre os agentes: {1} jogo não cooperativo ou adversário é aquele no qual os agentes não podem estabelecer compromissos garantidos e normalmente atuam como antagonistas; e {2} jogo cooperativo é aquele no qual os agentes podem estabelecer compromissos garantidos visando um melhor resultado. Geralmente (não necessariamente) em um jogo cooperativo os agentes são aliados (OSBORNE, 2004; TAN; LIE; MEMBER, 2001).

Brandenburger (2007) chama a atenção para o fato de que em jogos não cooperativos também há espaço para cooperação entre os jogadores. Por mais contraditória que possa parecer, em muitos jogos não cooperativos as decisões individuais dos jogadores podem levar a resultados que signifiquem o bem de todos os outros agentes (DIXIT, 2004). Por outro lado, em um jogo cooperativo também pode haver conflitos, por exemplo, em processos de negociação durante barganhas (BRANDENBURGER, 2007; DIXIT, 2004).

2.2.2 Jogos Não Cooperativos

Uma característica dos jogos não cooperativos é que eles dependem diretamente dos procedimentos que são adotados nas negociações, por isso são rotulados como procedurais (STUART JR, 2001; BRANDENBURGER, 2007).

Uma das formas sugeridas para a análise das interações entre agentes de jogos não cooperativos é a criação de matrizes de recompensas. Samieh e Wahba (2007) definem recompensa ou incentivo como um número que mede a preferência do jogador por um determinado resultado. Uma matriz de recompensas mostra os resultados (também chamados de ganhos, recompensas ou incentivos) de todas as estratégias que podem ser adotadas pelos agentes (BIERMAN; FERNANDEZ, 2011). Formalmente, Gmytrasiewicz e Durfee (1995) definem uma matriz de recompensa P_i de um agente g_i como uma tripla (G, S, U) .

A forma básica de uma matriz de recompensas $P_{(2 \times 2)}$ pode ser vista no exemplo exposto na tabela 1:

Tabela 1 - Forma básica de uma matriz de recompensas

		Agente2	
		Ação1	Ação2
Agente1	Ação1	(7 , 7)	(-1 , 8)
	Ação2	(8 , 1)	(1 , 1)

Fonte: o autor.

A tabela 1 mostra as estratégias puras (*ação1* ou *ação2*) que dois agentes podem tomar em uma determinada situação e as recompensas (números) obtidas para cada escolha. Percebe-se que os ganhos dos agentes estão condicionados às estratégias escolhidas por ambos. Os ganhos do agente1 são contabilizados verticalmente, de acordo com as escolhas feitas pelo agente2. No caso, quando o agente2 escolhe *ação1* a melhor recompensa para o agente1 ocorre na opção *ação2* (*recompensa=8*). Da mesma forma, quando o agente2 escolhe *ação2* a melhor escolha do agente1 ocorre também na opção *ação2* (*recompensa=1*).

Observa-se que a racionalidade busca a melhor recompensa e que uma decisão estratégica pode ser julgada melhor que a outra desde que ofereça uma recompensa também melhor. Uma solução, portanto, pode ser expressa na forma: $v_1(s_1) > v_2(s_2)$ se e somente se $u_1(s_1) > u_2(s_2)$ (OSBORNE, 2004). De acordo com Sartini et al. (2004) há dois conceitos inter-relacionados bastante usados para solucionar jogos estratégicos: dominância e Equilíbrio de Nash (EN).

Se for analisado o perfil de estratégias $S_i = \{S_{i1}, S_{i2}, \dots, S_{in}\}$ de um determinado jogador g_i enquanto as estratégias escolhidas pelos outros jogadores permanecem fixas, é possível encontrar casos em que algumas estratégias do jogador g_i sejam racionalmente melhores que outras. Neste caso, mediante o perfil de estratégia dos outros jogadores, há uma dominância de algumas estratégias do jogador g_i sobre outras. Sendo assim, se for denotada uma decisão estratégica para cada jogador, exceto o jogador g_i na forma: $s_{-i} = (s_{1k1}, \dots, s_{(i-1)ki-1}, s_{(i+1)ki+1}, \dots, s_{nkn}) \in S_{-i} = S_1 \times \dots \times S_{i-1} \times S_{i+1} \times \dots \times S_n$, é possível estabelecer um perfil de estratégias para todos os jogadores, exceto o jogador g_i na forma: $s = (s_{iki}, s_{-i}) = (s_{1k1}, \dots, s_{(i-1)ki-1}, s_{iki}, s_{(i+1)ki+1}, \dots, s_{nkn})$ (SARTINI et al., 2004).

Há diversos tipos de dominância. Estratégia estritamente dominante é aquela que domina estritamente todas as outras estratégias do jogador. Racionalmente será a estratégia escolhida. A estratégia s_1 domina estritamente a estratégia s_2 de um jogo se, dada qualquer coleção de estratégias que poderiam ser adotadas pelos demais jogadores, a adoção de s_1 resulta em uma recompensa estritamente mais alta para este jogador do que adotar s_2 . Um jogador racional nunca adotará uma estratégia estritamente dominada nem esperará que um jogador racional a adote (BIERMAN; FERNANDEZ, 2011). O perfil de estratégias $s_i = \{s_1, s_2, \dots, s_n\}$ é um equilíbrio de

estratégia estritamente dominante se, para qualquer jogador i , s_i for uma estratégia estritamente dominante (BIERMAN; FERNANDEZ, 2011).

Formalmente, uma estratégia pura $s_{ik} \in S_i$ do jogador $g_i \in G$ é estritamente dominada pela estratégia $s_{ik'} \in S_i$ se $u_i(s_{ik'}, s_{-i}) > u_i(s_{ik}, s_{-i})$, para todo $s_{-i} \in S_{-i}$ (SARTINI et al., 2004).

Estratégia fracamente dominante é aquela que domina fracamente todas as outras estratégias do jogador. Racionalmente será a estratégia escolhida. A estratégia s_1 domina fracamente a estratégia s_2 de um jogo se, dada qualquer coleção de estratégias que poderiam ser adotadas pelos demais jogadores, a adoção de s_1 nunca resulta em uma recompensa mais baixa para este jogador do que adotar s_2 e, em pelo menos um caso, s_1 der ao jogador uma recompensa estritamente mais alta do que s_2 . A diferença entre uma estratégia estritamente dominante e uma fracamente dominante é que a primeira possui uma recompensa **maior** que todas as outras, enquanto a última tem uma recompensa cujo valor é **maior ou igual** a outras estratégias. Um jogador racional raramente adotará uma estratégia fracamente dominada. O perfil de estratégias $s = \{s_1, s_2, \dots, s_n\}$ é um equilíbrio de estratégia fracamente dominante se, para qualquer jogador i , a estratégia s_i for fracamente dominante (BIERMAN; FERNANDEZ, 2011). Formalmente a estratégia $s_{ik} \in S_i$ é fracamente dominada pela estratégia $s_{ik'} \in S_i$ se $u_i(s_{ik'}, s_{-i}) \geq u_i(s_{ik}, s_{-i})$, para todo $s_{-i} \in S_{-i}$ (SARTINI et al., 2004).

Uma estratégia estritamente dominada iterada ocorre depois que a matriz de recompensas sofreu pelo menos uma iteração (eliminação de alguma estratégia dominada). Formalmente essa situação ocorre se estratégia em questão for a única em $\hat{S}_i$, onde $\hat{S}_i$ é a intersecção da seguinte sequência de conjuntos aninhados de estratégias: $\{1\}$ S_i^1 consiste de todas as estratégias do jogador g_i que não são estritamente dominadas; $\{2\}$ para $n > 1$, S_i^n consiste das estratégias S_i^{n-1} que não são estritamente dominadas quando restringimos os outros $j \neq i$ jogadores às estratégias em S_j^{n-1} . O perfil de estratégias $(s_1, s_2, \dots, s_n)$ é um equilíbrio de estratégia estritamente dominante iterada se, para qualquer jogador g_i , s_i é uma estratégia estritamente dominada iterada (BIERMAN; FERNANDEZ, 2011).

O resultado O de um jogo é Pareto dominado se houver outro resultado O' , tal que: $\{1\}$ todo jogador prefere estritamente O' a O ou é indiferente entre O' e O ; e $\{2\}$

algum jogador prefere estritamente O' a O . Este tipo de comunicação pode levar um jogo ao resultado denominado ótimo de Pareto, bastante comum em jogos cooperativos. Bierman e Fernandez (2011) definem que um resultado é ótimo de Pareto se não for Pareto dominado por qualquer outro resultado do jogo. De forma básica, a situação ótimo de Pareto ocorre quando os jogadores não podem melhorar suas recompensas sem afetar negativamente as recompensas dos outros.

Como dito anteriormente, além da dominância, outro conceito usado para solucionar jogos estratégicos é o EN.

O EN é o estado de um jogo que leva ao mesmo resultado, independente do momento em que for jogado. Isso acontece porque o EN oferece a melhor recompensa que o jogador g_i pode obter mediante as decisões dos outros jogadores. Em suma, o EN ocorre quando uma escolha $s_{ik} \in S_i$ é realizada e os outros jogadores não podem obter uma recompensa melhor usando uma estratégia diferente – não há, portanto, motivação para a troca de estratégias dos jogadores (SARTINI et al., 2004; OSBORNE, 2004). Formalmente, um perfil de estratégias $s^* = (s^*_1, \dots, s^*_{(i-1)}, s^*_{(i+1)}, \dots, s^*_n) \in s$, é um EN se $u_i(s^*_i, s^*_{-i}) \geq u_i(s_{ik}, s^*_{-i})$, para todo $i=1, \dots, n$ e para todo $k=1, \dots, m_i$, com $m_i \geq 2$.

2.2.3 Jogos Cooperativos

Na definição de Stuart Jr (2001) quando há uma função que especifica qual valor pode ser gerado a partir de um subconjunto de jogadores, temos um jogo cooperativo. Isso não significa que os agentes de um jogo cooperativo não podem tomar decisões individuais. Segundo Brandenburger (2007), a realização de acordos que tendem a ser benéficos para todos os jogadores dá aos jogos cooperativos um caráter combinatório. Essas definições são importantes, pois revelam uma característica importante dos jogos cooperativos: a busca por acordos racionais que beneficiem todos os envolvidos, ainda que isso ocorra por meio de decisões individuais.

De acordo com Tan, Lie e Member (2001), em jogos cooperativos a ideia principal não é deixar que um determinado agente sacrifique seus próprios interesses em benefício de outros jogadores, mas sim que as ações sejam coordenadas no sentido

de beneficiamento mútuo, fato que significa também o interesse individual. De acordo com Khan e Ahmad (2009) os jogos cooperativos oferecem algumas vantagens em relação aos não cooperativos: {1} não é necessário especificar os movimentos e as sequências de decisões dos jogadores, já que se busca a cooperação mútua, a melhor sequência de decisões tende a prevalecer; e {2} costumam ser mais equilibrados, já que convergem para uma solução adequada aos jogadores.

A análise de jogos estratégicos, mesmo os cooperativos, pode ser bastante complexa. Um ponto de tensão são as tentativas individuais de maximização dos ganhos que podem resultar em perdas coletivas. De acordo com Samieh e Wahba (2007) há pelo menos três momentos nos quais a análise de um jogo cooperativo deve ser cuidadosa: {1} quando há problemas de cooperação: jogos onde a pior opção é a busca pelo ganho individual, por exemplo, o jogo clássico do dilema do prisioneiro⁹; {2} quando há jogos de soma zero e de diferenciação: jogos nos quais um jogador escolhe uma estratégia necessariamente diferente do outro jogador; por exemplo, o jogo clássico da galinha¹⁰; e {3} quando há jogos de coordenação: jogos nos quais a possibilidade de obter ganhos com a cooperação é baixa e a opção pela busca da individualidade é viável.

2.2.4 Situações de Barganha

Um tópico proveniente da Teoria dos Jogos Cooperativos, bastante interessante e útil para o desenvolvimento deste trabalho é o conceito de barganha. Uma barganha entre dois jogadores é uma negociação na qual cada um visa obter a melhor vantagem possível, tendo em vista que o aumento da vantagem de um jogador, na

⁹ Um conhecido jogo de estratégia que conta com dois prisioneiros cúmplices que devem tomar decisões sem se comunicarem. Se um deles confessar (traí o companheiro) e o outro não, o que confessou fica livre enquanto o segundo cumpre 10 anos de pena. Se ambos não confessarem (colaboram um com o outro) ambos pegam apenas um ano de pena. Se os dois confessarem, a pena de ambos será de cinco anos. Dias (2005) afirma que há um resultado que é melhor para ambos (ninguém confessa), mas existe um incentivo para desvio dessa estratégia e o resultado é um EN pior para ambos. Em jogos infinitamente repetidos pode-se escapar desse EN pior.

¹⁰ Um conhecido jogo de estratégia no qual dois jovens dirigem seus carros um contra o outro. A galinha é o primeiro que desviar, evitando a colisão. Se ambos mantiverem o curso, ambos são prejudicados pela colisão. Se ambos desviarem o curso, ambos perdem o jogo.

maior parte das vezes, implica na diminuição da vantagem do outro. Apesar da ênfase na negociação e na busca por vantagens, trata-se de um jogo cooperativo.

Uma situação de barganha entre dois jogadores diz respeito a: {1} colaboração mútua para benefício de ambos; {2} há um conflito sobre a conclusão deste acordo que tende a beneficiá-los; e {3} nenhum dos lados pode impor condições sem aprovação (NASH JR, 1950; OSBORNE; RUBINSTEIN, 1990). Nada impede, porém, que aconteçam jogos cooperativos sem a ocorrência de barganhas.

Para que uma situação de barganha em um jogo cooperativo ocorra, pelo menos três exigências são essenciais: {1} a racionalidade; {2} condições iguais de barganha; e {3} o total conhecimento de preferências entre os jogadores (OSBORNE; RUBINSTEIN, 1990). Note que essas três exigências são perfeitamente aplicáveis em um ambiente onde haja controles de segurança gerenciados por um administrador.

Dois conceitos bastante importantes que servem para formalizar situações de barganha são as antecipações e as funções utilidade.

Uma antecipação é um estado de expectativa que algumas contingências podem gerar. Alguns destes estados podem ser estabelecidos como certeza, enquanto alguns são probabilísticos (NASH JR, 1950). Por exemplo, um cliente está determinado a comprar um carro ou uma moto. Nessa situação, há uma antecipação de $\frac{1}{2}$ carro e $\frac{1}{2}$ moto. De acordo com Nash Jr (1950), se A e B representam duas antecipações e há uma probabilidade p para a antecipação A e uma probabilidade $(1-p)$ para a antecipação B , onde $0 \leq p \leq 1$, então a antecipação pode ser definida por $pA + (1-p)B$, que representa uma combinação de probabilidades das duas antecipações.

A partir do conceito de antecipações (NASH JR, 1950), as propriedades da teoria da utilidade para um único indivíduo podem ser definidas da seguinte forma:

- Um indivíduo com duas possíveis antecipações pode decidir qual é preferível ou se ambas são desejáveis;
- A ordem produzida é transitiva: se A é melhor que B e B é melhor que C , então A é melhor que C ;

- As combinações de probabilidades de estados desejados iguais são equivalentes;
- Continuidade: se A , B e C tem ordem transitiva, isto é A é melhor que B e C , então há uma combinação de probabilidades de A e C que é tão desejável quanto C ;
- Se $0 \leq p \leq 1$ e A e B são igualmente desejáveis, então $pA + (1-p)C$ e $pB + (1-p)C$ são igualmente desejáveis. Também se A e B são igualmente desejáveis, A pode ser substituído por B em qualquer outra relação que seja satisfeita por B .

Essas propriedades mostram a existência de uma Função Utilidade Transferível¹¹ (FUT) com um número real definido para cada antecipação. De acordo com Nash Jr (1950), com as letras maiúsculas representando antecipações e as minúsculas representando números reais, uma FUT tem as seguintes propriedades:

- Não é única, sendo u uma função, $au + b$ (para $a > 0$) também é uma função utilidade;
- $u(A) > u(B)$ equivale a: A é mais desejável que B ;
- A linearidade da função utilidade funciona tanto para os casos que envolvem uma pessoa como uma dupla: se $0 \leq p \leq 1$, então $u[pA + (1-p)B] = pu(A) + (1-p)u(B)$.

O uso da função utilidade é importante quando não há cooperação entre os indivíduos em uma barganha e os jogadores assumem antecipações zeradas. Nessas situações as funções utilidade individuais são multiplicadas por números reais positivos e passam a ser passíveis de escolha¹² (NASH JR, 1950).

Apesar da negociação sobre um determinado valor de recompensa envolver uma divisão¹³, na qual cada jogador busca obter o ganho máximo (ganho que implica na diminuição da recompensa do outro jogador), as barganhas não são jogos de soma

¹¹ Alguns autores denotam a FUT simplesmente como função utilidade.

¹² Nash Jr (1950) sugere a criação de um gráfico que mostre as duas antecipações escolhidas pelos jogadores. Este gráfico deve conter todas as antecipações possíveis, ser compacto e convexo. Convexo: uma antecipação marcada em uma posição qualquer de um segmento de reta que esteja entre dois pontos do conjunto pode sempre ser obtida pela combinação de probabilidade destas duas antecipações.

Compacto: esta propriedade indica que os pontos devem estar ligados dentro de um plano quadrado e que qualquer função de utilidade contínua assume o valor máximo em algum ponto do conjunto.

¹³ Essa divisão envolve o montante total gerado no jogo. Este montante indica o valor máximo que pode ser gerado em um jogo por meio das negociações entre os agentes. Para que um processo de barganha seja bem sucedido este montante deve ser maior ou igual a soma das recompensas individuais obtidas pelos jogadores envolvidos (DIXIT, 2004). Este assunto é desenvolvido de forma mais completa na Seção 2.3.3.

zero. O motivo disso é que há um incentivo fortíssimo para a negociação ser bem sucedida: **se a barganha (negociação) falhar, nenhum deles vai obter ganho** (DIXIT, 2004).

Os conflitos em processos de barganha podem se manifestar por meio da {1} não aceitação; {2} da incomparabilidade de critérios; e {3} pela incerteza. A condição mais natural para ocorrência de conflitos em um ambiente de gerenciamento de controles de segurança é a incerteza, já que envolve a falta de entendimento comum sobre as consequências de uma ação. Casos como esse podem ocorrer, por exemplo, quando dois controles de segurança apresentam diagnósticos diferentes do ambiente (AZEVEDO, 2001).

Problemas de incomparabilidade, apesar de mais raros, podem acontecer quando, por exemplo, os resultados dos controles de segurança não são passíveis de comparação. Já a condição de não aceitação praticamente inexistente em um ambiente de segurança gerenciado por um administrador.

Para que a definição de barganha seja completamente entendida é importante definir o conceito de equilíbrio perfeito em subjogos. Em um jogo no qual todas as decisões são conhecidas (jogo de informação perfeita), um subconjunto dos nós (estratégias) e ramos do jogo original, que constitui um jogo em si, é denominado um subjogo. Quando um perfil de estratégias de EN continua a ser um EN quando aplicado em um subjogo, tem-se então o equilíbrio perfeito em subjogos (BIERMAN; FERNANDEZ, 2011). Duas definições complementares são fornecidas por Osborne (2004): {1} sendo h um trecho de um jogo, um subjogo subsequente é justamente a parte que resta depois que h ocorre; e {2} um equilíbrio perfeito em subjogos corresponde a um perfil de estratégias s^* que não contém subjogos nos quais o jogador i possa ter uma recompensa maior a partir da escolha de uma estratégia diferente s_i^* , dado que o outro jogador usa uma estratégia s_j^* . Portanto, o equilíbrio perfeito em subjogos é um conjunto de estratégias que leva ao EN em todos os subjogos.

A figura 1 apresenta um exemplo, baseado no original de Osborne (2004), que ilustra bem uma situação de equilíbrio perfeito em subjogos.

Figura 1 - Exemplo de equilíbrio perfeito em subjogos

Fonte: Baseado em Osborne (2004).

A figura 1 mostra as recompensas disponíveis a dois jogadores rivais. O JOGADOR 1 é um desafiante, que pode invadir a área do JOGADOR 2 (estratégia DENTRO) ou simplesmente não invadir (estratégia FORA). O JOGADOR 2 também tem duas estratégias: CONCORDA ou LUTA. A figura 1 mostra que a sequência (DENTRO, FORA) é um equilíbrio perfeito em subjogos, pois as escolhas do jogador 1 são otimizadas a partir das estratégias do jogador 2 – tanto no jogo completo, como no subjogo escolhido a partir da opção DENTRO.

Porém, ainda em relação à figura 1, o EN (FORA, LUTA) não é um equilíbrio perfeito em subjogos porque a partir da estratégia DENTRO, a estratégia LUTA não é a melhor escolha.

Situações de barganha podem envolver negociações longas com muitas rodadas. Nestes cenários as negociações podem ser realizadas com ou sem paciência (BIERMAN; FERNANDEZ, 2011).

A barganha com paciência ocorre quando os jogadores não se importam em estender a negociação por diversas rodadas. Costuma ser influenciada pela quantidade e pela ordem das contraofertas oferecidas na negociação (BIERMAN; FERNANDEZ, 2011). Como um exemplo, imagine uma situação de barganha na qual o jogador $J1$ está disposto a pagar no máximo \$300 (preço de reserva de $J1$) por um serviço realizado pelo jogador $J2$, que não aceita como pagamento um valor abaixo de \$200 (preço de reserva de $J2$). A diferença de \$100 entre os preços de reserva deve ser dividida entre os jogadores por meio de uma negociação. Se $J1$ fizer a primeira oferta ($\$200 \leq P1 \leq \300) e $J2$ aceitar a barganha acabou com $P1$ aceito. Se $J2$ recusar $P1$ e fizer uma contraoferta $P2$ ($\$200 \leq P1 < P2 \leq \300), cabe ao jogador $J1$ aceitá-la, interrompendo a barganha, ou recusá-la dando continuidade a negociação. A figura 2 mostra um exemplo de barganha com paciência.

Figura 2 - Exemplo de barganha com paciência

Fonte: Baseado em Bierman e Fernandez (2011).

Na figura 2 está ilustrada a dinâmica do jogo. Usando indução retroativa¹⁴ percebe-se que o último a fazer uma contraoferta tem vantagens.

Em muitos processos de barganha há um montante (custo ou desconto) que recai sobre cada jogador a cada contraoferta oferecida. Este montante costuma ser associado ao tempo necessário para a realização de uma nova rodada de negociações. Como esse custo é imposto a cada rodada de negociação os jogadores tem grande interesse em limitar a quantidade de rodadas. Tem-se nesses casos situações de barganha com impaciência. Para que a contabilização deste montante seja realizada da melhor forma possível, algumas informações, como o conhecimento sobre quem iniciou a negociação, a quantidade de contraofertas possíveis e os limites mínimos e máximos que podem ser negociados por cada jogador, são essenciais (BIERMAN; FERNANDEZ, 2011).

Sendo t uma rodada de negociações, q o ganho do jogador e x o desconto (ou custo) que este jogador tem ao negociar mais uma rodada, verifica-se que os ganhos de uma negociação futura $t+1$ tem um percentual de desconto (ou custo) de $\delta = (1 - x)\%$ que recai sobre q .

Dessa forma, em uma barganha entre dois jogadores C e V , na qual: {1} o jogador C é o iniciante; {2} a quantidade de contraofertas alternadas é ilimitada; {3} os jogadores tem um percentual de desconto (ou custo) de $0 < \delta < 1$ em relação à q ; e {4} está definido que ambos aceitam as contraofertas no momento em que para eles é indiferente a aceitação ou rejeição das mesmas, o teorema de Rubinstein determina que nesta barganha haja um único equilíbrio perfeito em subjogos, onde C oferece uma contraoferta prontamente aceita por V : $\frac{\delta_V(1-\delta_C)}{1-\delta_V\delta_C}$ do excedente é oferecido a V , enquanto C retém $\frac{1-\delta_V}{1-\delta_V\delta_C}$ para si (BIERMAN; FERNANDEZ, 2011).

É interessante notar que uma situação de barganha só vai ocorrer em um jogo cooperativo se um montante ou valor for criado a partir das interações que ocorrem entre os jogadores participantes. Esse é um ponto fundamental para que se entendam os conceitos associados a situações de UCB.

¹⁴ Esta análise ocorre de baixo para cima. Na definição de Osborne (2004) a indução retroativa ocorre quando um jogador que precisa fazer um movimento deduz, para cada um dos seus possíveis movimentos, as ações que outros jogadores (e ele próprio) vão tomar e consequentemente ele acaba por escolher o seu movimento preferido.

2.2.5 Unrestricted Cooperative Bargains (UCB)

O Total Value Created (TVC) corresponde ao valor total (máximo) que pode ser gerado como recompensa em um jogo, sem que sejam levadas em conta as estratégias dos jogadores. Em um jogo cooperativo o TVC é contabilizado de forma vertical, ou seja, envolve toda a cadeia de jogadores. Um exemplo bastante esclarecedor é mostrado no trabalho de Brandenburger e Stuart Jr (1996), que aborda a relação entre fornecedores, firmas, e compradores e pode ser visto na figura 3:

Figura 3 - TVC e o conceito de compartilhamento de valores

Fonte: Baseado em Brandenburger e Stuart Jr (1996).

O exemplo da figura 3 ilustra um jogo cujo TVC representa o valor máximo que pode ser gerado em uma transação entre custo (fornecedores), preço (firmas), e Willingness To Pay (WTP) (compradores)¹⁵. No jogo estratégico traçado, o processo

¹⁵ Cada jogador possui uma forma individual para geração de valor em um jogo cooperativo. Por exemplo, o WTP indica o valor máximo que os compradores estão dispostos a pagar por produtos oferecidos pelas firmas, enquanto que a diferença entre o preço e o custo indica o valor mínimo que as firmas aceitam receber pelos produtos. Da mesma forma, a diferença entre o custo e o custo de oportunidade indica o valor mínimo que os fornecedores aceitam receber no fornecimento de recursos.

de UCB procura definir: $\{1\}$ como acontece a divisão de valores entre os jogadores; e $\{2\}$ como os jogadores alcançam determinados valores.

A descrição formal de TVC é dada por Brandenburger e Stuart Jr (1996) e por Stuart Jr (2001). Seja FUT uma função utilidade de um jogo cooperativo composto por um conjunto finito N , mapeado por $v:2^N \rightarrow \mathbb{R}$, onde N representa o conjunto de jogadores e v representa uma função característica especificada para o conjunto de jogadores. Sendo assim, o TVC gerado entre os jogadores de um conjunto N é denotado por $v(N)$ enquanto que para um conjunto SS , onde $SS \subseteq N$, o TVC é denotado por $v(SS)$.

De acordo com Stuart Jr (2001), um jogo cooperativo (N, v) apresenta como resultado um vetor $x \in \mathbb{R}^{|N|}$ que especifica o TVC e como ele é dividido. O componente x_i indica o valor capturado pelo jogador i .

Definição 1: entende-se, portanto, que para qualquer $SS \subseteq N$ o Individual Value Created (IVC) pode ser denotado por $\sum_{i \in SS} x_i = x(N) = v(N)$. Sendo assim, o núcleo (do inglês *core*) de uma FUT em um jogo cooperativo (N, v) é o conjunto que satisfaz $x(N) = v(N) \wedge SS \subseteq N, x(SS) \geq v(SS)$ (STUART JR, 2001). Para que essa divisão de valores ocorra é necessário compreender o conceito de valor agregado (*added value*). O valor agregado ou Contribuição Marginal (CM) corresponde ao valor gerado por um determinado jogador em um jogo. De acordo com Brandenburger e Stuart Jr (1996) o termo foi criado com o objetivo de funcionar como um indicador corporativo para medir o valor com que a empresa contribuirá para a economia em um futuro próximo.

Definição 2: segundo Brandenburger (2007) a CM de um subconjunto $SS \subset N$ é definido como $v(N) - v(N \setminus \{i\})$, onde o termo $v(N \setminus \{i\})$ corresponde ao valor criado em um jogo sem a presença do jogador i (BRANDENBURGER, 2007; STUART JR, 2001).

Definição 3: o Princípio da Contribuição Marginal define que se em um jogo um jogador i obtiver menos valor do que obteria sozinho, o ideal para i é que ele nem entre no jogo. Percebe-se que a ausência do jogador i , nestas circunstâncias é benéfica para os outros jogadores que tendem a ter um montante maior para ser dividido: $x_i > v(N) - v(N \setminus \{i\})$.

Definição 4: o Princípio da Contribuição Marginal também define que jogador algum pode capturar mais valor que a CM que ele gera: $\forall i \in N, x_i \leq v(N) - v(N \setminus \{i\})$, if $x(N) = v(N)$ (STUART JR, 2001).

A divisão do TVC entre os jogadores recebe o nome de **alocação** e é formalmente definida por Brandenburger (2007):

Definição 5: dado um jogo cooperativo (N, v) uma alocação corresponde a uma coleção de números $(x_1, x_2, \dots, x_n)$.

Definição 6: uma alocação é **individualmente racional** se para todo i , $x_i \geq v(\{i\})$. Ou seja, um jogador deve obter no mínimo o mesmo valor que receberia sem interagir com os outros jogadores.

Definição 7: uma alocação é eficiente se $\sum_{i=1}^n x_i = v(N)$. Ou seja, todo o TVC que puder ser criado será de fato criado.

Definição 8: uma alocação individual, racional e eficiente satisfaz o Princípio da Contribuição Marginal se $x_i \leq CM_i$ para todo i .

Para que um jogador se aproprie de sua CM, esta deve ser positiva, ou seja, deve haver assimetrias positivas ¹⁶ entre um jogador e seus concorrentes (BRANDENBURGER; STUART JR, 1996) - desde que seja respeitado, como especificado no Princípio da Contribuição Marginal, que não é possível para um jogador capturar valores acima da CM que ele gera.

Depois de definidos os valores que os jogadores criam na cadeia vertical, a barganha entre eles determina o quanto cada um vai capturar¹⁷ ao longo do jogo.

De acordo com Brandenburger e Stuart Jr (1996), quando a soma das CM iguala o TVC do jogo, então a barganha é determinada. Porém, quando a soma das CM ultrapassa o TVC do jogo, então se tem a barganha indeterminada. É o que ocorre no exemplo a seguir, ilustrado na figura 4.

¹⁶ São os motivos que levam diferentes jogadores a se apropriar de diferentes valores dentro de um jogo. Por exemplo, um produto ou serviço diferenciado oferecido por um jogador, que eleva o preço de venda do produto dele e consequentemente acaba gerando mais valor a este jogador no jogo.

¹⁷ Em um cenário com muitos jogadores nem sempre a CM depende inteiramente da situação de barganha. Isso ocorre porque alguns valores são gerados, em parte, por expectativas criadas fora do jogo. Por exemplo, uma determinada assimetria como propaganda boca a boca que leva um produto a ser mais caro que outro, nem sempre está dentro do jogo. O trabalho de Brandenburger e Stuart (1996) apresenta algumas práticas que podem minimizar ou mesmo eliminar esses efeitos externos e levar a solução em situações de barganha. Estas práticas, contudo, não são mostradas aqui.

Seja um jogo entre dois fornecedores, duas firmas e um comprador. Trata-se de uma cadeia vertical: o fornecedor entrega a matéria prima às firmas; as firmas transformam a matéria prima em produto; o comprador compra o produto das firmas. Cada fornecedor só pode realizar transações com uma única firma. Da mesma forma, o comprador só pretende comprar de uma única firma. O custo de cada fornecedor é de \$10, ou seja, não há diferenças na escolha dos fornecedores. O comprador está disposto a pagar pelos produtos das firmas 1 e 2 respectivamente \$100 e \$150. A figura 4 ilustra a divisão de valores desse jogo.

Figura 4 - TVC e o conceito de valor compartilhado em uma barganha indeterminada

Fonte: Baseado em Brandenburger e Stuart Jr (1996).

No exemplo ilustrado na figura 4 o TVC (valor total que pode ser gerado no jogo) é a subtração entre o máximo que o comprador está disposto a pagar (\$150) e o custo de qualquer um dos fornecedores (\$10): $\$150 - \$10 = \$140$. A transação do comprador deve acontecer com a firma2, pois esta gera mais valor no jogo (\$150) que a firma1 (\$100).

A alocação ou divisão dos valores mostrados na figura 4 tem como base a CM, (vide as definições 2, 3 e 4). Sendo assim: {1} a CM de cada fornecedor é \$0; {2} a CM da firma1 é \$0, pois não realiza transações com o comprador; {3} a CM da firma2 é \$50, que corresponde à diferença entre os valores gerados pela firma1 (\$100) e pela firma2 (\$150). Isso indica que dependendo da barganha que deve ocorrer no jogo, a firma2 é responsável por geral um valor entre \$0 e \$50; e {4} a CM

do comprador é \$140, o que indica que a partir da barganha, o comprador deve gerar um valor entre $(\$140 - \$50) = \$90$ e \$140. **Portanto, é importante ressaltar que em uma situação UCB, o jogador que tiver a maior CM é o vencedor da negociação.**

Na verdade nem sempre um jogador vai ter como CM o valor que ele cria, porque isso depende do resultado de barganhas irrestritas entre os jogadores (BRANDENBURGER; STUART Jr, 1996). Sendo assim, a CM funciona como um limite máximo que pode ser alcançado por um determinado jogador.

Em um jogo cooperativo, todas as vezes que os jogadores estão em busca de transações favoráveis, ativamente envolvidos na criação e divisão de valores, ao mesmo tempo em que nenhum destes jogadores pode individualmente determinar o valor de preços e custos, temos uma situação de UCB (STUART JR, 2001). Em resumo, uma situação de UCB se consolida quando a divisão dos valores gerados em um jogo acontece especificamente pela negociação. Ao contrário, em situações de barganhas restritas, nas quais não se observam os princípios encontrados em situações de UCB, há dificuldades no estabelecimento da CM que pode não representar a correta divisão de valores (por exemplo, conflito em jogos nos quais o TVC é superior à soma do IVC) (BRANDENBURGER; STUART JR, 1996). Portanto, situações de barganhas restritas não são ideais para cenários nos quais a negociação deve prevalecer na divisão de valores.

O conceito de UCB é tratado de forma bastante clara nos trabalhos de Brandenburger e Stuart Jr (1996, 2007), Stuart Jr (2001) e Brandenburger (2007) por meio da definição de dois princípios bastante próximos e complementares: o Princípio do Core (núcleo) e o já citado Princípio da Contribuição Marginal. Ambos os princípios podem ser usados em conjunto e levam a uma mesma solução de barganha.

O Princípio do Core considera a importância da coalizão entre os jogadores: se um jogador obtém sozinho mais valor que obteria no jogo corporativo, o ideal é que este jogador nem entre no jogo: $x(SS) < v(SS)$. Da mesma forma, se um jogador i for receber mais do que pode gerar: $x_i > v(N) - v(N \setminus \{i\})$ o ideal, para os outros jogadores, seria excluí-lo do jogo (STUART JR, 2001) (veja o princípio da Contribuição Marginal (BRANDENBURGER, 2007)).

Como dito anteriormente, uma alocação $(x_1, x_2, \dots, x_n)$ está no núcleo (*core*) do jogo se for individual, racional e eficiente e para qualquer subconjunto SS de N , $x(SS) \geq v(SS)$. Além disso, uma alocação eficiente está no *core* se, e apenas se, para todo subconjunto SS de N , $x(SS) \leq CM_s$ ¹⁸.

Em geral, ao invés de especificar um único valor, o Princípio do Core especifica uma faixa de valores que um jogador pode obter como ganho em um jogo. A menor faixa representa o mínimo que este jogador pode obter, enquanto a amplitude até o valor máximo representa a barganha residual: possíveis valores que podem ser divididos ou capturados pelo jogador durante a negociação (STUART JR, 2001).

O Princípio da Contribuição Marginal baseia-se mais em um determinado valor fixo mínimo. Se for possível ao jogador aumentar este valor, a diferença entre o mínimo e o quanto se obteve de aumento é justamente a barganha residual.

Pode-se perceber, portanto, que ambos os Princípios (Core e Contribuição Marginal) usados em UCB são complementares e chegam a um mesmo resultado. Para ilustrar os conceitos sobre as diversas formas possíveis da divisão de valores em jogos cooperativos, um último conceito deve ficar claro: os jogos de atribuição de três lados.

Os jogos de atribuição de três lados são jogos corporativos ideais para demonstração de exemplos dos Princípios do Core e da Contribuição Marginal porque representam bem parte da cadeia vertical. Empresas podem ser compradoras e vendedoras ao mesmo tempo. Fazendo uma analogia ao agrupamento de mecanismos de proteção, controles de segurança podem ser benéficos para alguns recursos do sistema ao mesmo tempo em que podem ser maléficos para outros recursos. Por exemplo, o aumento de uma chave criptográfica melhora a confidencialidade, porém, em contrapartida, há um consumo maior de energia.

De acordo com Stuart Jr (2001), a formalização de um jogo de atribuição de três lados pode ser denotada da seguinte forma:

Definição 9: sejam três conjuntos isolados de jogadores $N1, N2, N3$ e uma matriz tridimensional A com dimensões $n1, n2, n3$, que representam os respectivos

¹⁸ A prova de conceito está em Brandenburger (2007) e não é demonstrada aqui.

determinantes: $n1=|N1|$, $n2=|N2|$, $n3=|N3|$. Uma interação (*matching*) é uma tripla ijk consistindo de um jogador i do conjunto $N1$, um jogador j do conjunto $N2$ e um jogador k do conjunto $N3$.

Definição 10: o elemento a_{ijk} da matriz A é interpretado como o IVC pela correspondência ijk .

Definição 11: duas correspondências $i^a j^a k^a$ e $i^b j^b k^b$ são distintas se $i^a \neq i^b$, $j^a \neq j^b$, $k^a \neq k^b$.

Definição 12: uma atribuição de tamanho r é um conjunto de r correspondências distintas.

Definição 13: na construção de jogos com m -lados: {1} a criação de valor é determinada apenas por correspondências distintas; e {2} o valor criado deve ser o maior possível. A prova segue na sequência (STUART JR, 2001):

- Seja N o conjunto formado por $N1 \cup N2 \cup N3$. A função característica v é definida por $v(\emptyset) = 0$ e para todo $T \subseteq N$;
- Nenhum valor é gerado quando não há interação entre os jogadores: $v(T) = 0$ se $T \cap N_m = \emptyset, m \in \{1,2,3\}$;
- O valor gerado por *matching* será: $v(\{i, j, k\}) = a_{ijk} \forall i \in N_1, j \in N_2, k \in N_3$;
- O TVC por uma dada coalizão (conjunto de jogadores) é calculado pelo arranjo de jogadores em um conjunto de *matchings* distintos que contém os valores mais altos: $v(T) = \max(a_{i^1 j^1 k^1} + \dots + a_{i^r j^r k^r}), r, AS_r$, onde $r \leq \min \{|T \cap N1|, |T \cap N2|, |T \cap N3|\}$, sendo que AS é o conjunto de atribuições de tamanho r construídos a partir do conjunto T .

Para ilustrar os conceitos expostos até aqui, exemplos de jogos de atribuição de três lados e cálculo UCB podem ser vistos nos APÊNDICES A e B.

Como visto, para muitas situações os conceitos relacionados a UCB podem ser úteis. Majoritariamente em situações onde há necessidade de negociações cooperativas, como as que podem acontecer nos processos de agrupamento de controles de segurança, tais conceitos podem ter bastante utilidade.

3 TRABALHOS RELACIONADOS

Neste capítulo alguns detalhes sobre os principais trabalhos referenciados são apresentados e discutidos com o objetivo de justificar a linha de pesquisa adotada na proposta do SGA.

Alguns trabalhos endossam a dificuldade existente na integração de diferentes controles de segurança voltados para a detecção de eventos maliciosos e adotam conceitos estratégicos para melhorar a precisão dos resultados. O trabalho de Zhu et al. (2009) é um exemplo, pois lembra que mecanismos de detecção individuais estão sujeitos à falhas e por isso nem sempre transmitem informações honestas. Este comportamento, contudo, tende a acontecer com menor frequência em um ambiente no qual os mecanismos de detecção trabalham de forma cooperativa. Apesar disso, de acordo com Zhu et al. (2009), quando um sistema de cooperação é habilitado ele pode não ser efetivo, já que alguns componentes ajudam mais que outros. Os autores propõe um esquema de incentivos no qual a ajuda que um componente recebe de outro deve ser equivalente à ajuda que o primeiro par pode oferecer, além de um algoritmo que converge as decisões para o EN com rapidez. Percebe-se que a pesquisa de Zhu et al. (2009) caminha bem próxima do conceito de agrupamento pelo equilíbrio dos métodos adotado no SGA.

A restrição de recursos também é um problema que pode afetar a qualidade da detecção de eventos. A fim de proteger o nó mais vulnerável de uma rede de sensores com restrições de energia, memória e processamento, Agah et al. (2004) criam um jogo de soma não zerada usando dois cenários: {1} a adoção de processos decisórios de Markov para prever qual o nó mais vulnerável; e {2} a proteção do nó com maior atividade. Os resultados da simulação mostram que a proposta utilizada pode melhorar a chance de detecção de intrusão.

A diversidade de recursos, configurações e opções também pode se tornar um complicador em alguns casos. No trabalho de Chen e Leneutre (2009) tenta-se resolver o problema de detecção de intrusões em redes heterogêneas (com diferentes itens de segurança não correlacionados). Para entender qual o comportamento esperado das estratégias do defensor e do atacante, é proposto um

jogo não-cooperativo voltado para detecção de intrusões, cuja análise do EN e das implicações relacionadas ajuda a entender o comportamento racional dos atacantes, os recursos mínimos de monitoramento e a estratégia otimizada do defensor.

Há diversos trabalhos que procuram realizar o agrupamento de mecanismos de detecção com o intuito de minimizar os problemas decorrentes dos processos de detecção de eventos e servem para corroborar alguns conceitos sugeridos no SGA. Por exemplo, a pesquisa conduzida por Srinivas, Andrew e Ajith (2005) confirma que o uso conjunto de técnicas de agrupamento de mecanismos de detecção pode melhorar os resultados individuais e a pré-classificação de ataques. Outro exemplo pode ser visto no trabalho de Sagale e Kale (2014) que agrupa algoritmos. Porém, ambos os trabalhos tratam apenas de mecanismos de detecção.

Um fator importante nos processos de agrupamento de controles de segurança é a igualdade de pesos. Segundo Armstrong (2001), além desta igualdade ser um ponto de partida interessante quando se está combinando muitos métodos, a diferença entre pesos é útil quando se sabe qual o mecanismo mais preciso e há um bom domínio tecnológico sobre este método. O assunto também é abordado por Huang e Lee (2010) que fazem uma distinção interessante entre combinação de previsões e combinação de informações:

1. Combinação de previsões corresponde ao agrupamento formal de mecanismos de previsão, com vários esquemas de ponderação, que se ocupam de divisões de um mesmo conjunto de dados;
2. Combinação de informações corresponde à junção de informações para a construção de um modelo único e completo de previsão (nesta abordagem os autores minimizam os problemas de proliferação de parâmetros e de componentes principais).

Por meio de simulações e análises, Huang e Lee (2010) comparam técnicas e chegam à conclusão que a metodologia de combinação de previsões com pesos próximos teve um desempenho melhor que a combinação de informações. No trabalho de Lin et al. (2013) diferentes SDI/SPI são agrupados tendo como base a atribuição de pesos a diferentes tipos de ataques. O objetivo dos autores é diminuir os índices de FP e FN. Diante dessas constatações, é importante ressaltar que o SGA suporta a igualdade e a diferença de pesos durante a análise estratégica.

Partindo de outro tipo de abordagem, os trabalhos realizados por Giacinto e Roli (2002, 2003) têm como base o reconhecimento de padrões para agrupar mecanismos de detecção de intrusão baseados em uso indevido (assinaturas) com mecanismos baseados na análise de anomalias. Diversas informações sobre o fluxo de dados são levantadas, categorizadas e por fim analisadas em uma matriz de recompensas. Em uma evolução do trabalho anterior, Giacinto; Roli E Didaci (2003) usam novamente o reconhecimento de padrões para agrupar múltiplos classificadores e afirmam que a generalização que este tipo de ferramenta oferece pode ser útil no desenvolvimento de SDI/SPI avançados. Um alerta interessante contido no referido trabalho refere-se à dificuldade na criação de assinaturas que combinem mais de um tipo de controle de segurança, reafirmando as dificuldades existentes neste tipo de processo. Contudo, ao contrário da proposta sugerida no presente trabalho, que analisa os resultados individuais dos controles de segurança, agrupando-os de forma estratégica, a metodologia adotada por Giacinto e Roli (2002) tem um caráter mais genérico que analisa os dados de forma agrupada.

Na verdade, há uma grande variedade de trabalhos que focam os dados de origem dos controles de segurança e buscam agrupar métodos de classificação para detecção de tráfego malicioso. Boa parte destes trabalhos utiliza métodos de aprendizado de máquina ou técnicas para criação de *cluster* de dados.

Entre os exemplos que usam aprendizado de máquinas podem ser citados os trabalhos de Govindarajan (2014) e Kuang e Zhang (2014). Govindarajan (2014) usa a metodologia de combinação (*ensemble*) em conjunto com Support Vector Machines (SVM) e técnicas de função de base radial para combinar classificadores homogêneos e heterogêneos. No trabalho de Kuang e Zhang (2014) SVM e técnicas de Kernel Principal Component Analysis (KPCA) são combinadas para detectar intrusões. Outro exemplo é o trabalho de Zhang et al. (2013) que procura detectar e combinar fluxos de informações desconhecidas com tráfego real por meio de correlação, suprimindo assim um dos principais problemas dos métodos de aprendizado de máquina: a necessidade de informações em aplicações semi supervisionadas. Duas variações interessantes nessa linha de pesquisa são os trabalhos de Hung; Hung e Lin (2014) que agrupa SVM e técnicas de médias móveis para prever saques em dinheiro em caixas eletrônicos; e o trabalho de Prasad e Ramakrishna (2014) que usa um sistema inteligente de aprendizado de máquinas

para reconhecer imagens aéreas e prever picos de tráfego urbano por meio de árvores de decisão. Porém, nenhum dos trabalhos referenciados procura combinar mecanismos de detecção com mecanismos de previsão com o objetivo de melhorar a precisão dos resultados obtidos, da forma como este trabalho sugere.

Em relação aos trabalhos que utilizam a combinação de técnicas e dados para formação de *clusters* podem ser citados os trabalhos de Juvonen e Sipola (2013) que combina mapas de difusão e técnicas de extração de regras para a criação de um detector anômalo baseado em *clusters*; Gao e Wang (2014) que combina a entropia das informações com o algoritmo *k-means* a fim de conseguir mais precisão na definição dos *clusters* centrais iniciais; Song; Zhu e Price (2014) que compõe *clusters* por meio da combinação da Teoria da Informação Mútua com o algoritmo Fuzzy C; e finalmente Yao (2014) que agrupa métodos de classificação por meio da criação de análise de evidências e da criação de *clusters*, com a intenção detectar intrusões. Da mesma forma que nos trabalhos referenciados sobre aprendizado de máquinas, nas pesquisas sobre combinações para a criação de *clusters* não há menção à união de mecanismos de detecção e de previsão.

É fato, portanto, que o cuidado com os dados analisados deve ser levado em conta quando se pretende agrupar controles de segurança. Diante disso, o trabalho de Chebrolu; Abraham e Thomas (2005) vai um pouco além, focando no importante conceito que nem todos os dados analisados no processo de detecção são importantes, pois podem não ter correlação uns com os outros, ser redundantes ou mesmo desprezíveis. Os autores afirmam não haver modelos formais criados para esta tarefa e propõem a seleção dos dados por meio de técnicas de mineração (*data mining*). Dois métodos de detecção, redes Bayesianas e árvores de regressão e classificação são agregados com resultados positivos.

Ainda mostrando a importância que os dados e configurações de origem têm no resultado final de um trabalho de análise e agrupamento, Lipovetsky e Conklin (2001) mostram que variáveis independentes com viés podem levar a resultados falsos. Os autores também chamam a atenção para a multicolinearidade existente entre as variáveis independentes analisadas. Na verdade, a colinearidade não prejudica a previsibilidade, mas pode influir negativamente nos resultados estatísticos. Da mesma forma, a pesquisa realizada por Ekel et al. (2011) trata da

multiplicidade de decisões tomadas sobre incertezas presentes nas informações. Os autores utilizam-se da abordagem de Bellman-Zadeh com modelagem Fuzzy, além de matrizes de recompensas para melhorar os resultados. Segundo os autores estas matrizes fazem parte de uma abordagem clássica para análise estratégica em modelos nos quais há incertezas sobre as informações, pois permitem combinações de diferentes soluções e estados.

Mesmo de forma marginal, o SGA aborda o problema da importância dos dados dedicando um módulo para verificação de influência que as variáveis independentes exercem nas variáveis resposta.

Há alguns trabalhos que trazem conceitos interessantes, mas que por focarem exclusivamente na estatística ou nas técnicas de previsão são mais limitados que o SGA. Entre os exemplos, pode ser citado o trabalho de Ghasemigol; Ghaemi-Bafghi e Takabi (2016) que apresenta uma metodologia para previsão de eventos indesejados baseada em grafos. O método proposto lida com fatores como incerteza das probabilidades dos ataques usando informações adicionais como alertas sobre intrusões e respostas ativas. Já o trabalho realizado por Beauquier, e Hu (2007) adota algumas técnicas estatísticas e combinam diferentes métodos de detecção usando a correlação de Pearson para medir a distância entre os comportamentos analisados. A premissa utilizada indica que quanto mais distante um comportamento estiver do comportamento normal, maior a probabilidade de se tratar de uma intrusão. Apesar de um ligeiro aumento no índice de alertas FP, a análise de medidas de estatísticas de dissimilaridade melhorou os índices de detecção. Já o trabalho realizado por Lin et al. (2013) se encontra no limite entre a simples detecção e o agrupamento de controles de segurança. Os autores usam um esquema de votações baseadas em pesos para diminuir a incidência de FP e FN de vários SDI/SPI.

Há diferenças sutis entre trabalhos voltados para o agrupamento de mecanismos de detecção e de previsão. De forma geral, técnicas e metodologias que buscam o agrupamento de controles de segurança voltados para a previsão tendem a empregar conceitos e ferramentas bastante formais. Como um exemplo, partindo da constatação que a combinação de mecanismos de previsão tende a melhorar o desempenho individual, Timmermann (2005) analisa diversas variáveis para aferir

quais são os fatores que contribuem para este aumento de desempenho: o grau de correlação entre o erro de previsão, o tamanho relativo das variâncias dos erros dos modelos de previsão individuais, a falta de especificação, a não estacionaridade e as regras de combinações de pontos, entre outros fatores.

Outras abordagens para combinar exclusivamente métodos de previsão são sugeridas, como a Lógica Fuzzy (STUBER; GEMMAR; GREVING, 2000); combinações de análise de séries temporais, análise estatística e técnicas de mineração de dados (SEHUN et al., 2008); a computação evolutiva (RIEDEL; GABRYS, 2005) ou mesmo combinações de metodologias como na pesquisa realizada por Palit e Popovic (2000) que envolve redes neurais e Lógica Fuzzy na combinação de previsões de forma não linear. Outros exemplos fora da área de TI incluem os trabalhos de Wei e Yang (2012) e de Barrow e Kourentzes (2016).

Apesar de focar mais na melhoria da classificação dos dados, um dos poucos trabalhos que utiliza a previsão para melhorar os resultados finais de detecção é a pesquisa realizada por Shin (2013). Os autores usam cadeias de Markov para modelar os estados de eventos indesejados e assim formar *clusters* e classificar dados com base em um conceito chamado de *outlier-factor*. Contudo, como o trabalho não possui um nível estratégico, é bastante difícil adaptá-lo em outros cenários.

Dentro da literatura referenciada também há pesquisas nas quais os autores usam conceitos oriundos da Teoria dos Jogos para melhorar os índices dos controles de segurança. Há estudos focados em conceitos totalmente opostos aos propostos neste trabalho, como o de Kumar e Kumar (2012) que usa conceitos como dominância de estratégias para analisar redes neurais artificiais em conjunto com algoritmos genéticos e métodos de votação na detecção de intrusões. Na pesquisa desenvolvida por Liu, Comaniciu e Man (2006), os autores usam um SDI/SPI e propõem um modelo de jogo para analisar as interações entre nós atacantes e defensores visando uma formulação Bayesiana. Formulações Bayesianas de jogos oferecem condições para que o defensor selecione suas estratégias com base em suas crenças sobre o tipo de oponente. Os resultados mostram que quando as expectativas são baixas, jogos Bayesianos estáticos levam a estratégias mistas conhecidas por Equilíbrio de Nash Bayesiano. Por outro lado, em jogos Bayesianos

dinâmicos a tendência é ter estratégias mistas chamadas de Equilíbrio Bayesiano Perfeito. O equilíbrio de estratégias tende a gerar economia de energia e melhora de desempenho da rede. Justamente por seguir abordagens não-cooperativas, tanto a proposta de Kumar e Kumar (2012) como de Liu, Comaniciu e Man (2006) têm limitações já que dependem de um conhecimento acentuado sobre as intenções do oponente.

Porém, também há pesquisas que sugerem a construção de estratégias focadas em jogos cooperativos. É interessante notar que mesmo se tratando de cooperação, em geral, pesquisas como estas procuram maximizar os ganhos dos jogadores por meio do equilíbrio entre os parâmetros analisados. Este tipo de equilíbrio é necessário porque apesar da cooperação, os parâmetros analisados podem ser antagônicos. Sendo assim, um jogo cooperativo pode ser criado para equilibrar ganhos relacionados a parâmetros opostos. Um exemplo pode ser visto no trabalho de Khan e Ahmad (2009), onde é proposta uma forma de aperfeiçoamento do consumo de energia e da alocação de tarefas em sistemas computacionais com o uso de jogos cooperativos baseados no conceito de Solução de Barganha de Nash (SBN). Segundo o autor, as máquinas chegam a uma decisão coletiva (formalizada matematicamente) que alcança a otimização de Pareto. Um dos méritos do trabalho de Khan e Ahmad (2009) é a criação de um algoritmo que realiza o SBN com menos iterações que o processo tradicional criado por Nash Jr (1950). Porém, o modelo proposto pelos autores não tem a amplitude de uma arquitetura como o SGA, que pode ser aplicada em recursos díspares como mecanismos de detecção e de previsão.

Em outro exemplo, Khan e Ahmad (2009) criam um jogo cooperativo e formalizam matematicamente uma análise estratégica que parte de dois recursos que se influenciam de forma linear e inversa: consumo de energia e processamento em um cluster de máquinas. A partir da análise dos ganhos individuais de cada máquina calculam-se os ganhos coletivos por meio de diversas interações entre as máquinas. Os ganhos de cada interação são organizados de forma que seja possível verificar em quais pontos acontecem os ganhos coletivos máximos, chamados de pontos de barganha. Por fim, a identificação dos pontos de barganha permite que se alcance o SBN, no qual a análise dos ganhos coletivos converge para a otimização de Pareto. Cenários como este, proposto por Khan e Ahmad (2009), nos quais se busca o

equilíbrio entre os recursos, são comuns quando se analisam jogos cooperativos e são cenários ideais para aplicação do SGA.

Portanto, percebe-se que há na literatura referenciada farto material sobre os benefícios da prevenção da ação de eventos maliciosos em sistemas cibernéticos por meio de levantamentos estatísticos e análises estratégicas. Este material, além de corroborar as hipóteses elaboradas para cumprimento dos objetivos deste trabalho, também abre espaço para o desenvolvimento de pesquisas que preencham algumas lacunas sobre o assunto.

Com o intuito de servir como comparativo, a tabela 2 mostra algumas características dos principais trabalhos referenciados que tratam diretamente sobre a combinação de mecanismos de detecção e previsão de eventos maliciosos em sistemas cibernéticos.

Tabela 2 - Comparativo entre o SGA e trabalhos relacionados

Trabalhos referenciados	Características de comparação										
	A	B	C	D	E	F	G	H	I	J	
Giacinto; roli e Didaci (2003)	X									X	
Agah et al. (2004)		X		X			X			X	
Srinivas; Andrew e Ajith (2005)	X							X			
BeauquieR, e Hu (2007)	X							X		X	
Sehun et al. (2008)		X						X	X	X	
Zhu et al. (2009)	X			X	X	X	X			X	
Khan e Ahmad (2009)				X	X	X	X	X	X	X	
Chen e Leneutre (2009)	X			X		X	X			X	
Lin et al. (2013)	X					X				X	
Shin (2013)	X							X		X	
SGA	X	X	X	X	X	X	X	X	X	X	

Legenda

- A - Agrupa mecanismos de detecção
- B - Agrupa mecanismos de previsão
- C - Agrupa mecanismos de detecção e previsão
- D - Análise estratégica
- E - Emprego de jogos cooperativos
- F - Emprego de recompensas e incentivos
- G - Busca de equilíbrio estratégico ou processos de barganha
- H - Emprego de modelos estatísticos
- I - Facilidade na adição de outros mecanismos à combinação estabelecida
- J - Criação de métricas ou pesos que avaliam a combinação

Fonte: o autor.

4 PROPOSTA: STRATEGIC GROUPING ARCHITECTURE (SGA)

Neste capítulo todos os aspectos e detalhes da arquitetura proposta são apresentados. A Seção 4.1 A Arquitetura – tem como foco a explanação do SGA e é dividida em duas subseções para fins de organização: 4.1.1 Nível Técnico – que lida com os aspectos técnicos, planejamento e captura dos dados; e 4.1.2 Nível Estratégico – que apresenta o cálculo UCB realizado nos dados coletados no nível técnico. A Seção 4.2 Discussão sobre a Arquitetura Proposta – traz algumas conclusões sobre a proposta do SGA.

4.1 A Arquitetura

Na literatura de TI há muitas definições sobre o termo arquitetura. Jaakkola e Thalheim (2010) definem uma arquitetura de *software* como um conjunto que engloba as formas usadas para comunicação entre *scripts*, programas e sistemas, além das pré-condições necessárias para configuração de componentes e interfaces. Entre os objetivos está a geração de benefícios durante o desenvolvimento e reuso de *software*.

Uma arquitetura de TI diz respeito ao conjunto de estruturas essenciais de *hardware* e *software* de um sistema. Envolve também o relacionamento, as propriedades e a documentação destas estruturas (BACHMAN et al., 2010). Entre os exemplos de propriedades dos elementos de uma arquitetura, podem ser citados serviços, características, falhas, dados de desempenho e uso de recursos compartilhados, entre outros (BASS; CLEMENTS; KAZMAN, 2003).

O SGA é uma arquitetura dividida em dois níveis: {1} nível técnico; e {2} nível estratégico. Esta divisão facilita o entendimento e permite que quaisquer novos módulos possam ser adicionados sem interferir com outros (figura 5).

Figura 5 - Níveis do SGA

Fonte: o autor.

4.1.1 Nível Técnico

No nível técnico os dados dos controles de segurança que fazem parte do SGA são coletados e preparados para serem usados no nível estratégico. De forma geral o nível técnico pode ser dividido em duas partes principais: {1} *Planejamento*; e {2} *Coleta e preparação dos dados*.

Como pode ser visto na figura 6, cada divisão do nível técnico é composta por diversos módulos que executam funções específicas, geram variáveis e séries históricas de dados que alimentam os outros módulos da arquitetura. Os módulos 1.6 e 1.7 são opcionais, pois a arquitetura pode funcionar mesmo se eles não forem executados. Além disso, é importante notar que algumas funções são automatizadas, enquanto outras requerem o acompanhamento de um administrador.

Figura 6 - Módulos do nível técnico

Fonte: o autor.

Os módulos da divisão de *Planejamento* (1.1 a 1.4) correspondem ao planejamento necessário para o funcionamento do SGA. Nestes módulos ainda não há geração de dados.

No módulo 1.1 são definidos os controles de segurança que fazem parte da arquitetura. Estes controles podem ser mecanismos de detecção e/ou de previsão.

Um vetor com o nome das variáveis independentes é definido no módulo 1.2. Este vetor contém as configurações (apenas os nomes) usadas nos controles de segurança que serão avaliadas. Essas configurações podem ser representadas de duas formas: {1} pelas configurações dos controles de segurança (por exemplo, o conteúdo de um campo do cabeçalho TCP/IP que é bloqueado por um SDI/SPI); ou {2} as configurações escolhidas em um experimento (por exemplo, a quantidade de passos a frente escolhida para ser usada em um processo de previsão).

Seguindo a orientação sobre a importância da escolha de variáveis que não sejam redundantes ou desprezíveis, presente no trabalho de Chebrolu; Abraham e Thomas (2005), é importante que sejam escolhidas variáveis independentes que realmente exercem influência nas variáveis resposta. Por isso essa escolha deve ser criteriosa e baseada na experiência do administrador do sistema.

Outro vetor com o nome das variáveis resposta é definido no módulo 1.3 e contém os resultados (apenas os nomes) gerados pelos controles de segurança que serão avaliados, como o conteúdo de alertas de segurança, *logs* ou cálculos realizados sobre séries históricas de dados.

Os vetores contendo os nomes das variáveis resposta e das variáveis independentes são repassados ao módulo 1.4, responsável pela definição do tamanho das amostras usadas na geração das séries históricas que passam por análise. Essa definição é importante principalmente se o método de previsão utilizado for baseado em estatística, pois muitos cálculos estatísticos não são precisos com amostras de tamanho reduzido.

Ao término da divisão de *Planejamento* tem início a divisão de *Coleta e preparação dos dados*. O módulo 1.5 é responsável pela coleta de dados diretamente dos controles de segurança definidos previamente. Uma *matriz de*

dados é gerada com os valores das variáveis resposta e das variáveis independentes definidas nos módulos 1.2 e 1.3.

A *matriz de dados* é repassada ao módulo 1.6, responsáveis por estabelecer limites (máximos e/ou mínimos) para as variáveis resposta. Em seguida, as variáveis resposta cujos valores estejam fora dos limites são cortadas da *matriz de dados*. Por exemplo, podem ser eliminadas eventuais medições não confiáveis coletadas por um controle de segurança com pouca energia nas baterias. Por fim, como resultado, é gerada uma matriz (*matriz limite*), composta pela *matriz de dados* sem os valores que ultrapassaram os limites definidos no módulo 1.6.

Apesar de ser um processo importante que filtra eventuais valores extremos da *matriz de dados* evitando que a análise do nível estratégico seja prejudicada, este módulo é opcional. Não há impedimento de aplicar o SGA sem a definição destes limites, que podem ser definidos de diversas formas. Por exemplo, por meio de valores estabelecidos por padrões, por experimentos prévios ou ainda gerando valores extremos que servem como faixa de corte.

O módulo 1.7 recebe a *matriz limite* do módulo 1.6 e é encarregado de levantar quais são as variáveis independentes que mais exercem influência (relevância) nas variáveis resposta. Apesar de opcional, o objetivo dessa prática é eliminar dos cálculos de UCB eventuais variáveis independentes que não tenham influência nos valores alcançados pelas variáveis resposta. O resultado deste módulo é uma matriz que contém apenas as variáveis independentes relevantes e as variáveis resposta já limitadas, chamada de *matriz relevante*. A forma para verificação da relevância das variáveis independentes é arbitrária. Apesar da experiência do administrador de TI ser fundamental na escolha destas variáveis independentes, há alternativas que podem ser adotadas. Um exemplo é levantar por meio de regressão linear quais são as variáveis independentes que mais exercem influência nas variáveis resposta.

O final do nível técnico consiste no envio da *matriz relevante* ao nível estratégico.

4.1.2 Nível Estratégico

O nível estratégico é responsável por agrupar os controles de segurança que compõe o SGA a partir dos dados gerados no nível técnico. Para cumprir este objetivo, o nível estratégico verifica, entre os controles de segurança envolvidos, quais configurações apresentam mais vantagens quando funcionam em conjunto. Dentro deste contexto, e até por utilizar os resultados finais de cada controle de segurança, **o nível estratégico cumpre uma função que busca equilibrar os resultados e as configurações utilizadas.**

Como discutido no Capítulo 3 Trabalhos Relacionados, destacar essa função de equilíbrio que o SGA busca alcançar é de extrema importância para o entendimento deste trabalho.

O nível estratégico se divide em quatro módulos principais, como pode ser visto na figura 7:

Figura 7 - Detalhes sobre os módulos do nível estratégico

Fonte: o autor.

O módulo 2.1 é responsável pela criação de uma matriz com os jogadores que vão fazer parte da UCB. De forma geral, **os jogadores são os valores das variáveis independentes associados aos respectivos valores de variáveis resposta definidos no nível técnico**. Por exemplo, se na *matriz relevante* houver duas variáveis independentes (Indep 1 e Indep 2) com três valores diferentes em cada uma delas (Valor 1, Valor 2 e Valor 3), ligados a dois vetores de variáveis resposta (Resp1 e Resp2), são formados nove jogadores (3^2), cada um deles associado a um valor do vetor Resp1 e a um valor do vetor Resp2, como pode ser visto na tabela 3.

Tabela 3 - Exemplo de formação de jogadores

Jogadores			Vetor Resp 1	Vetor Resp 2
Número do jogador	Indep 1	Indep 2		
Jogador 1	Valor 1	Valor 1	111	211
Jogador 2	Valor 1	Valor 2	112	212
Jogador 3	Valor 1	Valor 3	113	213
Jogador 4	Valor 2	Valor 1	121	221
Jogador 5	Valor 2	Valor 2	122	222
Jogador 6	Valor 2	Valor 3	123	223
Jogador 7	Valor 3	Valor 1	131	231
Jogador 8	Valor 3	Valor 2	132	232
Jogador 9	Valor 3	Valor 3	133	233

Fonte: o autor.

Ajustes podem ocorrer caso o menor valor possível seja o valor ideal que as variáveis resposta podem assumir. Essa ação ocorre porque o cálculo UCB lida sempre com variáveis resposta que tem como valor ideal os maiores valores possíveis. Em outras palavras, para que a barganha (UCB) ocorra, os jogadores devem sempre buscar os maiores valores. Sendo assim, a variável resposta negociada na barganha precisa estar indexada por ordem crescente (maior valor é o ideal) para funcionar como WTP. Variáveis resposta cujos valores ideais sejam os menores, devem passar por um processo de inversão. Por exemplo, uma variável resposta como energia economizada poderia ser utilizada como WTP, pois valores

altos são desejáveis nessa variável. O inverso ocorre se a variável usada for energia gasta, que tem os baixos valores como desejáveis. Para este último caso o ajuste de inversão é a recomendação. Como outro exemplo, considere dois vetores de variáveis resposta $A=\{1, 2, 3, 4, 5\}$ e $B=\{98\%, 97\%, 96\%, 95\%, 94\%\}$ cujos valores ideais são os menores possíveis. A inversão no vetor A ocorre multiplicando-se seus valores por $-1 \Rightarrow -1*A=\{-1, -2, -3, -4, -5\}$. Para o vetor B a inversão ocorre subtraindo-se o vetor de $100 \Rightarrow 100-B=\{2\%, 3\%, 4\%, 5\%, 6\%\}$.

O próximo passo, realizado no módulo 2.2, é padronizar as variáveis resposta da *matriz dos jogadores* a partir da fórmula:

$$Z_{J_i} = \frac{J_i - \mu_i}{\sigma_i^2} \quad (19)$$

onde: J = vetor com as variáveis resposta da *matriz dos jogadores*;

$i = 1, 2, 3, \dots, n$;

n = quantidade de elementos em J ;

μ = média aritmética de J ;

σ^2 = variância de J .

O processo de padronização permite que eventuais cálculos entre os vetores sejam realizados. Por exemplo, se um determinado vetor contiver valores percentuais enquanto outro possuir conteúdo escalar, após a padronização estes vetores podem passar por operações aritméticas. A saída do módulo 2.2 é chamada de *matriz estratégica*.

No módulo 2.3, a partir da *matriz estratégica*, os vetores de variáveis resposta são classificados em dois tipos: {1} com função WTP¹⁹; e {2} com função de custo. Dentro do cálculo de UCB os primeiros são usados para definir os valores gerados no jogo estratégico cooperativo, enquanto os segundos são subtraídos dos primeiros e servem como contraponto.

¹⁹ Apesar dos jogos cooperativos propostos na validação do SGA não se tratarem de jogos de atribuição de três lados, termos como WTP, custo e CM são mantidos da forma como são utilizados originalmente nos artigos de Stuart Jr (2001); Brandenburger e Stuart (1996); Brandenburger e Stuart (2007); Brandenburger e Dekel (1987); Brandenburger e Nalebuff (1995) e Brandenburger (2007).

A arquitetura comporta o uso de vários vetores de ambos os tipos. Note, porém, que a presença de vetores com função de custo não é uma premissa obrigatória no SGA. É possível realizar os cálculos da UCB usando apenas os vetores com função de WTP. Contudo, o uso de custos no processo estratégico pode se traduzir em maior precisão na análise, principalmente se houver correlação entre os vetores de WTP e de custo. Além disso, em casos onde haja dois ou mais vetores de custo, o processo de padronização permite que os custos sejam somados.

Um exemplo da flexibilidade do uso de vetores de custo contendo variáveis independentes é a possibilidade da aplicação de pesos a estes vetores, como sugerido no trabalho de Huang e Lee (2010). Como os vetores passam pelo processo de padronização é possível, por exemplo, multiplicar (ou dividir) o vetor de custo de acordo com o peso que o mesmo exerce no vetor WTP.

O módulo 2.4 realiza o cálculo e a análise da UCB. Para cada jogador os seguintes campos são calculados: {1} a soma ou os valores individuais de cada WTP representando as variáveis resposta; {2} a soma ou os valores individuais de cada custo representando também as variáveis resposta; {3} o IVC ou $v(N)$ que representa o valor criado individual gerado por um jogador i em um jogo; {4} o $v(N \setminus \{i\})$ que representa o valor criado em um jogo sem a presença do jogador i ; {5} a CM; {6} o valor mínimo garantido ao jogador i ; {7} as faixas mínimas e máximas garantidas ao jogador i ; e {8} a barganha residual do jogador i . **O jogador (na verdade, o conjunto de variáveis independentes) mais apropriado para uso é o que tem o maior valor de CM.**

Percebe-se que o cálculo da UCB, apesar de ser aplicado sobre toda a *matriz estratégica* e gerar resultados sobre todos os jogadores, mostra efetivamente como resultado qual jogador individual²⁰ é o mais apropriado para uso. Em geral, ambientes de TI onde se analisam controles de segurança tem como característica certa elasticidade em relação ao nível de segurança utilizado. Em outras palavras, cada ambiente, cada administrador, cada sistema, requer um nível de segurança

²⁰ A primeira impressão é que o cálculo de CM focada em cada jogador individual não atende a um dos princípios básicos da construção de jogos de atribuição de m-lados: a criação de valores só acontece quando há interação entre os jogadores (STUART JR, 2001). Contudo, é possível notar que cada jogador é formado por um conjunto de diferentes variáveis independentes. Dessa forma, cada variável independente pode ser entendida como um jogador individual pertencente a um subconjunto SS cujos membros interagem entre si.

diferente configurado nos controles de segurança. Isso é importante, pois evita que o resultado de uma arquitetura voltada para o agrupamento de controles de segurança, como o SGA, seja muito rígida, levando o administrador a usar os resultados obtidos sem flexibilidade alguma. Para resolver essa questão, assim que realizado pela primeira vez, o cálculo da UCB é repetido sem a presença do jogador com maior valor de CM (mais apropriado), com o objetivo de mostrar qual seria o próximo jogador indicado pela UCB. Essa repetição é realizada até o último jogador e o resultado final, denominado *matriz UCB*, é uma sequência fidedigna ²¹ sobre quais são os jogadores mais apropriados para serem usados.

4.2 Discussão Sobre a Arquitetura Proposta

Uma das principais ideias defendidas neste trabalho é baseada na afirmação de que uma situação de UCB ocorre quando está em andamento um processo de agrupamento de controles de segurança. O termo UCB foi criado com o objetivo de funcionar como um indicador corporativo para medir o valor com que a empresa contribuirá para a economia em um futuro próximo (BRANDENBURGER; STUART JR, 1996). Esta definição encontra ressonância nessa tese, já que parte dos resultados dos controles de segurança é usada na composição da CM.

Como já discutido, o SGA tem como função principal equilibrar os jogadores (configurações e resultados gerados pelos controles de segurança). Portanto, as **potencialidades da arquitetura são mais bem aproveitadas quando existe o contraponto entre WTP e custo**. Em situações nas quais o incremento de uma mesma variável independente influencia duas variáveis resposta de forma oposta, a aplicação do SGA é bastante interessante.

Essa constatação é muito importante, pois abre a possibilidade de adaptação e uso do SGA para resolução de problemas fundamentais de busca por equilíbrio entre diferentes variáveis em áreas diferentes da área de TI.

²¹ O resultado alcançado com apenas uma execução do cálculo UCB é ordenado em ordem decrescente de CM. Contudo, o segundo jogador mais apropriado pode não corresponder necessariamente ao próximo jogador mais apropriado em uma segunda repetição.

5 RESULTADOS E VALIDAÇÃO DO TRABALHO

Este capítulo implementa de forma prática a arquitetura proposta, já que discute os resultados da aplicação do SGA em dois experimentos. Em cada um deles um problema específico relacionado à segurança da informação é contextualizado e discutido. As propostas de solução apresentadas para cada um dos experimentos são independentes e ratificam a variedade de casos onde o SGA pode ser aplicado. Ao final há um comparativo entre os experimentos e a literatura atual que funciona como uma validação dos resultados alcançados.

Na Seção 5.1 Primeiro Experimento é descrita uma MANET na qual métodos de detecção e de previsão são utilizados para verificar se há atividade maliciosa na troca de informações sobre o posicionamento dos nós. O SGA é usado para determinar qual configuração é mais adequada para o uso.

A Seção 5.2 Segundo Experimento formula um *ranking* que serve como preditor em um ambiente com dois SDI/SPI. A medida que o *ranking* aumenta, mais suspeito se torna o tráfego. O SGA é usado para determinar o nível mais adequado de uso deste *ranking*.

Por fim, como forma de validação, após uma transformação na categoria do jogo (NUDELMAN et al., 2004), a Seção 5.3 Validação busca encontrar nos resultados do SGA dos dois experimentos os mesmos equilíbrios (EN e ótimo de Pareto) alcançados nos trabalhos realizados por Wu et al. (2010), Nadkar et al. (2012), Ma; Rao e Yau (2011), Khan e Ahmad (2009), e Agah et al. (2004).

5.1 Primeiro Experimento

Muitas aplicações voltadas para MANET dependem do posicionamento informado pelos nós. Entre os exemplos podem ser citadas as aplicações relacionadas às Vehicular Ad-hoc NETworks (VANET) (AL-SULTAN et al., 2014; ZEADALLY et al., 2012) ou às Redes de Sensores Sem Fio (RSSF) (BOSMAN et al., 2016; BHUIYAN;

WANG; VASILAKOS, 2015) . Portanto, a localização dos nós em redes como essas é um processo crítico que requer confiabilidade. Uma estratégia bastante comum para a obtenção deste tipo de informação é capturar pacotes de dados trocados periodicamente entre os nós. Esses pacotes, chamados de pacotes *beacon*, contêm diversos parâmetros que podem ser utilizados para se determinar a localização dos nós.

Além de outras informações, em geral, pacotes *beacon* carregam dados sobre a energia do nó de origem e sobre o Received Signal Strength Indicator (RSSI), uma medida da força do sinal. Quando combinadas a outros parâmetros como características dos nós, topologia da rede, informações sobre protocolos usados e processos sendo executados é possível estimar o posicionamento dos nós em uma rede e a distância entre eles (YANG; LIU, 2012; GUO et al., 2010).

Uma abordagem alternativa é a adoção de modelos preditivos para determinar a localização a partir do padrão de movimentação dos nós (LIU et al., 2008; YE et al., 2005). Porém, ainda é bastante difícil alcançar bons índices de confiabilidade, precisão e segurança por meio deste tipo de abordagem. Em relação à segurança, uma RSSF está sempre sujeita a ação de tráfego falso gerado por nós maliciosos (NADKAR ET AL., 2012). Além disso, em uma MANET a mobilidade dos nós é um fator complicador para aplicação de medidas de segurança (GUO; SCHILDT; WOLF, 2013). A economia de energia dos nós também é outro problema que deve ser sempre levado em conta quando se abordam nós em RSSF (BOSMAN et al., 2016; BHUIYAN; WANG; VASILAKOS, 2015; LUO; CHANG, 2015; NIRMALA; MANJUNATHA, 2015; YE et al., 2005).

Na busca por soluções alguns trabalhos tem como foco a análise de diversos parâmetros a fim de melhorar a confiabilidade e a segurança das localizações (GUO; SCHILDT; WOLF, 2013; RAYMOND et al., 2009). Seguindo uma linha bastante parecida, a proposta deste Primeiro Experimento procura aliar a ação de métodos de detecção e previsão para garantir que as informações de localização trafegadas em uma MANET não sejam maliciosas, ao mesmo tempo em que busca otimizar o uso da energia. Para isso, a partir da detecção de séries históricas de posicionamentos, dois métodos de previsão – tendência polinomial (SU, 2010; HUANG et al., 2010) e (GM(1,1) (KAYACAN; ULUTAS; KAYNAK, 2010; TSENG; YU; TZENG, 2001) - são

usados para prever as posições nas quais os nós móveis estarão depois de alguns passos. A análise da variabilidade das previsões provê indícios sobre a existência de informações falsas de posicionamento com baixo processamento. Essas vantagens permitem que a metodologia apresentada possa ser utilizada em conjunto com outros métodos voltados para detecção de posicionamento presentes na literatura (LIU et al., 2008; RAYMOND et al., 2009; GUO; SCHILDT; WOLF, 2013). A figura 8 mostra a topologia da MANET:

Figura 8 - Topologia da MANET usada no Experimento 1

- Nós armazenadores (0-9): nós imóveis que recebem pacotes beacon dos nós controlados
- Nós controlados (10-14): nós com movimento aleatório que enviam pacotes para os nós armazenadores
- Nós fixos (20-29): nós imóveis que servem apenas como *multi-hop*
- Nós de viagem (15-19): nós com movimento aleatório que servem apenas como *multi-hop*

Fonte: o autor.

Apesar das vantagens do método, a troca constante de pacotes *beacon* eleva o consumo de energia. A fim de equilibrar os limites de variabilidade permitidos para as séries previstas e o gasto de energia proporcionado pela troca de pacotes *beacon* o SGA é aplicado aos resultados.

Há muitos trabalhos que abordam o tema da localização de nós e podem servir de comparação com a proposta deste Primeiro Experimento. Entre as diversas abordagens utilizadas podem ser citadas, o treinamento de máquinas (NAKAYAMA et al., 2009) e metodologias que analisam informações diversas como topologia, características dos sinais RSSI e aspectos relacionados aos protocolos (LIU et al., 2008). Outro exemplo é o trabalho de Yao et al. (2015) que utiliza o resultado de um algoritmo de localização centróide baseado no RSSI para criar um método que adota pesos na determinação do posicionamento dos nós. Abordagens parecidas utilizam métodos de localização sem a presença de nós âncoras, como a média da distância entre os nós (PADMANABHAN; KAMALAKKANNAN; SHANTHARAJAH, 2013); otimização por meio de *least square suport vector machine* – uma variação das tradicionais SVM - e algoritmos inteligentes como o *fischer fishing* (YU; LV, 2016). Estes trabalhos, porém, não definem formas de avaliar a confiabilidade das informações trocadas entre os nós.

Há alguns trabalhos que focam aspectos de segurança durante o processo de localização dos nós. Como exemplos, o trabalho de He et al. (2014) cria um esquema de reputação para classificar as RSSF de acordo com a confiabilidade; já o trabalho desenvolvido por Miao et al. (2015) adota a colaboração entre nós para avaliar a confiabilidade da informação do posicionamento deles. Porém, estes trabalhos tem foco na detecção em detrimento da prevenção gerada por técnicas de previsão.

As propostas que sugerem métodos de previsão de séries históricas em RSSF e MANET tem como foco a redução de energia por meio da diminuição da redundância de informações na troca de pacotes (LUO; CHANG, 2015; LE BORGNE; SANTINI; BONTEMPI, 2007; YANN-AEL; GIANLUCA, 2005; KAANICHE; KAMOUN, 2010; MUPPIRISETTY; SVENSSON; WYMEERSCH, 2016). Contudo, a confiabilidade dos pacotes acaba ficando em um segundo plano. Entre os poucos trabalhos que levam em conta os aspectos de segurança e confiabilidade das informações na estimação das posições dos nós, podem ser citados: Liu et al. (2011) que procuram determinar regiões da topologia onde as falhas tendem a ocorrer por meio de análise de confiabilidade baseada em probabilidade; Yang e Liu (2012) que desenvolvem um mecanismo estatístico para validar as mensagens trocadas na rede; e Ye et al. (2005) que propõem um algoritmo baseado em

regressão polinomial e cadeias de Markov para reduzir o *hand-off* vertical em redes sem fio heterogêneas. Uma limitação comum nestes trabalhos é a falta de uma abordagem estratégica que permita o balanceamento dos resultados de confiabilidade, segurança e consumo de energia.

Por fim, o trabalho mais próximo da proposta deste Primeiro Experimento é desenvolvido por Basilico et al. (2014) que utiliza: {1} multilateração verificável²² para estimar o posicionamento dos nós; e {2} jogos não cooperativos para otimizar os resultados. Entretanto, segundo os próprios autores, o método não é tão efetivo se houver conluio entre os nós para o envio de pacotes em intervalos e quantidades pré-determinadas. A tabela 4 traz um comparativo com alguns dos principais trabalhos relacionados com o Primeiro Experimento.

Tabela 4 - Comparativo entre o Primeiro experimento e trabalhos relacionados

Trabalhos referenciados	Características de comparação							
	A	B	C	D	E	F	G	H
Miao et al. (2015)						X		
Nakayama et al. (2009)			X	X	X	X		
Liu et al. (2008)				X	X	X		
Yao et al. (2015)*								
Yu e Lv (2016)*					X			
He et al. (2014)		X		X	X	X		
Basilico et al. (2014)				X	X	X		X
Luo e Chang (2015)*	X				X		X	
Le Borgne; Santini e Bontempi (2007)*	X				X		X	
Yann-Ael e Gianluca (2005)*	X				X		X	
Kaaniche e Kamoun (2010)*	X	X	X		X			
Muppirisetty; Svensson e Wymeersch (2016)*	X		X		X	X		
Liu et al. (2011)	X				X	X		
Ye et al. (2005)			X	X	X	X	X	
Primeiro Experimento	X	X	X	X	X	X	X	X

Legenda

- A - Usa mecanismos de previsão
- B - Pode ser usado em conjunto com outras técnicas de previsão
- C - Funciona com mobilidade
- D - Identifica um nó que esteja previamente infectado
- E - Analisa diversos parâmetros
- F - Realiza testes de segurança
- G - Também foca em energia
- H - Usa abordagem estratégica
- * Artigos não relacionados à segurança

Fonte: o autor.

²² Multilateração verificável e suas variantes (bilateração e trilateração) estão entre os métodos mais usados para localização de nós em RSSF e MANET (NIRMALA; MANJUNATHA, 2015).

Os trabalhos referenciados na tabela 4 confirmam a relevância das pesquisas voltadas para localização de nós em MANET e mostram que há poucos trabalhos que usam abordagens estratégicas ou podem ser usados em conjunto com alguma outra técnica de previsão/detecção já estabelecida.

5.1.1 Conceitos Fundamentais do Primeiro Experimento

Há dois métodos de previsão cujos conceitos são importantes para o entendimento da proposta do Primeiro Experimento: tendência polinomial e Grey Theory. Ambos são tratados nas seções subsequentes.

5.1.1.1 Tendência Polinomial

Uma série temporal é uma sequência de observações de uma variável em intervalos regulares de tempo (SU, 2010; KAYACAN; ULUTAS; KAYNAK, 2010). Em uma MANET, quando sensores enviam diferentes valores de posicionamento, cada conjunto de observações destes valores representa uma amostra da série temporal (YANG; LIU, 2012). A previsão da trajetória dos nós pode ser denotada como uma curva no gráfico das séries temporais (MORETTIN; TOLOI, 1987)

Considerando que uma série temporal de posicionamentos pode ser definida como Z , o valor da série no instante (t) pode ser descrito como Z_t ($t=1,2,...,n$). Quando a série temporal em questão apresenta tendências, representadas por T_b , a tendência polinomial pode ser usada para prever valores futuros na forma:

$$T_t = \beta_0 + \beta_1 + \dots + \beta_m t^m \quad (2)$$

Onde $m < n$; n é o número de observações; t é o tempo; e m é grau do polinômio. Neste Primeiro Experimento os valores de posicionamento dos nós da MANET são os valores observados. Sendo assim, um polinômio de primeiro grau é adequado para representar a curva de previsões, que pode ser denotada na forma (MORETTIN; TOLOI, 1987):

$$Z_t = \beta_0 + \beta_1 t + e_t \quad (3)$$

Onde β_0 e β_1 são parâmetros a serem estimados e e_t indica erro aleatório com variância σ_e^2 .

Minimizando a soma dos quadrados dos resíduos:

$$f(\beta_0, \beta_1) = \sum_{t=1}^N (Z_t - \beta_0 - \beta_1 t)^2 \quad (4)$$

Obtêm-se as equações normais (MORETTIN; TOLOI, 1987):

$$\hat{\beta}_0 = \bar{Z} - \hat{\beta}_1 \bar{t} \quad (5)$$

$$\hat{\beta}_1 = \frac{\sum_{t=1}^N t Z_t - \left[\frac{(\sum_{t=1}^N t)(\sum_{t=1}^N Z_t)}{N} \right]}{\sum_{t=1}^N t^2 - \frac{(\sum_{t=1}^N t)^2}{N}} \quad (6)$$

Onde $\bar{Z}$ e $\bar{t}$ são respectivamente a média amostral das N observações e a média amostral dos intervalos de tempo usados ($t=1, 2, \dots, N$).

Porém, como o grau do polinômio usado é de primeiro grau ($m=1$), o modelo polinomial se reduz a uma regressão linear. Assume-se aqui, portanto, que os dados (valores de posicionamento dos nós ao longo do tempo) possuem os pré-requisitos exigidos por uma regressão linear: resíduos que possuem distribuição normal, com média zero e variância constante (MORETTIN; TOLOI, 2006; DRAPER; SMITH, 1998).

5.1.1.2 Grey Theory

A Grey Theory é um conjunto de conceitos²³ cujo objetivo é a resolução de sistemas com pouca ou incompleta informação. Neste trabalho o Grey forecasting Model (GM) é explorado como uma técnica de previsão. Ao contrário do modelo linear (polinomial com grau 1) usado como método de previsão, o GM não é um modelo regressivo estatístico e não depende de modelos estocásticos (TSENG; YU; TZENG, 2001). Essa é a principal razão da escolha desse modelo para fazer oposição ao modelo de previsão por tendência polinomial (DENG, 1989).

Em um modelo $GM(n,m)$, m representa o número de variáveis, enquanto n representa a ordem da equação diferencial. GM considera uma frequência fixa para

²³ A Grey Theory envolve os conceitos de Grey Forecasting Model, Relational Space, Grey Generating Space, Grey Forecasting, Grey Decision Making e Grey Control

as amostras e dados positivos. Três operações básicas compõe um modelo GM (DENG, 1989; KAYACAN; ULUTAS; KAYNAK, 2010): {1} Accumulated Generation (AGO), uma operação que visa resolver a equação diferencial GM(1,1), um alisamento da aleatoriedade para determinar o próximo valor n -passos à frente; {2} Inverse Accumulated Generation (IAGO), uma operação que visa a previsão dos dados a partir da determinação da AGO; e {3} a operação de modelagem. Na sequência segue o roteiro para o cálculo de GM(1,1) extraído de Kayacan; Ulutas e Kaynak (2010) com valores de exemplo:

$$X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)), \quad n \geq 4 \quad (7)$$

Onde $X^{(0)}$ é uma sequência de tempo não negativa (na sigla $GM(n,m)$ enquanto $n=1$ representa a ordem da equação diferencial, $m=1$ a quantidade de variáveis) e n é o tamanho da amostra.

Por exemplo, seja uma sequência numérica qualquer $X^{(0)} = [820; 840; 835; 850; 890; 893]$.

$AGO(X^{(0)}) = X^{(1)}$, indica o resultado da operação AGO em $X^{(0)}$.

$$X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)), \quad n \geq 4 \quad (8)$$

$$X^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), \quad k = 1, 2, \dots, n \quad (9)$$

Seguindo o exemplo, $X^{(1)} = [820; 1660; 2495; 3345; 4235; 5128]$.

$Z^{(1)}$ é a sequência média gerada de $X^{(1)}$ e $z^{(1)}(k)$ é o valor médio dos dados adjacentes.

$$Z^{(1)} = (z^{(1)}(1), z^{(1)}(2), \dots, z^{(1)}(n)), \quad n \geq 4 \quad (10)$$

$$Z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1), \quad \dots \quad k = 2, \dots, n \quad (11)$$

Seguindo o exemplo, $Z^{(1)} = [410; 1240; 2487,5; 4160; 6277,5; 8841,5]$.

A equação diferencial de GM(1,1) é estimada pelos mínimos quadrados.

$$x^{(0)}(k) + ax^{(1)}(k) = b \quad (12)$$

Cuja alternativa é:

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b \quad (13)$$

A solução de $x^{(1)}(t)$ no tempo k :

$$x_p^{(1)}(k+1) = [x^{(0)}(1) - \frac{b}{a}]e^{-ak} + \frac{b}{a} \quad (14)$$

Onde $[a, b]^T$ é uma sequencia de parâmetros definida como:

$$[a, b]^T = (B^T B)^{-1} B^T Y \quad (15)$$

Onde:

$$Y = [x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)]^T \quad (16)$$

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix} \quad (17)$$

Seguindo o exemplo, $Y = [840; 835; 850; 890; 893]$ e $B = \begin{bmatrix} -1240,1 & 1 \\ -2487,5 & 1 \\ -4160 & 1 \\ -6277,5 & 1 \\ -8841,5 & 1 \end{bmatrix}$.

A operação IAGO é usada para estabilizar o modelo Grey e obter o valor predito a partir dos dados originais no tempo $(k+1)$ e $(k+H)$:

$$x_p^{(0)}(k+1) = [x^{(0)}(1) - \frac{b}{a}]e^{-ak}(1 - e^a) \quad (18)$$

$$x_p^{(0)}(k+H) = [x^{(0)}(1) - \frac{b}{a}]e^{-a(k+H-1)}(1 - e^a) \quad (19)$$

Seguindo o exemplo, os próximos valores (preditos) são denotados por $x_p^{(0)}(k+H) = [869,2109; 876,6860; 884,2254; 891,8296; 899,4993; 907,2349]$.

5.1.2 Formalização do Primeiro Experimento

O método aqui proposto prevê séries temporais (chamadas de séries previstas) a partir de séries históricas (chamadas de séries reais) que detectam as posições dos nós em uma MANET. Como séries históricas contendo falsas localizações tendem a

ter alta variabilidade²⁴, o principal objetivo deste experimento é analisar o desvio padrão das séries previstas para prevenir nós maliciosos de enviar dados falsos. Sendo assim, nós suspeitos, com alta variabilidade nas medições de posicionamento podem ser descobertos antecipadamente. Em resumo, as séries previstas antecipam a variabilidade da próxima série histórica e indicam se um nó apresenta comportamento indesejado. Por exemplo, um administrador pode avaliar uma MANET em tempo real com o processo de previsão e tomar as ações necessárias quando um limite máximo de variabilidade é alcançado.

Considera-se que todos os nós estão dentro da área de cobertura da RSSF e tem as mesmas capacidades de comunicação. Além disso, diversas variáveis independentes que tem influência no consumo de energia e/ou na variabilidade das séries históricas são analisadas: {1} intervalo de envio de pacotes *beacon*; {2} velocidade dos nós; {3} passos à frente na previsão; e {4} tamanho dos pacotes *beacon*. Algumas configurações do experimento, chamadas aqui de variantes, também são levadas em conta:

- O método de previsão: como dito anteriormente há dois métodos de previsão: {1} tendência polinomial; e {2} GM(1,1)²⁵;
- O tipo de posicionamento: tendo como base o RSSI, os pacotes *beacon* carregam as coordenadas de posicionamento (x,y). A partir destas coordenadas dois tipos de posicionamento são analisados: {1} o coeficiente angular (M) é um número usado para descrever tanto a direção como a inclinação de uma linha; e {2} cálculo normalmente usado na combinação de erros, a raiz da soma dos quadrados ($M2$) é uma medida que pode ser usada para lidar com a incerteza das coordenadas (x,y). Ambos os cálculos podem ser observados na figura 9 e nas fórmulas (20) a (23);

²⁴ Um trabalho prévio do autor desta tese (SILVA et al., 2014) mostra que altas variabilidades podem ser prejudiciais, pois podem indicar falsos posicionamentos dos nós.

²⁵ A sequência de tempo não negativa requerida pelo método GM(1,1) é representada pelo posicionamento dos nós.

Figura 9 - Representação de M e $M2$, usados para determinar a posição de um nó

Fonte: o autor.

$$M2_b = \sqrt{\frac{X_b^2 + Y_b^2}{2}}, \text{ onde } X_b \text{ e } Y_b \text{ são as coordenadas do ponto B da figura 9} \quad (20)$$

$$M2_c = \sqrt{\frac{X_c^2 + Y_c^2}{2}}, \text{ onde } X_c \text{ e } Y_c \text{ são as coordenadas do ponto C da figura 9} \quad (21)$$

$$M_b = \frac{Y_b - Y_a}{X_b - X_a}, \text{ onde } X_a \text{ e } Y_a \text{ são as coordenadas do ponto A da figura 9} \quad (22)$$

$$M_c = \frac{Y_c - Y_a}{X_c - X_a} \quad (23)$$

- O estimador de variabilidade: a variabilidade é medida por meio do desvio padrão. Dois estimadores são usados: {1} a amplitude entre o maior e o menor desvio padrão das séries históricas previstas; e {2} a média aritmética dos desvios padrão das séries históricas previstas;
- Os percentuais de variabilidade na geração de dados falsos: a fim de determinar qual limite de variabilidade pode ser considerado nocivo, séries maliciosas aleatórias são geradas com diferentes percentuais de variação nos valores de posicionamento.

A simulação da MANET é realizada usando o simulador NS-2 (<http://www.isi.edu/nsnam/ns/>) e o protocolo de Vetor de Distância Ad-hoc On-demand Distance Vector (AODV). Ambos, o simulador NS-2 e protocolo AODV são muito usados em experimentos em RSSF. A simulação é realizada usando uma

máquina virtual Linux com sistema operacional Centos OS 5.4, kernel 2.6.18 em um *notebook* Dell-Vostro com processador Intel Core i5 M-520 com 2,40 Giga-Hertz (GHz) e 4 Gigabyte (GB) Random Access Memory (RAM). Os arquivos gerados estão em formato texto e os seguintes *softwares* são usados na análise de dados:

- Gerenciador de banco de dados: Microsoft Visual FoxPro;
- *Software* matemático: Mathlab (<https://www.mathworks.com/products/matlab.html>);
- Simulador de processamento: Avrora (<http://compilers.cs.ucla.edu/avrora/>);
- Planilha eletrônica: Microsoft Excel;
- *Software* estatístico: R-free/R-studio (<https://www.r-project.org/>).

Os parâmetros gerais do Primeiro Experimento estão resumidos na tabela 5:

Tabela 5 - Parâmetros gerais do Primeiro Experimento

Parâmetros	Valores
Simulador	NS-2
Interface de rede	MAC IEEE 802.11
Tipo de link	LL
Modelo de antena	Omnidirectional
Número de nós	30
Protocolo de roteamento	AODV
Dimensão da topologia	1000 m x 1000 m
Modelo de energia	$rxPower = 35.28e-3 W$ $txPower = 31.32e-3 W$ $idlePower 712e-6 W$ $sleepPower 144e-9 W$
Comunicação entre nós (só nós armazenadores ímpares são analisados)	Nó 0 ← Nó 10 → Nó 1 Nó 2 ← Nó 11 → Nó 3 Nó 4 ← Nó 12 → Nó 5 Nó 6 ← Nó 13 → Nó 7 Nó 8 ← Nó 14 → Nó 9
Velocidade dos nós	5, 3 e 1 m/s
Tamanho dos pacotes <i>beacon</i>	200, 100, 50, 25 e 12 bytes
Intervalo de envio de pacotes	0.05 – 9.85 s (passo of 0.1 s)
Passos à frente na previsão	3, 6 e 12 passos à frente
Tempo de simulação	150 s
Tipo de tráfego	CBR
Pacotes <i>beacon</i>	UDP

Fonte: o autor.

Todos os nós da topologia podem servir como *multi-hop*²⁶. Por exemplo, os *nós de viagem* circulam aleatoriamente pela rede e tem majoritariamente a função de servir como comutadores na comunicação. Os *nós controladores* enviam pacotes *beacon* para os *nós armazenadores* usando o protocolo User Datagram Protocol (UDP) com Constant BitRate (CBR). A troca de pacotes *beacon* entre os *nós controladores* (10-14) ocorre sempre com um par exclusivo de *nós armazenadores* (0-9). Por exemplo, o nó 12 envia pacotes *beacon* exclusivamente para os nós 4 e 5, gerando duas séries estocásticas de dados (série real). Como apenas uma única série é suficiente, apenas a comunicação com os *nós armazenadores* ímpares é considerada nos experimentos.

5.1.2.1 Intervalo de Envio de Pacotes

Aumentar o intervalo de envio de pacotes *beacon* tende a diminuir a quantidade de pacotes enviada por um nó. A tendência é que esta ação diminua o consumo de energia, mas seja prejudicial para a variabilidade da previsão, já que quando o intervalo é grande um nó móvel pode se movimentar mais tempo, gerando uma série real com maior variabilidade. Como a série real é usada como base de cálculo da série prevista, a variabilidade da previsão também tende a aumentar. O intervalo de envio de pacotes varia de 0.05 s a 9,85 s, com passo de 0,1 s – Há, portanto, 99 séries históricas com diferentes intervalos de envio de pacotes.

5.1.2.2 Velocidade dos Nós

Um *nó controlado* que se move com velocidade maior que outro tende a ter um aumento na variabilidade de sua série prevista de posicionamentos. Contudo, a energia não é afetada. As baixas velocidades adotadas (1, 3 e 5 m/s) são mais difíceis de serem detectadas que as velocidades usadas por Silva et al. (2014). Há,

²⁶ Em uma MANET os nós *multi-hop* são os intermediários entre uma origem e um destino que estejam fora de alcance. Nós *multi-hop* são caracterizados por servir de caminho para a informação, atuando como comutadores, mas também podem funcionar como origem ou destino de informações.

portanto, três diferentes velocidades avaliadas, que em conjunto com o intervalo de envio de pacotes geram 297 séries históricas para avaliação.

5.1.2.3 Passos à Frente na Predição

O posicionamento de um nó é baseado a partir das coordenadas (x,y) enviadas pelos *nós controladores* (série real). A série prevista contém previsões baseadas na série real previamente detectada usando três janelas de previsão: 3, 6 e 12 passos à frente. Por exemplo, se a quantidade de *passos à frente*=3, a série prevista é calculada usando-se as três últimas leituras da série real. A série prevista vai ter então, três valores. A quantidade de passos à frente tende a influenciar a variabilidade da série prevista sem afetar o consumo de energia. Com o acréscimo desta variável independente, chegam a 891 as séries históricas para avaliação.

5.1.2.4 Tamanho dos Pacotes

O tamanho dos pacotes tende a influir diretamente no consumo de energia. Cinco tamanhos de pacotes são avaliados: 200, 100, 50, 25 e 12 bytes. Dessa forma, o total de séries históricas para avaliação chega a 4.455.

5.1.2.5 Cálculo da Variabilidade

O cálculo da variabilidade da série prevista usa os valores de M e $M2$ calculados para cada par de coordenadas (x,y) e leva em conta a amplitude ou a média do desvio padrão das séries previstas.

Por exemplo, suponha um conjunto de variáveis independentes: *velocidade*=5 m/s; *intervalo*=4,05 pacotes/s; *tamanho*=25 bytes; *passos à frente*=6. Com base na quantidade de passos à frente são gerados dois vetores com seis valores de M e seis valores de $M2$, esta é a série prevista. Os dois métodos de previsão são aplicados aos dois vetores e o desvio padrão da série prevista é calculado. Eventuais medidas corretivas, baseadas nos valores de desvio padrão podem ser tomadas a partir deste momento. É possível notar que outros vetores como esse são

gerados antes do término do experimento. Ao término da simulação (150 s – vide tabela 5), todos os valores de desvio padrão calculados para as séries previstas são colocados em um vetor e a amplitude e a média do desvio padrão deste vetor são calculadas. Dessa forma, para cada simulação há um valor de amplitude e um valor de média do desvio padrão, que são usados no nível técnico para cálculo da *matriz UCB*. De forma geral, pode-se definir que a variabilidade (amplitude e a média dos desvios padrão) são resultados do processo de previsão.

Não são gerados cálculos a respeito da variabilidade de energia. Os valores sobre a energia restante nos *nós controladores* são enviados nos pacotes *beacon* e aparecem na análise estratégica como contraponto à análise de variabilidade. Também aqui, é possível definir que a energia restante nos nós é um dos resultados do processo de detecção.

Como citado, os métodos de previsão (GM(1,1) e tendência polinomial), o tipo de posicionamento (*M* e *M2*) e o estimador de variabilidade (amplitude e média) não são considerados variáveis independentes, mas sim variantes analisadas separadamente neste Primeiro Experimento. Dessa forma, as 4.455 séries históricas são geradas oito vezes, uma para cada variante. Ao final são geradas 35.640 séries históricas para avaliação de energia e variabilidade.

5.1.2.6 Definição de Limites

Seis percentuais são escolhidos para envenenar os valores de posicionamento da série real: 1%, 2%, 5%, 10%, 15% e 20%. Esses percentuais são aplicados em um algoritmo gerador de números pseudoaleatórios, chamado MRG32K3a (L'ECUYER et al., 1999), para criar faixas aleatórias que representam limites mínimos e máximos para geração de falsos valores. A ideia é criar séries aleatórias falsas e compará-las com as séries previstas. Dessa forma é possível avaliar se as mudanças pontuais em cada variável independente (intervalo de envio de pacotes, velocidade dos nós, passos à frente na previsão e tamanho dos pacotes) ultrapassam os limites de variabilidade estabelecidos nas séries falsas.

Por exemplo, suponha um par de coordenadas ($x=16,04$ e $y=307,25$) exposto a 5% de aleatoriedade. Para este exemplo, as faixas limites são:

- *Limite inferior $x = 16,04 * (-5\%) = 15,238$;*
- *Limite superior $x = 16,04 * (+5\%) = 16,842$;*
- *Limite inferior $y = 307,25 * (-5\%) = 291,8875$;*
- *Limite superior $y = 307,25 * (+5\%) = 322,6125$.*

Assim, as falsas coordenadas podem ser quaisquer valores compreendidos entre as faixas mostradas. Quando uma determinada série, com uma combinação de variáveis independentes específica gerar um valor qualquer de variabilidade, este valor pode ser comparado com as séries falsas para que se saiba se a variabilidade destas ultrapassou a variabilidade das séries falsas de 1%, 2%, 5%, 10%, 15% ou 20%. Este tipo de informação é muito importante, pois a análise estratégica pode indicar em qual nível de variabilidade nociva uma determinada combinação de variáveis independentes pode funcionar. Assim, de forma geral, estes valores funcionam como limites para o SGA. Não há geração de valores falsos para a energia - a opção foi determinar um esquema com quartis de energia mínimos para trabalho (vide Seção 5.1.3 Resultados do Primeiro Experimento).

Com a geração de cinco séries falsas geradas com todas as variantes e variáveis independentes anteriormente citadas, o total de séries históricas para avaliação chega a 178.200. Ao invés de avaliar todos os cinco nós armazenadores (1, 3, 5, 7 e 9) e ter uma repetição de resultados, optou-se por analisar apenas um deles. Arbitrariamente o escolhido foi o nó 5.

5.1.2.7 Definição de Relevância

Apesar de ser um item opcional, como especificado na proposta do SGA, a relevância das variáveis independentes é verificada neste Primeiro Experimento. Na realidade as variáveis independentes (velocidade dos nós, intervalo de envio de pacotes, passos à frente nas previsões e tamanho dos pacotes) foram escolhidas com base na experiência – já se sabia previamente que essas variáveis tinham alguma influência nas variáveis resposta (variabilidade e/ou energia). Contudo, a fim de demonstrar a viabilidade do uso de um modelo para indicar a relevância das variáveis independentes, estas foram ajustadas por meio de um modelo linear ao fim dos experimentos.

Modelos lineares de primeira ordem apresentam a forma (DRAPER; SMITH, 1998):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + e \quad (24)$$

Onde: $x_1, x_2, \dots, x_k$ são variáveis independentes que podem ser observadas e/ou controladas e cujos valores podem exercer alguma influência em uma variável resposta denotada como y .

$e = \text{erro aleatório} \Rightarrow Y_i \sim N(\mu, \sigma^2) \Rightarrow$ com distribuição Normal, variância constante (σ^2) e esperança ($E(Y) = 0$).

$\beta_0, \beta_1, \dots, \beta_k$ são parâmetros desconhecidos que podem ser estimados usando o Método dos Mínimos Quadrados observado em Draper e Smith (1998).

É importante lembrar os modelos lineares possuem alguns pré-requisitos: resíduos que possuem distribuição normal, com média zero e variância constante (MORETTIN; TOLOI, 2006; DRAPER; SMITH, 1998). Porém, um simples ajuste nos dados usando regressão linear, **desde que não aconteçam inferências a respeito dos resultados do modelo**, pode ser realizado sem que se observem os pré-requisitos citados. O programa de regressão linear dos dados, escrito para funcionar com o aplicativo estatístico R-free/R-studio (<https://www.r-project.org/>) está disponível no APÊNDICE C.

5.1.2.8 Aplicação do SGA no Primeiro Experimento

O SGA é aplicado para que um equilíbrio entre os resultados de variabilidade e de energia seja alcançado a partir do processo de detecção (série real) e de previsão (série prevista). Na Tabela 6 e na Tabela 7 os módulos do nível técnico e do nível estratégico são descritos.

Tabela 6 - Nível técnico: configurações do Primeiro Experimento

Módulos	Descrição	Detalhes
1.1 Seleciona controles de segurança	Detecção: a série real Previsão: a série prevista	Detecção: a energia Previsão: a variabilidade
1.2 Define variáveis independentes	Intervalo de envio de pacotes (s) Velocidade de movimentação dos nós (m/s) Passos à frente na previsão Tamanho dos pacotes (<i>bytes</i>)	
1.3 Define variáveis resposta	Desvio padrão Energia restante	Variantes: Dois métodos de previsão: GM(1,1) e tendência polinomial Dois tipos de posicionamento: <i>M</i> e <i>M2</i> Dois Estimadores de variabilidade: amplitude e média
1.4 Define o tamanho das amostras	Simulação de 150 s gerando por volta de 300 amostras no menor intervalo de envio de pacotes (0,05 s)	Para energia, a simulação é de 1500 s
1.5 Obtêm os resultados e configurações dos controles de segurança	<i>Matriz de dados</i> : série prevista gerada a partir da série real	
1.6 Define limites para as variáveis resposta	<i>Matriz limite</i> (variabilidade): geração de cinco séries com valores de posicionamento falsos: 1%, 2%, 5%, 10%, 15% e 20% <i>Matriz limite</i> (energia): valores mínimos aceitáveis acima do primeiro quartil	Para valores mínimos de energia, vide Seção 5.1.3 Resultados do Primeiro Experimento
1.7 Define relevância	<i>Matriz relevante</i> (variabilidade): intervalo, velocidade e passos à frente na previsão <i>Matriz relevante</i> (energia): intervalo e tamanho dos pacotes	

Fonte: o autor.

Tabela 7 - Nível estratégico: configurações do Primeiro Experimento

Módulos	Descrição	Detalhes
2.1 Variáveis independentes → jogadores	<i>Matriz dos jogadores:</i> variáveis independentes ligadas às respectivas variáveis resposta	
2.3 Padroniza as variáveis resposta a partir da série limite	<i>Matriz estratégica:</i> todas as variáveis resposta são padronizadas	Não são usados pesos
2.4 Define WTP e custo	<p>WTP variante 1: amplitude do desvio padrão de M com tendência polinomial</p> <p>WTP variante 2: amplitude do desvio padrão de $M2$ com tendência polinomial</p> <p>WTP variante 3: média do desvio padrão de M com tendência polinomial</p> <p>WTP variante 4: média do desvio padrão de $M2$ com tendência polinomial</p> <p>WTP variante 5: amplitude do desvio padrão de M com GM(1,1)</p> <p>WTP variante 6: amplitude do desvio padrão de $M2$ com GM(1,1)</p> <p>WTP variante 7: média do desvio padrão de M com GM(1,1)</p> <p>WTP variante 8: média do desvio padrão de $M2$ com GM(1,1)</p> <p>Custo: energia restante nos nós</p>	
2.5 Cálculo e análise UCB	<i>Matriz UCB:</i> Geração de oito matrizes UCB: uma para cada variante	Os cálculos UCB são repetidos para todos os registros das amostras

Fonte: o autor.

No total são geradas oito matrizes UCB, uma para cada variante. Com isso, além da análise realizada sobre as variáveis independentes, também é possível analisar como os resultados do SGA funcionam para cada variante.

Para um melhor entendimento a tabela 8 ilustra alguns valores das variáveis independentes ligados aos respectivos valores das variáveis resposta (x_{11} a x_{n9}), nas oito variantes possíveis de variabilidade e na energia.

Tabela 8 - Valores das variáveis independentes e das variantes das variáveis resposta.

Jogador (Variáveis Independentes)				Variáveis resposta								Energia
				GM(1,1)				Tendência Polinomial				
				Amplitude		Média		Amplitude		Média		
Interv.	Vel.	Passos	Tam.	M	M2	M	M2	M	M2	M	M2	
0,05	1	3	12	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}
0,05	1	3	25	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	x_{27}	x_{28}	x_{29}
:	:	:	:	:	:	:	:	:	:	:	:	:
9,85	5	12	200	x_{n1}	x_{n2}	x_{n3}	x_{n4}	x_{n5}	x_{n6}	x_{n7}	x_{n8}	x_{n9}

Fonte: o autor.

As quatro primeiras colunas da tabela 8 mostram valores assumidos pelas variáveis independentes intervalo de envio de pacotes, velocidade dos nós, passos à frente na previsão e tamanho dos pacotes. As colunas subsequentes mostram os valores para as variáveis resposta de variabilidade (com todas as variantes possíveis: método de previsão, estimador de variabilidade e tipo de posicionamento) e energia restante no nó

É importante notar que a energia armazenada nos nós não precisa ser gerada para cada variante, pois apenas um valor de energia para cada jogador é suficiente para denotar quanta energia é economizada.

5.1.3 Resultados do Primeiro Experimento

Os gráficos das figuras 10 e 11 (para o método GM(1,1)) e das figuras 12 e 13 (para o método de tendência polinomial) ilustram de forma bastante clara a

tendência de subida da variabilidade gerada pela combinação das variáveis independentes presentes na *matriz de dados*. Percebe-se que há um aumento linear da variabilidade diretamente proporcional ao aumento do intervalo de tempo no envio de pacotes. Observando-se os valores dos eixos das coordenadas de todos os gráficos (valores aumentam quando a velocidade aumenta), pode-se perceber que a velocidade dos nós também influencia positivamente a variabilidade.

Figura 10 - Resultados do primeiro experimento - Amplitude do desvio padrão - GM(1,1)

(a)

(b)

(c)

(d)

(e)

(f)

Fonte: o autor.

Figura 11 - Resultados do primeiro experimento - Média do desvio padrão - GM(1,1)

(a)

(b)

(c)

(d)

(e)

(f)

Fonte: o autor.

Figura 12 - Resultados do primeiro experimento - Amplitude do desvio padrão - Tendência polinomial

(a)

(b)

(c)

(d)

(e)

Fonte: o autor.

Figura 13 - Resultados do primeiro experimento - Média do desvio padrão - Tendência polinomial

(c)

(d)

(e)

(f)

Fonte: o autor.

Nos gráficos das figuras 10 a 13 também se pode perceber que em relação ao tipo de posicionamento, *M2* mostra um ajuste levemente melhor, com linhas mais suaves que *M*. Já em relação ao método de previsão, fica bem claro que o ajuste da tendência polinomial é mais suave (e linear!) que o conseguido pelo método GM(1,1). O principal motivo é a falta do caráter estatístico do método GM(1,1), mais apropriado para previsões com poucos dados de entrada.

Estas tendências se repetem em maior ou menor grau em todas as séries previstas. Nas séries maliciosas esta tendência é bem menos acentuada, pois devido a alta variabilidade a curva não tem um aspecto tão linear (vide figuras 10 e 11).

Figura 14 - Comparativo entre a variabilidade de séries previstas e falsas com 1% de aleatoriedade para o método GM(1,1)

Fonte: o autor.

Figura 15 - Comparativo entre a variabilidade de séries previstas e falsas com 1% de aleatoriedade para o método de tendência polinomial

Fonte: o autor.

Os gráficos das figuras 14 e 15 mostram dois exemplos de contraponto entre as curvas de variabilidade das séries previstas e as curvas das séries falsas. Os dados ilustrados são referentes a um conjunto de variáveis independentes e variantes

específicas ordenados por intervalo de envio de pacotes (velocidade de 1 m/s; 6 passos à frente; tipo de posicionamento M ; estimador de variabilidade pela média do desvio padrão). Nas duas figuras percebe-se a maior variabilidade e linearidade praticamente nula das séries falsas, mesmo com o percentual de aleatoriedade no menor valor avaliado (1%).

Para servir de comparativo, a tabela 9 apresenta a taxa de aumento entre o menor percentual da série falsa (1%) com a série prevista.

Tabela 9 - Percentual de aumento no desvio padrão

Velocidade = 1 m/s					
1% Aleatoriedade – Percentual de aumento do desvio padrão					
Método previsão	Passos à frente	% Aumento na amplitude		% Aumento na média	
		M	$M2$	M	$M2$
GM(1,1)	3	226,55	99,64	313,37	214,92
GM(1,1)	6	49,33	11,96	96,67	43,2
GM(1,1)	12	5,91	2,65	22,12	5,96
Tendência Polinomial	3	18,35	5,27	96,07	18,52
Tendência Polinomial	6	2,95	0,81	11,80	1,49
Tendência Polinomial	12	0,07	0,17	0,15	0,18

Fonte: o autor.

O objetivo da tabela 9 é mostrar o percentual de aumento da variabilidade na comparação entre a série prevista e a série falsa com 1% de aleatoriedade, ambas com velocidade dos nós de 1 m/s (o menor valor). A primeira coluna mostra os dois métodos de previsão usados (tendência polinomial e GM(1,1); a segunda coluna mostra a quantidade de passos à frente na previsão; as próximas colunas mostram os percentuais de aumento na amplitude e na média do desvio padrão de M e $M2$ (tipo de posicionamento) da série falsa em relação à série prevista. Pode-se perceber aumentos significativos das amplitudes e das médias dos desvios padrão de M e $M2$. Os aumentos são mais acentuados quando a série adota o método de previsão GM(1,1) – corroborando os resultados dos gráficos das figuras 10 a 13 que mostram um melhor ajuste para o método de tendência polinomial. Como um exemplo extremo, pode-se notar na tabela 9 que o uso de coordenadas falsas com

1% de aleatoriedade (o menor percentual avaliado) gera, quando o estimador é a média, um aumento de 313,37% para *M* e 214,92% para *M2* (dados marcados em negrito na primeira linha da tabela 9). Também é possível notar que os percentuais de aumento de ambos os estimadores (amplitude e média) são inversamente proporcionais à quantidade de passos à frente da previsão. Isso acontece porque a tendência é que quanto maior a quantidade de passos à frente da previsão, maior é a variabilidade. Consequentemente uma maior janela de previsão aproxima a série prevista da série falsa.

Ao mostrar que, para determinadas combinações de variáveis independentes da *matriz de dados*, uma série composta por falsos valores de posicionamento com 1% de aleatoriedade pode ser detectada como maliciosa por causa da variabilidade, a capacidade de identificação de nós maliciosos do método fica evidente.

Em relação à geração da *matriz relevante*, parte dos resultados da regressão linear aplicada às variáveis independentes (velocidade, intervalo de envio de pacotes, passos à frente na previsão e tamanho dos pacotes) em relação à cada uma das variáveis resposta (amplitude do desvio padrão, média do desvio padrão e energia restante no nó) estão expostos na tabela 10 e confirmam as tendências anteriormente citadas para as variáveis independentes – uma versão mais completa da tabela 10 com os resultados de relevância pode ser vista na tabela 29 no APÊNDICE D. É importante lembrar dois detalhes importantes deste processo: {1} as variáveis independentes são escolhidas tendo como base conhecimentos prévios sobre a relevância das mesmas; e {2} os dados são ajustados em modelos lineares, mas não passam por inferências, fato que livra o modelo da obrigação da verificação da normalidade da distribuição.

Tabela 10 - Relevância de parte das variáveis independentes com base em um modelo de regressão linear.

(a) Variabilidade - Amplitude - M – GM(1,1)

Var. Indep.	Estimativas	p-value
Velocidade	1,41E-003	0
Intervalo	7,73E-004	0
Passos à frente	-1,67E-004	3,14E-153
Tamanho	4,39E-020	1

(b) Energia restante

Var. Indep.	Estimativas	p-value
Velocidade	1,12E-014	1
Intervalo	1,28	0
Passos à frente	7,57E-015	1
Tamanho	-9,89E-003	1,35E-015

Fonte: o autor.

A tabela 10 (a) é relativa ao método de previsão GM(1,1), ao estimador amplitude, ao tipo de posicionamento M e traz como variável resposta a variabilidade. A primeira coluna mostra as variáveis independentes. A segunda coluna mostra as estimativas resultantes da regressão linear. A terceira coluna mostra o nível de significância (p -value) de cada variável independente. P -values menores que 0,05 indicam probabilidades máximas de erro para as estimativas em 5%. Estimativas positivas indicam a influência linear direta (o aumento/diminuição da variável independente aumenta/diminui a variável resposta). Por outro lado, estimativas negativas indicam influência linear inversa (o aumento/diminuição da variável independente diminui/aumenta a variável resposta). Percebe-se, a partir dos baixos p -values, que a velocidade, o intervalo e os passos à frente têm significância na variabilidade, enquanto o tamanho dos pacotes não.

Para avaliar qual registro (combinação de variáveis independentes) da *matriz de dados* consome mais energia, uma simulação um pouco mais extensa é realizada. Os valores de energia são recolhidos para todas as séries previstas por 1500 s. As duas variáveis independentes que influenciam a economia de energia são o

tamanho e o intervalo de envio de pacotes. Ao final de 1500 s a energia restante nos nós é medida. Tem-se, portanto, uma medida de energia restante para cada registro (combinação de variáveis independentes).

Sendo assim, a tabela 10 (b) mostra os resultados do modelo de regressão linear para a energia restante nos nós. Há apenas uma tabela com a economia das baterias porque não há diferenças de energia se as variantes (métodos de previsão, estimador e tipo de posicionamento) forem alteradas. Uma única leitura de energia na simulação de 1500 s é suficiente.

Algumas pequenas conclusões podem ser tomadas a partir da análise da relevância das variáveis independentes do Primeiro Experimento presentes na tabela 10:

- As estimativas positivas para a variabilidade e negativas para a energia denotam que o aumento no intervalo de envio de pacotes influencia positivamente a variabilidade e negativamente o consumo de energia;
- O aumento na velocidade dos nós influencia positivamente a variabilidade;
- O aumento nos passos à frente da previsão influencia positivamente a variabilidade quando o método de previsão é a tendência polinomial. Para o método de previsão GM(1,1) as estimativas negativas indicam que o aumento nos passos à frente diminui a variabilidade. Trata-se de um resultado inesperado para o método GM(1,1), porém, mostra que um método estatístico como a tendência polinomial pode ser mais adequado para este tipo de previsão (baseada em valores anteriores de posicionamento);
- O aumento no tamanho dos pacotes influencia positivamente o consumo de energia. Porém, é possível notar que as estimativas são baixas, mostrando que a influência desta variável independente no consumo de energia não é tão enfática (vide figura 16).

Uma versão mais completa da tabela 10 pode ser vista na tabela 29 (APÊNDICE D) que apresenta a relevância de todas as variáveis independentes.

A figura 16 mostra a distribuição de energia para cada tamanho *versus* intervalos de envio de pacotes.

Figura 16 - Distribuição de energia ao longo de 1500 s para diferentes tamanhos e intervalos de envio de pacotes

Fonte: o autor.

A figura 16 mostra a média da energia restante dos nós (eixo das coordenadas) a partir do aumento do tempo de intervalo de envio de pacotes (eixo das abscissas) em relação aos cinco tamanhos de pacotes testados: 200, 100, 50, 25 e 12 *bytes*. Pode-se perceber que nos primeiros intervalos, quando o envio de pacotes é intenso, a energia que resta nos nós é baixa (o consumo por envio de pacotes é maior). Porém, em relação ao tamanho dos pacotes, o comportamento das curvas é bastante parecido. O consumo é levemente afetado pelo tamanho, com pacotes maiores consumindo um pouco mais (vide as baixas estimativas para o tamanho dos pacotes em relação ao consumo de energia na tabela 10).

Para definição dos limites mínimos de energia, os valores de energia restante nos nós de todas as combinações de variáveis independentes, separadas pelo tamanho do pacote, são armazenados em um vetor chamado de *vetor de energia*. Em seguida, calculam-se os quartis deste vetor. Arbitariamente foi definido que apenas

os nós cujos valores de energia restante fossem maiores que os valores do primeiro quartil do *vetor de energia* seriam aceitos. Dessa forma, quaisquer combinações de variáveis independentes da *matriz de dados* cujos valores de energia forem mais baixos que o primeiro quartil do *vetor de energia* são eliminadas dos cálculos UCB. Para melhor entendimento a tabela 11 mostra os quartis obtidos a partir do *vetor de energia* para os cinco tamanhos de pacote:

Tabela 11 - Vetor de energia com os quartis obtidos em simulação de 1500 s

Percentual de energia restante nos nós para intervalo de envio de pacotes: 0,05 a 9,85					
Quartis	Tamanho dos pacotes				
	200 bytes	100 bytes	50 bytes	25 bytes	12 bytes
0º (0%) – corte	53,05%	54,43%	54,36%	55,01%	54,76%
1º (25%) – corte	94,82%	95,80%	96,20%	96,56%	96,66%
2º (50%)	96,77%	97,28%	97,57%	97,61%	97,70%
3º (75%)	97,41%	97,77%	97,94%	97,99%	98,05%
4º (100%)	97,82%	98,04%	98,19%	98,23%	98,26%

Fonte: o autor.

Na tabela 11 podem ser observados os valores médios de energia que restaram nos nós. Esses valores estão separados (na primeira coluna) em quartis de 0%, 25%, 50%, 75% e 100%. Dependendo do tamanho dos pacotes especificados nas colunas dois a seis, todos os nós cuja energia restante seja menor que os valores do quartil de 25% são descartados. De acordo com a tabela 11, um registro (combinação de variáveis independentes) da *matriz de dados* cujo tamanho do pacote seja 200 bytes precisa ter mais de 94,82% de energia restante no nó (medida marcada em negrito na segunda linha da tabela 11), caso contrário é cortado da *matriz de dados*.

Sendo assim, a *matriz limite* é composta pelos valores de energia menores ou iguais aos valores do primeiro quartil do *vetor de energia* e pelos valores de variabilidade encontrados nas cinco séries falsas.

Tendo em vista que os jogadores que possuem os maiores valores de CM são os beneficiados em uma barganha irrestrita, a composição da *matriz UCB* segue o que está disposto nas tabelas 12 e 13, que mostram para os dois métodos de previsão apenas os cinco primeiros registros com maior CM com a amplitude como estimador e *M* como tipo de posicionamento.

Tabela 12 - Cinco registros com maior CM da matriz UCB para o método de tendência polinomial.

Método: Tendência polinomial – Estimador: Amplitude – Tipo: M						% Aleatoriedade permitida					
Jogador *	WTP	Custo	V(N)	V(N\{i})	CM	1%	2%	5%	10%	15%	20%
1-6,05-3-12	398,739	319,527	79,211	-318,525	397,737	X	X	☺	☺	☺	☺
1-5,75-3-25	396,634	319,529	77,104	-318,525	395,630	☺	☺	☺	☺	☺	☺
1-5,95-3-12	396,422	319,527	76,894	-318,525	395,420	☺	☺	☺	☺	☺	☺
1-5,85-3-25	396,272	319,528	76,743	-318,525	395,269	☺	☺	☺	☺	☺	☺
1-6,15-3-12	395,914	319,527	76,386	-318,525	394,912	☺	☺	☺	☺	☺	☺

* Jogador= velocidade dos nós – intervalo de envio de pacotes – passos à frente na previsão – tamanho dos pacotes.

Fonte: o autor.

Tabela 13 - Cinco registros com maior CM da matriz UCB para o método GM(1,1).

Método: GM(1,1) – Estimador: Amplitude – Tipo: M						% Aleatoriedade permitida					
Jogador *	WTP	Custo	V(N)	V(N\{i})	CM	1%	2%	5%	10%	15%	20%
1-5,75-12-25	1197,195	947,357	249,837	-946,356	1196,194	☺	☺	☺	☺	☺	☺
1-5,75-6-25	1193,998	947,357	246,640	-946,356	1192,996	☺	☺	☺	☺	☺	☺
1-5,95-12-12	1192,003	947,356	244,647	-946,356	1191,004	☺	☺	☺	☺	☺	☺
1-6,05-12-12	1188,215	947,356	240,858	-946,356	1187,215	☺	☺	☺	☺	☺	☺
1-5,85-12-25	1188,177	947,357	240,820	-946,356	1187,177	X	☺	☺	☺	☺	☺

* Jogador= velocidade dos nós – intervalo de envio de pacotes – passos à frente na previsão – tamanho dos pacotes.

Fonte: o autor.

As tabelas 12 e 13 trazem respectivamente o método de previsão tendência polinomial e o método GM(1,1). Ambas usam a amplitude como estimador de variabilidade e *M* como o tipo de posicionamento. A primeira coluna mostra o jogador formado pelas variáveis independentes (velocidade dos nós, intervalo de envio de pacotes, passos à frente na previsão e tamanho dos pacotes). A segunda coluna *WTP* indica o valor que um jogador pode usar no processo de barganha, representado pelo desvio padrão de cada jogador (o valor passou pelo processo de padronização). Na terceira coluna *Custo* está o valor que recai sobre o jogador no processo de barganha, representado pelo gasto de energia (valor da energia restante invertido e padronizado). A quarta coluna *V(N)* é a subtração das duas

colunas anteriores e mostra o TVC (valor total gerado no jogo). A quinta coluna $V(N\{i\})$ mostra valor gerado no jogo caso o jogador em questão saia da negociação. A sexta coluna *CM* mostra a Contribuição Marginal (CM) cujo alto valor é desejável e mostra o jogador mais preparado para vencer a negociação – no caso deste Primeiro Experimento, o jogador mais adequado em termos de baixa variabilidade nas medições de posicionamento dos nós e economia de energia.

Alguns detalhes podem ser destacados observando-se as tabelas 30 e 31 do APÊNDICE E, que são versões completas das tabelas 12 e 13, com os cinco primeiros registros com maior CM de todas as variantes.

- A velocidade ideal é 1 m/s , a menor possível e a que proporciona as menores amplitudes e médias para os valores de desvio padrão;
- O intervalo de envio de pacotes dos melhores resultados, salvo algumas exceções, se situa entre $5,75$ e $8,35\text{ pacotes/s}$. Como já observado, o tempo de intervalo influencia tanto a variabilidade quanto o consumo de energia e baixos tempos no envio de pacotes estão diretamente ligados aos baixos valores de variabilidade. Este resultado é justamente interessante porque mostra que há uma zona no tempo dos intervalos onde ocorre o equilíbrio entre bons valores de variabilidade e bons valores de consumo de energia;
- A janela de passos à frente apresenta um comportamento diferente para cada método de previsão – fato observado anteriormente e que mostra coerência, visto que os métodos de previsão usados apresentam resultados um pouco diferentes ao longo do trabalho. Na previsão realizada pelo método de tendência polinomial, cuja curva de previsão foi mais suave e linear que o método GM(1,1) (vide figuras 10 a 13), a quantidade de passos à frente dos melhores resultados é sempre 3, indicando que uma janela de previsão menor tende a mostrar menor variabilidade nos resultados de posicionamento dos nós. Porém, na previsão realizada pelo método GM(1,1), para os melhores resultados a janela de previsão alterna entre os valores 3, 6 e 12 passos à frente;
- O tamanho dos pacotes indica que os pacotes menores tem sua importância no consumo de energia. O equilíbrio dos melhores resultados está entre 12 e 25 bytes (os menores possíveis). Há duas exceções nas quais jogadores chegam até o tamanho de 50 bytes .

Um resultado adicional que pode ser observado nas tabelas 12 e 13, nas últimas colunas (*% Aleatoriedade permitida*), é a comparação dos resultados de variabilidade dos cinco melhores resultados com a variabilidade das séries falsas. As colunas marcadas com 1%, 2%, 5%, 10%, 15% e 20% estão preenchidas com **X** quando a variabilidade do respectivo jogador é **maior** que a série falsa. Isso indica que as configurações indicadas no jogador, por melhores que sejam os resultados, não podem ser diferenciadas da variabilidade da série falsa em questão. O inverso ocorre quando aparece nas colunas de percentuais o símbolo ☺ (sucesso). Por exemplo, na primeira linha da tabela 12 (*jogador=1-6,05-3-12*) os percentuais de 1% e 2% estão marcados com **X**. Isso indica que as configurações (variáveis independentes) deste jogador, apesar de terem gerado o melhor resultado de CM, geram um resultado de amplitude do desvio padrão para *M* que não pode ser diferenciado de uma variabilidade gerada por uma série falsa com 1% ou 2% de aleatoriedade. Sendo assim, fica a critério de um administrador de segurança utilizar ou não a configuração do jogador. É fato, porém, que nas tabelas 30 e 31 do APÊNDICE E a grande maioria dos percentuais estão preenchidos pelo símbolo ☺, indicando que mesmo para a menor aleatoriedade (1%) o método é capaz de identificar a presença de séries falsas.

5.1.4 Discussões sobre o Primeiro Experimento

Este Primeiro Experimento mostra com resultados satisfatórios, a viabilidade da aplicação de uma metodologia de previsão de séries históricas de posicionamento de nós em uma MANET em conjunto com uma metodologia de detecção de energia. Ambas as metodologias são consideradas respectivamente controles de previsão e de detecção. O principal objetivo é a identificação de posicionamentos maliciosos dos nós considerando aspectos de segurança, métodos de previsão, variabilidade, coordenadas de localização, consumo de energia, estimadores e simulação de falsos posicionamentos dos nós. Também são levadas em conta as configurações dos controles de segurança (variáveis independentes): velocidade dos nós, intervalo de envio de pacotes, passos à frente na previsão e tamanho dos pacotes, além dos resultados (variáveis resposta) destes mesmos controles: amplitude e média do desvios padrão das séries previstas e a energia restante nos nós.

Para avaliar todos estes parâmetros a opção é o agrupamento dos controles de segurança voltados para detecção e previsão por meio da aplicação do SGA.

Os resultados são bastante satisfatórios e mostram faixas de trabalho para configurações que correspondem aos menores valores de variabilidade das séries previstas (obtidos por previsão) em conjunto com os maiores valores de economia de energia nos nós (obtidos por detecção). Com a aplicação do SGA é possível ainda identificar, entre os resultados mais otimizados, quais faixas de variabilidade indicam comportamento malicioso por parte dos nós da rede. Na maioria dos casos, variações de apenas 1% na aleatoriedade de séries históricas maliciosas podem ser identificadas como falsas.

Há alguns aspectos que podem ser citados como exemplos de limitações deste Primeiro Experimento:

- A simulação do experimento não permitiu a presença de variações aleatórias reais entre os resultados da série real. Por exemplo, flutuações nos valores ao longo do tempo ou interferência de ruídos na captura dos dados. Em termos estatísticos o ideal seria se pequenas variações as quais um experimento prático está sujeito pudessem ter sido simuladas;
- Alguns valores usados nas variáveis independentes, como velocidade dos nós (1, 3 e 5 m/s), passos à frente na previsão (3, 6 e 12) foram escolhidos com o intuito de refinar a análise e são baseados em valores publicados em um trabalho anterior (SILVA et al., 2014). Porém, a variável independente tamanho dos pacotes (200, 100, 50, 25 e 12 bytes) e a variante aleatoriedade das séries falsas (1%, 2%, 5%, 10%, 15%, 20%) foram escolhidas arbitrariamente.
- Apesar da antecipação gerada pelos métodos de previsão empregados, a análise estratégica foi realizada ao final das simulações. Ou seja, não houve análise estratégica em tempo real, somente no fim do experimento. Isso viabiliza o uso das configurações otimizadas indicadas nos resultados do SGA apenas para um próximo experimento e não para uso em tempo real.

5.2 Segundo Experimento

Um dos problemas na implementação de controles de segurança voltados para previsão na análise de tráfego malicioso é a dificuldade de se encontrar um método preditivo que funcione em tempo real. Em geral, isso acontece porque a previsão de tráfego tem como base a análise de séries históricas, ou seja, depende muito do processo de detecção (PAPAGIANNAKI, 2005; CORTEZ et al., 2012). Na verdade, se a previsão for eficiente, essa dependência não chega a ser necessariamente um problema, pois um processo preditivo pode até funcionar baseado em parâmetros que não são oriundos do processo de detecção (ABDULLAH et al., 2015). Contudo, como já relatado, o agrupamento de controles de segurança de detecção e de previsão em ambientes que exigem análise de tráfego malicioso ainda apresenta desafios.

Neste Segundo Experimento um *site* de comércio eletrônico passa por ataques cibernéticos variados. Como pode ser visto na figura 17, a topologia do *site* permite uma análise bastante precisa do tráfego de rede já que é composta por dois controles de segurança (*Filtro 1* e *Filtro 2*) voltados para a detecção e que trabalham em sequencia:

Figura 17 - Topologia de rede com dois controles de segurança: *Filtro 1* e *2*

Fonte: o autor.

O *Filtro 1* é um Web Application Firewall (WAF) que inspeciona o cabeçalho dos pacotes na camada de aplicação do protocolo TCP/IP e reconhece uma série de ataques sofisticados como injeção de Standard Query Language (SQL) e Distributed Denial of Service (DDoS). O *Filtro 2* é um aparelho de *firewall* que bloqueia tráfego maliciosos de acordo com regras pré-definidas. A topologia ilustrada na figura 17 mostra que o acesso ao *site* de comércio eletrônico passa necessariamente pelos dois filtros - apenas o tráfego liberado pelo *Filtro 1* chega ao *Filtro 2*. É importante notar que os filtros detectam o tráfego malicioso de forma diferente um do outro.

Apesar de todo o aparato de segurança, buscam-se formas de melhorar a eficiência do sistema por meio da prevenção dos ataques em tempo real. O objetivo traçado é expandir a atuação do *Filtro 1* bloqueando previamente tráfego malicioso detectado apenas pelo *Filtro 2*. Além da antecipação da descoberta de ataques, a consequência direta dessa ação é a diminuição da carga de trabalho do *Filtro 2*. A topologia peculiar da rede, com os dois filtros em sequência, até contribui para a implantação desse esquema de prevenção. Contudo, a dificuldade reside no fato do

Filtro 1 não detectar o tráfego malicioso que o *Filtro 2* detecta – apesar de ambos analisarem o mesmo fluxo de dados, cada filtro detecta diferentes anomalias no tráfego. Vale mencionar que um esquema de prevenção nos moldes citados deve ter o menor índice de bloqueios FP possível, já que o corte de pacotes válidos é prejudicial ao *site*.

Para resolução dos problemas listados sugere-se a análise do tráfego de rede que passa pelos filtros em amostras com 500000 pacotes cada uma, agrupadas por configurações (campos da camada de aplicação do protocolo TCP/IP coletados pelo *Filtro 1*) e pelo número IP de origem. O tamanho das amostras foi arbitrariamente escolhido simplesmente por ser passível de manipulação por editores de texto comuns. A partir da análise do tráfego propõe-se a criação de um *ranking* crescente que indique níveis de suspeição para as configurações e possa ser usado como preditor para a próxima amostra. Dessa forma, a função principal desse *ranking* é barrar as configurações suspeitas quando estas passarem pelo *Filtro 1*, antes portanto, da chegada ao *Filtro 2*. Supõe-se que o efeito dessa antecipação diminua a carga e melhore o desempenho do *Filtro 2*.

O *ranking* é formado por um conjunto de dados recolhidos das amostras. De forma geral, as seguintes variáveis são levadas em conta: {1} o percentual de tráfego malicioso detectado para cada configuração; {2} a representatividade percentual da quantidade de ataques detectados por configuração na amostra e na totalidade de dados analisados; {3} o percentual de tráfego malicioso detectado para cada par IP + configuração; {4} a representatividade percentual da quantidade de ataques detectados para cada par IP + configuração, na amostra atual e no total dos dados analisados; e {5} um percentual que indica se há na configuração analisada, campos que influem na geração de ataques. De acordo com a relevância, cada uma dessas variáveis é multiplicada por pesos pré-determinados.

Em uma segunda etapa, como a coleta de dados funciona em tempo real, existe a necessidade de se estabelecer qual o nível ideal de uso para o *ranking* – qual valor deve ser usado - o SGA é implementado para esse fim.

A ideia de usar um *ranking* ou um esquema de votações baseado em pesos para classificar eventos indesejados não é nova e está presente em diversas abordagens. Entre os exemplos de diversidade, pode ser citado o trabalho de Gupta; Kumar e

Abraham (2013) onde são criados perfis para diferentes máquinas virtuais para detecção de intrusões com base em um *ranking*. Outro exemplo é o trabalho de Hock e Kappes (2014) que usa treinamento de dados baseado em um esquema de votações para construir um SDI/SPI baseado em anomalia.

Em alguns casos a abrangência desse tipo de metodologia se aproxima da proposta do SGA. Por exemplo, um modelo matemático baseado em um esquema de votações é proposto por Mitchell e Chen (2013) para balancear a detecção de intrusões e o consumo de energia de nós móveis.

Em um trabalho bastante representativo, e já referenciado no Capítulo 3 Trabalhos Relacionados, Lin et al. (2013) procuram diminuir a ocorrência de FP e FN em vários SDI/SPI com funcionamento simultâneo. Inicialmente a credibilidade de cada mecanismo de detecção é avaliada pelo monitoramento das respostas para vários tipos de ataque. Em seguida pesos são atribuídos a cada tipo de ataque tendo como base a ocorrência de FP e FN. Por fim as saídas de cada SDI/SPI são agrupadas por meio de um esquema de votações e pesos. Os autores provam que o uso de pesos provê um desempenho melhor que esquemas de votação ordinária.

As pesquisas referenciadas servem para justificar a adoção de um *ranking* baseado em pesos na proposta deste Segundo Experimento.

5.2.1 Formalização do Segundo Experimento

Para a realização do Segundo Experimento a seguinte configuração é utilizada:

- Computador com 4 núcleos com 16 GB de RAM e 1 TeraByte (TB) de disco, com máquina virtual com servidor Windows Server 2012 R2 para armazenamento e transformação dos dados;
 - Gerenciador de banco de dados Microsoft SQL Server 2014 Standard (<https://www.microsoft.com/pt-br/server-cloud/products/sql-server/default.aspx>) para classificação, ordenamento e transformação dos dados no formato texto;

- *Software* WinSCP (<https://winscp.net/eng/download.php>) para o estabelecimento de conexões Secure SHell (SSH) entre os filtros 1 e 2;
- *Notebook* Dell-Vostro com processador Intel Core i5 M-520 com 2,40 Giga-Hertz (GHz) e 4 GB Random Access memory (RAM) para análise dos dados;
 - Microsoft Visual FoxPro (banco de dados);
 - Planilha eletrônica: Microsoft Excel;
 - Aplicativo estatístico: R-free/R-studio (<https://www.r-project.org/>).

Há quatro processos distintos envolvidos na condução do Segundo Experimento: {1} detecção; {2} criação do *ranking*; {3} previsão; e {4} agrupamento.

5.2.1.1 Detecção

O processo de detecção é caracterizado pela coleta dos pacotes de tráfego que passam pelos filtros e segue o seguinte escopo:

- Os filtros $F_i = \{f_1, f_2\}$ detectam características diferentes no tráfego de rede;
- A classificação dos campos dos filtros está na tabela 14. Cada conjunto de campos com diferentes combinações forma uma configuração: $C_j = \{c_1, \dots, c_n\}$;
- O par formado pelo IP + configuração também é agrupado na forma $IP_j = \{ip_1, \dots, ip_n\}$;
- Há dois níveis de segurança $S_k = \{s_1, s_2\}$ com os quais o *Filtro 1* pode funcionar:
 - $S_1=400$ (aberto): pacotes cujo campo $STATUS < 400$ são liberados pelo *Filtro 1*. É um nível menos restritivo que aprova mais pacotes;
 - $S_2=201$ (fechado): pacotes cujo campo $STATUS < 201$ são liberados pelo *Filtro 1*. É um nível mais restritivo que bloqueia mais pacotes;
- Um pacote é considerado malicioso pelo *Filtro 1* quando o valor do campo $STATUS > 0$;
- Um pacote é considerado malicioso pelo *Filtro 2* quando o valor do campo $REGRA > 0$.
- Tráfego malicioso detectado no *Filtro 1* não chega ao *Filtro 2*

- O tráfego de rede é capturado em amostras de 500000 pacotes: $A_l = \{a_1, \dots, a_r\}$;
- Pesos são atribuídos aos componentes do *ranking*: P_m , onde m é o componente analisado;
- Todas as amostras passam pelo processo de detecção.

Tabela 14 - Descrição da classificação dos campos detectados pelos filtros 1 e 2

Campo	Conteúdo	Classificação
ACAO	ACCEPT	0
ACAO	DROP	1
ACAO	DETECT	2
COOKIE	NÃO	0
COOKIE	SIM	1
DOMINIO	COM	0
DOMINIO	NET	1
DOMINIO	ORG	2
DOMINIO	GOV	3
DOMINIO	EDU	4
DOMINIO	INDEFINIDO	5
METODO	GET	0
METODO	HEAD	1
METODO	POST	2
METODO	PUT	3
METODO	OPTI	4
METODO	PROP	5
METODO	COPY	6
METODO	MKCO	7
METODO	TRAC	8
SERVICO		0
SERVICO	HTTP	1
SERVICO	HTTPS	2
SERVICO	FTP	3
SERVICO	DOMAI	4
SERVICO	IKE	5
SERVICO	NBNAM	7
SERVICO	33638/447	8
SERVICO	SMTP	9
SERVICO	SSH	10
SERVICO	TELNET	11
USERAGENT	NÃO	0
USERAGENT	SIM	1
PAIS	BR	0
PAIS	US	1
PAIS	INDEFINIDO	2
PAIS	OUTROS	3
BYTES	MÉDIA DE <i>BYTES</i> POR CONEXÃO	
TEMPO	MÉDIA DE TEMPO EM SEGUNDO POR CONEXÃO	

Fonte: o autor.

A tabela 14 mostra a estruturação dos dados capturados pelos filtros. A primeira coluna indica o nome dos campos, a segunda coluna mostra o conteúdo possível destes campos e a terceira coluna mostra um número de classificação utilizado para codificar cada valor. Por exemplo, no tráfego capturado pelo *Filtro 1* consta como informação o uso de *cookies*. Caso a informação seja verdadeira, o campo *COOKIE* é preenchido com a classificação $1=SIM$, caso seja falsa o campo é preenchido com o valor $0=NÃO$.

O processo de detecção é dividido em três análises: {1} análise das configurações; {2} análise dos IP; e {3} análise regressiva.

5.2.1.1.1 Análise das Configurações

A análise das configurações é uma quintupla $Q_h = \{F_i, C_j, S_k, A_l, P_m\}$, onde as configurações são agrupadas de acordo com os campos indicados na tabela 14: *METODO*, *URI*, *URL*, *USERAGENT*, *BROWSER*, *COOKIE*, *DOMINIO*, *PAIS*, *BYTES*, *TEMPO* e *SERVICO*. Para cada configuração algumas variáveis são calculadas e servem para o processo de criação do *ranking*.

A variável *QSTATUS* indica o total de pacotes de uma mesma configuração que passam pelo *Filtro 1*.

$$QSTATUS = F_1(\sum_{j=1}^n C_j (\sum_{k=1}^2 S_k (\sum_{l=1}^r A_l))) \quad (25)$$

O total de pacotes que passam pelo *Filtro 2* é armazenado na variável *QREGRA*.

$$QREGRA = F_2(\sum_{j=1}^n C_j (\sum_{k=1}^2 S_k (\sum_{l=1}^r A_l))) \quad (26)$$

Na variável *SSTATUS* é armazenado o total de ataques detectados pelo *Filtro 1* por configurações.

$$SSTATUS = \{F_1(\sum_{j=1}^n C_j (\sum_{k=1}^2 S_k (\sum_{l=1}^r A_l))), se STATUS > 0 \quad (27)$$

O conteúdo da variável *SREGRA* representa o total de ataques detectados pelo *Filtro 2*.

$$SREGRA = \{F_2(\sum_{j=1}^n C_j (\sum_{k=1}^2 S_k (\sum_{l=1}^r A_l))), se REGRA > 0 \quad (28)$$

$MSTATUS$ representa o percentual de ataques detectados pelo *Filtro 1* na amostra.

$$MSTATUS = (SSTATUS * 100) / QSTATUS \quad (29)$$

O percentual de ataques detectados pelo *Filtro 2* é representado pela variável $MREGRA$.

$$MREGRA = (SREGRA * 100) / QREGRA \quad (30)$$

Supõe-se que o percentual de ataques seja um preditor de importância, pois indica o nível de suspeição de uma determinada configuração c_j analisada pelo *Filtro 2*. Sendo assim, o peso atribuído para $MREGRA$ é 100%.

$$P_{MREGRA} = 1$$

A variável $TACUM$ representa a contagem acumulada das configurações da primeira amostra a_1 até a amostra atual a_i .

$$TACUM = \sum_1^i QREGRA \quad (31)$$

A variável $RREGRA$ indica o percentual de quanto a variável $QREGRA$ representa dentro do total acumulado das configurações C da primeira amostra a_1 até a amostra atual a_i . O valor de $RREGRA$ indica a relevância quantitativa de uma determinada configuração c_j no *Filtro 2* em relação ao total de configurações analisadas até o momento. O valor de $RREGRA$ não é contabilizado quando não há ataques registrados no *Filtro 2*.

$$RREGRA = \left\{ \frac{QREGRA * 100}{TACUM}, se \ MREGRA > 0 \right. \quad (32)$$

Como um exemplo, supondo-se que $TACUM=20.000.000$ e há duas configurações $C=\{c_1, c_2\}$ com os seguintes totais: $QREGRA_1=1$, $SREGRA_1=1$, $MREGRA_1=100\%$, $QREGRA_2=181.579$, $SREGRA_2=181.579$ e $MREGRA_2=100\%$. Como ambas as configurações (c_1 e c_2) tem $MREGRA=100\%$, a comparação entre $RREGRA_1= 0,000005$ e $RREGRA_2= 0,907895$ indica que c_2 é a configuração a mais relevante em termos quantitativos – apareceu 181.579 vezes em 20.000.000, enquanto c_1 apareceu apenas uma vez dentro da mesma quantidade acumulada.

O peso atribuído à variável $RREGRA$ é 100% quando há ataques e 0% quando não há ataques detectados pelo *Filtro 2* – supõe-se que quando há ataques a

quantidade de ocorrências deve ser levada em conta já que a configuração c_j tem um histórico suspeito. De forma inversa, quando não há ataques não há razão para se contabilizar o valor de $RREGRA$ para a configuração c_j .

$$P_{RREGRA} = \begin{cases} 1, & \text{se } MREGRA > 0 \\ 0, & \text{se } MREGRA \leq 0 \end{cases}$$

Enquanto $RREGRA$ é um indicativo acumulado, a variável $R500F2$ tem um caráter restrito à amostra atual a_i . A variável $R500F2$ indica o percentual de quanto $QREGRA$ representa dentro do total de configurações C que passaram pelo *Filtro 2* na amostra atual $a_i=500000$. O valor de $R500F2$ indica a relevância quantitativa de uma determinada configuração c_j em relação à amostra atual a_i . O valor de $R500F2$ não é contabilizado quando não há ataques registrados no *Filtro 2*.

$$R500F2 = \left\{ \frac{QREGRA \cdot 100}{500000}, \text{se } MREGRA > 0 \right. \quad (33)$$

O peso atribuído à variável $R500F2$ é 100% quando há ataques no *Filtro 2* pelos mesmos motivos alegados na variável $RREGRA$.

$$P_{R500F2} = \begin{cases} 1, & \text{se } MREGRA > 0 \\ 0, & \text{se } MREGRA \leq 0 \end{cases}$$

Por exemplo, supondo-se que dentro de uma amostra a_i com 500000 pacotes, há duas configurações $C=\{c_1, c_2\}$ com os seguintes totais: $QREGRA_1=1$, $SREGRA_1=1$, $MREGRA_1=100\%$, $QREGRA_2=181.579$, $SREGRA_2=181.579$ e $MREGRA_2=100\%$. Como ambas as configurações tem $MREGRA=100\%$, a comparação entre $R500F2_1=0,0002$ e $R500F2_2=36,3158$ indica que c_2 é a configuração mais relevante em termos quantitativos recentes – apareceu 181.579 vezes em 500000, enquanto c_1 apareceu apenas uma vez na amostra a_i .

5.2.1.1.2 Análise dos IP

A análise do par IP_j também é uma quintupla $Q_h = \{F_i, IP_j, S_k, A_l, P_m\}$. Para cada par IP_j são agrupados o IP de origem e as configurações indicadas na tabela 14: *METODO*, *URI*, *URL*, *USERAGENT*, *BROWSER*, *COOKIE*, *DOMINIO*, *PAIS*, *BYTES*, *TEMPO* e *SERVICO*. Para o processo de criação do *ranking* as seguintes variáveis são calculadas:

O total de pacotes de cada par IP_j que passa pelo *Filtro 1* na amostra a_i é indicado pela variável $QIP1$. Da mesma forma, $QIP2$ indica a quantidade de pacotes que passa pelo *Filtro 2*.

$$QIP1 = F_1(\sum_{j=1}^n IP_j (\sum_{k=1}^2 S_k (\sum_{l=1}^r A_l))) \quad (34)$$

$$QIP2 = F_2(\sum_{j=1}^n IP_j (\sum_{k=1}^2 S_k (\sum_{l=1}^r A_l))) \quad (35)$$

As variáveis $EIP1$ e $EIP2$ indicam o total de ataques do par IP_j detectados na amostra a_i respectivamente pelos *Filtros 1 e 2*.

$$EIP1 = \{F_1(\sum_{j=1}^n IP_j (\sum_{k=1}^2 S_k (\sum_{l=1}^r A_l))), se STATUS > 0 \quad (36)$$

$$EIP2 = \{F_2(\sum_{j=1}^n IP_j (\sum_{k=1}^2 S_k (\sum_{l=1}^r A_l))), se REGRA > 0 \quad (37)$$

Os percentuais de ataque do par IP_j detectados na amostra a_i pelos *Filtros 1 e 2* são representados respectivamente pelas variáveis $PIP1$ e $PIP2$, cujos valores são contabilizados apenas quando há tráfego malicioso.

$$PIP1 = \left\{ \frac{EIP1 \cdot 100}{QIP1}, se EIP1 > 0 \quad (38)$$

$$PIP2 = \left\{ \frac{EIP2 \cdot 100}{QIP2}, se EIP2 > 0 \quad (39)$$

O peso atribuído à variável $PIP1$ é 25% quando há ataques detectados pelo *Filtro 1* no par IP_j . A justificativa é que apesar do tráfego malicioso ter sido bloqueado pelo *Filtro 1* e não ser propagado para o *Filtro 2*, o IP_j é considerado suspeito.

Para atribuição de peso para a variável $PIP2$ alguns fatores adicionais devem ser levados em conta. A variável lógica $ATTACKIPS$ é verdadeira (valor 1) quando pelo menos um ataque é atribuído ao IP analisado - leva em conta apenas o IP (não se baseia no par IP_j) e contabiliza desde a amostra inicial a_1 até a amostra atual a_i . A variável $ATTACKIPS$ é falsa (valor 0) quando não há tráfego malicioso atribuído ao IP analisado desde a amostra inicial a_1 até a amostra atual a_i . O peso atribuído à variável $PIP2$ é 100% quando o par IP_j tem tráfego malicioso detectado pelo *Filtro 2*. A justificativa é que a detecção de ataques no par IP_j tem bastante importância na determinação de um *ranking* preditivo. Quando o par IP_j não tem ocorrência de ataques, mas o IP analisado possui histórico de ataques em outras configurações

($ATTACKIPS=1$), o peso atribuído à variável $PIP2$ é 50% - considera-se que o histórico de ataques positivo em outras configurações torna o IP suspeito. Quando o IP analisado não possui ocorrência de ataques em outras configurações, o peso atribuído a $PIP2$ é 0 (neste caso, pouco importa o peso pois $PIP2=0$).

$$P_{PIP1} = \begin{cases} 0.25, & \text{se } REGRA > 0 \\ 0, & \text{se } REGRA \leq 0 \end{cases}$$

$$P_{PIP2} = \begin{cases} 1, & \text{se } REGRA > 0 \\ \text{se } REGRA \leq 0 \begin{cases} 0.5, & \text{se } ATTACKIPS = 1 \\ 0, & \text{se } ATTACKIPS = 0 \end{cases} \end{cases}$$

A variável $TIPACUM$ representa a contagem acumulada do par IP_j desde a amostra inicial a_1 até a amostra atual a_i .

$$TIPACUM = \sum_1^i QREGRA \quad (40)$$

O percentual de quanto $QIP2$ representa dentro de $TIPACUM$ é armazenado na variável $RIP2$. O valor de $RIP2$ indica a relevância quantitativa de um determinado par IP_j em relação ao total de configurações C analisadas até o momento e não é contabilizado quando não há detecção de tráfego malicioso..

$$RIP2 = \left\{ \frac{QIP2 * 100}{TIPACUM}, \text{se } EIP2 > 0 \right. \quad (41)$$

Por exemplo, supondo-se que $TACUM=20.000.000$ e há duas configurações $C=\{c_1, c_2\}$ com os seguintes totais: $QIP2_1=1$, $EIP2_1=1$, $PIP2_1=100\%$, $QIP2_2=181.579$, $EIP2_2=181.579$ e $PIP2_2=100\%$. Como ambas as configurações tem $RIP2=100\%$, a comparação entre $RIP2_1=0,000005$ e $RIP2_2= 0,907895$ indica que a c_2 é a configuração mais relevante em termos quantitativos – apareceu 181.579 vezes em 20.000.000, enquanto a c_1 apareceu apenas uma vez dentro da mesma quantidade acumulada.

O peso atribuído à variável $RIP2$ é 100% quando há detecção de tráfego malicioso no par IP_j e 0% quando não há detecção pelo *Filtro 2*. A justificativa para esta atribuição de pesos é que o total de ocorrências maliciosas em um IP identificado como suspeito é um indicativo que deve ser levado em conta.

$$P_{RIP2} = \begin{cases} 1, & \text{se } EIP2 > 0 \\ 0, & \text{se } EIP2 \leq 0 \end{cases}$$

Enquanto $RIP2$ representa um indicativo acumulado, a variável $RIP500F2$ diz respeito ao percentual de quanto $QIP2$ representa dentro do total de configurações C detectadas pelo *Filtro 2*, presentes na amostra atual $a_i=500000$. O valor de $RIP500F2$ indica a relevância quantitativa de um determinado par IP_j em relação à amostra atual a_i . O valor de $RIP500F2$ não é contabilizado quando não há ataques.

$$RIP500F2 = \left\{ \frac{QIP2 \cdot 100}{500000}, \text{ se } EIP2 > 0 \right. \quad (42)$$

O peso atribuído à variável $RIP500F2$ é 100% quando há ataques no *Filtro 2* pelos mesmos motivos alegados na variável $RIP2$. Como justificativa, entende-se que o total de ocorrências de um par IP_j identificado como suspeito, dentro da amostra atual a_i é um indicativo que deve ser levado em conta.

$$P_{RIP500F2} = \begin{cases} 1, & \text{se } EIP2 > 0 \\ 0, & \text{se } EIP2 \leq 0 \end{cases}$$

Como exemplo, pode-se supor que em uma amostra a_i com 500000 pacotes, há dois pares IP_1 e IP_2 com os seguintes totais: $QIP2_1=1$, $EIP2_1=1$, $PIP2_1=100\%$, $QIP2_2=181.579$, $EIP2_2=181.579$ e $PIP2_2=100\%$. Como ambas os pares tem $RIP2=100\%$, a comparação entre $RIP500F2_1=0,0002$ e $RIP500F2_2=36,3158$ indica que o par IP_2 é mais relevante em termos quantitativos recentes – apareceu 181.579 vezes em 500000, enquanto o par IP_1 apareceu apenas uma vez na amostra a_i .

5.2.1.1.3 Análise de Relevância pelo Ajuste de um Modelo

Regressivo Linear

Um modelo linear indica a influência de variáveis independentes em uma variável resposta. De forma geral, quando uma variável influi linearmente em outra, o aumento da primeira tende a causar acréscimo ou decréscimo na segunda (DRAPER; SMITH, 1998). Como discutido anteriormente (vide Seção 5.1.2.7 Definição de Relevância), dadas duas variáveis x (independente) e y (resposta), um modelo linear pode ser descrito na forma: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + e$, onde e indica um erro aleatório e $\beta_0, \beta_1, \dots, \beta_k$ são parâmetros desconhecidos estimados com o Método dos Mínimos Quadrados observado em Draper e Smith (1998). Uma das exigências de um modelo linear é que os resíduos (diferença entre o modelo ajustado e o modelo real) tenham distribuição normal, variância constante e

esperança ($E(Y)=0$). Porém, **como se trata de um simples ajuste nos dados para a observação de relevância para uso no ranking e como não acontecem inferências sobre os dados resultantes do modelo linear, esses requisitos não são observados.**

Um modelo de regressão linear, construído no aplicativo estatístico R-free/R-studio (<https://www.r-project.org/>), é utilizado para verificar quais são os campos analisados pelo *Filtro 1* que mais influem na detecção de tráfego malicioso pelo *Filtro 2*. Os campos *METODO*, *URI*, *URL*, *USERAGENT*, *BROWSER*, *COOKIE*, *DOMINIO*, *PAIS*, *BYTES* e *TEMPO* são modelados como variáveis independentes em uma regressão linear cuja variável resposta é o percentual de ataques detectados no *Filtro 2* (*MREGRA*).

Inicialmente define-se a variável $MAX-P-VALUE=0.05$ como o limite máximo para os *p-values* (indica probabilidade máxima de erro para as estimativas em 5%). A partir do resultado da regressão linear, Os campos cujos coeficientes sejam positivos e cujos *p-values* sejam **menores** que $MAX-P-VALUE$ são chamados de $CAMPOS_{INFLUENTES}$. Esses campos são os que mais influenciam a ocorrência de um ataque no *Filtro 2*. O coeficiente positivo de maior valor é gravado na variável *QRPOSITIVO*. Este valor é usado na composição do *ranking*.

$$QRPOSITIVO = \max (coeficiente > 0 \text{ e } p - value < 0.05) \quad (43)$$

O peso atribuído à variável *QRPOSITIVO* pode ter três valores:

- 0% quando uma configuração não possui valores coincidentes com os $CAMPOS_{INFLUENTES}$.

$$P_{QRPOSITIVO} = \{0, \text{ se } \left\{ \begin{array}{l} CAMPOS_{INFLUENTES} \not\subset c_j \\ \text{ou} \\ (\sum CAMPOS_{INFLUENTES}) = 0 \end{array} \right. \}$$

- 50% quando uma configuração possui valores coincidentes com **todos** os $CAMPOS_{INFLUENTES}$ e desde que a quantidade de $CAMPOS_{INFLUENTES}$ seja maior que 1. A justificativa é que uma configuração que possui valores coincidentes com todos os $CAMPOS_{INFLUENTES}$ é considerada suspeita. O valor de 50% indica que a influência pode estar dividida entre vários

$CAMPOS_{INFLUENTES}$ e o ideal é que o peso da variável $QRPOSITIVO$ seja diluído.

$$P_{QRPOSITIVO} = \{0.5, se \begin{cases} (\sum CAMPOS_{INFLUENTES} > 1) \\ e \\ all(CAMPOS_{INFLUENTES} \subset c_j) \end{cases};$$

- 10% quando uma configuração possui um ou mais (mas não todos) valores coincidentes com os $CAMPOS_{INFLUENTES}$. Supõe-se que a ocorrência de valores coincidentes em poucos campos da configuração não seja um indicador forte de suspeição.

$$P_{QRPOSITIVO} = \{0.1, se \begin{cases} (\sum CAMPOS_{INFLUENTES} \geq 1) \\ e \\ at\ least\ one(CAMPOS_{INFLUENTES} \subset c_j); \\ e \\ at\ least\ one(CAMPOS_{INFLUENTES} \not\subset c_j) \end{cases};$$

5.2.1.2 Criação do *Ranking*

Como citado anteriormente, a criação do *ranking* envolve a leitura de amostras $A_l = \{a_1, \dots, a_r\}$ compostas por 500000 registros²⁷ ($R_t = \{r_1, \dots, r_{500000}\}$) que passam pelo processo de detecção.

Uma amostra a_l , que já passou pelo processo de detecção, serve de base para o cálculo de parte do *ranking* e é chamada de *ANTERIOR*.

O *ranking* é a soma das variáveis coletadas em cada registro multiplicadas pelos respectivos pesos:

$$\begin{aligned} \text{Ranking} = & (MREGRA * P_{MREGRA}) + (RREGRA * P_{RREGRA}) + (QRPOSITIVO * \\ & P_{QRPOSITIVO}) + (PIP1 * P_{PIP1}) + (PIP2 * P_{PIP2}) + (RIP2 * P_{RIP2}) + (R500F2 * P_{R500F2}) + \\ & (RIP500F2 * P_{RIP500F2}) \end{aligned} \quad (44)$$

O *ranking* se aplica aos registros individuais - dessa forma, cada **registro lido** possui um *ranking* de suspeição. **Quanto mais alto o *ranking*, mais suspeito é um registro r_t .** O valor do *ranking* atribuído a um **registro r_t lido** deve ser armazenado e

²⁷ Pacotes coletados a partir do tráfego de rede têm como característica a leitura e análise dos campos do próprio cabeçalho. Como os dados analisados pelo *ranking* são mais abrangentes que a leitura dos campos do cabeçalho TCP/IP – um registro é composto pelo par IP + configuração e pelos relacionamentos citados no processo de detecção. Sendo assim, para um melhor entendimento, a partir desse momento, a palavra registro é usada no lugar da palavra pacote.

utilizado para julgar se um **novo registro** r_{t+1} deve ou não ser barrado pelo *Filtro 1* – antes de chegar ao *Filtro 2*. Sendo assim, apesar de ser calculado sobre a amostra $a_t=ANTERIOR$, a aplicação do *ranking* acontece na próxima amostra a_{t+1} , chamada de *POSTERIOR*.

A cada vez que um registro r_t é lido, o *ranking* passa por uma pequena atualização, em tempo real, da relevância quantitativa da amostra $a_t=POSTERIOR$:

$$Ranking = ranking + (R500F2 * PR500F2) + (RIP500F2 * PRIP500F2) \quad (45)$$

5.2.1.3 Previsão

Como dito anteriormente, a previsão é realizada sobre uma amostra $a_t=POSTERIOR$. Cada novo registro r_t que passa pelo *Filtro 1* deve ser bloqueado ou aceito a partir de um limite máximo para o *ranking* de suspeição, definido na leitura da amostra $a_t=ANTERIOR$. Um resumo do processo de previsão pode ser visto no detalhamento a seguir:

1. Uma amostra inicial $a_1=ANTERIOR$ é lida e o *ranking* de cada registro é armazenado. Essa amostra inicial $a_1=ANTERIOR$ é a única que não passa pelo processo de previsão;
2. Um valor de limite é definido a partir da aplicação do SGA nos resultados da amostra $a_1=ANTERIOR$ (vide Seção 5.2.1.4 Agrupamento);
3. Uma nova amostra $a_t=POSTERIOR$ começa a ser lida;
4. Verifica-se se o registro r_t da amostra $a_t=POSTERIOR$ já possui um *ranking*:
 - a. Se $ranking(r_t) \geq limite$, o registro r_t é bloqueado antes de chegar ao *Filtro 2*;
 - b. Se $ranking(r_t) < limite$ ou caso o registro r_t ainda não possua um registro cadastrado, r_t é liberado para o *Filtro 2*;
5. O valor do *ranking* do registro r_t é atualizado a cada nova leitura de registro realizada na amostra $a_t=POSTERIOR$;
6. Ao fim da leitura, a amostra $a_t=POSTERIOR$ passa a ser considerada $a_t=ANTERIOR$ e repetem-se os passos a partir do passo 2.

A melhor forma de medir a precisão das previsões é justamente comparar os dados previstos com os dados reais. A fim de contabilizar o desempenho preditivo

do *ranking*, o total de ataques confirmados pelo *Filtro 2* é **contabilizado para cada valor inteiro do *ranking*** de duas formas:

- O percentual de registros marcados como suspeitos no *ranking* e que foram pegos no *Filtro 2* como ataques é contabilizado como VP²⁸: $VP = \frac{PA(a_i) * 100}{QREGRA(a_i)}$, onde:

- $PA(a_i)$ =quantidade de registros previstos $\geq$ limite;
- $QREGRA(a_i)$ =quantidade de registros que passaram pelo *Filtro 2* na amostra a_i ;

- O percentual **total** de registros não marcados como suspeitos no *ranking* e que foram pegos no *Filtro 2* como ataques é contabilizado como VN:

$$VN = \frac{[(SREGRA(a_i) - PA(a_i)) * 100]}{QREGRA(a_i)}.$$

Onde:

- $SREGRA(a_i)$ =total de ataques detectados pelo *Filtro 2*.

Neste ponto cabe uma observação interessante sobre a topologia de rede deste Segundo Experimento composta por dois controles de segurança sequenciais: o processo que prevê se um registro r_t é malicioso ou não depende da leitura do mesmo, o que corresponde a um processo de detecção. Porém, como a previsão é relacionada ao *Filtro 2*, o *ranking* funciona como um preditor acoplado ao *Filtro 1*. Sendo assim, a partir da leitura da amostra a_1 é possível montar um histórico dos registros r_t suspeitos para comparação com novas leituras. Um dos problemas relacionados a esse comportamento é que eventuais registros maliciosos que ainda não estão ranqueados podem passar pelo *Filtro 1* e chegar ao *Filtro 2*. Para sobrepor este tipo de imprecisão, o percentual de VN formado por registros novos também é contabilizado - na variável Verdadeiro Negativo Novo (VNN). Como é acoplado ao VN, o VNN serve como referência sobre a precisão do método de previsão, pois, indica qual percentual do VN é relativo à registros novos (não contabilizados). Por exemplo, se o $VNN=100$, isso significa que 100% do valor de VN é relativo a registros novos, que por não serem ranqueados não passam por previsão na primeira leitura.

²⁸ O percentual de registros marcados como suspeitos no *ranking* e que não foram pegos no *Filtro 2* é denotado como FP. Como o índice FP é exatamente o inverso de VP, o índice FP não é contabilizado na criação da UCB.

Toda essa contabilização é realizada em uma tabela onde a quantidade de registros é agrupada a partir dos valores unitários do *ranking* (valor máximo do *ranking*=200). Em tese, apesar do *ranking* de um determinado registro r_t poder alcançar valores superiores a 200, na prática esses valores dificilmente ultrapassam a marca de 120. Sendo assim, a tabela 15 mostra um exemplo de como é estruturada a contabilização dos percentuais de VP, VN e VNN para uma determinada amostra:

Tabela 15 - Duas linhas da tabela com contabilização dos registros da amostra A003 com $S_t=400$.

RANKING	QREGRA(a_i)	SREGRA(a_i)	PA(a_i)	VP	VN	VNN	BYTES	TEMPO
31	498318	721	712	99,30	1,24	100	13353	59
30	498318	721	717	98,48	0,55	100	14868	69

Fonte: o autor.

Na tabela 15 as colunas QREGRA(a_i), SREGRA(a_i) e PA(a_i) indicam respectivamente a quantidade de registros que passam pelo *Filtro 2*, a quantidade de ataques detectados pelo *Filtro 2* e a quantidade de ataques prevista. Todos os valores dizem respeito à quantidade de registros que possuem valor de *ranking* **maior ou igual** ao indicado na coluna RANKING. As duas últimas colunas da tabela 15 – BYTES e TEMPO – representam respectivamente as médias de *bytes* e de tempos (s) acumuladas por *ranking*. Essas duas variáveis são usadas no cálculo de UCB. Os valores de VP e VN variam dependendo do valor do *ranking*.

É importante notar que os valores de VP e VN contabilizados nada mais são que comparações entre as previsões e as detecções executadas pelo *Filtro 2*. Um percentual de VP próximo de 100% significa que a previsão está muito próxima da detecção real, indicando alta precisão do processo preditivo. Da mesma forma, percentuais de VN próximos de 0% indicam que o processo preditivo é capaz de impedir quase que a totalidade da incidência de ataques no *Filtro 2*. Contudo, um valor ideal de uso para o *ranking* deve ser escolhido com cautela. *Rankings* altos são restritivos (bloqueiam muitas conexões) e apresentam altos índices de VN. Porém, a vantagem dos *rankings* altos são os bons percentuais de VP. Por outro lado, devido à pouca restrição no bloqueio de conexões, *rankings* baixos

apresentam como vantagem baixos percentuais de VN e como desvantagens baixos percentuais de VP.

A fim de determinar qual o limite dentre os valores de *ranking* que maximiza os benefícios – baixos percentuais de VN em conjunto com altos percentuais de VP, baixa quantidade de BYTES e TEMPO em cada amostra a_i – durante o processo de agrupamento a arquitetura SGA é aplicada aos resultados.

5.2.1.4Agrupamento

Para aplicação do SGA os módulos do nível técnico e do nível estratégico são configurados de acordo com os detalhes descritos nas tabelas 16 e 17.

Tabela 16 - Nível técnico: configurações do Segundo Experimento.

Módulos	Descrição
1.1 Seleciona controles de segurança	Detecção: o processo de detecção Previsão: o processo de previsão
1.2 Define variáveis independentes	Percentual do <i>ranking</i> e nível de segurança S_k
1.3 Define variáveis resposta	VP VN BYTES: pacotes menores consomem menos recursos TEMPO: pacotes que ocupam menos tempo consomem menos recursos
1.4 Define o tamanho das amostras	500000 registros/amostra: de um total de 78 (a primeira é descartada), sobram 77 amostras
1.5 Obtêm os resultados e configurações dos controles de segurança	<i>Matriz de dados</i> : obtida dos processos de detecção, criação do <i>ranking</i> e previsão
1.6 Define limites para as variáveis resposta	<i>Matriz limite</i> : corte: $VP \leq 30\%$ (mínimo), $VN \geq 70\%$ (máximo)
1.7 Define relevância	<i>Matriz relevante</i> : manteve as variáveis independentes S_k e <i>ranking</i> a partir de um modelo de regressão linear

Fonte: o autor.

Tabela 17 - Nível estratégico: configurações do Segundo Experimento.

Módulos	Descrição	Detalhes
2.1 Variáveis independentes → jogadores	<i>Matriz dos jogadores:</i> variáveis independentes ligadas às respectivas variáveis resposta	Valor máximo adotado para o <i>ranking</i> : 200
2.3 Padroniza as variáveis resposta a partir da série limite	<i>Matriz Estratégica:</i> VP, VN, BYTES, TEMPO	BYTES e TEMPO funcionam como pequenas penalidades em relação aos percentuais de <i>ranking</i> alcançados. Sendo assim, ambos possuem um peso de $\frac{1}{10}$, aplicado após o processo de padronização
2.4 Defi7ne WTP e custo	WTP = VP Custo = VN, BYTES e TEMPO	
2.5 Cálculo e análise UCB	<i>Matriz UCB</i>	Uma <i>matriz UCB</i> é gerada para cada amostra Os cálculos UCB são repetidos para todos os registros das amostras

Fonte: o autor.

Neste Segundo Experimento a definição sobre quais são as variáveis independentes relevantes para as variáveis resposta VP, VN, BYTES e TEMPO acontece por meio da experiência do administrador de segurança. O percentual de *ranking* e o nível de segurança s_k do *Filtro 1* estão diretamente ligados aos percentuais de VP e VN, já era sabido, portanto, que essas variáveis independentes eram significativas. Porém, com o intuito de validar essa questão e verificar a relevância para as variáveis resposta BYTES e TEMPO, um modelo de regressão linear é gerado ao fim da leitura de cada amostra para que se observem quais são as variáveis independentes mais relevantes.

Como pode ser observado nas configurações descritas na tabela 16, o único WTP é o percentual de VP (valores altos são considerados ótimos), enquanto os percentuais de VN e os valores das variáveis BYTES e TEMPO são considerados como custo (valores baixos são considerados ótimos). Além disso, um limite é calculado para cada uma das 77 amostras. Sendo assim, para cada amostra é gerada uma *matriz UCB*.

5.2.2 Resultados do Segundo Experimento

Todos os resultados são gerados levando-se em conta as variáveis independentes que formam um jogador: percentual do *ranking* e nível de segurança S_k . Apesar das variáveis terem sido escolhidas porque já se sabia previamente da influência das mesmas, o modelo de regressão linear é ajustado para todas as amostras. A tabela 18 mostra a média dos *p-value* das 77 amostras (a primeira amostra a_1 não é utilizada). Não é usual usar a média dos *p-values*, porém, a ideia aqui é apenas mostrar se as variáveis realmente tem significância para as variáveis respostas avaliadas (VP, VN, BYTES e TEMPO). Na tabela 32 (APÊNDICE F) o resultado da regressão linear das 77 amostras pode ser visualizado.

Tabela 18 – Média dos *p-value* dos modelos lineares das 77 amostras.

Var. Indep.	VP (<i>p-value</i>)	VN (<i>p-value</i>)	BYTES (<i>p-value</i>)	TEMPO (<i>p-value</i>)
S_k	0,003754	0,261255	0,002567	0,022315
<i>Ranking</i>	1,5929e-27	5,2103e-70	0,006740	0,010479

Fonte: o autor.

Os baixos valores de *p-value* exibidos na tabela 18 indicam que todas as amostras apresentam relevância tanto no percentual de *ranking* como no nível de segurança S_k . Na média o nível de segurança sk se mostra um pouco menos significativo para a variável resposta VN, porém, sua significância esteve adequada para VP, BYTES e TEMPO. Portanto, ambas as variáveis são significativas e podem ser usadas na composição do jogador. Contudo, para um melhor entendimento, os próximos resultados desta seção são apresentados separados por nível de segurança. Por exemplo, as tabelas 19 e 20 mostram em detalhes a contabilização dos *rankings* das amostras a_{002} e a_{003} para os dois níveis de segurança $S_1=400$ e $s_2=201$.

Tabela 19 - Contabilização dos *rankings* das amostras a_{002} e a_{003} com nível de segurança $s_1=400$

Amostra a_{002} – nível de segurança $s_1=400$					
RANKING	VP	VN	VNN	BYTES	TEMPO
200-101	100,00	55,01	100,00	5505	21
100-25	100,00	0,14	100,00	6634	35
24-19	99,57	0,14	100,00	9017	62
18-14	98,72	0,14	100,00	8749	60
13	59,21	0,14	100,00	8499	58
12	37,85	0,14	100,00	8822	56
11	34,88	0,14	100,00	8877	57
10	27,64	0,14	100,00	8683	55
9	17,84	0,14	100,00	10833	66
8	17,04	0,14	100,00	12908	77
7	15,25	0,14	100,00	15333	94
6	14,89	0,14	100,00	16829	102
5	14,20	0,14	100,00	17717	106
4	11,32	0,14	100,00	17961	108
3	10,11	0,14	100,00	19188	124
2	7,95	0,14	100,00	22043	173
1	5,48	0,14	100,00	21879	169
0	1,99	0	99,85	26021	259

Amostra a_{003} – nível de segurança $s_1=400$					
RANKING	VP	VN	VNN	BYTES	TEMPO
200-107	100,00	68,37	100,00	9662	55
106-105	100,00	66,85	100,00	8291	46
104-102	100,00	58,66	100,00	7312	40
101	100,00	53,39	100,00	7956	38
100-50	99,44	1,24	100,00	10110	46
49-31	99,30	1,24	100,00	13353	59
30-25	98,48	0,55	100,00	14868	69
24-16	97,81	0,55	100,00	15027	67
15-13	96,89	0,55	100,00	14024	64
12-11	94,21	0,55	100,00	15571	86
10-9	93,35	0,55	100,00	15284	84
8	56,45	0,55	100,00	14705	83
7	35,95	0,55	100,00	15989	97
6	15,48	0,55	100,00	17488	103
5	13,11	0,55	100,00	17621	102
4	10,07	0,55	100,00	20745	135
3	8,81	0,55	100,00	22357	179
2	5,96	0,55	100,00	23351	192
1	4,77	0,41	100,00	23726	198
0	1,38	0,00	100,00	27053	275

Fonte: o autor.

Tabela 20 - Contabilização dos *rankings* das amostras a_{002} e a_{003} com nível de segurança $s_2=201$

Amostra a_{002} – nível de segurança $s_2=201$					
RANKING	VP	VN	VNN	BYTES	TEMPO
200-104	100,00	55,09	100,00	5505	21
103-102	100,00	54,81	100,00	7007	27
101	100,00	54,66	100,00	9949	55
100-29	100,00	0,14	100,00	6634	35
28-25	73,34	0,14	100,00	7625	40
24-23	43,10	0,14	100,00	30161	303
22	43,02	0,14	100,00	30238	303
21	38,84	0,14	100,00	30878	319
20	33,90	0,14	100,00	31545	343
19	26,90	0,14	100,00	32624	398
:	:	:	:	:	:
4	1,91	0,14	100,00	28514	307
0	1,38	0,00	99,86	28189	301

Amostra a_{003} – nível de segurança $s_2=201$					
RANKING	VP	VN	VNN	BYTES	TEMPO
200-113	100,00	75,25	100,00	9662	55
112-107	100,00	66,78	100,00	16759	73
106-105	100,00	64,41	100,00	15757	66
104	100,00	54,41	100,00	14734	61
103	100,00	54,07	100,00	13839	57
102	100,00	54,07	100,00	13839	57
101	100,00	47,97	100,00	13064	54
100	99,32	1,36	100,00	9575	44
99-32	99,32	1,36	100,00	9575	44
31	99,32	1,19	100,00	9299	43
30-25	98,33	0,51	100,00	11824	57
24	51,63	0,51	100,00	28065	264
23	51,54	0,51	100,00	28170	264
22	50,09	0,51	100,00	29197	332
21	48,88	0,51	100,00	30104	380
20	42,75	0,51	100,00	30991	429
19	28,13	0,51	100,00	32723	519
:	:	:	:	:	:
4	1,46	0,00	100,00	29205	335
0	0,81	0,00	100,00	28507	300

Fonte: o autor.

Nas tabelas 19 e 20 os *rankings* são agrupados por faixa na coluna RANKING. Por exemplo, é possível notar que na tabela 19, para a amostra a_{002} , dentro da faixa de *rankings* compreendida entre 200-101, os valores de VP, VN, VNN, BYTES e TEMPO são exatamente os mesmos. As linhas riscadas indicam que os limites ($VP \leq 30\%$ e $VN \geq 70\%$) foram ultrapassados. Também é possível perceber que apesar das amostras representadas serem sequenciais, há diferenças nos percentuais de VP, VN e nos valores de BYTES e TEMPO de acordo com as variações de *ranking*. São justamente essas diferenças entre as amostras que justificam o uso do SGA para definição dos valores mais ajustados de limite que devem ser usados na previsão e bloqueio dos registros suspeitos da próxima amostra. Outra característica interessante, que indica o alto nível de precisão do método de previsão, são os percentuais de VNN que ficaram entre 99,85% e 100% nas duas amostras, indicando que praticamente todas as ocorrências de VN acontecem devido a novos registros ainda não ranqueados. De forma geral, os resultados de VNN são muito parecidos para as outras amostras, independentemente do nível de segurança analisado.

Na tabela 21 há um exemplo da aplicação do processo de padronização para os valores vistos anteriormente na tabela 19:

Tabela 21 - Valores de WTP e custo após o processo de padronização - Amostras a_{002} e a_{003} com nível de segurança $s_1=400$

Amostra a_{002} – nível de segurança $s_1=400$		
RANKING	WTP=VP	Custo=VN+BYTES+TEMPO
200-101	1,1582	1,181363
100-25	1,1582	1,108679
24-19	1,1575	1,108707
18-14	1,1562	1,108705
13	1,0928	1,108705
12	1,0585	1,108701
11	1,0538	1,108702
10	1,0422	1,108700
9	1,0264	1,108713
8	1,0252	1,108725
7	1,0223	1,108744
6	1,0217	1,108752
5	1,0206	1,108757
4	1,0160	1,108759
3	1,0140	1,108777
2	1,0106	1,108829
1	1,0066	1,108825
0	1,0010	1,108822

Amostra a_{003} – nível de segurança $s_1=400$		
RANKING	WTP=VP	Custo=VN+BYTES+TEMPO
200-107	1,1788	1,191877
106-105	1,1788	1,190670
104-102	1,1788	1,184065
101	1,1788	1,179764
100-50	1,1788	1,137370
49-31	1,1775	1,137381
30-25	1,1761	1,136790
24-16	1,1748	1,136789
15-13	1,1732	1,136786
12-11	1,1684	1,136804
10-9	1,1668	1,136802
8	1,1003	1,136801
7	1,0633	1,136813
6	1,0264	1,136818
5	1,0222	1,136818
4	1,0167	1,136845
3	1,0144	1,136880
2	1,0093	1,136890
1	1,0071	1,136795
0	1,0010	1,136458

Fonte: o autor.

Na tabela 21 as colunas de WTP representam o valor de VP depois do processo de padronização, enquanto as colunas CUSTO representam a soma dos valores padronizados de VN, BYTES e TEMPO. Como estão todos na mesma ordem (valores menores são desejáveis) e já passaram pelo processo de padronização, a soma dos custos facilita os cálculos de UCB.

Ordenado em ordem decrescente de CM, a tabela 22 agrupa as faixas do *ranking* que possuem um mesmo valor de CM na *matriz UCB*. Os dados dizem respeito à amostra a_{002} com nível de segurança $s_1=400$ (os mesmos dados das tabelas 19 e 21):

Tabela 22 - Matriz UCB em ordem decrescente de CM, calculados para a amostra a_{002} com nível de segurança $s_1=400$

Amostra a_{002} – nível de segurança $s_1=400$					
RANKING	WTP	Custo	V(N)	V(N\{i})	CM
100-25	1.15820000	1.10867900	0.04952100	-0.10886800	0.15838900
24-19	1.15750000	1.10870700	0.04879300	-0.10886800	0.15766100
18-14	1.15620000	1.10870500	0.04749500	-0.10886800	0.15636300
13	1.09280000	1.10870300	-0.01590300	-0.10886800	0.09296500
200-101	1.15820000	1.18136300	-0.02316300	-0.10886800	0.08570500
12	1.05850000	1.10870100	-0.05020100	-0.10886800	0.05866700
11	1.05380000	1.10870200	-0.05490200	-0.10886800	0.05396600
10	1.04220000	1.10870000	-0.06650000	-0.10886800	0.04236800
9	1.02640000	1.10871300	-0.08231300	-0.10886800	0.02655500
8	1.02520000	1.10872500	-0.08352500	-0.10886800	0.02534300
7	1.02230000	1.10874400	-0.08644400	-0.10886800	0.02242400
6	1.02170000	1.10875200	-0.08705200	-0.10886800	0.02181600
5	1.02060000	1.10875700	-0.08815700	-0.10886800	0.02071100
4	1.01600000	1.10875900	-0.09275900	-0.10886800	0.01610900
3	1.01400000	1.10877700	-0.09477700	-0.10886800	0.01409100
2	1.01060000	1.10882900	-0.09822900	-0.10886800	0.01063900
1	1.00660000	1.10882500	-0.10222500	-0.10886800	0.00664300
0	1.00100000	1.10882200	-0.10782200	-0.10886800	0.00104600

Fonte: o autor.

Na *matriz UCB* retratada na tabela 22 as colunas RANKING, WTP e Custo indicam respectivamente uma faixa com *rankings* que possuem os mesmos valores padronizados. As outras colunas seguem as definições dadas na Seção 2.2.5 Unrestricted Cooperative Bargains (UCB): $V(N)$ indica o TVC dos jogadores, $V(N \setminus \{i\})$ corresponde ao valor criado no jogo sem a presença do jogador i e $CM = v(N) - v(N \setminus \{i\})$ é o valor agregado ou Contribuição Marginal - **os valores mais altos de CM são desejáveis**. A tabela 22 mostra apenas as colunas necessárias para o entendimento do processo de UCB; colunas com os valores mínimos garantidos, faixas mínimas, faixas máximas e barganha residual não são apresentadas. Na tabela 22 os cálculos de UCB mostram que os *rankings* compreendidos dentro da faixa **ideal** de 100-25 possuem os mais altos valores de CM para a amostra a_{002} (marcado em negrito). O valor máximo da faixa **ideal + 1** é usado para definição do limite da próxima amostra. No exemplo, se a amostra $a_{002} = \text{ANTERIOR}$ possui uma faixa ideal de 100-25, o limite adotado para a próxima amostra é 101, indicando que registros com valores **iguais ou superiores** a este valor serão bloqueados durante a leitura da amostra $a_{003} = \text{POSTERIOR}$.

Para verificar se a aplicação do SGA resulta em limites apropriados para prever a próxima amostra, a tabela 23 traz as faixas ideais e a contabilização dos valores de VP, VN, VNN, BYTES e TEMPO das primeiras amostras com nível de segurança $s_1 = 400$. Uma versão mais completa da tabela 23, com as 77 amostras para os dois níveis de segurança S_k pode ser vista na tabela 23, no APÊNDICE G.

Tabela 23 - Faixas ideais para algumas amostras do nível de segurança $s_1=400$.

AMOSTRA	NÍVEL DE SEGURANÇA: $S_1=400$							
	RANKING		REAL			BASEADO NO MIN		
	MIN	MAX	VP	VN	VNN	VP	VN	VNN
A002	25	100	100,00	0,14	100,00	-	-	-
A003	50	100	99,44	1,25	100,00	98,49	0,55	100,00
A004	25	50	100,00	2,12	100,00	100,00	2,12	100,00
A005	59	100	100,00	0,10	100,00	81,41	0,10	100,00
A015	45	103	100,00	1,66	100,00	100,00	1,66	100,00
A024	28	101	99,63	0,42	100,00	99,63	0,42	100,00
A032	40	102	100,00	0,03	100,00	98,97	0,03	100,00
A033	29	102	99,97	0,09	100,00	99,97	0,09	100,00
A044	49	103	99,37	0,00	100,00	99,37	0,00	100,00
A057	30	103	99,95	0,00	100,00	99,95	0,00	100,00
A069	38	104	100,00	0,00	100,00	100,00	0,00	100,00
A077	62	120	100,00	0,00	100,00	99,80	0,00	100,00
A078	46	119	100,00	0,00	100,00	100,00	0,00	100,00

Fonte: o autor.

Na tabela 23 a coluna RANKING mostra as faixas ideais, ou seja, valores mínimos (MIN) e máximos (MAX) calculadas pelo SGA para cada amostra. As faixas ideais contêm os *rankings* que obtiveram os maiores valores de CM durante a aplicação do SGA. Por exemplo, para a amostra a_{003} qualquer *ranking* com valor entre 50 e 100 é considerado ideal. A coluna REAL contém os resultados mais otimizados possíveis para os percentuais de VP, VN e VNN, obtidos justamente por meio do uso destes valores ideais de *ranking*. Porém, esses percentuais otimizados só podem ser aplicados se o SGA for aplicado depois que a amostra for lida. O problema deste procedimento é que a previsão não acontece em tempo real. A fim de solucionar essa dificuldade, o valor **mínimo de *ranking*** de uma amostra $a_j=ANTERIOR$ é utilizado como o *ranking* ideal de uma amostra $a_i=POSTERIOR$. Ou seja, os percentuais de VP, VN e VNN da amostra $a_i=POSTERIOR$ são calculados tendo como base valores de *ranking* da amostra $a_j=ANTERIOR$. Por exemplo, na tabela 23 a coluna BASEADO NO MIN, relativa à amostra a_{003} , mostra os percentuais de $VP=98,49\%$, $VN=0,55\%$ e $VNN=100\%$ calculados com base no *ranking*=25, obtido da amostra $a_{002}=ANTERIOR$. Os registros que contêm valores de *ranking* mais baixos²⁹ são menos suspeitos, por isso são adotados. Na média os

²⁹ Apesar de realmente serem menos suspeitos, os registros com *ranking* menor, quando usados na previsão das próximas amostras, representam os piores casos para os cálculos de VP, pois um valor de limite muito baixo obtido em uma amostra $a_i=ANTERIOR$ pode representar valores baixos de VP em uma amostra $a_j=POSTERIOR$.

percentuais de VNN estão bem altos (em torno de 99,99%) e servem para atestar a precisão da metodologia usada na criação do *ranking* e do processo de previsão.

Para uma visualização mais adequada, os gráficos das figuras 18 e 19 ilustram a comparação entre os resultados reais e os resultados baseados nos valores mínimos de limite para os percentuais de VP mostrados na tabela 23.

Figura 18 - Comparação entre VP real e VP baseado no valor mínimo de *ranking* ($S_1=400$).

Fonte: o autor.

Figura 19 - Comparação entre VP real e VP baseado no valor mínimo de *ranking* ($S_2=201$).

Fonte: o autor.

O gráfico ilustrado na figura 18 mostra que para o nível de segurança $s_1=400$, ao longo das amostras $A=\{a_1, \dots, a_{78}\}$ os valores reais de VP (linha cheia que corresponde às conexões barradas pelo *Filtro 2*) ficaram todos muito próximos ou iguais a 100% - fato que também pode ser observado na tabela 23. A previsão de VP tendo como base o uso do valor mínimo de *ranking* da amostra $a_i=ANTERIOR$ (linha tracejada), apesar de algumas variações e de um *outlier* de 81,41% na amostra a_{003} , coincidiu com os valores reais. Para o nível de segurança $s_2=201$ (figura 19) há quatro *outliers* bem mais acentuados nas amostras $A=\{a_{004}, a_{005}, a_{010}, a_{067}\}$. O nível de segurança $s_2=201$, justamente por ser mais restritivo, tende a ter uma variação maior na quantidade de registros maliciosos detectados a cada amostra. Essa tendência se apresenta como o principal motivo para os *outliers* encontrados no nível de segurança $s_2=201$.

Os gráficos das figuras 20 e 21 mostram a comparação entre os resultados reais (detectados pelo *Filtro 2*) e os resultados baseados nos valores mínimos de limite (previstos pelo *ranking*) para os percentuais de VN:

Figura 20 - Comparação entre VN real e VN baseado no valor mínimo ($S_1=400$)

Fonte: o autor.

Figura 21 - Comparação entre VP real e VP baseado no valor mínimo de ranking ($S_2=201$)

Fonte: o autor.

A coincidência entre as linhas dos percentuais de VN nos gráficos das figuras 20 e 21 mostram que em relação ao percentual de VN, o agrupamento pelo valor

mínimo de limite teve um excelente resultado, com a previsão espelhando exatamente os resultados reais.

Um fato que poderia ser encarado como defeito em outros ambientes, mas que funciona de forma bastante satisfatória em um ambiente voltado para a segurança de TI é a liberdade na escolha dos valores mais equilibrados proporcionada pelas faixas resultantes do processo de agrupamento de mecanismos de segurança realizado pelo SGA. De forma geral, a escolha do menor valor de *ranking* como limite representa o pior caso para a previsão dos percentuais de VP. Essa constatação pode ser comprovada pelos gráficos das Figuras 22 e 23 que mostram respectivamente as faixas ideais de *ranking* calculadas pelo UCB para os níveis de segurança $s_1=400$ e $s_2=201$.

Figura 22 - Valores mínimos e máximos de limite de cada amostra ($S_1=400$)

Fonte: o autor.

Figura 23 - Valores mínimos e máximos de limite de cada amostra ($S_2=201$)

Fonte: o autor.

Nos gráficos das figuras 22 e 23 as linhas verticais indicam as faixas ideais calculadas por UCB para cada amostra. A ponta de cada linha indica os valores máximos e mínimos sugeridos como limites para a próxima linha (amostra). Nas faixas mais altas (entre 70 e 100 para $S_1=400$ e entre 80 e 100 para $S_2=201$) é possível traçar linhas horizontais que cruzam quase todas as linhas verticais. Isso significa que valores mais altos de limite e que estejam dentro da faixa ideal podem ser usados para prever as próximas amostras com uma dose de precisão altíssima. Se esses valores fossem adotados como limites teriam níveis de previsão muito bons, superiores aos valores mínimos adotados nos gráficos das Figuras 18 e 19.

A tabela 24 mostra os percentuais finais reais e baseados nos valores mínimos ideais:

Tabela 24 - Percentuais finais reais e baseados nos valores mínimos ideais

Total de conexões que chegaram ao <i>Filtro 1</i> : 38.868.309					
Nível de segurança	Conexões maliciosas detectadas – <i>Filtro 2</i>	% conexões maliciosas <i>Filtro 2</i>	Previamente barradas pelo <i>ranking</i>	Eficácia do <i>ranking</i> mínimo	Média de VN
$S_1=400$	161977	0,41%	161916	99,96%	0,19%
$S_2=201$	141516	0,36%	141456	99,95%	0,23%

Fonte: o autor.

Na tabela 24 os valores previstos usando-se o *ranking* no valor mínimo ideal (pior caso) estão muito próximos dos valores reais obtidos pela detecção. Percentuais de VP próximos à 100% indicam a proximidade com os valores reais detectados pelo *Filtro 2*, enquanto percentuais de VN próximos à 0% indicam a capacidade do método (*ranking*) de não bloquear eventos legítimos antes de chegarem ao *Filtro 2*. Também é possível perceber que a carga de conexões maliciosas que incide sobre o *Filtro 2* diminuiu consideravelmente, mesmo com o uso do *ranking* no valor mínimo ideal (pior caso). Para o nível de segurança $S_1=400$, das 161977 conexões maliciosas detectadas pelo *Filtro 2*, 161916 conexões foram barradas previamente (99,96%). Da mesma forma, com o nível de segurança $S_2=201$ das 141516 conexões maliciosas, 141456 conexões foram barradas previamente (99,95%). Esses resultados demonstram a eficácia do uso do SGA no agrupamento dos métodos detecção e previsão em relação aos percentuais de VP e VN.

A contabilização dos percentuais de previsões FP, em comparação com a detecção efetuada pelo *Filtro 2*, fica em torno de 0,05% (inverso dos percentuais de VP). Porém, o grande ganho de todo este processo é a diminuição na carga do *Filtro 2*. A tabela 24 mostra que aconteceram 38.868.309 conexões no dia do Segundo Experimento. As conexões maliciosas detectadas pelo *Filtro 2* em um dia inteiro correspondem a 0,41% (161916 conexões maliciosas para nível de segurança $S_1=400$) e 0,36% (141456 conexões maliciosas para o nível de segurança $S_2=201$). Com a aplicação do *ranking* esse percentual praticamente zero, já que 99,95% das conexões maliciosas são barradas antes de chegarem ao *Filtro 2*. Em um único dia, entre 161.916 ($S_1=400$) e 141.456 ($S_2=201$) conexões maliciosas não chegam ao *Filtro 2*.

5.2.3 Discussões sobre o Segundo Experimento

O método baseado em *ranking* e pesos adotado neste Segundo Experimento obteve resultados bastante satisfatórios. O *ranking* serviu como um índice prévio entre o *Filtro 1* e 2, permitindo uma antecipação que diminuiu a carga do *Filtro 2* entre 0,41% e 0,36% (conexões maliciosas) com uma eficácia em torno de 99,95%, um percentual de FP em torno de 0,05% e um percentual de VN muito próximo de 0%. A métrica usada para avaliar o desempenho do método foi a contabilização dos

índices de VP e VN de acordo com dois níveis de restrição adotados pelo *Filtro 1* e em um segundo plano a quantidade de *bytes* e o tempo (s) usados por cada pacote.

A aplicação do SGA funcionou como um complemento do mecanismo de detecção originalmente usado, pois permitiu que os resultados dos mecanismos de detecção de ambos os filtros, mas principalmente do *Filtro 1*, fossem agrupados com os resultados obtidos pelo mecanismo de previsão, no caso o *ranking*. Na verdade, a aplicação do SGA como forma de equilibrar os resultados trouxe como resultado adicional a possibilidade do uso de faixas com diferentes *rankings* e mesmos valores de VP e VN. A grande vantagem do uso das faixas é que um administrador pode escolher os valores de *ranking* que mais se adequam ao seu uso ou ao nível de previsibilidade que ele pretende obter. A prova mais veemente disso é a observação dos valores previstos de VP e VN baseados no valor mínimo do *ranking* (pior caso) descritos na tabela 24.

Entre as limitações do trabalho que podem ser citadas está o fato do peso (10%) usado pelas variáveis resposta BYTES e TEMPO ter sido baseado na experiência adquirida na administração da rede usada no experimento. O peso utilizado teve por definição um baixo valor (10%) porque tanto BYTES como TEMPO tem uma importância muito pequena na definição das melhores conexões. Por outro lado, os pesos usados também reforçam a flexibilidade do método proposto neste Segundo Experimento – se fossem usados pesos mais altos, valores finais diferentes seriam gerados pelo SGA.

Também deve ser levado em conta que os resultados foram alcançados devido a topologia da rede, formada por dois controles de segurança sequenciais em linha. Apesar do uso de camadas de controles de segurança ser uma recomendação de proteção de perímetro reconhecida por profissionais de segurança, eventuais reproduções ou novos desenvolvimentos devem ter em conta este aspecto. Além disso, os bons resultados obtidos pelo método de previsão proposto mostram que os dados analisados em uma determinada amostra possuem relação com os dados analisados em uma amostra anterior, fato que pode não se confirmar em outros ambientes.

5.3 Validação

Diversas pesquisas referenciadas neste trabalho buscam alcançar o EN por meio da construção de jogos não-cooperativos (WU et al., 2010; NADKAR ET AL., 2012; MA; RAO; YAU, 2011; KHAN; AHMAD, 2009; AGAH et al., 2004). Uma forma bastante plausível de comparar os resultados alcançados pelo SGA é verificar se o equilíbrio gerado pelos cálculos UCB corresponde a um EN. Apesar da proposta do SGA ser totalmente baseada em jogos cooperativos, é possível transformar um jogo cooperativo (ou parte dele) em não-cooperativo (NUDELMAN et al., 2004). Além disso, as negociações que envolvem o processo de UCB podem ser consideradas **um subjogo não-cooperativo dentro de um jogo cooperativo maior**, pois os jogadores (variáveis independentes) buscam as melhores recompensas individuais sem que fique descaracterizado o caráter cooperativo do jogo.

Sendo assim, a fim de validar os resultados, parte das variáveis independentes é usada na formação de dois jogadores antagonistas de um subjogo não-cooperativo. Uma segunda parte das variáveis independentes é transformada no conjunto de estratégias S de cada jogador. As variáveis resposta são os valores das estratégias. Busca-se então o EN por meio do corte de estratégias com base em dominância iterada. Apesar de haver outras formas matemáticas para se obter o EN (AUMANN; BRANDENBURGER, 1995), a técnica escolhida é um dos meios mais simples e práticos para se alcançar o EN. A dominância iterada se caracteriza por cortar da matriz de recompensas as estratégias que são fortemente dominadas (vide Seção 2.2.2 Jogos Não-Cooperativos). Além disso, o resultado desse processo tende a ser também um estado ótimo de Pareto - quando um jogador não consegue mudar sua estratégia para um resultado melhor sem prejudicar os ganhos do oponente. Como na verdade esses oponentes fazem parte de um jogo maior cooperativo, onde as estratégias são pensadas para benefício geral, a possibilidade da obtenção de um estado ótimo de Pareto é grande (vide Seção 2.2.2 Jogos Não-Cooperativos).

Nas próximas seções a criação de subjogos não-cooperativos e a utilização de dominância iterada são reproduzidas para os resultados dos dois experimentos com o objetivo de verificar se estes resultados correspondem a situações de EN e ótimo de Pareto.

5.3.1 Validação do Primeiro Experimento

O Primeiro Experimento é caracterizado por uma grande variedade de pequenos experimentos. A necessidade de transformar uma parte das variáveis independentes em jogadores não-cooperativos e outra parte no conjunto de estratégias *S*, além da divisão das variáveis resposta em diversas variantes, torna o uso da dominância iterada uma tarefa bastante complexa. O principal problema está na quantidade de dados e matrizes de recompensas que devem ser geradas. A utilização de mais de um jogador para demonstrar a ocorrência do EN também não facilita a tarefa (AUMANN; BRANDENBURGER, 1995). Dessa forma, a opção é a validação de apenas um dos resultados mostrados no quadro 1, que mostra que os resultados de EN obtidos pela dominância iterada convergem para os mesmos resultados do SGA.

A variante escolhida arbitrariamente é referente ao método de previsão por tendência polinomial, amplitude como estimador da variabilidade e tipo de posicionamento *M2*. Os valores desta variante, mostrados de forma mais completa na tabela 30 (APÊNDICE E), são reproduzidos na sequência na tabela 25:

Tabela 25 - Os cinco registros com CM mais alto (valores reproduzidos da tabela 30 (APÊNDICE E)).

Método: Tend. Polin. – Estimador: Amplitude – Tipo: M2			
Jogador *	WTP	Custo	CM
1-5,75-3-25	2,1200	1,8280	1,1195
1-5,85-3-25	2,1179	1,8275	1,1179
1-5,95-3-12	2,1155	1,8267	1,1163
1-6,05-3-12	2,1129	1,8267	1,1137
1-6,15-3-12	2,1108	1,8267	1,1116

* Jogador= velocidade dos nós – intervalo de envio de pacotes – passos à frente na previsão – tamanho dos pacotes.

Fonte: o autor.

Percebe-se na tabela 25 que os dois jogadores com maiores valores de CM diferem apenas no intervalo de envio entre pacotes (5,75 s para o primeiro e 5,85 s para o segundo). A principal ideia é formar um subjogo não cooperativo tendo como jogadores adversários os dois primeiros colocados da tabela 25. A base para criação desse novo subjogo é o intervalo de envio de pacotes (única variável independente

com valores diferentes). Sendo assim, dois novos jogadores adversários são formados levando-se em conta o intervalo de envio de pacotes de 5,75 s para o chamado **jogador azul** e 5,85 s para o chamado **jogador vermelho**. O conjunto de estratégias S de cada um deles é formado pela velocidade dos nós, pela quantidade de passos à frente e pelo tamanho dos pacotes. O objetivo final é verificar se após o processo de dominância iterada o EN deste conjunto de estratégias converge para os valores de intervalo de envio de pacotes escolhidos.

Apesar desta simplificação a quantidade de estratégias ainda é muito grande: 3 (*velocidades*) $\times$ 3 (*passos*) $\times$ 5 (*tamanho*) = 45 estratégias para cada jogador, dentro da variante especificada. Dessa forma, o quadro 1 mostra apenas parte do conjunto de estratégias e valores da tabela 25 que passam pelo processo de dominância iterada:

Quadro 1 - Parte do processo de dominância iterada com EN do Primeiro Experimento.

		Jogador: intervalo=5,85 s																			
	Estratégias*	1-3-200				1-3-100				1-3-50				1-3-25				1-3-12			
		Ampl. M2	Ampli. M2	Energia	Energia	Ampl. M2	Ampli. M2	Energia	Energia	Ampl. M2	Ampli. M2	Energia	Energia	Ampl. M2	Ampli. M2	Energia	Energia	Ampl. M2	Ampli. M2	Energia	Energia
Jogador: intervalo = 5,75 s	1-3-200	0,9285	0,9583	97,122	97,073	0,9285	0,9583	97,122	97,535	0,9285	0,9583	97,122	97,735	0,9285	0,958	97,122	97,9131	0,9285	0,9583	97,122	97,824
	1-3-100	0,9285	0,9583	97,458	97,073	0,9285	0,9583	97,458	97,535	0,9285	0,9583	97,458	97,735	0,9285	0,958	97,458	97,9131	0,9285	0,9583	97,458	97,824
	1-3-50	0,9285	0,9583	97,458	97,073	0,9285	0,9583	97,458	97,535	0,9285	0,9583	97,458	97,735	0,9285	0,958	97,458	97,9131	0,9285	0,9583	97,458	97,824
	1-3-25	0,9285	0,9583	97,458	97,073	0,9285	0,9583	97,458	97,535	0,9285	0,9583	97,458	97,735	0,9285	0,958	97,458	97,9131	0,9285	0,9583	97,458	97,824
	1-3-12	0,9285	0,9583	97,458	97,073	0,9285	0,9583	97,458	97,535	0,9285	0,9583	97,458	97,735	0,9285	0,958	97,458	97,9131	0,9285	0,9583	97,458	97,824
	1-6-200	1,7485	0,9583	97,122	97,073	1,7485	0,9583	97,122	97,535	1,7485	0,9583	97,122	97,735	1,7485	0,958	97,122	97,9131	1,7485	0,9583	97,122	97,824
	1-6-100	1,7485	0,9583	97,458	97,073	1,7485	0,9583	97,458	97,535	1,7485	0,9583	97,458	97,735	1,7485	0,958	97,458	97,9131	1,7485	0,9583	97,458	97,824
	1-6-50	1,7485	0,9583	97,762	97,073	1,7485	0,9583	97,762	97,535	1,7485	0,9583	97,762	97,735	1,7485	0,958	97,762	97,9131	1,7485	0,9583	97,762	97,824
	1-6-25	1,7485	0,9583	97,892	97,073	1,7485	0,9583	97,892	97,535	1,7485	0,9583	97,892	97,735	1,7485	0,958	97,892	97,9131	1,7485	0,9583	97,892	97,824
	1-6-12	1,7485	0,9583	97,795	97,073	1,7485	0,9583	97,795	97,535	1,7485	0,9583	97,795	97,735	1,7485	0,958	97,795	97,9131	1,7485	0,9583	97,795	97,824
	1-12-200	2,8649	0,9583	97,122	97,073	2,8649	0,9583	97,122	97,535	2,8649	0,9583	97,122	97,735	2,8649	0,958	97,122	97,9131	2,8649	0,9583	97,122	97,824
	1-12-100	2,8649	0,9583	97,458	97,073	2,8649	0,9583	97,458	97,535	2,8649	0,9583	97,458	97,735	2,8649	0,958	97,458	97,9131	2,8649	0,9583	97,458	97,824
	1-12-50	2,8649	0,9583	97,762	97,073	2,8649	0,9583	97,762	97,535	2,8649	0,9583	97,762	97,735	2,8649	0,958	97,762	97,9131	2,8649	0,9583	97,762	97,824
	1-12-25	2,8649	0,9583	97,892	97,073	2,8649	0,9583	97,892	97,535	2,8649	0,9583	97,892	97,735	2,8649	0,958	97,892	97,9131	2,8649	0,9583	97,892	97,824
	1-12-12	2,8649	0,9583	97,795	97,073	2,8649	0,9583	97,795	97,535	2,8649	0,9583	97,795	97,735	2,8649	0,958	97,795	97,9131	2,8649	0,9583	97,795	97,824
	3-3-200	2,7578	0,9583	97,122	97,073	2,7578	0,9583	97,122	97,535	2,7578	0,9583	97,122	97,735	2,7578	0,958	97,122	97,9131	2,7578	0,9583	97,122	97,824
	3-3-100	2,7578	0,9583	97,458	97,073	2,7578	0,9583	97,458	97,535	2,7578	0,9583	97,458	97,735	2,7578	0,958	97,458	97,9131	2,7578	0,9583	97,458	97,824

* Estratégias = velocidade dos nós – passos à frente na previsão – tamanho dos pacotes.

Fonte: o autor.

O quadro 1 mostra a amplitude em relação ao tipo de posicionamento *M2* para o jogador com *intervalo=5,75 s* (em azul na horizontal) e para o jogador com *intervalo=5,85 s* (em vermelho na vertical). A dominância iterada acontece da seguinte forma: quando os valores de amplitude (valores baixos são desejáveis) ou de energia (valores altos são desejáveis) de uma linha (estratégia do jogador azul com *intervalo=5,75 s*) são dominados por outra linha, toda a linha é riscada. O mesmo procedimento acontece para o jogador vermelho com *intervalo=5,85 s* em relação às colunas. É possível notar que a estratégia 1-3-25 (velocidade-passos-tamanho) contém os mesmos valores de amplitude e energia, exceto quando o *tamanho do pacote=200 bytes*, cuja linha está riscada por causa do aumento do consumo de energia causado pelo maior tamanho do pacote (primeira linha do quadro 1). O *tamanho de pacote=25* se sobressai (está marcado em negrito) justamente por conter o maior CM – informação que não consta no quadro 1.

O EN alcançado coincide com a *matriz UCB* gerada pelo SGA para as variantes escolhidas mostradas na tabela 25. Apesar de ter atuado em apenas uma fração de todos os jogadores, estratégias e variantes possíveis, os resultados da dominância iterada servem de validação para os resultados do SGA no Primeiro Experimento.

Em relação ao estado ótimo de Pareto, no qual o jogador não consegue melhorar sua recompensa sem afetar negativamente as recompensas dos outros jogadores, para o jogador azul há quatro estratégias diferentes para tamanho de pacote com os mesmos valores: 100, 50, 25 e 12 bytes. A primeira vista não há como garantir qual dessas estratégias (todas iguais) será usada pelo jogador azul. Há inclusive uma tendência que ele utilize a estratégia com *tamanho de pacote = 12 bytes*, justamente por ser o menor tamanho possível. Entretanto, deve ser levado em conta o valor do CM que não aparece no quadro 1, e indica que a melhor opção para ambos os jogadores é realmente o *tamanho de pacote = 25 bytes*. Além disso, a escolha do jogador azul não tem potencial para prejudicar o jogador vermelho, pois além dos valores das estratégias serem iguais, há uma situação de EN que direciona a escolha do *tamanho de pacote = 25 bytes*. Sendo assim, além do EN, a dominância iterada também traz embutida um estado ótimo de Pareto para o resultado.

Percebe-se que em linhas gerais a análise cooperativa proporcionada pelo SGA é mais ajustada que a análise não-cooperativa, pois ordena as estratégias de forma

precisa sem deixar de mostrar as outras estratégias que podem ser usadas. Como dito anteriormente, deve-se lembrar de que ao contrário da abordagem realizada em outros ramos da ciência, faixas opcionais de configurações são bastante desejáveis em ambientes de TI voltados para segurança.

5.3.2 Validação do Segundo Experimento

No Segundo Experimento o jogador é formado por duas variáveis independentes: $\{1\}$ o nível de segurança S_k e o percentual de *ranking*. Para criação do subjogo não-cooperativo são criados dois jogadores³⁰ representando o nível de segurança S_k : $S_1=400$ e $S_2=201$. Os percentuais de *ranking* = (0 a 200) constituem o conjunto E ³¹ de estratégias de cada jogador.

Na avaliação de cada uma das 77 amostras usadas no Segundo Experimento o resultado foi o mesmo: o cálculo de UCB levou a um EN para o subjogo não-cooperativo e conseqüentemente a um estado ótimo de Pareto. O quadro 2 mostra o processo de dominância iterada de parte³² da amostra A_{002} .

³⁰ Essa é a melhor escolha possível, pois há dois níveis de segurança e 200 percentuais de *ranking*. O inverso implicaria na criação de 200 jogadores (*ranking*) com apenas duas estratégias cada (nível de segurança).

³¹ Normalmente chamado de S , aqui o conjunto de estratégias é chamado de E para não confundir com os níveis de segurança $S_1=400$ e $S_2=201$.

³² Devido ao tamanho a opção foi mostrar apenas a parte que mostra o EN do quadro 2. Todo o resto não mostrado corresponde a trechos riscados.

Quadro 2 - Parte do processo de dominância iterada e faixas com EN do Segundo Experimento.

		Jogador: S ₂ =201																							
	Estratégias	104-200				103				102				101				100-29				28-25			
		VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN
Jogador: S ₁ =400	104-200	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	103	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	102	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	101	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	100-29	100	100	0,1433	55,0933	100	100	0,1433	54,8063	100	100	0,1433	54,8063	100	100	0,1433	54,6628	100	100	0,1433	0,1435	100	73,3404	0,1433	0,1435
	28-25	100	100	0,1433	55,0933	100	100	0,1433	54,8063	100	100	0,1433	54,8063	100	100	0,1433	54,6628	100	100	0,1433	0,1435	100	73,3404	0,1433	0,1435
	24	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	23	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	22	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	21	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	20	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	19	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	18	98,7252	100	0,1433	55,0933	98,7252	100	0,1433	54,8063	98,7252	100	0,1433	54,8063	98,7252	100	0,1433	54,6628	98,7252	100	0,1433	0,1435	98,7252	73,3404	0,1433	0,1435

Fonte: o autor.

O quadro 2 contém as recompensas de VP e VN para os dois jogadores ($S_1=400$ em azul na horizontal e $S_2=201$ em vermelho na vertical) no cruzamento de cada estratégia usada. Faixas que tiveram valores iguais de VP e VN dentro do mesmo nível de segurança estão agrupadas para facilitar a visualização. Há quatro recompensas em cada estratégia: $S_1\{E_{(VP;VN)}\}$ e $S_2\{E_{(VP;VN)}\}$ e todas são levadas em conta na exclusão, lembrando que para VP o maior valor é desejável, enquanto para VN ocorre o inverso. Caso apenas uma das quatro recompensas seja fortemente dominada por outra, toda a estratégia é excluída. Por exemplo, o cruzamento entre as estratégias $S_1\{28-25_{(100;0,1433)}\}$ e $S_2\{28-25_{(73,304;0,1435)}\}$ é dominada fortemente em relação ao cruzamento que contém o EN (estratégia $S_1\{28-25_{(100;0,1433)}\}$ e $S_2\{100-29_{(100;0,1435)}\}$) apenas na recompensa de S_2 para VP (73,304 é dominado por 100). Mesmo assim todo o cruzamento entre as estratégias $\{28-25_{(VP;VN)}\}$ e $S_2\{28-25_{(VP;VN)}\}$ é excluído.

O quadro 2 possui dois cruzamentos de estratégias com EN - não há incentivo para mudanças por parte dos jogadores. Os cruzamentos também representam estados com ótimo de Pareto, visto não há melhora para os jogadores, mesmo se essa melhora representasse piora para o adversário. Dessa forma, o resultado final mostra que para a amostra A_{002} com nível de segurança $S_1=400$ a faixa ideal de trabalho compreende os percentuais de *ranking* de 100 a 25, enquanto que para o nível de segurança $S_2=201$ (mais fechado que o anterior) os percentuais de *ranking* são um pouco mais restritos na faixa de 100 a 29. Exatamente os mesmos valores alcançados pelos cálculos UCB do SGA e mostrados na tabela 23 na Seção 5.2.2 Resultados do Segundo Experimento, confirmando assim a confiabilidade da arquitetura proposta.

No APÊNDICE H pode ser vista uma sequência de iterações realizadas passo a passo para os dados do Segundo Experimento.

6 CONCLUSÕES

Neste capítulo é apresentado um pequeno resumo sobre o trabalho e os resultados alcançados pelos experimentos. Também são discutidas as conclusões finais, as limitações e as sugestões para trabalhos futuros em conjunto com uma sequência contendo as publicações geradas a partir dessa pesquisa.

6.1 Resumo e Conclusão

Este trabalho se debruça sobre a ação de eventos indesejados sobre ambientes de TI e a dificuldade que existe no agrupamento de resultados gerados por diferentes controles de segurança, principalmente se estes forem mecanismos de detecção e previsão. Para isso, neste trabalho é proposta uma arquitetura denominada de Strategic Grouping Architecture (SGA) cujo principal objetivo é agrupar controles de segurança de detecção e/ou previsão. A hipótese principal da proposta é que um equilíbrio entre os resultados e configurações destes controles de segurança é uma forma válida de agrupamento, já que indica os resultados mais otimizados que podem ser alcançados. No SGA este equilíbrio é alcançado por meio da divisão da análise dos eventos (legítimos e maliciosos) em dois níveis de abstração: o técnico e o estratégico.

O nível técnico corresponde ao planejamento e operação dos controles de segurança envolvidos. Neste nível também são coletados resultados, definidos limites que não devem ser ultrapassados e estabelecidos níveis de relevância das configurações usadas pelos diferentes controles de segurança.

O nível estratégico corresponde à análise dos resultados e configurações levantados no nível técnico. Toda a análise estratégica é precedida por uma padronização estatística dos dados e tem como base cálculos de Unrestricted Cooperative Bargains (UCB), conceito proveniente da Teoria dos Jogos. Por meio do UCB é realizado um jogo estratégico cooperativo que leva em conta os valores das

configurações de cada mecanismo de proteção (variáveis independentes) e os resultados destes mesmos mecanismos (variáveis resposta).

O resultado gerado pelo SGA é uma matriz (*matriz UCB*) que indica quais configurações e resultados são mais equilibrados e devem consequentemente ser utilizados.

Para explorar a capacidade da arquitetura proposta de agrupar mecanismos de previsão e detecção e funcionar em diferentes ambientes sob a ação de eventos indesejados, dois experimentos são realizados como prova de conceito.

No Primeiro Experimento métodos de previsão e detecção são agrupados para a identificação de ação maliciosa no envio de informações de posicionamento dos nós em uma Mobile Ad-hoc NETwork (MANET). Extensivas análises tendo como base diversas configurações, variantes e resultados gerados pelos controles de segurança são realizadas. A partir da verificação de que altos valores de variabilidade das previsões de posicionamento são indicativos de atividade maliciosa, o SGA é utilizado para otimizar as melhores configurações de trabalho levando-se em conta o consumo de energia. Com a aplicação do SGA foi possível identificar atividade maliciosa dentro de séries históricas com variações aleatórias de posicionamento a partir de 1%.

No Segundo Experimento um ambiente de comércio eletrônico real com dois controles de segurança de detecção colocados sequencialmente é analisado por meio de um *ranking* que tem como base diversos aspectos da rede. Este *ranking* funciona como um método preditivo em tempo real, que se antecipa à análise do segundo controle de segurança (*Filtro 2*) e indica quais são as configurações suspeitas que devem ser bloqueadas. As métricas por trás do desempenho do método são os percentuais de Verdadeiro Positivo (VP), que indicam eventos indesejados previstos e que foram corretamente bloqueados pelo *Filtro 2*, os percentuais de Verdadeiro Negativo (VN), que indicam eventos indesejados não previstos e que foram bloqueados pelo *Filtro 2*, além da quantidade de *bytes* e do tempo em segundos de cada conexão. O SGA é utilizado para determinação das faixas de trabalho mais adequadas para o *ranking* e para os percentuais de VP e VN. Os resultados do Segundo Experimento mostraram que a eficiência do método proposto, usando o pior caso para aplicação do SGA, ficou em torno de 99,95%.

Esses resultados significaram que no dia do Segundo Experimento, das 38.868.309 conexões, entre 161.916 (0,41%) e 141.456 (0,36%) foram conexões maliciosas. Destas, em torno de 99,95% foram corretamente preditas e barradas antes de chegarem ao *Filtro 2*. Apenas 0,05% das conexões maliciosas passaram indevidamente pelo ranking – na comparação com as conexões detectadas pelo *Filtro 2* este índice corresponde ao percentual de FP. A previsão gerada também apontou baixíssimos índices de VN quando comparada com as conexões detectadas: em torno de 0,19% e 0,23%, dos quais 99,99% aconteceram porque as conexões ainda não estavam registradas no *ranking* (novas conexões).

Tanto no Primeiro, como no Segundo Experimento os resultados alcançados com o SGA foram bastante satisfatórios, agrupando os controles de segurança utilizados e indicando limites de uso e faixas de trabalhos otimizadas para as configurações dos mecanismos de proteção.

Parte das *matrizes UCB* do Primeiro e do Segundo Experimento foram validados por meio da: {1} transformação do jogo cooperativo usado em um jogo não cooperativo; e {2} a aplicação da técnica de dominância iterada que elimina as estratégias fortemente dominadas em busca do estado de Equilíbrio de Nash (EN). Os resultados das validações de ambos os experimentos confirmaram a otimização alcançado pelo SGA ao mostrar que os resultados alcançados pela *matriz UCB* correspondem a estados de EN e ótimo de Pareto em subjogos não-cooperativos.

Portanto, os resultados e validações dos experimentos vão de encontro às contribuições originais sugeridas na Seção 1.3, demonstrando que o equilíbrio entre resultados e configurações dos controles de segurança proporcionado pelo uso do SGA pode ser usado como forma de agrupamento de controles de segurança, inclusive quando há mecanismos de detecção e de previsão envolvidos. O equilíbrio trouxe ainda como vantagem adicional o conhecimento dos limites e das configurações mais adequadas de trabalho de cada controle de segurança.

A UCB mostra-se como uma alternativa interessante para prover equilíbrio em ambientes onde ocorre o gerenciamento de controles de segurança, já que algumas características necessárias para o funcionamento de uma barganha cooperativa, como a racionalidade e o conhecimento dos resultados estão presentes nestes ambientes. Sendo assim, fica demonstrado que uma abordagem cooperativa

estratégica - baseada nos melhores resultados, como a UCB - que funciona em um nível de abstração acima do técnico pode ser utilizada para prover o equilíbrio e agrupar controles de segurança. O fato de a análise estratégica funcionar em um nível de abstração diferente credencia o SGA a ser usado em qualquer outro ambiente que contiver ao menos uma variável resposta que é influenciada de forma oposta por pelo menos duas variáveis independentes. Na verdade, esta é uma das grandes contribuições adicionais do SGA: poder ser utilizado, dentro de ambientes de TI, para resolução de uma categoria de problemas que envolva a busca por equilíbrio a partir do relacionamento entre configurações (variáveis independentes) e resultados (variáveis resposta).

Outro ponto que surge como diferencial no uso da UCB neste trabalho é sequência de vários resultados na *matriz UCB* que servem como opções de uso para quem gerencia os controles de segurança. Essa sequência de resultados otimizados é conseguida pela execução continuada dos cálculos e pela eliminação contínua do jogador mais otimizado a cada repetição.

6.2 Limitações e Trabalhos Futuros

Mesmo com os bons resultados o trabalho também apresenta algumas limitações, discutidas brevemente a seguir:

- Alguns pontos importantes do SGA, como a escolha das variáveis independentes e a forma de definição dos limites de uso para os controles de segurança, não seguem uma metodologia fixa. Isso confere certa liberdade para a adaptação da arquitetura para diversos ambientes. Porém, este fato também pode trazer incertezas para usuários menos experientes que precisam implementar o SGA;
- Claramente o SGA é indicado para ambientes nos quais as variáveis independentes influenciam uma mesma variável resposta de forma oposta. Justamente nessas situações a busca pelo equilíbrio é importante. Em ambientes onde as variáveis independentes influenciam as variáveis resposta em uma mesma direção, o SGA pode ser implementado normalmente, porém,

a necessidade de equilíbrio entre as configurações e resultados em situações como essas é bem menor;

- O SGA não foi testado em ambientes nos quais o relacionamento entre as variáveis (independentes e resposta) não é muito evidente, possui viés, ou mesmo nos quais a relevância das variáveis é baixa;
- Testes em ambientes refratários às condições de um processo de barganha eficiente, como pouca racionalidade, desconhecimento de resultados ou falta de cooperação entre os jogadores não foram realizados. Ainda que sejam contrários ao senso comum que indica que essas características devem estar presentes em um ambiente de segurança de TI controlado, desenvolvimentos do SGA nestas condições seriam úteis para verificar a eficácia da arquitetura em ambientes com informações falsas ou cujos controles de segurança estejam sob ataque malicioso;
- Mesmo seguindo as recomendações sugeridas no nível técnico do SGA, o cruzamento de variáveis (independentes e resposta) de métodos de previsão e detecção ainda é uma tarefa extremamente difícil. A literatura mostra que entre as poucas implementações de métodos de previsão em ambientes de segurança de TI, muitos funcionam para casos e abordagens específicas. Além disso, por se tratar de métodos bastante diferentes, encontrar relacionamentos entre as configurações e resultados de mecanismos de detecção e de previsão é uma tarefa bastante difícil;
- O uso do SGA em ambientes diferentes dos testados nessa tese é perfeitamente possível, pois a abordagem estratégica embutida na arquitetura permite ampliar o uso para outras áreas. Porém, adequações são necessárias. Por exemplo, um ambiente no qual a coleta de dados do nível técnico seja problemática, pode afetar significativamente os resultados do nível estratégico.

Como trabalhos futuros a pretensão é desenvolver o SGA para que seja aplicado em outras áreas da TI que não envolvam a segurança da informação, visto que o potencial da arquitetura pode ser utilizado em uma categoria de problemas onde exista a necessidade por equilíbrio entre configurações e resultados. Como exemplo de ambientes que podem se beneficiar do uso do SGA podem ser citadas as redes cognitivas e a identificação de padrões biométricos comportamentais.

Desenvolvimentos que envolvam a definição mais formal das variáveis envolvidas no agrupamento também são necessários.

Os valores gerados na *matriz UCB* como barganha residual também precisam ser mais bem pesquisados. Em situações de barganha esses resíduos representam valores cuja divisão entre os jogadores não está definida. O desenvolvimento de mais pesquisas sobre o tratamento dado à barganha residual pode ser útil durante a aplicação do SGA em ambientes que requerem precisão absoluta, como redes industriais ou sistemas críticos. Uma sugestão para trabalhos futuros que pode ajudar na definição da divisão da barganha residual é expandir e desenvolver melhor a questão da relevância das variáveis independentes, verificando, por exemplo, o tipo de relacionamento que elas possuem.

Desenvolvimentos que incluam jogos cooperativos que envolvam mais de uma variável resposta ainda são necessários. Pesquisas deste tipo envolvem situações de barganha entre três ou mais jogadores e abrem caminho para o agrupamento de controles de segurança com outros componentes do ambiente de TI, como o sistema operacional ou sistemas de monitoramento.

Estudos também podem ser feitos para dividir a análise estratégica do SGA para que alguns módulos possam ser aplicados separadamente. A intenção é permitir que o processo de barganha fique mais flexível e possa ser aplicado em tempo real. Por exemplo, em processos de agrupamento nos quais os dados estejam bem ajustados para uso no SGA, ao invés da execução de todos os módulos sugeridos no nível estratégico, pode haver necessidade apenas do cálculo em tempo real da UCB. Estudos como este podem abrir espaço para que o esforço computacional de execução do SGA seja medido e compartilhado entre módulos que funcionem de forma independente.

Por fim, pesquisas com variáveis independentes que influenciem uma variável resposta em três ou mais direções diferentes são desenvolvimentos interessantes que podem somar qualidades ao SGA. Por exemplo, em situações nas quais uma variável resposta tem valores qualitativos nominais (como cores ou termos como rápido, médio e lento) ao invés de ordinais (como valores quantitativos) podem exigir novas formulações de jogos cooperativos.

6.3 Publicações Relacionadas

O desenvolvimento desta tese gerou algumas publicações. Algumas são diretamente originárias dos desenvolvimentos descritos neste trabalho outras foram geradas de forma indireta. Ainda assim, muitos dos trabalhos listados são baseados na linha de pesquisa relacionada à segurança de TI e prevenção de eventos indesejados que seguimos.

Artigo completos publicados em periódicos:

- PONTES, E., SILVA, A. A. A., GUELFÍ, A. E., KOFUJI, S. T. Security in Cloud Manufacturing: Forecasting and Multicorrelation Techniques for Dealing with Cyber attacks. **Journal of Information Security Research**. v.2, p.129 148, 2011.
- Anderson Silva; Fen Zhou; Elvis Pontes; Marcos A. Simplicio; Rui Aguiar; Adilson Guelfi; Sergio Kofuji. Energy-Efficient Node Position Identification through Payoff Matrix and Variability Analysis. **Aceito para publicação em Outubro de 2016 no Journal of Telecommunication Systems (TELS)** - ISSN: 1018-4864 (print version) ISSN: 1572-9451 (electronic version) - manuscrito: TELS-D-16-00114R2.

Trabalhos completos publicados em anais de eventos:

- Avanço, L.; Guelfi, A. E.; Pontes, Elvis; Silva, A. A. A.; KOFUJI, S. T.; ZHOU, F. An Effective Intrusion Detection Approach for Jamming Attacks on RFID Systems **In: The 5th International EURASIP Workshop on RFID 2015 Technology**, Rosenheim - Alemanha, 2015. <http://www.eurasip-rfid.org/>.
- A. A. A. Silva; PONTES, E.; GUELFÍ, A. E.; CAPRONI, I. M.; AGUIAR, R.; ZHOU, F.; KOFUJI, S. T. Predicting Model for Identifying the Malicious Activity of Nodes in MANETs **In: 20th IEEE Symposium on Computers and Communication (ISCC)**, Lanarca. 20th IEEE Symposium on Computers and Communication (ISCC). , 2015. p.475 482. <http://iscc2015.ieee-iscc.org>.
- SILVA, A. A. A.; PONTES, E.; ZHOU, F.; KOFUJI, S. T. Grey Model and Polynomial Regression for Identifying Malicious Nodes in MANETs **In: IEEE Conference and Exhibition Global Telecommunications (GLOBECOM)**, Austin, TE, USA, 2014, p.168-174. <http://www.ieeeglobecom.org/2014/>.

- SILVA, A. A. A.; Pontes, E.; GUELF, A. E.; ZHOU, F.; KOFUJI, S. T. PRBS/EWMA Based Model for Predicting Burst Attacks (Brute Force, DoS, DDoS). in **International Conference on Digital Information Management (ICDIM 2014)**. Bangkok, Tailândia, 2014. <http://www.icdim.org/icdim2014>.
- FORAIN, I.; GUELF, A. E.; PONTES, E.; SILVA, A. A. A. Detecção de Intrusão Utilizando Análise de Séries Temporais com Modelos ARMAX/GARCH In: **XIII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais (SBSeg)**, Manaus, 2013. v.1. p.100 113. <http://sbseg2013.icomp.ufam.edu.br>.
- Pontes, E.; GUELF, A. E.; KOFUJI, S. T.; SILVA, A. A. A.; GUELF, A. E. Applying multicorrelation for improving forecasting in cyber security In: **2011 IEEE Sixth International Conference on Digital Information Management (ICDIM)**. Melbourne. Austrália, 2011. pp. 179-186. <http://www.icdim.org/icdim2011>.

Trabalhos resumidos publicados em anais de eventos:

- Bongiovanni, W.; GUELF, A. E.; Pontes, E.; Silva, A. A. A.; ZHOU, F.; KOFUJI, S. T. Viterbi Algorithm for Detecting DDoS Attacks. In: **The 40th IEEE Conference on Local Computer Networks (LCN)**, Florida, EUA, 2015. <http://ieeeln.org>.

Capítulo de livros publicados:

- Pontes, E.; Silva, A. A. A.; GUELF, A. E.; KOFUJI, S. T. Earthquake Prediction: Analogy with Forecasting Models for Cyber Attacks in Internet and Computer Systems In: **Earthquake Research and Analysis Statistical Studies, Observations and Planning**.1 ed.Rijeka : InTech, 2012, v.1, p. 101-124. Editado por Sebastiano D'Amico, ISBN 978-953-51-0134-5, 470 pgs. Publicado por InTech em 2 de Março de 2012 (CC BY 3.0 License). DOI: 10.5772/2461.
- Pontes, E.; E., GUELF, A. E.; Silva, A. A. A.; KOFUJI, S. T. A Comprehensive Risk Management Framework for Approaching the Return on Security Investment (ROSI) In: **Risk Management in Environment, Production and Economy**.1 ed.Rijeka : InTech, 2011, v.2, p. 20-46. Editado por Matteo

Savino, ISBN 978-953-307-313-2, 226 pgs. Publicado por InTech em 12 de Setembro de 2011 (CC BY –NC-AS 3.0 License). DOI: 10.5772/965.

Editoração e prefácio de livro científico

- Silva, A. A. A.; Pontes, E.; E., GUELF, A. E.; KOFUJI, S. T. **E-Learning - Engineering, OnJob Training and Interactive Teaching**. Rijeka:Intech, 2012. (Editoração, Prefácio, Posfácio).
- Pontes, E.; E., GUELF, A. E.; Silva, A. A. A.; KOFUJI, S. T. **E-Learning - LongDistance and Lifelong Perspectives**. Rijeka:Intech, 2012. (Editoração, Prefácio, Posfácio).
- Pontes, E.; E., GUELF, A. E.; Silva, A. A. A.; KOFUJI, S. T. **E-learning - Organizational Infrastructure and Tools for Specific Areas**. Rijeka:Intech, 2012. (Editoração, Prefácio, Posfácio).
- Pontes, E.; E., GUELF, A. E.; Silva, A. A. A.; KOFUJI, S. T. **Methodologies, Tools and New Developments for E-Learning**. Rijeka:Intech, 2012. (Editoração, Prefácio, Posfácio).

REFERÊNCIAS

- ABDULLAH, A. et al. Intrusion Detection Forecasting Using Time Series for Improving Cyber Defence. **International Journal of Intelligent Systems and Applications in Engineering**, 2015, v. 3, n. 1, p. 28–33.
- AGAH, A. et al. Intrusion Detection in Sensor Networks: a Non-cooperative Game Approach. **Third IEEE International Symposium on Network Computing and Applications NCA**. Cambridge, EUA, 2004, p.343–346.
- AL-SULTAN, S. et al. A Comprehensive Survey on Vehicular Ad Hoc Network. **Journal of Network and Computer Applications**. Elsevier, 2014, v. 37, n. 1, p. 380–392.
- ARMSTRONG, J. Combining forecasts. **Principles of Forecasting: A Handbook for Researchers and Practitioners**, 2001, p. 1–19.
- AUMANN, R.; BRANDENBURGER, A. Epistemic Conditions for Nash Equilibrium. **Econometrica**, 1995, v. 7, n. 11, p. 1161–1180.
- AXELSSON, S. Intrusion Detection Systems : A Survey and Taxonomy. **Computer Engineering**, 2000, p. 1–27.
- AZEVEDO, P. F. DE. Informação e Barganha: Implicações Estratégicas em Arranjos Verticais. **Seminário Brasileiro da Nova Economia Institucionalista**, 2001, v. 0.
- BACHMANN, F. et al. **Documenting Software Architectures: Views and Beyond** 2nd Edition. EUA: Addison-Wesley Professional, 2010, 592 p.
- BARROW, D. K.; KOURENTZES, N. Distributions of Forecasting Errors of Forecast Combinations: Implications for Inventory Management. **International Journal of Production Economics**, 2016, v. 177, p. 24–33.
- BASILICO, N. et al. Security Games for Node Localization Through Verifiable Multilateralation. **IEEE Transactions on Dependable and Secure Computing**, 2014, v. 11, n. 1, p. 72–85.
- BASS, L.; CLEMENTS, P.; KAZMAN, R. **Software Architecture in Practice**. Pearson Education, Inc., 2003, 592 p.
- BASS, T. Intrusion Detection Systems & Multisensor Data Fusion : Creating Cyberspace Situational Awareness. **Communications of the ACM**, 2000, n. 43.4, p. 100–105.
- BEAUQUIER, J.; HU, Y. Intrusion Detection Based on Distance Combination. **International Journal of Computer, Information Science and Engineering**, 2007, v. 1, n. 7, p. 172–180.
- BHUIYAN, Z. A.; WANG, G.; VASILAKOS, A. V. Local Area Prediction-Based Mobile Target Tracking in Wireless Sensor Networks. **IEEE Transactions on Computers**, 2015, v. 64, n. 7, p. 1968–1982.
- BIDARKOTA, P. V. The Comparative Forecast Performance of Univariate and

Multivariate Models: an Application to Real Interest Rate Forecasting. **International Journal of Forecasting**, 1998, v. 14, n. 4, p. 457–468.

BIERMAN, H. S.; FERNANDEZ, L. **Teoria dos Jogos - Game Theory With Economic Applications**. Brazil/EUA: Pearson, 2011.

BORGNE, Y.-A. LE; SANTINI, S.; BONTEMPI, G. Adaptive Model Selection for Time Series Prediction in Wireless Sensor Networks. **Signal Processing**, 2007, v. 87, n. 12, p. 3010–3020.

BOSMAN, H. et al. Spatial Anomaly Detection in Sensor Networks Using Neighborhood Information. **Information Fusion**, 2016, v. 33, n. January, p. 41–56.

BRANDENBURGER, A. Cooperative Game Theory: Characteristic Functions, Allocations, Marginal Contribution. **Stern School of Business. New York University**, 2007, v. 4, n. 1, p. 1–9.

BRANDENBURGER, A.; DEKEL, E. Rationalizability and Correlated Equilibria. **Econometrica**, 1987, v. 55, n. 6, p. 1391–1402.

BRANDENBURGER, A. M.; NALEBUFF, B. J. The Right Game: Use Game Theory to Shape Strategy. **Harvard Business Review**, July-August 1995, p. 57–71.

BRANDENBURGER, A. M.; STUART JR, H. W. Value Based Business Strategy. **Journal of Economics & Management Strategy**, 1996, v. 5, n. 1, p. 5-24.

BRANDENBURGER, A.; STUART JR, H. Biform Games. **Management Science**, 2007, v. 53, n. 4, p. 537–549.

CHEBROLU, S.; ABRAHAM, A.; THOMAS, J. P. Feature Deduction and Ensemble Design of Intrusion Detection Systems. **Computers and Security**, 2005, v. 24, n. 4, p. 295–307.

CHEN, L. C. L.; LENEUTRE, J. A Game Theoretical Framework on Intrusion Detection in Heterogeneous Networks. **IEEE Transactions on Information Forensics and Security**, 2009, v. 4, n. 2, p. 165–178.

CHRISTOFFERSEN, P. F. Evaluating Interval Forecasts. **International Economic Review**, 1998, v. 39, n. 4, p. 841–862 .

CLEMENTS, M. P.; SMITH, J. Evaluating Multivariate Forecasting Densities: a Comparison of Two Approaches. **International Journal of Forecasting**, 2002, v. 18, p. 397–407.

CORTEZ, P. et al. Multiscale Internet Traffic Forecasting Using Neural Networks and Time Series Methods. **Expert Systems**, 2012, v. 29, n. 2, p. 143–155.

DENG, J. L. Introduction to Grey System Theory. **In The Journal of Grey System**, 1989, n. 1, p. 1–24.

DIAS, M. A. R. Opções Reais Híbridas com Aplicações em Petróleo. **Educação & Sociedade**, 2005, v. 25, n. 88, p. 509.

DIXIT, A. K.; SKEATH, S. **Games of Strategy**. 2nd ed. New York, EUA: Norton, 2004, 665 p.

DRAPER, N. R.; SMITH, H. **Applied Regression Analysis**. 3rd ed. EUA: Wiley-Interscience, 1998, 706 p.

EKEL, P. et al. Multicriteria analysis Based on Constructing Payoff Matrices and Applying Methods of Decision Making in Fuzzy Environment. **Optimization and Engineering**, 2011, v. 12, n. 1–2, p. 5–29.

FEITOSA, E.; SOUTO, E.; SADOK, D. H. An Orchestration Approach for Unwanted Internet Traffic Identification. **Computer Networks**, 2012, v. 56, n. 12, p. 2805-2831.

FIANI, R. **Teoria dos Jogos - Com Aplicações em Economia, Administração e Ciências Sociais**. 2a. ed. São Paulo, Brazil: Elsevier, 2006, 394 p.

GAO, M.; WANG, N. A Network Intrusion Detection Method Based on Improved K-means Algorithm. **Advanced Science and Technology Letters**, 2014, v. 53, n. AITS 2014, p. 429–433.

GARCIA-TEODORO, P. et al. Anomaly-based Network Intrusion Detection: Techniques, Systems and Challenges. **Computers & Security**, 2009, v. 28, n. 1, p. 18–28.

GHASEMIGOL, M.; GHAEMI-BAFGHI, A.; TAKABI, H. A Comprehensive Approach for Network Attack Forecasting. **Computers and Security**. Elsevier. 2016, v. 58, p. 83–105.

GIACINTO, G.; ROLI, F. Intrusion Detection in Computer Networks by Multiple Classifier Systems. In **Proc. of the IEEE 16th International Conference on Pattern Recognition (ICPR)**. Quebec Canada, 2002, v. 2, p. 390–393.

GIACINTO, G.; ROLI, F.; DIDACI, L. Fusion of Multiple Classifier for Intrusion Detection in Computer Networks. **Pattern Recognition Letters**, 2003, v. 24, p. 1795–1803.

GMYTRASIEWICZ, P. J.; DURFEET, H. A Rigorous, Operational Formalization of Recursive Modeling. **First International Conference on Multiagent Systems**, San Francisco, EUA, 1995, p. 125–132.

GONZÁLEZ-RIVERA, G.; YOLDAS, E. Autocontour-based Evaluation of Multivariate Predictive Densities. **International Journal of Forecasting**. Elsevier, 2012, v. 28, n. 2, p. 328–342.

GOVINDARAJAN, M. Hybrid Intrusion Detection Using Ensemble of Classification Methods. **International Journal of Computer Network and Information Security**, 2014, v. 2, n. January, p. 45–53.

GUO, Z. et al. Perpendicular Intersection: Locating Wireless Sensors With Mobile Beacon. **IEEE Transactions on Vehicular Technology**, 2010, v. 59, n. 7, p. 3501–3509.

GUO, Y.; SCHILDT, S.; WOLF, L. Detecting Blackhole and Greyhole Attacks in Vehicular Delay Tolerant Networks. In: **IEEE Fifth International Conference on Communication Systems and Networks (COMSNETS)**. Bangalore, India, 2013, p. 1-7.

- GUPTA, S.; KUMAR, P.; ABRAHAM, A. A Profile Based Network Intrusion Detection and Prevention System for Securing Cloud Environment. **International Journal of Distributed Sensor Networks**, 2013, v. 9, n. 3, 12 p.
- HASLUM, K.; ABRAHAM, A.; KNAPSKOG, S. Fuzzy Online Risk Assessment for Distributed Intrusion Prediction and Prevention Systems. **In 10th IEEE International Conference on Computer Modelling and Simulation EUROSIM/UKSim2008**, EUA, 2008, p.216–223.
- HE, J. et al. Reputation-based Secure Sensor Localization in Wireless Sensor Networks. **The Scientific World Journal**, 2014, v. 2014, 10 p.
- HOCK, D.; KAPPES, M. A Self-Learning Network Anomaly Detection System Using Majority Voting. **Tenth International Network Conference (INC 2014)**, Plymouth, Reino Unido, 2014, p. 59–69.
- HODGE, V.; AUSTIN, J. A Survey of Outlier Detection Methodologies. **In Artificial Intelligence Review**, 2004, v. 22, n. 2, p. 85-126.
- HOFMEYR, S. A.; FORREST, S.; SOMAYAJI, A. Intrusion Detection Using Sequences of System Calls. **Journal of Computer Security**, 1998, v. 6, n. 3, p. 151–180.
- HUANG, H.; LEE, T.-H. To Combine Forecasts or to Combine Information? **Econometric Reviews**, 2010, v. 29, n. 56, p. 534–570.
- HUANG, L. et al. Predicting Execution Time of Computer Programs Using Sparse Polynomial Regression. **Advances in Neural Information Processing Systems**, 2010, p. 1–9.
- HUNG, C.; HUNG, C.; LIN, S. Predicting Time Series Using Integration of Moving Average and Support Vector Regression. **International Journal of Machine Learning and Computing**, 2014, v. 4, n. 6, p. 491–495.
- ISHIDA, C. et al. Forecast Techniques for Predicting Increase or Decrease of Attacks Using Bayesian Inference. **In: Communications, Computers and signal Processing - IEEE PACRIM**. Victoria, Canada, 2005, p. 450-453.
- JAAKKOLA, H.; THALHEIM, B. Architecture-Driven Modelling Methodologies. **Frontiers in Artificial Intelligence and Applications**, 2010, v. 225, p. 97–116.
- JANSEN, W.; GRANCE, T. **Guidelines on Security and Privacy in Public Cloud Computing**. Computer Security Division, Information Technology Laboratory, NIST, MD, USA, 2011, 80 p.
- JUVONEN, A.; SIPOLA, T. Combining Conjunctive Rule Extraction With Diffusion Maps for Network Intrusion Detection. **Proceedings - International Symposium on Computers and Communications**, 2013, p. 411–416.
- KAANICHE, H.; KAMOUN, F. Mobility Prediction in Wireless Ad Hoc Networks Using Neural Networks. **Journal of Telecommunications**, 2010, v. 2, n. 1, p. 95–101.
- KAYACAN, E.; ULUTAS, B.; KAYNAK, O. Grey System Theory-based Models in Time Series Prediction. **Expert Systems with Applications**. Elsevier, 2010, v. 37, n.

2, p. 1784–1789 .

KENT, K.; SOUPPAYA, M. **Guide to Computer Security Log Management**. Computer Security Division, Information Technology Laboratory, NIST, MD, EUA, 2006, 72 p.

KHAN, S. U.; AHMAD, I. A Cooperative Game Theoretical Technique for Joint Optimization of Energy Consumption and Response Time in Computational Grids. **IEEE Transactions on Parallel and Distributed Systems**, 2009, v. 20, n. 3, p. 346–360.

KO, S. I. M.; PARK, S. Y. Multivariate Density Forecast Evaluation: A Modified Approach. **International Journal of Forecasting**. Elsevier, 2013, v. 29, n. 3, p. 431–441.

KUANG, F.; XU, W.; ZHANG, S. A Novel Hybrid KPCA and SVM With GA Model for Intrusion Detection. **Applied Soft Computing Journal**. Elsevier, 2014, v. 18, p. 178–184.

KUMAR, G.; KUMAR, K. The Use of Multi-Objective Genetic Algorithm Based Approach to Create Ensemble of ANN for Intrusion Detection. **International Journal of Intelligence Science**, 2012, v. 2, n. October, p. 115–127.

L'ECUYER, P. et al. An Object-oriented Random-number Package with Many Long Streams and Substreams. **Operations Research**, 2002, v. 50, n. 6, p. 1073-1075.

LAI-CHENG, C. A. A High-efficiency Intrusion Prediction Technology Based on Markov Chain. In **IEEE International Conference on Computational Intelligence and Security Workshops CISW**. China, 2007, p.518–521.

LAZAREVIC, A. et al. Comparative Study of Anomaly Detection Schemes in Network Intrusion Detection. **Proceedings of the Third SIAM International Conference on Data Mining**, San Francisco, EUA, 2003, v. 3, p. 25–36.

LIN, Y. et al. Creditability-based Weighted Voting for Reducing False Positives and Negatives in Intrusion Detection. **Computers and Security**. Elsevier, 2013, v. 39, Part B, p. 460–474.

LIPOVETSKY, S.; CONKLIN, M. Analysis of Regression in Game Theory Approach. **Applied Stochastic Models in Business and Industry**, 2001, v. 17, n. 4, p. 319–330.

LIU, D. et al. Attack-resistant location estimation in wireless sensor networks. **ACM Transactions on Information and System Security**, 2008, v. 11, n. 4, p. 1–39.

LIU, J. et al. Reliability Assessment for Wireless Mesh Networks Under Probabilistic Region Failure Model. **IEEE Transactions on Vehicular Technology**, 2011, v. 60, n. 5, p. 2253–2264.

LIU, Y.; COMANICIU, C.; MAN, H. A Bayesian Game Approach for Intrusion Detection in Wireless Ad Hoc Networks. **Proceeding from the 2006 workshop on Game theory for communications and networks**, Pisa, Italia, 2006, 12 p.

LUO, X.; CHANG, X. A Novel Data Fusion Scheme Using Grey Model and Extreme

Learning Machine in Wireless Sensor Networks. **International Journal of Control, Automation and Systems**, 2015, v. 13, n. 3, p. 539–546.

MA, C. Y. T.; RAO, N. S. V; YAU, D. K. Y. A Game Theoretic Study of Attack and Defense in Cyber-physical Systems. **IEEE Conference on Computer Communications Workshops, INFOCOM WKSHPS**, Shangai, China, 2011, p. 708–713.

MANIKOPOULOS, C.; PAPAVALASSILIOU, S. Network intrusion and fault detection: A statistical anomaly approach. **IEEE Communications Magazine**, 2002, v. 40, n. 10, p. 76–82.

MELL, P.; KENT, K.; NUSBAUM, J. **Guide to malware incident prevention and handling**. Computer Security Division, Information Technology Laboratory, NIST, MD, USA, 2005, 110 p.

MIAO, C.; DAI, G.; YING, K.; CHEN, Q. Collaborative Localization and Location Verification in WSNs. **Sensors (Switzerland)**, 2015, v. 15, n. 5, p. 10631–10649.

MITCHELL, R.; CHEN, I. R. On Survivability of Mobile Cyber Physical Systems With Intrusion Detection. **Wireless Personal Communications**, 2013, v. 68, n. 4, p. 1377–1391.

MORETTIN, P. A.; BUSSAB, W. O. **Estatística Básica**. 8a. ed. São Paulo, Brazil: Saraiva, 2013, 526 p.

MORETTIN, P. A.; TOLOI, C. **Previsão de séries temporais**. São Paulo, Brazil: Atual, 1987, 356 p.

MORETTIN, P. A.; TOLOI, C. **Análise de séries temporais**. São Paulo, Brazil: Blucher, 2006, 538 p.

MUPPIRISSETTY, L. S.; SVENSSON, T.; WYMEERSCH, H. Spatial Wireless Channel Prediction under Location Uncertainty. **IEEE Transactions on Wireless Communications**, 2016, v. 15, n. 2, p. 1031–1044.

NADKAR, T. et al. Distributed Power Allocation for Secondary Users in a Cognitive Radio Scenario. **IEEE Transactions on Wireless Communications**, 2012, v. 11, n. 4, p. 1576–1586.

NAKAYAMA, H. et al. A Dynamic Anomaly Detection Scheme for AODV-Based Mobile Ad Hoc Networks. **IEEE Transactions on Vehicular Technology**, 2009, v. 58, n. 5, p. 2471–2481.

NASH JR, J. F Two-Person Cooperative Games. **Econometrica**, 1953 v. 21, n. 1, p. 128–140.

NASH JR, J. F. The Bargaining Problem. **Econometrica**, 1950, v. 18, n. 2, p. 155–162.

NIRMALA, M. B.; MANJUNATHA, A. S. Enhanced Voting based Secure Localization for Wireless Sensor Networks. **International Journal of Computer Network and Information Security**, 2015, v. 7, n. 12, p. 52–59.

NUDELMAN, E.; WORTMAN, J.; SHOHAM, Y.; LEYTON-BROWN, K. Run the GAMUT: A Comprehensive Approach to Evaluating Game-theoretic Algorithms. **Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS**, Nova York, EUA, 2004., p. 878–885.

O'Reilly, P. **Computer Security Division 2011 Annual Report**, Computer Security Division – NIST, EUA, 2012, 80 p.

OSBORNE, M. J. **An Introduction to Game Theory**. New York, EUA: Oxford University Press, 2004, 685 p.

OSBORNE, M. J.; RUBINSTEIN, A. **Bargaining and markets**. San Diego, EUA: Academic Press, Inc., 1990, 216 p.

PADMANABHAN, K.; KAMALAKKANNAN, P.; SHANTHARAJAH, S. P. Energy efficient Localization Technique in Wireless Sensor Network. **International Journal of Emerging Technology and Advanced Engineering**. 2013, v. 3, n. 11, p. 148–152.

PALIT, A. K.; POPOVIC, D. Nonlinear Combination of Forecasts Using Artificial Neural Network, Fuzzy Logic and Neuro-Fuzzy Approaches. **The Ninth IEEE International Conference of Fuzzy Systems**. San Antonio, EUA, 2000, v. 2, p.566–571.

PAPAGIANNAKI, K.; TAFT, N.; ZHANG, Z. L.; DIOT, C. Long-term Forecasting of Internet Backbone Traffic. **IEEE Transactions on Neural Networks**, 2005, v. 16, n. 5, p. 1110–1124.

PATCHA, A.; PARK, J. M. An overview of anomaly detection techniques: Existing solutions and latest technological trends. **Computer Networks**, v. 51, n. 12, p. 3448–3470, 2007.

PRASAD, K. S.; RAMAKRISHNA, S. An efficient traffic forecasting system based on spatial data and decision trees. **The International Arab Journey of Information Technology**, 2014, v. 11, n. 2, p. 186–194, 2014.

PREEZ, J. DU; WITT, S. F. Univariate versus multivariate time series forecasting: An application to international tourism demand. **International Journal of Forecasting**, 2003, v. 19, n. 3, p. 435–451.

RAYMOND, D. R.; MARCHANY, R. C.; MIDKIFF, S. F. Effects of Denial-of-Sleep Attacks on Wireless Sensor Network MAC Protocols. **IEEE Transactions on Vehicular Technology**, 2009, v. 58, n. 1, p. 367–380.

RIEDEL, S.; GABRYS, B. Evolving Multilevel Forecast Combination Models an Experimental Study. **NiSIS'2005 Symposium**. Albufeira, Portugal, 2005, 10 p.

SAGALE, A. D.; KALE, S. G. Hybrid Model For Intrusion Detection Using Naive Bayesian And Support Vector Machine. **International Journal of Computing and Technology**, 2014, v. 1, n. 3, p. 56–59.

SAMIEH, H. M.; WAHBA, K. Knowledge Sharing Behavior From Game Theory and Socio-psychology Perspectives. **Proceedings of the Annual Hawaii International Conference on System Sciences**, Hawaii, EUA, 2007, p. 1–10.

SARTINI, B. et al. Uma introdução a Teoria dos Jogos. **II Bienal da SBM - Universidade Federal da Bahia**, Bahia, Brasil, 2004, 64 p.

SCARFONE, K.; MELL, P. **Guide to Intrusion Detection and Prevention Systems (IDPS)**. Computer Security Division, Information Technology Laboratory, NIST, MD, USA, 2007, 127 p.

SCARFONE, K ; SOUPPAYA, M. **Guidelines for securing wireless local area networks (WLANs)**. Computer Security Division, Information Technology Laboratory, NIST, MD, USA, 2012, 24 p.

SEHUN, K. I. M. et al. Hybrid Intrusion Forecasting Framework for Early Warning System. **IEICE TRANSACTIONS on Information and Systems**, 2008, v. 91, n. 5, p. 1234-1241.

SHIN, S.; LEE, S.; KIM, H.; KIM, S. Advanced probabilistic approach for network intrusion forecasting and detection. **Expert Systems with Applications**, 2013, v. 40, n. 1, p. 315–322.

SILVA, A. A. A.; GUELF, A. E. Sistema para Identificação de Alertas Falso Positivos por meio de Análise de Correlacionamentos e Alertas Isolados. **In the 9th IEEE I2TS**, Rio de Janeiro, Brasil, 2010, 8 p.

SILVA, A. A. A.; PONTES, E.; ZHOU, F.; KOFUJI, S. T. Grey model and polynomial regression for identifying malicious nodes in MANETs. **IEEE Global Communications Conference, GLOBECOM**. Austin, EUA, 2014, p. 162–168.

SONG, J.; ZHU, Z.; PRICE, C. Feature Grouping for Intrusion Detection Based on Mutual Information. **Journal of Communications**, 2014, v. 9, n. 12, p. 987–993.

SRINIVAS, M.; ANDREW, H. S.; AJITH A. Intrusion Detection Using an Ensemble of Intelligent Paradigms. **Journal of Network and Computer Applications**. Elsevier Science, 2005, vol. 28, n. 2, p. 167-182.

STINE, K.; KISSEL, R.; BARKER, W. C.; LEE, A.; FASHLSING, J. **Appendices to Guide for Mapping Types of Information and Information Systems to Security Categories**. EUA, 2008a.

STUART JR, H. W. Cooperative Games and Business Strategy. **Game Theory and Business Applications**. 2001, p.189–213.

STÜBER, M.; GEMMAR, P.; GREVING, M. Machine Supported Development of Fuzzy-flood Forecast Systems. **European Conference on Advances in Flood Research**. Postdam, Alemanha, 2000, v. 2, p.504–515.

SU, L. Y. Prediction of Multivariate Chaotic Time Series with Local Polynomial Fitting. **Computers and Mathematics with Applications**. Elsevier, 2010, v. 59, n. 2, p. 737–744.

TAN, X. H.; LIE, T. T.; MEMBER, S. Allocation of Transmission Loss Cost Using Cooperative Game Theory in the Context of Open Transmission Access. **In IEEE Power Engineering Society Winter Meeting Conference**. Columbus, EUA, 2001 p.1215–1219.

- TIMMERMAN, A. Forecast Combinations. **Handbook of Economic Forecasting**. . Forthcoming in Elliott, G., C.W.J. Granger and A. Timmermann (Eds.), North Holland, 2005. pp. 135-196.
- TSENG, F. M.; YU, H. C.; TZENG, G. H. Applied Hybrid Grey Model to Forecast Seasonal Time Series. **Technological Forecasting and Social Change**, 2001, v. 67, n. 2–3, p. 291–302.
- WANG, Z.; BESSLER, D. A. Forecasting Performance of Multivariate Time Series Models with Full and Reduced Rank: An Empirical Examination. **International Journal of Forecasting**, 2004, v. 20, n. 4, p. 683–695.
- WEI, X.; YANG, Y. Robust Forecast Combinations. **Journal of Econometrics**, 2012, v. 166, n. 2, p. 224–236.
- WU, Q. et al. On Modeling and Simulation of Game Theory-based Defense Mechanisms Against DoS and DDoS Attacks. **ACM Spring Simulation Multiconference - SpringSim '10**. Orlando, EUA, 2010, 10 p.
- YANG, Z.; LIU, Y. Understanding Node Localizability of Wireless Ad-hoc and Sensor Networks. **IEEE Transactions On Mobile Computing**, 2012, v. 11, n. 8, p. 1249–1260.
- YANN-AEL, L.; GIANLUCA, B. Round Robin Cycle for Predictions in Wireless Sensor Networks. In **International Conference on Intelligent Sensors, Sensor Networks and Information Processing**. Melbourne, Australia, 2005, p.253–258.
- YAO, M. An Effective Intrusion Detection System Based on Multi-layers Mining Methods. **International Journal of Security and its Applications**, 2014, v. 8, n. 5, p. 311–322.
- YAO, Y. et al. A RSSI-based Distributed Weighted Search Localization Algorithm for WSNs. **International Journal of Distributed Sensor Networks**, 2015, v. 11, n. 4, 14 p.
- YE, F.; LUO, H.; LU, S.; ZHANG, L. Statistical En-route Filtering of Injected False Data in Sensor Networks. **Selected Areas in Communications**, 2004, v. 23, n. 4, p. 839–850.
- YU, A.; LV, S. Research into Positioning of the Least Square Support Vector Machine based on Fisher Fishing in WSN. **International Journal of Future Generation Communication and Networking**, 2016, v. 9, n. 6, p. 145–150.
- ZEADALLY, S. et al. Vehicular Ad hoc Networks (VANETS): Status, Results, and Challenges. **Telecommunication Systems**, 2012, v. 50, n. 4, p. 217–241.
- ZHANG, J.; CHEN, C.; XIANG, Y.; ZHOU, W.; VASILAKOS, A. V. An Effective Network Traffic Classification Method with Unknown Flow Detection. **IEEE Transactions on Network and Service Management**, 2013, v. 10, n. 2, p. 133–147.
- ZHANG, Y.; LEE, W. Intrusion Detection in Wireless Ad-Hoc Networks. In **6th ACM Annual International Conference on Mobile Computing and Networking**. Boston, EUA, 2000, p.275–283.

ZHANG, Y.; LEE, W.; HUANG, Y. Intrusion Detection Techniques for Mobile Wireless Networks. **In Proceedings of Mobile Networks and Applications**, EUA, 2003, v. 9, n. 5, p. 545-556.

ZHU, Q.; FUNG, C.; BOUTABA, R.; BAŞAR, T. A Game-theoretical Approach to Incentive Design in Collaborative Intrusion Detection Networks. **In International Conference on Game Theory for Networks, GameNets '09**. Istanbul, Turquia, 2009, p.384–392.

APÊNDICE A – Unrestricted Cooperative Bargain (UCB): Exemplo 1

Considere um jogo de atribuição de três lados, baseado no exemplo de Brandenburger (2007) e denotado por: $a_{ijk} = w_{jk} - c_{ij} \forall i \in N_1, j \in N_2, k \in N_3$, onde:

- N_1 = conjunto dos fornecedores: um fornecedor $N_1=\{i_0\}$;
- N_2 = conjunto das firmas: uma firma $N_2=\{j_1\}$;
- N_3 = conjunto dos compradores: dois compradores $N_3=\{k_2, k_3\}$;
- a_{ijk} = CM ou valor agregado (basicamente o preço-custo);
- $w_{j_1k_1}$ = preço = WTP do comprador k_2 interagindo com a firma $j_1 = \$9$;
- $w_{j_1k_3}$ = preço = WTP do comprador k_3 interagindo com a firma $j_1 = \$11$;
(percebe-se na tabela 26 que o fornecedor i_0 tem $CM=\$0$ e por isso está fora das interações);
- $c_{i_0j_1}$ = o custo de transação entre o fornecedor i_0 e a firma $j_1 = \$4$.

Portanto, para alcançar o TVC $v(N)$ do jogo é necessário subtrair o preço do custo (veja a terceira coluna da tabela 26).

- $v(\{i_0, j_1, k_2\}) = \$9 - \$4 = \$5$;
- $v(\{i_0, j_1, k_3\}) = \$11 - \$4 = \$7$ (o TVC é a principal transação do jogo, veja a segunda coluna da tabela 26);
- $v(\{k_2, k_3\}) = v(\{i_0, k_2\}) = v(\{i_0, k_3\}) = v(\{i_0, k_2, k_3\}) = \0 (não há valor gerado na interação entre estes jogadores);
- $v(\{i_0, j_1, k_2, k_3\}) = \7 (prevalece a transação mais vantajosa realizada com o comprador k_3);
- $v(\{j_1, k_2, k_3\}) = \$7$ (sem o fornecedor i_0 o custo é mantido no jogo porque está ligado ao preço do produto);

O CM ou valor agregado representa o montante que será removido do jogo se o jogador i sair do jogo (veja a quarta coluna da tabela 26).

- $CM_0 = v(\{i_0, j_1, k_2, k_3\}) - v(\{j_1, k_2, k_3\}) = \$7 - \$7 = \0 ;
- $CM_1 = v(\{i_0, j_1, k_2, k_3\}) - v(\{i_0, k_2, k_3\}) = \$7 - \$0 = \7 ;
- $CM_2 = v(\{i_0, j_1, k_2, k_3\}) - v(\{i_0, j_1, k_3\}) = \$7 - \$7 = \0 ;
- $CM_3 = v(\{i_0, j_1, k_2, k_3\}) - v(\{i_0, j_1, k_2\}) = \$7 - \$5 = \2 ;

O jogador com o mais alto valor de CM representa a melhor transação. Veja a tabela 26:

Tabela 26 - Cálculo de UCB do Exemplo 1

Jogador	Valor criado $v(N)$	$v(N \setminus \{i\})^{33}$	CM	Mínimo garantido³⁴	Faixas mínima e máxima³⁵	Barganha residual (max.-min.)
i_0	\$7	$v_{i_0}(\{j_1, k_2, k_3\}) = \7	$CM_0 = \$0$	\$0	\$0	\$0
j_1	\$7	$v_{j_1}(\{i_0, k_2, k_3\}) = \0	$CM_1 = \$7$	$CM_1 - CM_2 = \$5$	Min. = $\$9 - \$4 = \$5$ Max. = $\$11 - \$4 = \$7$	$\$7 - \$5 = \$2$
k_2	\$7	$v_{k_2}(\{i_0, j_1, k_3\}) = \7	$CM_2 = \$0$	\$0	\$0	\$0
k_3	\$7	$v_{k_3}(\{i_0, j_1, k_2\}) = \5	$CM_3 = \$2$	\$0	Min. = $\$9 - \$4 = \$5$ Max. = $\$11 - \$4 = \$7$	$\$7 - \$5 = \$2$

Fonte: o autor.

³³ O valor criado na transação sem o jogador i .

³⁴ O mais alto CM gerado por um jogador menos o menor valor maior que zero (leva-se em conta o custo).

³⁵ Faixa mínima de um jogador: o menor WTP de um jogador menos o custo. Faixa máxima de um jogador: o maior WTP de um jogador menos o custo.

APÊNDICE B – Unrestricted Cooperative Bargain (UCB): Exemplo 2

Considere um jogo de atribuição de três lados, baseado no exemplo de Stuart Jr (2001) e denotado por: $a_{sfb} = w_{fb} - c_{sf} \forall s \in N_1, f \in N_2, b \in N_3$, onde:

- $N1$ = conjunto de fornecedores: dois fornecedores $N1=\{s_1, s_2\}$;
- $N2$ = conjunto de firmas: duas firmas $N2=\{f_1, f_2\}$;
- $N3$ = conjunto de compradores: um comprador $N3=\{b_1\}$;
- $w_{f_1b_1}$ = preço = WTP do comprador b_1 interagindo com a firma $f_1 = \$100$;
- $w_{f_2b_1}$ = preço = WTP do comprador b_1 interagindo com a firma $f_2 = \$150$;
- $c_{s_1f_1} = c_{s_2f_1} = c_{s_1f_2} = c_{s_2f_2} = 10$ (já que o custo dos fornecedores é o mesmo, não importa qual deles fizer a transação com as firmas);
- $v(\{s_i, f_1, b_1\})$ = IVC da transação entre a firma f_1 e o comprador $b_1 = (w_{f_1b_1}) - c_i = \$100 - \$10 = \90 ;
- $v(\{s_i, f_2, b_1\})$ = IVC da transação entre a firma f_2 e o comprador $b_1 = (w_{f_2b_1}) - c_i = \$150 - \$10 = \140 (este é o TVC do jogo, veja a segunda coluna da tabela 27);
- $v(\{s_i, f_1, f_2\}) = 0$ (não há IVC se o comprador b_1 não estiver na transação);
- $CM_{b_1} = v(N) - v(\{s_i, f_1, f_2\}) = \$140 - \$0 = \140 (veja a quarta coluna da tabela 27);
- $CM_{f_1} = v(N) - v(\{s_i, f_2, b_1\}) = \$140 - \$140 = \0 (veja a quarta coluna da tabela 27);
- $CM_{f_2} = v(N) - v(\{s_i, f_1, b_1\}) = \$140 - \$90 = \50 (veja a quarta coluna da tabela 27);
- $CM_{s_1} = v(N) - v(N) = \0 .

Tabela 27 - Cálculo de UCB do Exemplo 2

Jogador	$v(N)$	$v(N \setminus \{i\})$	CM	Min. Garant.	Faixas min. e max.	Barganha residual (max.-min.)
b_1	\$140	$v_{b_1}(\{s_i, f_1, f_2\}) = \0	$CM_{b_1} = \$140$	$CM_{b_1} - CM_{f_2} = \$90$	Min. = \$100 - \$10 = \$90 Max. = \$150 - \$10 = \$140	\$140 - \$90 = \$50
f_1	\$140	$v_{f_1}(\{s_i, f_2, b_1\}) = \140	$CM_{f_1} = \$0$	\$0	\$0	\$0
f_2	\$140	$v_{f_2}(\{s_i, f_1, b_1\}) = \90	$CM_{f_2} = \$50$	\$0	Min. = \$100 - \$10 = \$90 Max. = \$150 - \$10 = \$140	\$140 - \$90 = \$50
s_i	\$140	$v(N) = \$140$	$CM_{s_i} = \$0$	\$0	\$0	\$0

Fonte: o autor.

Algumas considerações sobre os valores dispostos na tabela 27:

- \$0 é o IVC no jogo com a saída do comprador b_1 do jogo (é o principal jogador já que gera o TVC). Resta apenas o montante de \$10 da interação entre os fornecedores e as firmas. (tabela 27, segunda linha, terceira coluna);
- \$140 é o IVC pela presença do comprador no jogo: $a_{sifibi} = w_{fibi} - c_{si} \Rightarrow a_{sif1b1} - c_{si} \Rightarrow a_{sifibi} = \$150 - \$10 = \140 (tabela 27, segunda linha, quarta coluna);
- \$0 é o IVC no jogo, caso se mude o fornecedor, pois para ambos o custo $c = \$10$ (tabela 27, quinta linha, quarta coluna);
- \$50 é o IVC pela troca da firma f_1 pela f_2 : $CM_{sif2} = w_{sif2} - w_{sif1} \Rightarrow \$150 - \$100 = \50 . Na troca inversa da firma f_2 pela f_1 , não há geração de valor. O IVC é \$0 (tabela 27, terceira e quarta linhas, quarta coluna);
- \$90 é o menor IVC que o comprador b_1 tem direito. Do montante de \$140, que corresponde ao valor máximo gerado na transação entre a firma f_2 e o comprador b_1 , \$50 devem ser descontados ($w_{sif2} - w_{sif1}$). Não há garantias que o comprador b_1 vai gerar o valor mínimo de \$90, tendo em vista que ocorre uma barganha com a firma f_2 pelos \$50. Sendo assim o montante de \$90 pode aumentar e chegar a \$140. Percebe-se que a competição entre as duas firmas garante \$90 ao comprador b_1 e os \$50 restantes são divididos como uma barganha residual entre a firma f_2 e o comprador b_1 . Se o IVC do comprador b_1 chegar a \$140, que corresponde ao TVC do jogo, o IVC da firma f_2 será \$0 (tabela 27, segunda e quarta linhas, quinta coluna).

APÊNDICE C – programa de regressão linear (R-free)

Tabela 28 – Fonte do programa de regressão linear (R-free).

```
# caminho dos dados
caminho="c:/ipt/doutorado/pesquisa/progs"
setwd(caminho)
radio1=read.table(file="poly.csv",sep=";",dec=".",header=T) ## matriz de dados com desvio padrao
calculado por tendencia polinomial
radio2=read.table(file="grey.csv",sep=";",dec=".",header=T) ## matriz de dados com desvio padrao
calculado por GM(1,1)

# reg. linear radio1
rl1 = lm(formula = AMPLIPDM ~ SPEED+INTERV+STEPS+SIZE, data=radio1); rls1=summary(rl1)
rl2 = lm(formula = AMPLIPDM2 ~ SPEED+INTERV+STEPS+SIZE, data=radio1); rls2=summary(rl2)
rl3 = lm(formula = MEDIADMP ~ SPEED+INTERV+STEPS+SIZE, data=radio1); rls3=summary(rl3)
rl4 = lm(formula = MEDIADM2P ~ SPEED+INTERV+STEPS+SIZE, data=radio1); rls4=summary(rl4)
rl5 = lm(formula = ENERGY ~ SPEED+INTERV+STEPS+SIZE, data=radio1); rls5=summary(rl5)
# reg. linear radio2
rl6 = lm(formula = AMPLIPDM ~ SPEED+INTERV+STEPS+SIZE, data=radio2); rls6=summary(rl6)
rl7 = lm(formula = AMPLIPDM2 ~ SPEED+INTERV+STEPS+SIZE, data=radio2); rls7=summary(rl7)
rl8 = lm(formula = MEDIADMP ~ SPEED+INTERV+STEPS+SIZE, data=radio2); rls8=summary(rl8)
rl9 = lm(formula = MEDIADM2P ~ SPEED+INTERV+STEPS+SIZE, data=radio2); rls9=summary(rl9)
rl10 = lm(formula = ENERGY ~ SPEED+INTERV+STEPS+SIZE, data=radio2); rls10=summary(rl10)

# armazenamento e mostra os valores dos modelos lineares na tela
ctable1=coef(rls1);ctable1=array(c(ctable1[,1],ctable1[,4]),dim=c(5,2));colnames(ctable1)=c("Estimativ
a","p-value");rownames(ctable1)=c("Intercepto","Veloc","Interv","Passos","Tamanho"); ctable1
ctable2=coef(rls2);ctable2=array(c(ctable2[,1],ctable2[,4]),dim=c(5,2));colnames(ctable2)=c("Estimativ
a","p-value");rownames(ctable2)=c("Intercepto","Veloc","Interv","Passos","Tamanho"); ctable2
ctable3=coef(rls3);ctable3=array(c(ctable3[,1],ctable3[,4]),dim=c(5,2));colnames(ctable3)=c("Estimativ
a","p-value");rownames(ctable3)=c("Intercepto","Veloc","Interv","Passos","Tamanho"); ctable3
ctable4=coef(rls4);ctable4=array(c(ctable4[,1],ctable4[,4]),dim=c(5,2));colnames(ctable4)=c("Estimativ
a","p-value");rownames(ctable4)=c("Intercepto","Veloc","Interv","Passos","Tamanho"); ctable4
ctable5=coef(rls5);ctable5=array(c(ctable5[,1],ctable5[,4]),dim=c(5,2));colnames(ctable5)=c("Estimativ
a","p-value");rownames(ctable5)=c("Intercepto","Veloc","Interv","Passos","Tamanho"); ctable5
ctable6=coef(rls6);ctable6=array(c(ctable6[,1],ctable6[,4]),dim=c(5,2));colnames(ctable6)=c("Estimativ
a","p-value");rownames(ctable6)=c("Intercepto","Veloc","Interv","Passos","Tamanho"); ctable6
ctable7=coef(rls7);ctable7=array(c(ctable7[,1],ctable7[,4]),dim=c(5,2));colnames(ctable7)=c("Estimativ
a","p-value");rownames(ctable7)=c("Intercepto","Veloc","Interv","Passos","Tamanho"); ctable7
ctable8=coef(rls8);ctable8=array(c(ctable8[,1],ctable8[,4]),dim=c(5,2));colnames(ctable8)=c("Estimativ
a","p-value");rownames(ctable8)=c("Intercepto","Veloc","Interv","Passos","Tamanho"); ctable8
ctable9=coef(rls9);ctable9=array(c(ctable9[,1],ctable9[,4]),dim=c(5,2));colnames(ctable9)=c("Estimativ
a","p-value");rownames(ctable9)=c("Intercepto","Veloc","Interv","Passos","Tamanho"); ctable9
ctable10=coef(rls10);ctable10=array(c(ctable10[,1],ctable10[,4]),dim=c(5,2));colnames(ctable10)=c("E
stimativa","p-value");rownames(ctable10)=c("Intercepto","Veloc","Interv","Passos","Tamanho");
ctable10
```

APÊNDICE D – Regressão linear para relevância das variáveis independentes do Segundo Experimento

Tabela 29 – Regressão linear para ajuste dos dados - resultados de relevância das variáveis independentes do Primeiro Experimento

Variabilidade	Amplitude - M - Tend. Polin.		Amplitude - M – GM(1,1)	
Var. Indep.	Estimativas	p-value	Estimativas	p-value
Velocidade	3,72E-003	0	1,41E-003	0
Intervalo	2,22E-003	0	7,73E-004	0
Passos à frente	1,26E-003	0	-1,67E-004	3,14E-153
Tamanho	3,28E-019	1	4,39E-020	1

Variabilidade	Amplitude – M2 - Tend. Polin.		Amplitude – M2 – GM(1,1)	
Var. Indep.	Estimativas	p-value	Estimativas	p-value
Velocidade	1,29	0	0,54	0
Intervalo	0,77	0	0,31	0
Passos à frente	0,43	0	-0,04	1,19E-076
Tamanho	1,28E-016	1	2,21E-017	1

Variabilidade	Média – M - Tend. Polin.		Média – M – GM(1,1)	
Var. Indep.	Estimativas	p-value	Estimativas	p-value
Velocidade	6,96E-005	1,67E-133	2,96E-004	0
Intervalo	5,04E-005	4,08E-205	1,18E-004	0
Passos à frente	9,49E-005	0	-6,19E-005	0
Tamanho	1,69E-020	1	4,93E-021	1

Variabilidade	Média – M2 - Tend. Polin.		Média – M2 – GM(1,1)	
Var. Indep.	Estimativas	p-value	Estimativas	p-value
Velocidade	0,02	8,30E-166	0,06	0
Intervalo	0,01	4,42E-168	0,03	0
Passos à frente	0,03	0	-1,28E-003	3,01E-007
Tamanho	5,33E-018	1	2,03E-018	1

Energia	Tend. Polin.		GM(1,1)	
Var. Indep.	Estimativas	p-value	Estimativas	p-value
Velocidade	1,12E-014	1	1,12E-014	1
Intervalo	1,28	0	1,28	0
Passos à frente	7,57E-015	1	7,57E-015	1
Tamanho	-9,89E-003	1,35E-015	-9,89E-003	1,35E-015

Fonte: o autor.

APÊNDICE E – Matrizes UCB

Tabela 30 – Cinco registros com maior CM das matrizes UCB para o método de previsão tendência polinomial.

Método: Tendência Polinomial – Estimador: Amplitude – Tipo: M										% Aleatoriedade permitida					
Jogador *	WTP	Custo	V(N)	V(N\{i})	CM	Min. Garant.	Faixa Mín.	Faixa Máx.	Barg. Residual	1%	2%	5%	10%	15%	20%
1-6,05-3-12	398,7397	319,5278	79,2119	-318,5257	397,7376	397,7376	-318,5257	79,2119	397,7376	X	X	☺	☺	☺	☺
1-5,75-3-25	396,6340	319,5291	77,1049	-318,5257	395,6306	395,6306	-318,5257	77,1049	395,6306	☺	☺	☺	☺	☺	☺
1-5,95-3-12	396,4227	319,5278	76,8949	-318,5257	395,4206	395,4206	-318,5257	76,8949	395,4206	☺	☺	☺	☺	☺	☺
1-5,85-3-25	396,2723	319,5286	76,7437	-318,5257	395,2694	395,2694	-318,5257	76,7437	395,2694	☺	☺	☺	☺	☺	☺
1-6,15-3-12	395,9145	319,5278	76,3867	-318,5257	394,9124	394,9124	-318,5257	76,3867	394,9124	☺	☺	☺	☺	☺	☺
Método: Tendência Polinomial – Estimador: Amplitude – Tipo: M2										% Aleatoriedade permitida					
Jogador *	WTP	Custo	V(N)	V(N\{i})	CM	Min. Garant.	Faixa Mín.	Faixa Máx.	Barg. Residual	1%	2%	5%	10%	15%	20%
1-5,75-3-25	2,1200	1,8280	0,2920	-0,8275	1,1195	1,1195	-0,8275	0,2920	1,1195	☺	☺	X	☺	☺	☺
1-5,85-3-25	2,1179	1,8275	0,2904	-0,8275	1,1179	1,1179	-0,8275	0,2904	1,1179	☺	☺	☺	☺	X	☺
1-5,95-3-12	2,1155	1,8267	0,2888	-0,8275	1,1163	1,1163	-0,8275	0,2888	1,1163	X	☺	☺	☺	☺	☺
1-6,05-3-12	2,1129	1,8267	0,2862	-0,8275	1,1137	1,1137	-0,8275	0,2862	1,1137	X	X	☺	☺	☺	☺
1-6,15-3-12	2,1108	1,8267	0,2841	-0,8275	1,1116	1,1116	-0,8275	0,2841	1,1116	X	☺	☺	☺	☺	☺
Método: Tendência Polinomial – Estimador: Média – Tipo: M										% Aleatoriedade permitida					
Jogador *	WTP	Custo	V(N)	V(N\{i})	CM	Min. Garant.	Faixa Mín.	Faixa Máx.	Barg. Residual	1%	2%	5%	10%	15%	20%
1-5,75-3-25	8510,4977	7424,2878	1086,2099	-7423,2844	8509,4943	8509,4943	-7423,2844	1086,2099	8509,4943	☺	☺	☺	☺	☺	☺
1-5,85-3-25	8510,4288	7424,2873	1086,1415	-7423,2844	8509,4259	8509,4259	-7423,2844	1086,1415	8509,4259	☺	☺	☺	☺	☺	☺
1-8,35-3-12	8510,3607	7424,2832	1086,0775	-7423,2844	8509,3619	8509,3619	-7423,2844	1086,0775	8509,3619	☺	☺	☺	☺	☺	☺
1-8,35-3-25	8510,3607	7424,2833	1086,0774	-7423,2844	8509,3618	8509,3618	-7423,2844	1086,0774	8509,3618	☺	☺	☺	☺	☺	☺
1-8,35-3-50	8510,3607	7424,2856	1086,0751	-7423,2844	8509,3595	8509,3595	-7423,2844	1086,0751	8509,3595	☺	☺	☺	☺	☺	☺
Método: Tendência Polinomial – Estimador: Média – Tipo: M2										% Aleatoriedade permitida					
Jogador *	WTP	Custo	V(N)	V(N\{i})	CM	Min. Garant.	Faixa Mín.	Faixa Máx.	Barg. Residual	1%	2%	5%	10%	15%	20%
1-9,55-3-12	27,7185	24,2261	3,4924	-23,2338	26,7262	26,7262	-23,2338	3,4924	26,7262	☺	☺	☺	☺	☺	☺
1-8,55-3-12	27,7196	24,2273	3,4923	-23,2338	26,7261	26,7261	-23,2338	3,4923	26,7261	X	☺	☺	☺	☺	☺
1-7,85-3-12	27,7201	24,2279	3,4922	-23,2338	26,7260	26,7260	-23,2338	3,4922	26,7260	☺	☺	☺	☺	☺	☺
1-7,65-3-12	27,7200	24,2281	3,4919	-23,2338	26,7257	26,7257	-23,2338	3,4919	26,7257	☺	☺	☺	☺	☺	☺
1-9,35-3-12	27,7182	24,2263	3,4919	-23,2338	26,7257	26,7257	-23,2338	3,4919	26,7257	☺	☺	☺	☺	☺	☺

* Jogador= velocidade dos nós – intervalo de envio de pacotes – passos à frente na previsão – tamanho dos pacotes.

Fonte: o autor.

Tabela 31 – Cinco registros com maior CM das matrizes UCB para o método de previsão GM(1,1).

Método: GM(1,1) – Estimador: Amplitude – Tipo: M										% Aleatoriedade permitida					
Jogador *	WTP	Custo	V(N)	V(N\{i})	CM	Min. Garant.	Faixa Mín.	Faixa Máx.	Barg. Residual	1%	2%	5%	10%	15%	20%
1-5,75-12-25	1197,1954	947,3577	249,8377	-946,3566	1196,1943	1196,1943	-946,3566	249,8377	1196,1943	😊	😊	😊	😊	😊	😊
1-5,75-6-25	1193,9980	947,3577	246,6403	-946,3566	1192,9969	1192,9969	-946,3566	246,6403	1192,9969	😊	😊	😊	😊	😊	😊
1-5,95-12-12	1192,0038	947,3564	244,6474	-946,3566	1191,0040	1191,0040	-946,3566	244,6474	1191,0040	😊	😊	😊	😊	😊	😊
1-6,05-12-12	1188,2151	947,3564	240,8587	-946,3566	1187,2153	1187,2153	-946,3566	240,8587	1187,2153	😊	😊	😊	😊	😊	😊
1-5,85-12-25	1188,1779	947,3572	240,8207	-946,3566	1187,1773	1187,1773	-946,3566	240,8207	1187,1773	X	😊	😊	😊	😊	😊
Método: GM(1,1) – Estimador: Amplitude – Tipo: M2										% Aleatoriedade permitida					
Jogador *	WTP	Custo	V(N)	V(N\{i})	CM	Min. Garant.	Faixa Mín.	Faixa Máx.	Barg. Residual	1%	2%	5%	10%	15%	20%
1-5,75-3-25	3,9450	3,3448	0,6002	-2,3451	2,9453	2,9453	-2,3451	0,6002	2,9453	😊	😊	😊	😊	😊	😊
1-5,85-3-25	3,9360	3,3443	0,5917	-2,3451	2,9368	2,9368	-2,3451	0,5917	2,9368	😊	😊	😊	😊	😊	😊
1-5,85-6-25	3,9317	3,3443	0,5874	-2,3451	2,9325	2,9325	-2,3451	0,5874	2,9325	😊	😊	😊	😊	😊	😊
1-5,95-6-12	3,9251	3,3435	0,5816	-2,3451	2,9267	2,9267	-2,3451	0,5816	2,9267	😊	😊	😊	😊	😊	😊
1-5,75-6-25	3,9243	3,3448	0,5795	-2,3451	2,9246	2,9246	-2,3451	0,5795	2,9246	😊	😊	😊	😊	😊	😊
Método: GM(1,1) – Estimador: Média – Tipo: M										% Aleatoriedade permitida					
Jogador *	WTP	Custo	V(N)	V(N\{i})	CM	Min. Garant.	Faixa Mín.	Faixa Máx.	Barg. Residual	1%	2%	5%	10%	15%	20%
1-6,65-12-25	6300,5895	5170,1851	1130,4044	-5169,1832	6299,5876	6299,5876	-5169,1832	1130,4044	6299,5876	😊	😊	😊	😊	😊	😊
1-6,65-12-50	6300,5895	5170,1858	1130,4037	-5169,1832	6299,5869	6299,5869	-5169,1832	1130,4037	6299,5869	😊	😊	😊	😊	😊	😊
1-5,75-6-25	6276,1821	5170,1858	1105,9963	-5169,1832	6275,1795	6275,1795	-5169,1832	1105,9963	6275,1795	😊	😊	😊	😊	😊	😊
1-7,75-6-12	6273,2703	5170,1815	1103,0888	-5169,1832	6272,2720	6272,2720	-5169,1832	1103,0888	6272,2720	😊	😊	😊	😊	😊	😊
1-7,75-6-25	6273,2703	5170,1834	1103,0869	-5169,1832	6272,2701	6272,2701	-5169,1832	1103,0869	6272,2701	😊	😊	😊	😊	😊	😊
Método: GM(1,1) – Estimador: Média – Tipo: M2										% Aleatoriedade permitida					
Jogador *	WTP	Custo	V(N)	V(N\{i})	CM	Min. Garant.	Faixa Mín.	Faixa Máx.	Barg. Residual	1%	2%	5%	10%	15%	20%
1-5,75-3-25	30,1896	24,0141	6,1755	-23,0129	29,1884	29,1884	-23,0129	6,1755	29,1884	😊	😊	😊	😊	😊	😊
1-5,95-3-12	30,0826	24,0128	6,0698	-23,0129	29,0827	29,0827	-23,0129	6,0698	29,0827	😊	😊	😊	😊	😊	😊
1-6,15-3-12	30,0087	24,0128	5,9959	-23,0129	29,0088	29,0088	-23,0129	5,9959	29,0088	😊	😊	😊	😊	😊	😊
1-5,85-3-25	29,8588	24,0136	5,8452	-23,0129	28,8581	28,8581	-23,0129	5,8452	28,8581	X	😊	😊	😊	😊	😊
1-6,35-3-25	29,8451	24,0133	5,8318	-23,0129	28,8447	28,8447	-23,0129	5,8318	28,8447	😊	😊	😊	😊	😊	😊

* Jogador= velocidade dos nós – intervalo de envio de pacotes – passos à frente na previsão – tamanho dos pacotes.

Fonte: o autor.

APÊNDICE F – Regressão linear para relevância das variáveis independentes do Segundo Experimento

Tabela 32 – Regressão linear para ajuste dos dados - resultados de relevância das variáveis independentes do Primeiro Experimento.

Amostra	VAR	VP Estimate	VP p-value	VN Estimate	VN p-value	BYTES Estimate	BYTES p-value	TEMPO Estimate	TEMPO p-value
A002	CODMAX	0,02685726	0,01199576	-0,0001728918	0,9800339	-12,627066	5,762356E-006	-0,20588015	8,425421E-007
A002	INDICE	0,22291388	2,257538E-029	0,4097989172	3,114818E-122	-60,455887	8,350311E-032	-0,81149549	1,223638E-026
A003	CODMAX	0,03198224	0,001880757	-0,0143845246	0,09010627	-8,907548	0,0002434911	-0,23860597	0,00000278
A003	INDICE	0,19698472	1,188992E-025	0,5288073656	6,09445E-129	-54,627814	1,983489E-033	-0,76713536	1,775743E-017
A004	CODMAX	0,03991823	0,000103717	-0,0718349584	4,530453E-017	-9,081777	6,005209E-008	-0,23495587	3,52232E-007
A004	INDICE	0,20536883	1,241024E-027	0,4899648089	1,570041E-123	-58,911857	5,027593E-066	-1,05652726	8,197211E-035
A005	CODMAX	0,02900785	0,01667964	-0,0139790795	0,08630603	-7,494187	0,0001465517	-0,7959699	5,120817E-006
A005	INDICE	0,37338599	1,206425E-053	0,4925575762	6,494958E-125	-60,477082	1,229868E-053	-1,69890498	1,740436E-008
A006	CODMAX	0,03883843	0,002092879	-0,0139274332	8,968574E-005	-5,344584	0,01947739	-0,64136603	3,716202E-006
A006	INDICE	0,37984349	5,333488E-052	0,216986051	3,185772E-127	-60,111145	2,876754E-042	-1,49944658	4,43402E-010
A007	CODMAX	0,04267015	0,0006782481	-0,0069808345	0,1744351	-5,177604	0,003047441	-0,38833471	6,093069E-006
A007	INDICE	0,36907173	2,491199E-050	0,314442484	2,021211E-126	-23,763411	1,581636E-014	-0,87119772	4,484916E-009
A008	CODMAX	0,04823381	0,0001875145	-0,0065393935	6,758748E-005	-5,469262	0,003130727	-0,54351359	1,608887E-006
A008	INDICE	0,38476869	1,810428E-051	0,1021527774	6,663373E-130	-21,465909	3,744486E-011	-1,42696764	5,582933E-013
A009	CODMAX	0,05517751	0,000032307	-0,1150132503	1,106738E-045	-11,965774	0,01434689	-0,46058651	3,176842E-006
A009	INDICE	0,35826435	1,571832E-044	0,2604019682	4,011943E-068	165,237392	2,756182E-061	-0,71635511	2,278429E-005
A010	CODMAX	0,05368011	2,594884E-005	-0,0258286457	2,764632E-024	-15,282407	4,538577E-009	-0,61526538	1,417498E-006
A010	INDICE	0,39100266	8,520644E-054	0,1173927061	1,598691E-099	-89,34811	4,641682E-064	-1,83096128	3,890588E-016
A011	CODMAX	0,0543247	1,256021E-005	0,0049551239	0,5128085	-7,953149	0,002627444	-0,57888947	9,004873E-007
A011	INDICE	0,37256953	4,08745E-052	0,4522046804	3,675047E-123	-97,926924	1,108179E-069	-2,03563544	5,541844E-022
A012	CODMAX	0,055393	1,172641E-005	-0,0067957549	2,557186E-011	-4,623991	0,0007677655	-0,32855821	4,473497E-006
A012	INDICE	0,36547286	1,771184E-049	0,0588932069	1,569165E-122	3,808668	0,1042205	-0,72963647	3,85369E-009
A013	CODMAX	0,052454	2,519916E-005	-0,0207819495	2,154851E-016	-8,292107	2,25064E-006	-0,35028376	2,122259E-006
A013	INDICE	0,35838298	4,447461E-049	0,1300894461	1,540802E-109	-33,289993	1,220522E-025	-1,00273903	1,082943E-014
A014	CODMAX	0,05174754	3,060322E-005	-0,0151301208	5,641542E-013	-4,886647	2,634658E-005	-0,280057	1,528539E-006
A014	INDICE	0,34338917	3,123462E-046	0,1209283516	9,843953E-123	5,0772	0,01033649	-0,67990173	1,923626E-011
A015	CODMAX	0,04283717	0,0002843579	-0,0086129303	0,00231533	-2,872447	0,001661678	-0,29448236	4,673579E-006
A015	INDICE	0,32553647	7,812001E-046	0,1793265905	7,080857E-132	1,533013	0,3249903	-0,85254914	4,20442E-014

Amostra	VAR	VP Estimate	VP p-value	VN Estimate	VN p-value	BYTES Estimate	BYTES p-value	TEMPO Estimate	TEMPO p-value
A016	CODMAX	0,03671432	0,001756123	-0,0195808095	0,0002061289	-4,073777	0,007323286	-0,29785745	1,171374E-006
A016	INDICE	0,30479939	1,069826E-041	0,3116335875	8,284674E-123	55,511779	2,616172E-068	-0,50225949	1,738481E-006
A017	CODMAX	0,02778112	0,01594146	0,0087851196	0,293712	-5,258081	0,0001501041	-0,26083152	3,448306E-006
A017	INDICE	0,29223517	5,117331E-040	0,5248934962	1,034508E-129	29,616753	9,052879E-031	-0,44007758	4,893932E-006
A018	CODMAX	0,02615995	0,02615494	-0,0159734768	0,006444289	-3,177204	0,03203407	-0,25525638	3,956832E-006
A018	INDICE	0,30654856	1,027184E-041	0,3598102175	2,670391E-127	16,593369	1,748611E-010	-0,65093764	1,423471E-011
A019	CODMAX	0,03480529	0,0004886289	-0,0360611315	9,472036E-007	-7,241506	1,927686E-008	-0,1521538	2,491431E-006
A019	INDICE	0,21280954	1,262678E-030	0,4206174324	1,924846E-119	-48,260328	4,566728E-072	-0,63214201	6,440046E-027
A020	CODMAX	0,03441305	0,0005800845	-0,0101726093	0,06061941	-7,023626	3,567709E-005	-0,1399035	1,444746E-006
A020	INDICE	0,24435545	5,318229E-038	0,3255615624	4,089884E-124	-5,065249	0,07944971	-0,35211246	3,409996E-012
A021	CODMAX	0,04259603	5,411461E-005	0,0003439011	0,9585201	-11,198955	2,70686E-013	-0,2280057	1,587428E-008
A021	INDICE	0,2331761	1,4139E-032	0,4062899119	7,40575E-127	-47,627673	9,367739E-057	-0,66028447	2,860164E-020
A022	CODMAX	0,02521191	0,008560935	-0,0014110578	0,7532639	-4,157379	0,003961191	-0,11510288	6,841461E-006
A022	INDICE	0,20553713	1,005388E-030	0,2712673004	1,15082E-124	-93,601187	7,886408E-135	-0,6927176	7,179567E-045
A023	CODMAX	0,02881276	0,002729648	-0,0007437961	0,919788	-6,58989	2,042605E-006	-0,11950299	1,334536E-005
A023	INDICE	0,20938778	1,385064E-031	0,4508336731	5,644193E-126	-34,221337	5,702511E-039	-0,31893971	2,593816E-011
A024	CODMAX	0,02557168	0,007699657	-0,0155669142	0,008463774	-4,329183	0,000425707	-0,12822821	3,058789E-005
A024	INDICE	0,2018693	8,034236E-030	0,3588039574	8,607757E-126	-41,901765	2,19202E-062	-0,48789198	5,989446E-019
A025	CODMAX	0,02343851	0,0136294	0,0142028951	0,08889014	-5,213055	1,491823E-005	-0,13262832	5,741691E-006
A025	INDICE	0,19217511	5,948587E-028	0,5143079727	1,930235E-127	-64,536371	7,461124E-111	-0,58012192	1,711371E-027
A026	CODMAX	0,02144616	0,02047547	4,40111E-005	0,9935126	-2,564464	0,04350497	-0,12187805	0,0006648637
A026	INDICE	0,185798	1,341239E-027	0,3305667956	4,79279E-126	-86,856909	1,068495E-141	-0,82359761	1,502465E-034
A027	CODMAX	0,02560165	0,008130065	-0,0047114028	0,3178827	-6,20453	3,508262E-006	-0,13520338	1,874284E-005
A027	INDICE	0,19704665	2,657774E-028	0,2870973162	1,104919E-125	-67,153999	4,302287E-103	-0,58264815	2,438377E-024
A028	CODMAX	0,04849939	0,000008932	0,0028775219	0,6632591	-6,999375	3,431003E-010	-0,20160504	1,554707E-007
A028	INDICE	0,2601989	7,895038E-037	0,4033990775	5,351617E-126	-87,507151	1,321025E-164	-0,74912443	6,584948E-027
A029	CODMAX	0,02744007	0,005472428	-0,0021434136	0,5373641	-4,967299	1,100887E-005	-0,12942824	1,019214E-005
A029	INDICE	0,20346585	7,762275E-029	0,2190271379	2,836522E-130	-50,735033	4,386544E-090	-0,60144082	5,602991E-029
A030	CODMAX	0,02794662	0,003817093	0,0048211705	0,3678024	-5,079752	0,00107281	-0,0919273	4,586266E-005
A030	INDICE	0,21116483	9,362413E-032	0,3260502229	9,300722E-126	-34,346642	1,987146E-032	-0,1447244	0,0001783277
A031	CODMAX	0,02740395	0,004624088	-0,0194437861	5,709943E-006	-5,735843	0,0001685763	-0,13800345	7,38291E-006
A031	INDICE	0,19780274	1,705964E-028	0,244279203	9,452923E-119	-90,946951	4,034443E-124	-0,73405349	7,112795E-037

Amostra	VAR	VP Estimate	VP p-value	VN Estimate	VN p-value	BYTES Estimate	BYTES p-value	TEMPO Estimate	TEMPO p-value
A032	CODMAX	0,02282791	0,01858637	-0,0008965724	0,7989099	-4,46687	0,00235402	-0,0958774	2,695561E-005
A032	INDICE	0,21024416	3,02886E-031	0,2189124249	1,736061E-128	-103,996855	4,731168E-147	-0,70321856	3,560241E-054
A033	CODMAX	0,02159262	0,02437099	-0,0092119253	0,09948844	-4,591965	0,001526626	-0,08622716	7,728958E-005
A033	INDICE	0,20204632	8,018898E-030	0,342922367	8,108334E-127	-95,107906	1,323465E-136	-0,60354662	3,684679E-046
A034	CODMAX	0,02676796	0,005491882	-0,0020009525	0,7905648	-7,149804	9,031937E-005	-0,10195255	1,600427E-005
A034	INDICE	0,2017986	1,339161E-029	0,4466213042	3,782824E-122	-56,767778	8,067479E-055	-0,50108098	1,472028E-030
A035	CODMAX	0,02347128	0,01316882	-0,0003286432	0,9412273	-6,277507	4,523815E-005	-0,10940274	6,288752E-006
A035	INDICE	0,2010166	3,077381E-030	0,2777462156	1,042075E-128	-100,017432	8,539166E-136	-0,61441628	1,285097E-040
A036	CODMAX	0,02786958	0,005036448	0,0011122303	0,8325282	-5,377209	1,260947E-005	-0,10592765	3,774596E-006
A036	INDICE	0,21075191	3,081557E-030	0,3346379436	1,672694E-131	-67,595966	7,50754E-114	-0,55338555	1,112805E-037
A037	CODMAX	0,02678816	0,006281766	-0,0030342509	0,7025166	-4,725643	0,0003496286	-0,09540239	0,000022467
A037	INDICE	0,21513532	6,259055E-032	0,4877774164	9,195607E-127	-80,32188	1,84536E-126	-0,44756539	1,59226E-027
A038	CODMAX	0,03157725	0,001865114	0,0100294582	0,1252733	-5,840871	4,095419E-006	-0,10025251	1,271886E-006
A038	INDICE	0,22664395	6,769865E-033	0,404826993	5,569083E-128	-76,491609	3,426783E-126	-0,52187528	2,26104E-040
A039	CODMAX	0,02930333	0,002839014	-0,0010034951	0,863641	-3,093677	0,008166412	-0,08435211	2,198667E-005
A039	INDICE	0,22239848	1,241173E-033	0,370900613	3,402254E-131	-50,885816	8,656002E-086	-0,50897	4,052179E-041
A040	CODMAX	0,02792145	0,006516282	-0,0003551264	0,9573841	-4,850796	2,893354E-006	-0,08545214	5,897568E-007
A040	INDICE	0,23495032	3,761221E-034	0,4328644367	1,129293E-134	-56,41132	5,958956E-113	-0,42087483	6,594177E-039
A041	CODMAX	0,03030983	0,003012321	-0,0405422586	1,315609E-009	-4,827996	0,0001116937	-0,08455211	4,460238E-006
A041	INDICE	0,23483974	2,048433E-034	0,3983124169	7,052287E-126	-29,153706	1,867607E-035	-0,37203931	2,650645E-028
A042	CODMAX	0,06393103	3,535663E-008	0,019041501	0,001996731	-4,00205	4,705562E-006	-0,11460287	2,83381E-009
A042	INDICE	0,31284389	4,61576E-045	0,3592183889	8,87941E-121	-47,971046	6,071181E-114	-0,57567386	1,283464E-052
A043	CODMAX	0,04771324	1,300094E-005	0,0222227031	0,0004762009	-5,280757	0,000000024	-0,17017925	1,688174E-008
A043	INDICE	0,26403544	1,574654E-037	0,3909505182	6,508642E-128	-70,21096	1,152387E-155	-0,70124206	7,508705E-036
A044	CODMAX	0,05801283	2,097404E-007	-0,0201112803	0,001922697	-6,004	3,967191E-008	-0,16340409	1,26708E-007
A044	INDICE	0,29159859	1,096E-042	0,37926007	3,326474E-121	-50,246982	1,815695E-093	-0,58806709	7,547583E-026
A045	CODMAX	0,04976503	8,124458E-006	0,0028169279	0,5525228	-6,553114	6,510324E-012	-0,17575439	5,233142E-009
A045	INDICE	0,26308525	2,978884E-036	0,2871512175	6,02741E-125	-49,484607	5,222026E-109	-0,57668317	2,419388E-026
A046	CODMAX	0,04900798	9,828227E-006	0,0023765544	0,5303132	-11,706118	5,166855E-017	-0,22093052	4,648042E-009
A046	INDICE	0,25198422	3,508521E-034	0,2244895779	3,589829E-122	-97,51285	1,416477E-150	-0,76834786	4,01471E-029
A047	CODMAX	0,04510918	3,816249E-005	0,0115898847	0,08826384	-10,145854	1,0717E-017	-0,18990475	5,742554E-008
A047	INDICE	0,25092426	1,622647E-034	0,4021477992	4,296497E-122	-83,608616	2,10863E-152	-0,74226984	6,51876E-031

Amostra	VAR	VP Estimate	VP p-value	VN Estimate	VN p-value	BYTES Estimate	BYTES p-value	TEMPO Estimate	TEMPO p-value
A048	CODMAX	0,04625029	2,979382E-005	0,0029422111	0,7210113	-11,539288	1,161137E-020	-0,18520463	1,193374E-008
A048	INDICE	0,24978921	1,125357E-033	0,4904053372	1,090224E-122	-91,691085	1,400731E-160	-0,71701862	4,821278E-033
A049	CODMAX	0,0551052	9,130626E-007	0,0023756969	0,5064692	-8,226706	7,236862E-013	-0,15270382	0,000000053
A049	INDICE	0,27040835	1,247851E-037	0,2101012918	5,031082E-121	-74,324953	1,770192E-138	-0,6744717	1,093031E-037
A050	CODMAX	0,05054889	5,623064E-006	-0,0005883397	0,9340846	-9,049326	5,693127E-014	-0,16515413	2,442178E-007
A050	INDICE	0,27047066	5,443842E-038	0,4101145498	1,491644E-118	-92,599484	3,302134E-163	-0,70239175	1,424011E-032
A051	CODMAX	0,05218541	3,323915E-006	0,0069963474	0,1556284	-7,925523	3,428409E-012	-0,14650366	4,315994E-008
A051	INDICE	0,26277332	6,840184E-036	0,3075113442	4,301268E-129	-77,33864	1,176815E-144	-0,64090956	1,560277E-037
A052	CODMAX	0,05430788	1,497423E-006	-0,0135891747	0,009351058	-10,578614	4,808854E-016	-0,16340409	3,085229E-009
A052	INDICE	0,27349279	6,181603E-038	0,3188764121	2,015652E-126	-109,56154	2,549935E-177	-0,70185016	1,936816E-041
A053	CODMAX	0,06281779	7,111632E-008	0,000979502	0,7443058	-13,070527	2,18794E-014	-0,19495487	1,749484E-009
A053	INDICE	0,2998871	7,411519E-042	0,1644680528	4,628963E-112	-73,535405	5,326938E-088	-0,65398256	9,409576E-029
A054	CODMAX	0,06646393	1,841519E-008	0,02899809	0,0008591142	-13,314683	7,99109E-016	-0,16302908	2,232799E-007
A054	INDICE	0,29905198	6,067438E-041	0,4652090282	5,965707E-110	-92,739482	2,398385E-120	-0,71556007	1,973527E-034
A055	CODMAX	0,06539588	0,000000015	0,024919853	0,0005846946	-10,036951	1,245715E-014	-0,15510388	5,122869E-009
A055	INDICE	0,29168097	8,368155E-041	0,3895650986	1,158299E-110	-61,498082	8,765173E-099	-0,53914512	5,793743E-029
A056	CODMAX	0,06575176	1,0286E-008	0,0161678992	0,002028411	-11,472487	2,899332E-014	-0,160129	0,000000006
A056	INDICE	0,29694436	2,377006E-042	0,2767126933	2,577383E-108	-75,936923	3,685779E-106	-0,5618738	3,569091E-029
A057	CODMAX	0,072457	6,117821E-010	0,0190358034	0,000901442	-11,328458	4,019615E-017	-0,17397935	1,998981E-010
A057	INDICE	0,30421064	7,074956E-043	0,32580716	1,434344E-117	-94,877689	6,872449E-152	-0,65878528	3,080968E-038
A058	CODMAX	0,07149454	1,206163E-009	0,012879127	0,00246705	-12,888472	1,989085E-018	-0,16440411	7,24701E-013
A058	INDICE	0,32901653	6,65796E-048	0,2508944317	3,170174E-122	-103,75363	1,560741E-152	-0,80743018	1,532678E-067
A059	CODMAX	0,07035069	1,356917E-009	0,0160219255	0,001415199	-11,669592	1,816577E-018	-0,08325208	0,1331968
A059	INDICE	0,30140525	8,829748E-043	0,2724129004	1,049452E-111	-103,58864	6,478013E-167	-1,89397074	5,654366E-062
A060	CODMAX	0,07263759	3,854828E-010	0,0156164554	0,001851765	-10,704518	9,300588E-010	-0,15862897	1,152844E-009
A060	INDICE	0,31705784	2,478984E-046	0,277609351	4,009101E-114	-71,418207	3,928776E-081	-0,66511453	1,12339E-041
A061	CODMAX	0,07773424	2,461927E-011	0,0253723243	4,107854E-008	-12,099577	2,029758E-022	-0,17402935	0,000014484
A061	INDICE	0,3400853	2,011532E-051	0,2575473898	1,495984E-116	-88,141859	3,835734E-155	-1,38840328	5,347342E-064
A062	CODMAX	0,04105468	0,001068888	-0,0408492762	6,371097E-009	-7,636466	0,00000088	-0,20763019	0,0001996958
A062	INDICE	0,39312199	3,309314E-055	0,3005312372	1,211036E-085	-74,462061	7,350633E-098	-0,02322152	0,8066979
A063	CODMAX	0,03555884	0,005323247	0,026238561	0,0007120533	-7,820471	2,985941E-005	-0,23435586	6,145625E-005
A063	INDICE	0,3933896	8,570101E-054	0,4015657398	4,21903E-106	-82,815627	3,128937E-088	-1,22546845	6,700742E-030

Amostra	VAR	VP Estimate	VP p-value	VN Estimate	VN p-value	BYTES Estimate	BYTES p-value	TEMPO Estimate	TEMPO p-value
A064	CODMAX	0,03579297	0,004348458	0,0120855621	0,06519541	-9,388985	1,153678E-009	-0,28915723	8,775641E-006
A064	INDICE	0,38760746	6,096393E-054	0,3230165616	1,505666E-099	-101,737622	1,368179E-139	-1,34829171	1,712419E-029
A065	CODMAX	0,03887775	0,0017946	0,0194760969	0,005575262	-6,949749	0,000063813	-0,23955599	2,136568E-005
A065	INDICE	0,39724516	1,275736E-056	0,3479617049	1,98055E-100	-103,535985	4,361719E-124	-1,13571302	4,017424E-028
A066	CODMAX	0,04003757	0,001534366	0,0271553689	0,0003156225	-9,529663	7,741768E-006	-0,28185705	4,544282E-006
A066	INDICE	0,38792898	1,453495E-053	0,4050324048	1,091685E-110	-89,303604	1,109287E-082	-1,05150362	1,47809E-021
A067	CODMAX	0,03748435	0,002136629	0,016938941	0,002038921	-8,067477	7,972671E-005	0,0960774	0,6932504
A067	INDICE	0,39225265	3,459245E-057	0,2742315768	1,490664E-101	-94,68854	2,947347E-093	-4,19573962	2,427358E-021
A068	CODMAX	0,0451327	0,0003590436	0,0166773844	0,002392736	-9,534338	8,116439E-011	-0,02970074	0,8879895
A068	INDICE	0,39572435	3,311375E-055	0,3199812801	1,188559E-120	-105,133862	1,086655E-151	-3,80996601	4,325942E-023
A069	CODMAX	0,03701738	0,002727855	0,0140707543	0,02880258	-10,566239	7,642858E-008	-0,27890697	4,024848E-006
A069	INDICE	0,3816609	5,015677E-054	0,3167907027	1,685978E-099	-92,069961	1,609591E-095	-1,04096719	8,930762E-022
A070	CODMAX	0,03865231	0,001971358	0,0242502713	3,959484E-006	-10,825671	1,301852E-010	-0,22575564	1,288826E-005
A070	INDICE	0,38260095	2,239905E-053	0,2934551023	3,679448E-116	-94,066614	8,064869E-118	-0,99817201	3,262001E-026
A071	CODMAX	0,03845803	0,002189455	0,0107624291	0,006563133	-8,98125	5,551089E-010	-0,17437936	0,001332406
A071	INDICE	0,3866665	8,590035E-054	0,2126059525	3,487595E-110	-99,700514	1,038318E-145	-1,6080501	8,970763E-051
A072	CODMAX	0,0399596	0,001637406	-0,0136196155	0,01323931	-8,073252	0,000000001	-0,13490337	0,0009917933
A072	INDICE	0,38806559	2,857733E-053	0,2647795086	4,561226E-097	-111,301902	5,047486E-175	-1,41402172	2,398431E-063
A073	CODMAX	0,03492392	0,006568672	-0,0002691492	0,9699383	-7,278957	0,001020802	-0,1761794	7,880374E-006
A073	INDICE	0,41262343	4,931124E-057	0,3243729878	5,346994E-090	-90,422644	2,645315E-079	-0,9417312	5,653207E-037
A074	CODMAX	0,03380205	0,006932642	-0,0111681742	0,09939737	-5,59189	0,01511932	-0,19682992	4,266578E-005
A074	INDICE	0,40814038	2,832303E-058	0,3130341699	4,684445E-092	-99,158969	1,161936E-084	-1,18040934	1,857024E-038
A075	CODMAX	0,03509884	0,003754164	0,0125300133	0,05697317	-5,759319	1,286282E-005	-0,20170504	0,000000787
A075	INDICE	0,4347777	9,679464E-067	0,3558689959	9,91953E-111	-106,467667	1,087372E-166	-1,78498079	1,83232E-087
A076	CODMAX	0,03947865	0,001815156	-0,0139279632	0,02031698	-4,326258	0,03369058	-0,18322958	0,0001146009
A076	INDICE	0,42169129	3,199311E-060	0,2706629296	1,404269E-089	-35,403329	9,343502E-022	-1,11996971	4,7755E-036
A077	CODMAX	0,03785659	0,002081471	0,0030968799	0,6560456	-7,678567	0,0008458068	-0,18970474	8,529142E-007
A077	INDICE	0,43696488	7,243676E-066	0,3115677632	1,595442E-088	-88,579801	1,642404E-073	-1,34935126	2,192314E-065
A078	CODMAX	0,03983153	0,001431541	0,0240991075	0,005291107	-6,61259	2,598237E-006	-0,22458061	7,217889E-006
A078	INDICE	0,43959638	5,048946E-065	0,4116319414	6,401414E-096	-127,804478	7,745503E-185	-1,44692552	1,659843E-049

Fonte: o autor.

APÊNDICE G – Resultados da aplicação do SGA no para o Segundo Experimento

Tabela 33 – Faixas ideais e limites de *ranking* para as amostras do Segundo Experimento.

SAMPLES	NÍVEL DE SEGURANÇA: S ₁ =400									NÍVEL DE SEGURANÇA: S ₂ =201								
	RANKING		REAL			BASEADO NO MIN				RANKING		REAL			BASEADO NO MIN			
	MIN	MAX	VP	VN	VNN	VP	VN	VNN		MIN	MAX	VP	VN	VNN	VP	VN	VNN	
A002	25	100	100,00	0,14	100,00	-	-	-		29	100	100,00	0,14	100,00	-	-	-	
A003	50	100	99,44	1,25	100,00	98,49	0,55	100,00		31	31	99,32	1,19	100,00	98,33	0,51	100,00	
A004	25	50	100,00	2,12	100,00	100,00	2,12	100,00		34	50	100,00	3,18	100,00	78,85	3,18	100,00	
A005	59	100	100,00	0,10	100,00	81,41	0,10	100,00		63	104	100,00	0,12	100,00	44,99	0,12	100,00	
A006	44	99	100,00	0,74	100,00	100,00	0,74	100,00		46	101	100,00	0,83	100,00	100,00	0,83	100,00	
A007	42	103	100,00	0,11	100,00	100,00	0,11	100,00		45	105	100,00	0,00	100,00	100,00	0,00	100,00	
A008	59	104	100,00	0,00	100,00	96,83	0,00	100,00		59	106	100,00	0,00	100,00	95,96	0,00	100,00	
A009	41	103	100,00	0,08	100,00	100,00	0,08	100,00		43	104	100,00	0,50	100,00	100,00	0,50	100,00	
A010	42	103	100,00	0,79	100,00	98,62	0,79	100,00		50	106	100,00	0,55	100,00	28,19	0,55	100,00	
A011	41	102	100,00	0,10	100,00	100,00	0,10	100,00		50	104	100,00	0,18	100,00	100,00	0,18	100,00	
A012	41	103	100,00	0,00	100,00	100,00	0,00	100,00		49	105	100,00	0,00	100,00	100,00	0,00	100,00	
A013	30	103	100,00	0,00	100,00	100,00	0,00	100,00		79	104	100,00	0,00	100,00	97,74	0,00	100,00	
A014	62	103	100,00	0,00	100,00	99,36	0,00	100,00		65	104	100,00	0,00	100,00	100,00	0,00	100,00	
A015	45	103	100,00	1,66	100,00	100,00	1,66	100,00		79	103	100,00	2,09	100,00	96,89	2,09	100,00	
A016	41	103	100,00	0,10	100,00	100,00	0,10	100,00		47	103	100,00	0,19	100,00	100,00	0,19	100,00	
A017	44	103	100,00	0,67	100,00	99,07	0,67	100,00		47	103	100,00	0,77	100,00	100,00	0,77	100,00	
A018	48	103	100,00	0,90	100,00	98,88	0,90	100,00		50	103	100,00	1,05	100,00	98,79	1,05	100,00	
A019	50	98	100,00	0,31	100,00	99,79	0,31	100,00		50	99	100,00	0,42	100,00	100,00	0,42	100,00	
A020	39	103	100,00	0,11	100,00	100,00	0,11	100,00		47	103	100,00	0,12	100,00	100,00	0,12	100,00	
A021	43	99	100,00	0,72	100,00	99,19	0,72	100,00		58	103	100,00	0,73	100,00	99,21	0,73	100,00	
A022	33	101	100,00	0,08	100,00	100,00	0,08	100,00		47	101	100,00	0,09	100,00	100,00	0,09	100,00	
A023	42	101	99,53	0,75	100,00	98,59	0,75	100,00		47	101	99,52	0,75	100,00	99,52	0,75	100,00	
A024	28	101	99,63	0,42	100,00	99,63	0,42	100,00		47	101	99,57	0,43	100,00	99,57	0,43	100,00	
A025	40	102	100,00	0,13	100,00	98,60	0,13	100,00		48	102	100,00	0,12	100,00	95,70	0,12	100,00	
A026	34	102	100,00	0,47	100,00	100,00	0,47	100,00		47	102	100,00	0,47	100,00	100,00	0,47	100,00	
A027	30	103	100,00	0,09	100,00	100,00	0,09	100,00		47	102	100,00	0,09	100,00	100,00	0,09	100,00	
A028	36	102	100,00	0,16	100,00	99,44	0,16	100,00		59	103	100,00	0,16	100,00	99,47	0,16	100,00	

SAMPLES	NÍVEL DE SEGURANÇA: S ₁ =400									NÍVEL DE SEGURANÇA: S ₂ =201								
	RANKING		REAL			BASEADO NO MIN				RANKING		REAL			BASEADO NO MIN			
	MIN	MAX	VP	VN	VNN	VP	VN	VNN		MIN	MAX	VP	VN	VNN	VP	VN	VNN	
A029	29	102	100,00	0,08	100,00	100,00	0,08	100,00		47	102	100,00	0,08	100,00	100,00	0,08	100,00	
A030	38	102	100,00	0,03	100,00	99,00	0,03	100,00		47	102	100,00	0,04	100,00	100,00	0,04	100,00	
A031	29	102	100,00	0,00	100,00	100,00	0,00	100,00		47	102	100,00	0,00	100,00	100,00	0,00	100,00	
A032	40	102	100,00	0,03	100,00	98,97	0,03	100,00		48	102	100,00	0,00	100,00	95,49	0,00	100,00	
A033	29	102	99,97	0,09	100,00	99,97	0,09	100,00		48	70	99,97	0,10	100,00	99,97	0,10	100,00	
A034	69	102	100,00	0,03	100,00	99,69	0,03	100,00		69	103	100,00	0,03	100,00	99,97	0,03	100,00	
A035	67	103	100,00	0,15	100,00	100,00	0,15	100,00		67	103	100,00	0,12	100,00	100,00	0,12	100,00	
A036	56	103	100,00	0,08	100,00	100,00	0,08	100,00		56	103	100,00	0,12	100,00	100,00	0,12	100,00	
A037	46	103	100,00	0,00	100,00	100,00	0,00	100,00		48	103	100,00	0,00	100,00	100,00	0,00	100,00	
A038	59	103	99,96	0,04	99,96	99,91	0,04	99,96		59	103	99,95	0,05	99,95	99,90	0,05	99,95	
A039	59	103	100,00	0,00	100,00	100,00	0,00	100,00		59	103	100,00	0,00	100,00	100,00	0,00	100,00	
A040	60	103	100,00	0,05	100,00	99,45	0,05	100,00		60	103	100,00	0,11	100,00	99,45	0,11	100,00	
A041	58	103	100,00	0,47	100,00	100,00	0,47	100,00		58	103	100,00	0,86	100,00	100,00	0,86	100,00	
A042	49	103	99,63	0,30	100,00	99,63	0,30	100,00		65	119	99,63	0,74	100,00	94,74	0,74	100,00	
A043	49	103	99,97	0,00	100,00	99,97	0,00	100,00		49	104	99,96	0,00	100,00	99,96	0,00	100,00	
A044	49	103	99,37	0,00	100,00	99,37	0,00	100,00		65	104	99,21	0,00	100,00	97,20	0,00	100,00	
A045	47	103	100,00	0,00	100,00	100,00	0,00	100,00		49	104	100,00	0,00	100,00	100,00	0,00	100,00	
A046	47	104	100,00	0,00	100,00	100,00	0,00	100,00		49	104	100,00	0,00	100,00	100,00	0,00	100,00	
A047	46	103	100,00	0,07	100,00	100,00	0,07	100,00		49	104	99,96	0,07	100,00	99,96	0,07	100,00	
A048	47	103	100,00	0,00	100,00	99,83	0,00	100,00		80	104	100,00	0,00	100,00	99,97	0,00	100,00	
A049	46	103	100,00	0,09	100,00	100,00	0,09	100,00		49	104	100,00	0,10	100,00	100,00	0,10	100,00	
A050	46	103	100,00	0,00	100,00	100,00	0,00	100,00		62	104	100,00	0,00	100,00	98,45	0,00	100,00	
A051	39	103	100,00	0,00	100,00	100,00	0,00	100,00		96	104	100,00	0,00	100,00	99,70	0,00	100,00	
A052	30	103	100,00	0,04	100,00	100,00	0,04	100,00		52	104	100,00	0,05	100,00	100,00	0,05	100,00	
A053	30	103	100,00	0,03	100,00	100,00	0,03	100,00		65	120	100,00	0,04	100,00	95,02	0,04	100,00	
A054	30	104	100,00	0,06	100,00	100,00	0,06	100,00		66	121	100,00	0,06	100,00	96,43	0,06	100,00	
A055	35	104	100,00	0,00	100,00	99,66	0,00	100,00		66	121	100,00	0,00	100,00	100,00	0,00	100,00	
A056	30	103	100,00	0,04	100,00	100,00	0,04	100,00		66	120	100,00	0,04	100,00	100,00	0,04	100,00	
A057	30	103	99,95	0,00	100,00	99,95	0,00	100,00		66	120	100,00	0,00	100,00	100,00	0,00	100,00	
A058	46	103	99,91	0,43	100,00	99,83	0,43	100,00		66	120	99,91	0,09	100,00	99,91	0,09	100,00	
A059	46	103	100,00	0,00	100,00	100,00	0,00	100,00		66	121	100,00	0,00	100,00	100,00	0,00	100,00	
A060	47	103	99,79	0,00	100,00	99,73	0,00	100,00		75	120	99,79	0,00	100,00	99,71	0,00	100,00	

SAMPLES	NÍVEL DE SEGURANÇA: S ₁ =400									NÍVEL DE SEGURANÇA: S ₂ =201								
	RANKING		REAL			BASEADO NO MIN				RANKING		REAL			BASEADO NO MIN			
	MIN	MAX	VP	VN	VNN	VP	VN	VNN		MIN	MAX	VP	VN	VNN	VP	VN	VNN	
A061	47	103	100,00	0,09	100,00	100,00	0,09	100,00		66	120	100,00	0,10	100,00	100,00	0,10	100,00	
A062	46	104	99,82	0,18	100,00	99,82	0,18	100,00		65	120	99,59	0,41	100,00	99,59	0,41	100,00	
A063	46	104	100,00	0,13	100,00	100,00	0,13	100,00		67	120	100,00	0,07	100,00	97,34	0,07	100,00	
A064	46	104	100,00	0,00	100,00	100,00	0,00	100,00		67	121	100,00	0,00	100,00	100,00	0,00	100,00	
A065	46	100	100,00	0,00	100,00	100,00	0,00	100,00		67	120	100,00	0,00	100,00	100,00	0,00	100,00	
A066	46	104	100,00	0,00	100,00	100,00	0,00	100,00		66	120	100,00	0,00	100,00	100,00	0,00	100,00	
A067	46	104	99,95	0,00	99,95	99,95	0,00	99,95		67	121	99,95	0,00	99,95	88,52	0,00	99,95	
A068	46	104	99,95	0,00	100,00	99,95	0,00	100,00		66	120	99,93	0,00	100,00	99,93	0,00	100,00	
A069	38	104	100,00	0,00	100,00	100,00	0,00	100,00		66	121	100,00	0,00	100,00	100,00	0,00	100,00	
A070	46	104	100,00	0,00	100,00	98,60	0,00	100,00		66	121	100,00	0,00	100,00	100,00	0,00	100,00	
A071	46	104	100,00	0,00	100,00	100,00	0,00	100,00		66	121	100,00	0,00	100,00	100,00	0,00	100,00	
A072	46	104	100,00	0,00	100,00	100,00	0,00	100,00		66	121	100,00	0,00	100,00	100,00	0,00	100,00	
A073	63	119	100,00	0,00	100,00	99,37	0,00	100,00		67	120	100,00	0,00	100,00	96,28	0,00	100,00	
A074	46	120	100,00	0,00	100,00	100,00	0,00	100,00		67	121	100,00	0,00	100,00	100,00	0,00	100,00	
A075	46	119	100,00	0,43	100,00	100,00	0,43	100,00		67	120	100,00	0,38	100,00	100,00	0,38	100,00	
A076	46	61	99,92	0,00	100,00	99,92	0,00	100,00		67	120	100,00	0,00	100,00	100,00	0,00	100,00	
A077	62	120	100,00	0,00	100,00	99,80	0,00	100,00		67	121	100,00	0,00	100,00	100,00	0,00	100,00	
A078	46	119	100,00	0,00	100,00	100,00	0,00	100,00		67	120	100,00	0,00	100,00	100,00	0,00	100,00	

Fonte: o autor.

APÊNDICE H – Demonstração de dominância iterada na amostra A_{002} do Segundo Experimento.

Sem iterações

Amostra A_{002}

Jogador azul = $s_1=400$

Jogador vermelho = $s_2=201$

Quadro 3 – Processo de dominância iterada para a amostra A_{002} – SEM ITERAÇÕES.

		Jogador: $S_2=201$																							
	Estratégias	104-200				103				102				101				100-29				28-25			
		VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN
Jogador: $S_1=400$	104-200	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	103	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	102	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	101	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	100-29	100	100	0,1433	55,0933	100	100	0,1433	54,8063	100	100	0,1433	54,8063	100	100	0,1433	54,6628	100	100	0,1433	0,1435	100	73,3404	0,1433	0,1435
	28-25	100	100	0,1433	55,0933	100	100	0,1433	54,8063	100	100	0,1433	54,8063	100	100	0,1433	54,6628	100	100	0,1433	0,1435	100	73,3404	0,1433	0,1435
	24	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	23	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	22	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	21	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	20	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	19	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	18	98,7252	100	0,1433	55,0933	98,7252	100	0,1433	54,8063	98,7252	100	0,1433	54,8063	98,7252	100	0,1433	54,6628	98,7252	100	0,1433	0,1435	98,7252	73,3404	0,1433	0,1435

Fonte: o autor.

1ª iteração: VP do jogador azul: estratégia dominante = 100% → descarte das linhas 24 a 18

Amostra A_{002}

Jogador azul = $s_1=400$

Jogador vermelho = $s_2=201$

Quadro 4 – Processo de dominância iterada para a amostra A_{002} – 1ª ITERAÇÃO.

		Jogador: S ₂ =201																							
Estratégias	104-200				103				102				101				100-29				28-25				
	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	
Jogador: S ₁ =400	104-200	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	103	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	102	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	101	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	100-29	100	100	0,1433	55,0933	100	100	0,1433	54,8063	100	100	0,1433	54,8063	100	100	0,1433	54,6628	100	100	0,1433	0,1435	100	73,3404	0,1433	0,1435
	28-25	100	100	0,1433	55,0933	100	100	0,1433	54,8063	100	100	0,1433	54,8063	100	100	0,1433	54,6628	100	100	0,1433	0,1435	100	73,3404	0,1433	0,1435
	24	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	23	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	22	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	21	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	20	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	19	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
18	98,7252	100	0,1433	55,0933	98,7252	100	0,1433	54,8063	98,7252	100	0,1433	54,8063	98,7252	100	0,1433	54,6628	98,7252	100	0,1433	0,1435	98,7252	73,3404	0,1433	0,1435	

Fonte: o autor.

2ª iteração: VN do jogador vermelho: estratégia dominante = 0,1435% → descarte das colunas 104-200 a 101

Amostra A_{002}

Jogador azul = $s_1=400$

Jogador vermelho = $s_2=201$

Quadro 5 – Processo de dominância iterada para a amostra A_{002} – 2ª ITERAÇÃO.

		Jogador: $S_2=201$																							
	Estratégias	104-200				103				102				101				100-29				28-25			
		VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN
Jogador: $S_1=400$	104-200	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	103	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	102	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	101	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	100-29	100	100	0,1433	55,0933	100	100	0,1433	54,8063	100	100	0,1433	54,8063	100	100	0,1433	54,6628	100	100	0,1433	0,1435	100	73,3404	0,1433	0,1435
	28-25	100	100	0,1433	55,0933	100	100	0,1433	54,8063	100	100	0,1433	54,8063	100	100	0,1433	54,6628	100	100	0,1433	0,1435	100	73,3404	0,1433	0,1435
	24	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	23	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	22	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	21	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	20	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	19	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	18	98,7252	100	0,1433	55,0933	98,7252	100	0,1433	54,8063	98,7252	100	0,1433	54,8063	98,7252	100	0,1433	54,6628	98,7252	100	0,1433	0,1435	98,7252	73,3404	0,1433	0,1435

Fonte: o autor.

3ª iteração: VN do jogador azul: estratégia dominante = 0,1433% → descarte das linhas 104-200 a 101

Amostra A_{002}

Jogador azul = $s_1=400$

Jogador vermelho = $s_2=201$

Quadro 6 – Processo de dominância iterada para a amostra A_{002} – 3ª ITERAÇÃO.

		Jogador: S ₂ =201																							
	Estratégias	104-200				103				102				101				100-29				28-25			
		VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN
Jogador: S ₁ =400	104-200	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	103	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	102	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	101	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	100-29	100	100	0,1433	55,0933	100	100	0,1433	54,8063	100	100	0,1433	54,8063	100	100	0,1433	54,6628	100	100	0,1433	0,1435	100	73,3404	0,1433	0,1435
	28-25	100	100	0,1433	55,0933	100	100	0,1433	54,8063	100	100	0,1433	54,8063	100	100	0,1433	54,6628	100	100	0,1433	0,1435	100	73,3404	0,1433	0,1435
	24	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	23	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	22	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	21	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	20	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	19	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
18	98,7252	100	0,1433	55,0933	98,7252	100	0,1433	54,8063	98,7252	100	0,1433	54,8063	98,7252	100	0,1433	54,6628	98,7252	100	0,1433	0,1435	98,7252	73,3404	0,1433	0,1435	

Fonte: o autor.

4ª iteração: VP do jogador vermelho: estratégia dominante = 100% → descarte da coluna 28-25

Amostra A_{002}

Jogador azul = $s_1=400$

Jogador vermelho = $s_2=201$

Quadro 7 – Processo de dominância iterada para a amostra A_{002} – 4ª ITERAÇÃO.

		Jogador: S ₂ =201																							
Estratégias		104-200				103				102				101				100-29				28-25			
		VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN
Jogador: S ₁ =400	104-200	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	103	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	102	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	101	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	100-29	100	100	0,1433	55,0933	100	100	0,1433	54,8063	100	100	0,1433	54,8063	100	100	0,1433	54,6628	100	100	0,1433	0,1435	100	73,3404	0,1433	0,1435
	28-25	100	100	0,1433	55,0933	100	100	0,1433	54,8063	100	100	0,1433	54,8063	100	100	0,1433	54,6628	100	100	0,1433	0,1435	100	73,3404	0,1433	0,1435
	24	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	23	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	22	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	21	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	20	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	19	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
18	98,7252	100	0,1433	55,0933	98,7252	100	0,1433	54,8063	98,7252	100	0,1433	54,8063	98,7252	100	0,1433	54,6628	98,7252	100	0,1433	0,1435	98,7252	73,3404	0,1433	0,1435	

Fonte: o autor.

Resultado final

Amostra A_{002}

Estratégia dominante do jogador azul: 100-25

Estratégia dominante do jogador vermelho: 100-29

Quadro 8 – Processo de dominância iterada para a amostra A_{002} – RESULTADO FINAL.

		Jogador: S ₂ =201																							
Estratégias	104-200				103				102				101				100-29				28-25				
	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	VP	VP	VN	VN	
Jogador: S ₁ =400	104-200	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	103	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	102	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	101	100	100	55,0143	55,0933	100	100	55,0143	54,8063	100	100	55,0143	54,8063	100	100	55,0143	54,6628	100	100	55,0143	0,1435	100	73,3404	55,0143	0,1435
	100-29	100	100	0,1433	55,0933	100	100	0,1433	54,8063	100	100	0,1433	54,8063	100	100	0,1433	54,6628	100	100	0,1433	0,1435	100	73,3404	0,1433	0,1435
	28-25	100	100	0,1433	55,0933	100	100	0,1433	54,8063	100	100	0,1433	54,8063	100	100	0,1433	54,6628	100	100	0,1433	0,1435	100	73,3404	0,1433	0,1435
	24	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	23	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	22	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	21	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	20	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
	19	99,5714	100	0,1433	55,0933	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,8063	99,5714	100	0,1433	54,6628	99,5714	100	0,1433	0,1435	99,5714	73,3404	0,1433	0,1435
18	98,7252	100	0,1433	55,0933	98,7252	100	0,1433	54,8063	98,7252	100	0,1433	54,8063	98,7252	100	0,1433	54,6628	98,7252	100	0,1433	0,1435	98,7252	73,3404	0,1433	0,1435	

Fonte: o autor.