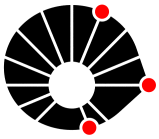
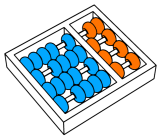
,





Anais do XIV Workshop de Teses, Dissertações e Trabalhos de Iniciação Científica do IC Unicamp

Technical Report - IC-19-09 - Relatório Técnico
November - 2019 - Novembro

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE COMPUTAÇÃO

Anais do XIV Workshop de Teses, Dissertações e Trabalhos de Iniciação Científica
Instituto de Computação - Unicamp

Apresentação

Este relatório técnico contém os resumos de 13 trabalhos cujos artigos foram autorizados a serem publicados no XIV Workshop de Teses, Dissertações e Trabalhos de Iniciação Científica (WTD)[1], do Instituto de Computação (IC) da Universidade Estadual de Campinas (Unicamp), edição 2019.

O XIV Workshop ocorreu entre os dias 27 e 28 de Novembro de 2019 e contou com cerca de 137 participantes, entre ouvintes, apresentadores de trabalhos e organizadores. Na ocasião houve 42 apresentações orais, 51 pôsteres, 32 lightning talk e 8 demonstrações. Aos alunos foi dada a possibilidade de escolher a forma de apresentação (oral, pôster ou lightning talk), bem como escolher se desejasse publicar ou não seu trabalho nos anais do evento. A publicação dos resumos, sob forma de relatório técnico, tem por objetivo divulgar os trabalhos em andamento e concluídos e registrar, de forma sucinta, o estado da arte da pesquisa do Instituto de Computação no ano de 2019.

Neste ano ocorreram duas palestras. A primeira, intitulada "Os Desafios do Processamento de Língua Portuguesa", foi proferida pelo Prof. Evandro Eduardo Seron Ruiz, Professor Associado do Departamento de Computação e Matemática do Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da USP, e atua como orientador no Programa de Pós-graduação em Computação Aplicada. A segunda, intitulada "O Jeito Hacker de Entender Ciência", foi proferida pelo Professor André Santanchè, Professor Associado do Departamento de Sistemas de Informação do Instituto de Computação da Unicamp. Também foi oferecido o minicurso intitulado "Introdução à Escrita Científica", ministrado pelo Professor Leonardo Montecchi, Professor Doutor do Departamento de Sistemas de Informação do Instituto de Computação da Unicamp, além de coordenador do Laboratório de Sistemas de Informação (LIS) e diretor do Museu Exploratório de Ciências da Unicamp.

Agradecemos aos alunos que participaram do evento, em particular àqueles que se dispuseram a apresentar seus trabalhos, seja oralmente ou na forma demonstrações e pôsteres, bem como aos orientadores que os incentivaram a fazê-lo. Agradecemos, também, aos professores, alunos de mestrado, doutorado e pós doutorado do IC que compuseram as bancas de avaliação dos trabalhos e aos colaboradores da secretaria que apoiaram a organização do evento. Agradecemos ao Professor Doutor Anderson de Rezende Rocha, diretor do IC, e a Professora Titular Cecília Mary Fischer Rubira, coordenadora da Pós-Graduação, pelo forte incentivo, apoio e patrocínio ao evento.

Agradecemos às empresas NeuralMind, ProFusion, Griaule, GoBots, Instituto Eldorado e Dairiki, que engrandeceram o evento como patrocinadoras.

Finalmente, agradecemos imensamente aos alunos do programa de Pós-Graduação do IC que efetivamente organizaram o evento e que são coeditores deste relatório – André Gomes Regino,

[1] https://www.ic.unicamp.br/wtd/2019/

Antonio Carlos Theophilo Costa Júnior, Eliane Zambon Victorelli Dias, Elisangela Silva, Enio de Jesus Pontes Monteiro, Francisco José Nardi Filho, Gustavo Caetano Borges, Gustavo Eloi de Paula Rodrigues, Heitor Soares Mattosinho, Helena de Almeida Maia, Letícia Bomfim, Luis Fillype Furtado Leite Fernandes, Marcos Felipe de Menezes Mota, Sheila Venero, Yusseli Lizeth Méndez Mendoza. A eles dedicamos o XIV Workshop de Teses, Dissertações e Trabalhos de Iniciação Científica do Instituto de Computação da Unicamp.

Prof. Julio Cesar Dos Reis
Profa. Esther Luna Colombini
Profa. Juliana Freitag Borin
Coordenadores do XIV WTD
Professores do Instituto de Computação - Unicamp

Sumário

1 Programação

Apresentamos a programação e algumas estatísticas do XIV Workshop de Teses, Dissertações e Trabalhos de Iniciação Científica (WTD) do Instituto de Computação (IC) da Unicamp.

Na edição deste ano tivemos a apresentação do minicurso intitulado "Introdução à escrita científica", que foi ministrado pelo Professor Doutor Leonardo Montecchi da Universidade Estadual de Campinas, Instituto de Computação. Contamos também com 3 sessões de Lightning Talks, a apresentação da empresa Neuralmind e 2 sessões de apresentação de poster no primeiro dia de evento.

Dia 1 (27/11/2019)				
Sala 351	*Minicurso: Introdução à escrita científica - Prof. Dr. Leonardo Montecchi - IC/Unicamp*			
08:30-10:00	*Empresa: Neuralmind*			
10:00-11:00	**Sessão 1 de Posters/Demos - Coffee break [Hall do IC 3]**			
Sala 351	**Aluno(a)**	**Trabalho**	**Orientador(a)**	**Área**
11:00-11:30	Juan Felipe Hernández Albarrac	Representação Isolada de Movimento para Transferência de Dinâmica entre	Adín Ramírez Rivera	**Sistemas de Informação**
11:30-12:00	Marcos Felipe de Menezes Mota	Modelagem Causal de Dados Clínicos para Treinamento Médico	André Santanchè	
12:00-12:30	Jacqueline Midlej do Espírito Sa	Exploring Semantics in Clinical Data Interoperability	Claudia Bauzer Medeir	
14:00-14:30	***Sessão 1 de Lightning Talk***			**Sistemas de Informação**
14:30-15:00	Anderson Rossanez	Generating Knowledge Graphs from Scientific Literature of Degenerative Dise	Julio Cesar dos Reis	
15:00-15:30	Eliane Zambon Victorelli	Participatory Evaluation of Human-Data Interaction Design Guidelines	Julio Cesar dos Reis	
15:30-16:30	**Sessão 2 de Posters/Demos - Coffee break [Hall do IC 3]**			
16:30-17:00	Eva Maia Malta	Exploring the Cost-benefit of AWS EC2 GPU Instances for Deep Learning Ap	Edson Borin	**Sistemas de Computação**
17:00-17:30	Gustavo Ciotto Pinton	CMP, ZO-CRS e OCT: aceleração de métodos de empilhamento de dados sís	Edson Borin	
17:30-18:00	Hayato Fujii	Fast AES implementation using ARMv8 ASIMD without Cryptography Extens	Julio López	

Figura 1: Programação do Primeiro Dia

ia 1 (27/11/2019)				
Sala 351				
08:30-10:00				
10:00-11:00				
Sala 351	**Chair**	**Banca Prof. 1**	**Banca Prof. 2**	**Banca Aluno**
11:00-11:30	Andressa Cristina dos Santos	Alexandre Mello	Julio Reis	Anderson Carlos Sousa e Santos
11:30-12:00				
12:00-12:30				
14:00-14:30	Stephane de Freitas Schwarz	Aurea Soriano	Ricardo Caceffo	Juan Felipe Hernández Albarracín
14:30-15:00				
15:00-15:30				
15:30-16:30				
16:30-17:00	Marcos Vinicius			Guilherme Andrino de Azevedo
17:00-17:30				
17:30-18:00				

Figura 2: Programação do Primeiro Dia

Auditório 1	Aluno(a)	Trabalho	Orientador(a)	Área
14:00-14:30	***Sessão 2 de Lightning Talk***			**Sistemas de Computação**
14:30-15:00	Leandro Negri Zanotto	High Performance Collision Cross Section (HPCCS) - Utilização de Técnicas	Guido Costa Souza de	
15:00-15:30	Jeferson R. Brunetta	Selecting Efficient Cloud Resources for HPC Workloads	Edson Borin	
15:30-16:30	**Sessão 2 de Posters/Demos - Coffee break [Hall do IC 3]**			
16:30-17:00	Fabíola Martins Campos de Oliv	Partitioning Convolutional Neural Networks to Minimize Communication and M	Edson Borin	**Sistemas de Computação e Teoria da Computação**
17:00-17:30	Yulle G. F. Borges	Precificação Inteligente de Energia para Gerenciamento Pelo Lado-da-Dema	Rafael Schouery	

Sala 352	Aluno(a)	Trabalho	Orientador(a)	Área
14:00-14:30	Leonardo Alves de Melo	Análise de Dados e Geolocalização de Veículos para Transporte Público/	Juliana Freitag Borin	**Sistemas da Computação e Teoria de Computação**
14:30-15:00	Italos Estilon da Silva de Souza	Stability Analysis of Hedonic Games	Rafael Schouery	
15:00-15:30	Matheus Jun Ota	Integer Programming Approaches to Balanced Connected k-Partition	Flávio Keidi Miyazawa	
15:30-16:30	**Sessão 2 de Posters/Demos - Coffee break [Hall do IC 3]**			
16:30-17:00	João Paulo Francisco da Silva, L	Criação e Aplicação de uma ontologia no contexto do portal de transparência	Julio Cesar dos Reis	**Apresentação Oral de Disciplina da Grad/Pós**
	Daniel de Souza Medina	Sistema de Recomendação de Filmes Baseado em Web-Semântica	Julio Cesar dos Reis	
17:00-17:30	Gabriel Capiteli Bertocco	Go To Goal and Avoid Obstacle applying Reinforcement Learning	Esther Colombini	
	Iury Cleveston	Implementing SLAM in VREP and extracting a 3D map	Esther Colombini	
17:30-18:00	Ciro Cavani	Aplicação de Reinforcement Learning em Robótica Móvel	Esther Colombini	
	Elisangela Silva dos Santos	Usar o algoritmo Q-learning para fazer o robô Pionner andar no ambiente ev	Esther Colombini	

Figura 3: Programação do Primeiro Dia

Auditório 1	Chair	Banca Prof. 1	Banca Prof. 2	Banca Aluno
14:00-14:30	Wellington Lobato		Maycon Peixoto	Giovane de Morais
14:30-15:00				
15:00-15:30				
15:30-16:30				
16:30-17:00	Joahannes Costa	Luiz Bittencourt	Fábio Usberti	Leonardo Alves de Melo
17:00-17:30				

Sala 352	Chair	Banca Prof. 1	Banca Prof. 2	Banca Aluno
14:00-14:30	Marcos Felipe de Menezes Mota	Lehilton	Vanessa Maike	
14:30-15:00				
15:00-15:30				
15:30-16:30				
16:30-17:00	Eliana Moreira			
17:00-17:30				
17:30-18:00				

Figura 4: Programação do Primeiro Dia

Sala 353	Aluno(a)	Trabalho	Orientador(a)	Área
11:00-11:30	José Dorivaldo Nascimento Sou	Environmental Labeling	Esther Colombini	Apresentação Oral de Disciplina da Grad/Pós
	Carlos Victor Dantas Araujo	GoToGol	Esther Colombini	
11:30-12:00	Felipe de Alcântara Monteiro	Simulação de Odometria e SLAM Visual e Inercial em ambiente Vrep	Esther Colombini	
	Junior Cupe Casquina	Mobile Robots Coordinated to form a soccer formation	Esther Colombini	
12:00-12:30	Gabriel Bianchin de Oliveira	Comparação entre odometria e odometria visual na localização de robôs móv	Esther Colombini	
	Gustavo Frederico Temple Pedr	DSL para geração de scaffolding Spring com RabbitMQ	Leonardo Montecchi	
14:00-14:30	Alexsandro Oliveira Alexandrino	Path Planning Optimization with Metaheuristics	Esther Colombini	Sistemas de Computação
14:30-15:00	Eduardo de Souza Gama	Futura Geração de Distribuição de Video Multiníveis para Cidades Inteligente	Luiz Fernando Bittenco	
15:00-15:30	***Sessão 3 de Lightning Talk***			
15:30-16:30	**Sessão 2 de Posters/Demos - Coffee break [Hall do IC 3]**			
16:30-17:00	José Ernesto Stelzer Monar	Análise de tráfego em internet das coisas usando virtualização	Luiz Fernando Bittenco	Sistemas de Computação
17:00-17:30	Maria Júlia Berriel de Sousa	Software update Over the Air for IoT devices	Juliana Freitag Borin	
17:30-18:00	Allan Mariano de Souza	Reliable Route Planning based Future Urban Dynamics	Leandro Aparecido Vill	

Figura 5: Programação do Primeiro Dia

Sala 353	Chair	Banca Prof. 1		
11:00-11:30		Esther Colombini		
11:30-12:00				
12:00-12:30				
		Banca Prof. 1	**Banca Prof. 2**	**Banca Aluno**
14:00-14:30		Esther Colombini	Flávio	André Gomes Regino
14:30-15:00				
15:00-15:30				
15:30-16:30				
16:30-17:00	Ademar Takeo Akabane	Breno de França	Eduardo Xavier	Francisco José Nardi Filho
17:00-17:30				
17:30-18:00				

Figura 6: Programação do Primeiro Dia

No dia 28 de novembro contamos com a apresentação de duas palestras, como exibido na tabela 7. A primeira palestra, intitulada “O Jeito Hacker de Entender Ciência”, foi realizada pelo Professor Doutor André Santanchè da Universidade Estadual de Campinas, Instituto de Computação. A segunda palestra, intitulada ”Os desafios do Processamento de Língua Portuguesa”, foi realizada pelo Professor Doutor Evandro Eduardo Seron Ruiz da FFCLRP, Universidade de São Paulo. Contamos também com 1 sessões de Lightning Talks, apresentações das empresas GoBots e Griaule e 2 sessões de apresentação de poster no segundo dia de evento.

O evento foi finalizado com a sessão de entrega de prêmios aos melhores trabalhos apresentados.

Dia 2 (28/11/2019)				
Auditório 1	*Palestra: O Jeito Hacker de Entender Ciência - Prof. Dr. André Santanchè - IC/Unicamp*			
08:30-10:00				
10:00-11:00	**Sessão 3 de Posters/Demos - Coffee break [Hall do IC 3]**			
Auditório 1	**Aluno(a)**	**Trabalho**	**Orientador(a)**	**Área**
11:00-11:30	Eldrey Seolin Galindo	Image Super-Resolution Improved by Edge Information	Hélio Pedrini	**Sistemas de informação**
11:30-12:00	Azael de Melo e Sousa	Asbestos-related Pleural Plaques Classification with Random Convolutior	Alexandre Xavier Falcão	
12:00-12:30	Amaury Mausbach Filho	Aprendizado profundo no projeto de anticorpos a partir de epítopos	João Meidanis	
12:30-13:00	Jose Luis Flores Campana	A comparative study of text localization and recognition methods	Ricardo da Silva Torres	
Auditório 1	*Palestra: Os desafios do Processamento de Língua Portuguesa - Prof. Dr. Evandro Eduardo Seron Ruiz - FFCLRP /USP*			
14:00-15:30	*Empresa: GoBots - Como aplicamos Pesquisa e Desenvolvimento em uma Startup.*			
15:30-16:30	**Sessão 4 de Posters/Demos - Coffee break [Hall do IC 3]**			
16:30-17:00	Klaus Rollmann	Comparing 4D Seismic and Reservoir Simulation Models through Deep L	Anderson Rocha	**Sistemas de informação**
17:00-17:30	Thiago Resek	Inferência de Localização Geográfica de Imagens	Anderson Rocha	
17:30-18:00	Bruno Malveira Peixoto	Detecção de Violência em Vídeos	Anderson Rocha	
18:00-18:30	***Sessão de Encerramento / Premiações***			

Figura 7: Programação do Segundo Dia

Dia 2 (28/11/2019)				
Auditório 1				
08:30-10:00				
10:00-11:00				
Auditório 1	**Chair**	**Banca Prof. 1**	**Banca Prof. 2**	**Banca Aluno**
11:00-11:30	Manuel Cordova	Marcos Cirne	Diego Addan	Fabrício Matheus Gonçalves
11:30-12:00				
12:00-12:30				
12:30-13:00				
Auditório 1				
14:00-15:30				
15:30-16:30	**Chair**	**Banca Prof. 1**	**Banca Prof. 2**	**Banca Aluno**
16:30-17:00	Danielle Dias	Edson Bollis	Manuel Cordova	Marcos Felipe de Menezes Mota
17:00-17:30				
17:30-18:00				
18:00-18:30				

Figura 8: Programação do Segundo Dia

Sala 353	Aluno(a)	Trabalho	Orientador(a)	Área
11:00-11:30	Túlio Brandão Soares Martins	Verificação de Consistência na Evolução da DBpedia	Julio Cesar dos Reis	IC/PFG
11:30-12:00	Raissa Cavalcante Correia	Revisão de Técnicas de Ciência de Dados para Sistemas de Recomenda	Esther Luna Colombini	
12:00-12:30				
14:00-14:30	*Empresa: Griaule - Big Data aplicado a soluções multibiométricas*			
14:30-15:00	***Sessão 4 de Lightning Talk***			
15:00-15:30				
15:30-16:30	**Sessão 4 de Posters/Demos - Coffee break [Hall do IC 3]**			
16:30-17:00	Lucas dos Santos Ramos	Geração de grafos de conhecimento a partir de textos não estruturados e	Julio Cesar dos Reis	IC/PFG
17:00-17:30	Willian Tadeu Beltrão / Raphael Pontes San	Análise de fluxo e fila dos restaurantes universitários	Juliana Freitag Borin	

Figura 9: Programação do Segundo Dia

Sala 353	Chair	Banca Prof. 1	Banca Prof. 2	Banca Aluno
11:00-11:30	Priscila Moraes	Julio Lopez		Gustavo Caetano Borges
11:30-12:00				
12:00-12:30				
14:00:14:30				
14:30:15:00	Chair	Banca Prof. 1	Banca Prof. 2	Banca Aluno
15:00-15:30	Priscila Moraes	Guilherme		Denis Contini
15:30-16:30				
16:30-17:00				
17:00-17:30				

Figura 10: Programação do Segundo Dia

A novidade nesta edição do WTD foi a apresentação de trabalhos em um curto período de tempo, os Lightning Talks. Os apresentadores tinham 3 minutos para descrever, de forma sucinta, os trabalhos em questão. Figuras 11 e 12 listam os trabalhos apresentados nessa modalidade.

Sessão 1 de Lightning Talk - dia 27 [14:00-14:30 na sala 351]			
Sala 351	**Aluno(a)**	**Trabalho**	**Orientador(a)**
14:00-14:03	Ramon Santos Nepomuceno	Integrating Multi-FPGA clusters into OpenMP Task Parallelism	Guido Costa Souza de Araújo
14:03-14:06	Lucas de Magalhães Araujo	Análise de Redes Neurais Convolucionais para Super-Resolução de Imagens Sísmicas	Edson Borin
14:06-14:09	Brenner Humberto Ojeda Rios	Survey of dynamic vehicle routing problems	Eduardo Candido Xavier
14:09-14:12	Gustavo Ciotto Pinton	CMP, ZO-CRS e OCT: aceleração de métodos de empilhamento de dados sísmicos na nuvem com CUD	Edson Borin
14:12-14:15	Oscar Jaime Ciceri Coral	Mecanismos para Gerenciamento de Banda Passante em Redes Ópticas Passivas Ethernet.	Nelson Luis Saldanha da Fonseca
14:15-14:18	Leonardo Yvens Schwarzstein	A Budget Balanced and Strategy-proof Auction for Multi-passenger Ridesharing	Rafael C. S. Schouery
14:18-14:21	Vinicius Loti de Lima	Improvements on the Arc-Flow Formulation for the Bin Packing Problem	Flávio Keidi Miyazawa
14:21-14:24	Ademar Takeo Akabane	Collaborative and Infrastructure-less Vehicular Traffic Rerouting for Intelligent Transportation Systems	Leandro Aparecido Villas
Sessão 2 de Lightning Talk - dia 27 [14:00-14:30 no auditório 1]			
Auditório 1	**Aluno(a)**	**Trabalho**	**Orientador(a)**
14:00-14:03	Allan Sapucaia Barboza	O Problema da Partição Convexa Mínima	Cid Carvalho de Souza
14:03-14:06	Luis Gustavo Lorgus Decker	Localização de textos em imagens de cena utilizando redes neurais convolucionais leves	Ricardo da Silva Torres
14:06-14:09	Joahannes Bruno Dias da Costa	Computação em Nuvem Veicular para Gerenciamento de Informação em Sistemas de Transporte Inteligi	Leandro Aparecido Villas
14:09-14:12	Guilherme Mendeleh Perrotta	Distributed stencil computations	Guido Araújo
14:12-14:15	Jônatas Trabuco Belotti	Problema de Reconfiguração de Redes Estocástico em 2 Estágios	Fábio Luiz Usberti
14:15-14:18	Felipe de Carvalho Pereira	Algoritmos Exatos e Heurísticos para o Problema Perfect Awareness	Pedro Jussieu de Rezende
14:18-14:21	William Villota Jácome	Request Admission Control for 5G Network Slicing	Nelson L. S. da Fonseca

Figura 11: Sessões 1 e 2 de Lightning Talks

Sessão 3 de Lightning Talk - dia 27 [15:00-15:30 na sala 353]			
Sala 353	**Aluno(a)**	**Trabalho**	**Orientador(a)**
15:00-15:03	Brenner Humberto Ojeda Rios	Survey of dynamic vehicle routing problems	Eduardo Candido Xavier
15:03-15:06	Daniela Maria Casas Velasco	Roteamento em Redes Definidas por Software com Aprendizado por Reforço	Nelson Luis Saldanha da Fonseca
15:06-15:09	Juliane Regina de Oliveira	Mecanismos para melhora da qualidade dos dados de aplicações de Internet das Coisas	Lucas Francisco Wanner
15:09-15:12	Giuliano Roberto Pinheiro	Sincronização de Vídeo para Reconstrução de Eventos	Anderson Rocha
15:12-15:15	Natanael Ramos	Heurísticas para Problemas de Poligonização de Área Ótima	Cid C. de Souza; Pedro J. de Reze
15:15-15:18	Beatriz Martins de Carvalho	Arquivos de Cabeçalho no Kernel Linux: uma análise automática da observância de boas práticas visanc	Islene Garcia
15:18-15:21	Frances Albert Santos	Urban Perception Extraction from Texts Shared onSocial Media: Framework and Applications	Leandro Aparecido Villas
Sessão 4 de Lightning Talk - dia 28 [14:30-15:00 na sala 353]			
Auditório 1	**Aluno(a)**	**Trabalho**	**Orienatdor(a)**
14:30-14:33	Julio Cesar Mendoza Bobadilla	Self-supervised depth and motion estimation in monocular videos	Helio Pedrini
14:33-15:36	Enoque Alves de Castro Neto	Algoritmos exatos e heurísticos para roteamentos de veículos com sincronização	Rafael C. S. Schouery
15:36-15:39	Juliane Regina de Oliveira	Criação de um modelo para a injeção automática de ruídos em dados IoT	Lucas Francisco Wanner
14:39-14:42	Vitoria Dias Moreira Pinho	OmpTracing: Profiling tools for OpenMP	Guido Araújo
14:42-14:45	Wellington Viana Lobato Junior	Nuvem Veicular Dinâmica para Suporte de Aplicações para Veículos Autônomos	Leandro Aparecido Villas
14:45-14:48	Jonathas Evangelista da Silveira	Detecção e correção de erros para dispositivos da Internet das Coisas usando processamento associativ	Lucas Wanner
14:48-15:51	Francisco Jhonatas Melo da Silva	Selfish Behaviour in Transportation Systems	Rafael C. S. Schouery
14:51-14:54	Matheus Ferraroni Sanches	Algoritmos Genéticos para Alocação de RSU	Leandro Aparecido Villas

Figura 12: Sessões 3 e 4 de Lightning Talks

2 Estatísticas

Apresentamos as estatísticas colhidas em relação ao teor dos trabalhos apresentados durante o evento. A Figura 13 mostra as áreas de concentração dos trabalhos. A Figura 14 mostra a divisão dos trabalhos dentre os 4 tipos. A Figura 15 mostra a quantidade de alunos de graduação, mestrado e doutorado que apresentaram seus trabalhos.

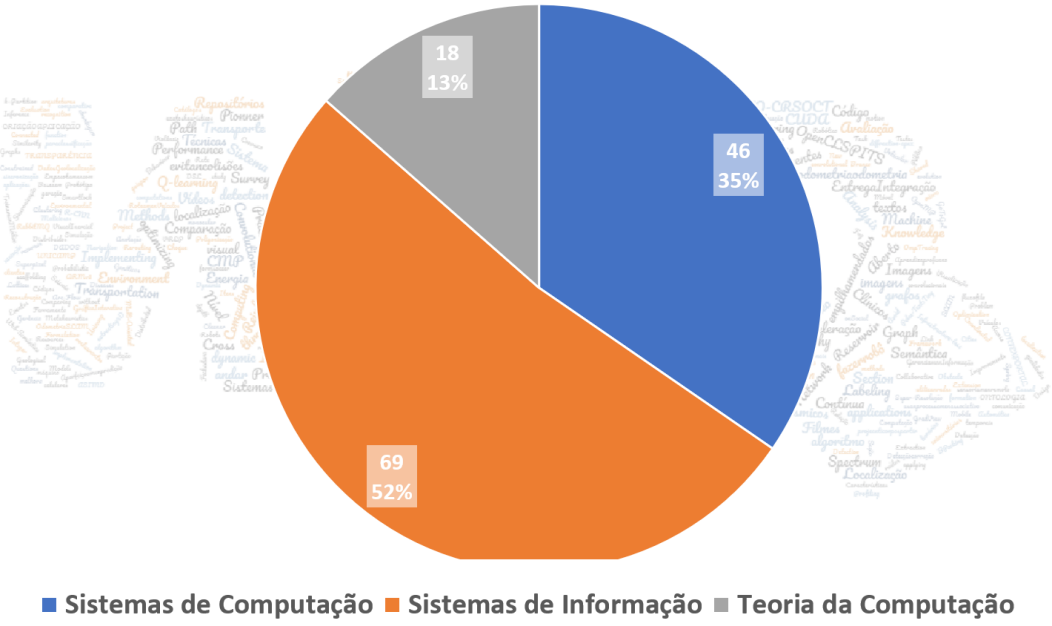


Figura 13: Áreas de Concentração

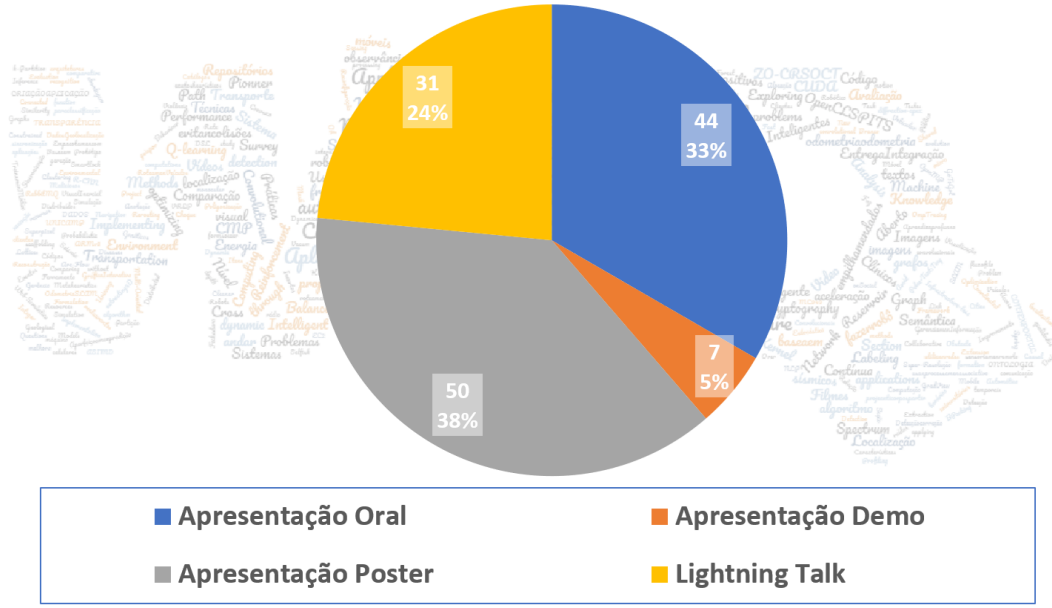


Figura 14: Tipos de Apresentações

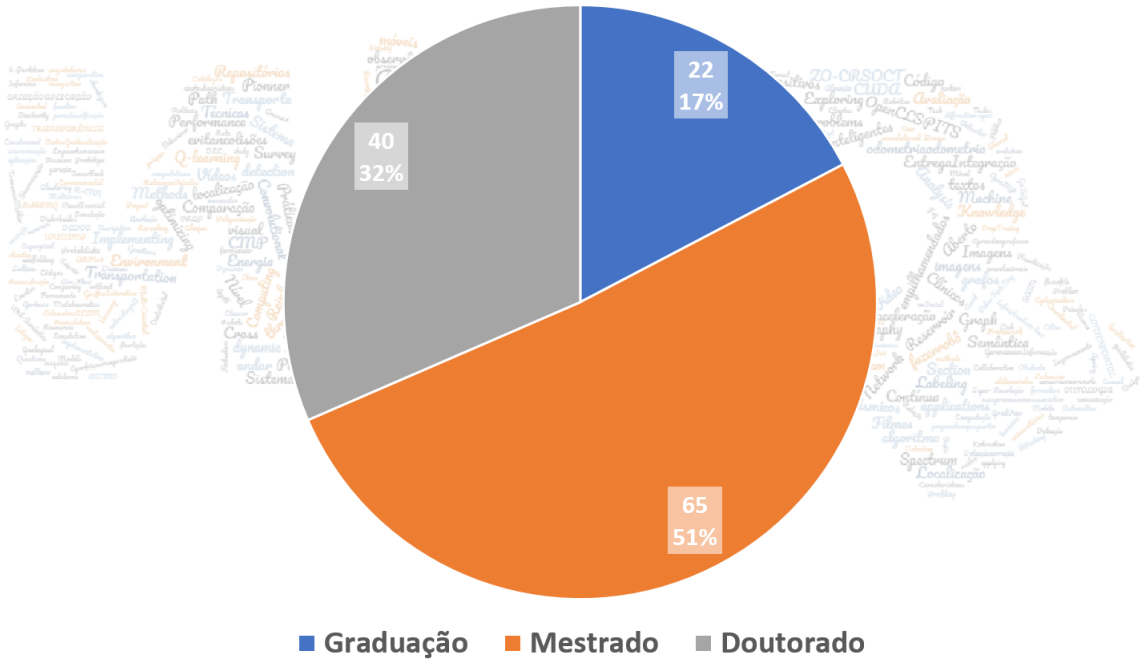


Figura 15: Titulação dos Alunos

A Figura 16 apresenta a nuvem com termos mais frequentes dentre os trabalhos. As Figuras 17, 18 e 19 mostram, respectivamente, a nuvem com os termos mais frequentes nas áreas de Sistemas de Informação, Sistemas de Computação e Teoria.

Figura 16: Termos Mais Frequentes

Figura 17: Termos Mais Frequentes - Sistemas de Informação

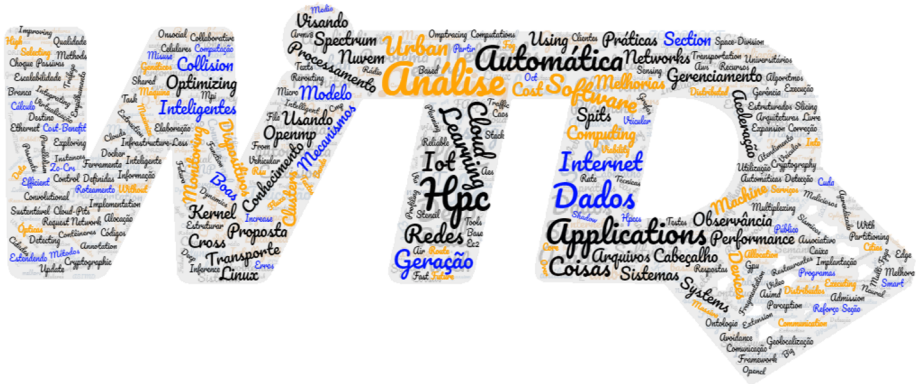


Figura 18: Termos Mais Frequentes - Sistemas de Computação

Figura 19: Termos Mais Frequentes - Teoria

3 Resumos Estendidos

High Performance Collision Cross Section (HPCCS) HPC Techniques to Accelerate the Collision Cross Section Calculation

Leandro Zanotto[1], Gabriel Heerdt[2], Guido Araújo[1]

[1]Institute of Computing – University of Campinas

[2]Department of Chemistry – University of Minas Gerais

leandro.zanotto@gmail.com, gheerdt@qui.ufmg, guido@ic.unicamp.br

Abstract. *Ion Mobility coupled to Mass Spectrometry technique (IM-MS) have been used since 2003 for research and analysis laboratories, when they were commercially introduced. It has been used as a tool for molecular separation, to obtain structural information for molecular ions. The interpretation of the resulting data is still a challenge, depending on collision cross section (CCS) calculation against a buffer gas. This work, presents a new software, High Performance Collision Cross Section - HPCCS, which is based on the trajectory method, using High Performance Computing techniques like parallelization, vectorization and optimization. By using HPCCS now calculate the CCS efficiently, from small organic molecules to protein complexes with a larger number of atoms. The results presented in this work when comparing to the state of the art software (MOBCAL), show an average speedup of 78 times on a cluster node with 24 cores and 48 threads.*

Resumo. *A técnica de Mobilidade Iônica junto com a Espectrometria de Massa (IM-MS) tem sido utilizada desde 2003 por laboratórios de pesquisa e análises. Ela é usada como uma ferramenta de separação molecular, para obter informação estrutural de íons moleculares. A interpretação dos dados obtidos ainda é um desafio, dependendo dos cálculos da seção de choque transversal (CCS) contra um gás de arraste. Este trabalho, apresenta um novo software, High Performance Collision Cross Section - HPCCS, que, baseado no método de trajetória, realiza os cálculos de CCS utilizando técnicas de High Performance Computing como paralelização, vetorização e otimização. Agora é possível calcular o CCS de maneira eficiente, desde para pequenas moléculas orgânicas até proteínas complexas com um número maior de átomos. Os resultados mostraram que, comparados com o software usado atualmente (MOBCAL), houve um ganho em média de 78 vezes em um nó de um cluster com 24 cores e 48 threads.*

1. Introduction

Mass spectrometry (MS) is an indispensable analytical tool in many related fields of science, like medicine. It is employed for example to explore single cells or objects from outer space, as a way to elucidate unknown substances. It has also been commonly used in forensics, quality control of drugs, foods and polymers analysis [Gross 2011].

An interest in ion mobility spectrometry (IMS) is increasing, when ion mobility is coupled with mass spectrometry. It presents an effective means of separating gaseous ions working as a chromatography technique. IMS can separate isobaric ions of different charge state, resulting from their distinct speed of propagation along the electric field of the ion mobility tube, or distinguish isobars of the same charge state by their steric properties [Gross 2011].

The rotationally-averaged collision cross-section represents the effective area for the interaction between an individual ion and the neutral gas, through which it is traveling. CCS is an important distinguishing characteristic of an ion, related to its chemical structure and three-dimensional conformation [Gross 2011].

Mobcal is an important software, widely used, for theoretical CCS calculation. It is based on three different treatments of the ion-buffer gas collisions: the projection approximation (PA), the exact hard sphere scattering (EHSS), and the trajectory method (TM). TM is the most accurate method, being the best choice for CCS estimates for highly charged macromolecules, such as proteins and proteins complexes [Mesleh et al. 1996]. Theoretical computations of CCS for biomacromolecular systems, under TM approximation, are inefficient with Mobcal, because of its outdated program language and technology, thus limiting its usage to studies of small proteins and organic molecules.

High Performance Computing (HPC) explores the computational resources like novel VLSI technology, parallelization algorithms and computer architectures to enable the solution of complex Engineering and Scientific applications in a feasible time. The time to process the results is reduced, but to achieve it, parallel algorithms are necessary to make an efficient usage of the new hardware [Hager and Wellein 2010].

HPCCS is a new software proposed, capable of performing CCS calculation for a large variety of molecular ions, ranging from small organic molecules to large protein complexes containing tens to hundreds thousand of atoms. It uses current processors features, like multicore processing and vectorization. It is based on Mobcal, and focused on the Trajectory Method. When the original Mobcal was written, processors were single core, so the CCS calculation was sequential [Zanotto et al. 2018].

2. Background

There is an analytical interest for compound identification from molecular masses, by mass spectrometry technique, as a mechanism to enable the mapping of the constituents of complex mixtures. Fields of application of MS are: Physics, Radiochemistry, Geochemistry, Organic chemistry, Polymer chemistry, Biochemistry, Physical chemistry, Thermochemistry, Quality control, Environmental analysis, medicine, etc.

3. Experimental Model – Mass Spectrometry and Ion Mobility

The basic principle of mass spectrometry is to generate ions, from either inorganic or organic compounds, by a suitable method, separating these ions by their mass-to-charge ratio (m/z) and detecting them qualitatively and quantitatively. The sample may be ionized thermally, by electric fields or by impacting on energetic electrons, ions or photons. Ions can be single ionized atoms, clusters, molecules or their fragments or associates. Ion separation can be affected by static or dynamic electric or magnetic fields [Price 1991].

As demonstrated with great success by the time-of-flight analyzer, ion separation by m/z can also be effected in field-free regions, providing the ions with well-defined kinetic energy at the entrance of flight path.

The most widespread developed and employed approach on IMS is the Drift-Time Ion Mobility Spectrometry (DTIMS), wich is the only IMS method providing a direct measure of CCS based in ion mobilities. Figure 1 shows a drift tube instrument, filled with inert buffer gas in a counter direction of the ion motion.

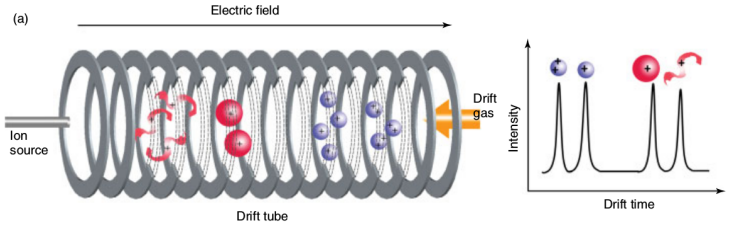


Figure 1. Scheme of a drift tube, filled by isobaric ions and being separated by their different chemical characteristics

4. Mobcal

Developed by Shvartsburg and Jarrold to calculate the theoretical CCS based on input coordinate file, called mfj, derived from X-ray crystallography, NMR studies or MD simulations [Mesleh et al. 1996].

The mobility of gas phase ion is a measure of how rapidly it moves through a buffer gas, under the influence of an electric field. The mobility depends on the average collision cross section, which it turn depends on the geometry [Laiko 2006].

It uses three different methods to CCS calculation: Projection Approximation (PA), Exact Hard Sphere Scattering (EHSS) and Trajectory Method (TM), each one is described below.

Projection Approximation

CCS is determined by averaging the area of projections on a plane, considering all possible orientations by rotations. However, this method ignores the long-distance interactions and the scattering process between the ion and buffer gas [Wyttenbach et al. 1997]. The calculation using PA is fast since it ignores the scattering process and long-range interactions between the ion and the gas [Bleiholder et al. 2011].

Exact Hard Sphere Scattering

The EHSS method calculates CCS by averaging the momentum transfer cross section over the relative velocity and collision geometry. It takes into account scattering and collision processes, but does not consider the effects of long range interactions. In summary, it is a simplification of the trajectory method, explained below [Zubarev et al. 2000]. It is commonly used on structural proteomics to examine them due to their large number of atoms.

Trajectory Method

TM is regarded as the most reliable and accurate method. It combines all the effects, including scattering events, long-range interactions and multiple collisions. The only weakness to consider is that time consuming, specially for macromolecules ions [Shvartsburg et al. 1998].

5. High Performance Collision Cross Section - HPCCS

Computers have become essential due to their ability to perform calculations, visualizations and general data processing at an incredible ever-increasing speed. Some decades ago the computers processors were built to run the software in serial using one processor with a single core. Today the computers are built with two cores at least on desktops or mobile devices, capable to do parallel processing. On clusters there are more than one processor per node, with much more cores on each one for scientific computation [Hager and Wellein 2010].

High Performance Computing (HPC) is a set of techniques that includes large scale computer cluster and parallelization and optimizations techniques that are used to deliver the performance that a single desktop cannot do, so as to solve large problems in science, engineering or business. Such machines process large amount of data, producing big data for analysis and calculations what would take months to have the results on a single computer.

Based on Mobcal trajectory method, HPCCS was re-written in the C/C++ computer language. HPC techniques were used to execute using only the necessary amount of memory, parallelization was applied to make use of all available processor cores and vectorization was used to speedup the potential calculation. HPCCS was designed in two versions: one using OpenMP for shared memory on a single node and MPI + OpenMP version using multiples nodes.

6. Experimental Results

The experimental results were calculated using the Center for Computing in Engineering & Sciences cluster (Kahuna) at the University of Campinas. FAPESP process 2012/24750-6, 2013/08293-7, 2016/04963-6.

There was an experiment using only one node running OpenMP version and the second experiment using the hybrid version MPI + OpenMP. They were executed using the nodes with the following configuration: Intel Xeon E5-2670 v3 with 2.30GHz and 48 threads. The goal was to measure the time spent and the speedup. Two groups of molecules were chosen to be simulated. The first group of molecules from Jurneczko and Barran [Jurneczko and Barran 2011], ranges from 432 to 4392 atoms in total. The other group, with protein complexes from Bush [Bush et al. 2010], contains 1674 to 32774 atoms. All experiments were compared with the original sequential non-optimized Mobcal version which was used as the speedup baseline.

7. Results using OpenMP

Below we present the results for all individual proteins from Perdita's group with the best speedup.

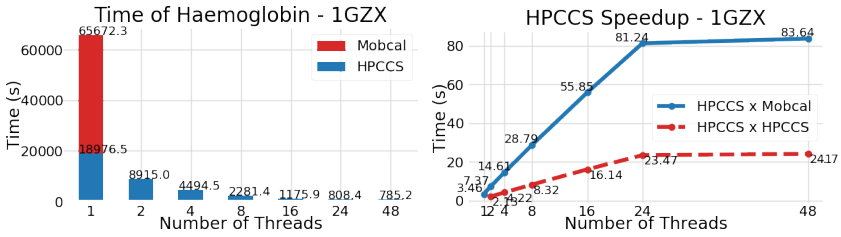


Figure 2. Time and Speedup of Haemoglobin - 1GZX with 4392 atoms.

Below we present the results for all individual proteins from Bush's group using the biggest molecule complex.

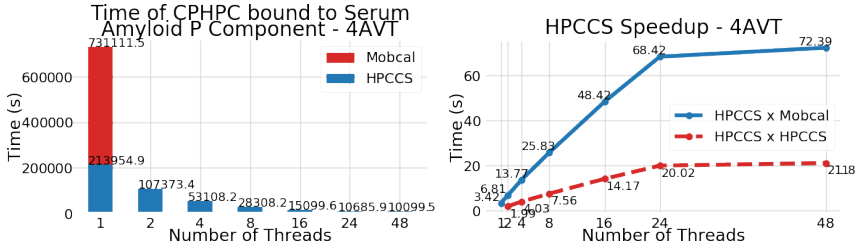


Figure 3. Time and Speedup of CPHPC bound to Serum Amyloid P Component - 4AVT with 32774 atoms.

8. Results using OpenMP + MPI

Below we present the results for all individual proteins from Bush's group using OpenMP + MPI for the best achieved speedup.

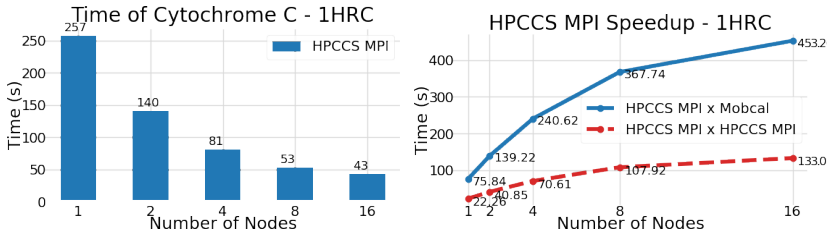


Figure 4. Time and Speedup of Cytochrome c - 1HRC with 1674 atoms.

9. Results and Conclusion

The sequential HPCCS is ≈ 3.5x faster than the Mobcal, as HPCCS is using many optimizations like vectorization and only the necessary amount of memory for each molecule simulation. In the OpenMP version, using hyper-thread, a technology that enables more than one thread per physical core, resulting in a speed up of ≈ 83x when compared to Mobcal. The time of simulations increases as the number of atoms of molecular complexity increases. Increasing the number of physical cores the simulation time decreases.

Mobcal Trajectory Method is the most accurate, it is expensive and requires many calculations. This was a problem to solve for this area, and HPCCS was written based on

Mobcal to solve it. To achieve that, it uses current HPC methods like modern processor architectures, code optimization, vectorization and other parallelization techniques.

HPCCS was able to process bigger molecules using multicore processors with vector units. Using a cluster with sixteen nodes with 24 cores each, the best speedup achieved was 453x when compared to Mobcal for 1HRC molecule with 1666 atoms.

Molecules can be calculate using the MPI + OpenMP version, on the other hand, even using a single desktop can compute CCS faster for small to medium molecules, than the traditional sequential Mobcal.

References

Bleiholder, C., Wyttenbach, T., and Bowers, M. T. (2011). A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (i). method. *International Journal of Mass Spectrometry*, 308(1):1 – 10.

Bush, M. F., Hall, Z., Giles, K., Hoyes, J., Robinson, C. V., and Ruotolo, B. T. (2010). Collision cross sections of proteins and their complexes: A calibration framework and database for gas-phase structural biology. *Analytical Chemistry*, 82(22):9557–9565.

Gross, J. H. (2011). *Mass Spectrometry*. Springer Berlin Heidelberg.

Hager, G. and Wellein, G. (2010). *Introduction to High Performance Computing for Scientists and Engineers*. CRC Press, Inc., Boca Raton, FL, USA, 1st edition.

Jurneczko, E. and Barran, P. E. (2011). How useful is ion mobility mass spectrometry for structural biology? the relationship between protein crystal structures and their collision cross sections in the gas phase. *Analyst*, 136:20–28.

Laiko, V. V. (2006). Orthogonal extraction ion mobility spectrometry. *Journal of the American Society for Mass Spectrometry*, 17(4):500–507.

Mesleh, M., Hunter, J., Shvartsburg, A., Schatz, G. C., and Jarrold, M. (1996). Structural information from ion mobility measurements: effects of the long-range potential. *The Journal of Physical Chemistry*, 100(40):16082–16086.

Price, P. (1991). Standard definitions of terms relating to mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 2(4):336–348.

Shvartsburg, A. A., Schatz, G. C., and Jarrold, M. F. (1998). Mobilities of carbon cluster ions: Critical importance of the molecular attractive potential. *The Journal of Chemical Physics*, 108(6):2416–2423.

Wyttenbach, T., Helden, G., Batka, J. J., Carlat, D., and Bowers, M. T. (1997). Effect of the long-range potential on ion mobility measurements. *Journal of the American Society for Mass Spectrometry*, 8(3):275–282.

Zanotto, L., Heerdt, G., Souza, P. C. T., Araujo, G., and Skaf, M. S. (2018). High performance collision cross section calculation-HPCCS. *Journal of Computational Chemistry*, 39(21):1675–1681.

Zubarev, R. A., Horn, D. M., Fridriksson, E. K., Kelleher, N. L., Kruger, N. A., Lewis, M. A., Carpenter, B. K., and McLafferty, F. W. (2000). Electron capture dissociation for structural characterization of multiply charged protein cations. *Analytical chemistry*, 72(3):563–573.

4 Resumos dos Pôsteres

Análise de Imagens Microtomográficas de Amostras de Reservatórios de Petróleo

Leticia S. Bomfim[1], Hélio Pedrini[1], Guilherme Avansi [2]

[1]Instituto de Computação – Universidade Estadual de Campinas (UNICAMP).

[2]Faculdade de Engenharia Mecânica – Universidade Estadual de Campinas (UNICAMP).

l209824@dac.unicamp.br, helio@ic.unicamp.br, avansi@unicamp.br

Resumo. *Amostras de rochas carbonáticas coletadas de reservatórios podem trazer muita informação sobre o ambiente de extração, com isso, propomos um método não invasivo que permite a análise de estruturas internas das rochas e de suas características morfológicas. Para tal, desenvolvemos uma ferramenta que analisa imagens de MicroCT em dois contextos: (i) os poros, extraindo-se características como porosidade, tamanho do poro e circularidade e (ii) as fraturas, analisando-se características como densidade, tamanho, angulação e tipo. A fim de identificar essas estruturas, utilizamos a segmentação pelo algoritmo de watershed e em seguida, através dos contornos extraídos, analisamos a geometria com base no bounding-box que engloba cada elemento da imagem. A ferramenta desenvolvida para a análise das imagens, vem como meio de auxiliar o pesquisador a reduzir as incertezas inerentes a caracterização e, melhorar a confiabilidade no poder de decisão ao longo da vida produtiva de um campo.*

1. Introdução

Amostras de rochas carbonáticas coletadas de reservatórios de petróleo podem trazer muita informação sobre o potencial de fluidos do ambiente analisado. Muitos desses reservatórios são formados em rochas porosas e fraturadas e, é esta condição que permite que haja a propagação e condução de substâncias químicas em seu interior pois, as conexões entre os espaços porosos que proporcionam a existência de escoamento**.** Com isso, conhecer o ambiente de produção e como o fluído se comportará, é de suma importância para uma otimização do processo de identificação de um reservatório em potencial. Para isso é necessário a aplicação de sondagens e amostragens que possibilitam o contato direto com o material que compõe o reservatório e de seus parâmetros macroscópicos. Como este processo de extração não é simples, e sua execução possui um alto custo, realizar análises que não agridam a integridade da amostra é a melhor forma de preservar e perpetuar o material colhido. Assim, a técnica de microtomografia computadorizada vem sendo empregada a fim de evitar que as características permo-porosas sejam modificadas a medida que experimentos laboratoriais são realizados, favorecendo a durabilidade e originalidade da amostra, além de possuir algumas vantagens, como a isotropia e resolução de pixels em escala micrométrica e, a possibilidade de criar modelos virtuais 3D.

2. Metodologia

A segmentação das imagens é realizada em duas etapas. A primeira etapa é para criar uma máscara que indica a região de interesse da imagem, enquanto a segunda etapa produz uma imagem referente ao conteúdo da rocha. Assim, pela combinação desses resultados, pode-se criar uma imagem somente com as estruturas internas que serão usadas para as próximas análises.

Após a segmentação da imagem, é necessário encontrar os contornos que foram extraídos da imagem original afim de analisar as suas características. Para isso, utiliza-se a função *findcountours* da biblioteca OpenCV, onde passamos a imagem segmentada como parâmetro e ela retorna os contornos que foram identificados. Como a análise dos poros e fraturas é feita de forma separada, torna-se necessário fazer a identificação de cada um desses grupos. As fraturas possuem corpo alongado similar a uma reta, então utilizamos esse critério para fazer a separação.

Características tais como área, altura, largura e diâmetro (no caso dos poros), podem ser obtidas através da mesma função que extrai o contorno, pois seu retorno comporta alguns dados a respeito da geometria do objeto. Com isso, as outras características são calculadas pela análise do corpo da estrutura presente no *bounding-box.*

4. Resultados

Como resultado da elaboração dos modelos de análise das estruturas internas de rochas, desenvolvemos uma interface que promove a interação do usuário com os dados obtidos. Dessa forma, foram criados dois diferentes ambientes para cada propósito. Nesses ambientes, todos os dados do processamento podem ser visualizados por meio de tabelas e gráficos que indicam métricas referentes a cada estrutura de forma individual, e informações do conjunto em análise. Sendo assim, os resultados podem ser analisados qualitativamente, e visualmente através do modelo 3D dos *slices* de entrada, onde o usuário pode interagir com sua amostra em diferentes ângulos.

5. Conclusões

Esta pesquisa visa apresentar uma ferramenta complementar para os pesquisadores que atuam com o manuseio de imagens advindas de microtomografia computadorizada de rochas. Nesse sentido, pode-se oferecer um ambiente em que os poros e as fraturas possam ser identificados e descritos, trazendo uma análise sobre seu comportamento e forma, a fim de contribuir para uma avaliação não-destrutiva de amostras que representam um conjunto em estudo e, complementar as análises feitas em laboratório.

Referências

Andrä, H., Combaret, N., Dvorkin, J., Glatt, E., Han, J., Kabel, M., ... and Marsh, M. Digital rock physics benchmarks—Part I: Imaging and segmentation. **Computers & Geosciences**, v.*50*, p.25-32, 2013.

Bieniek, Andreas; Moga, Alina. An efficient watershed algorithm based on connected components. **Pattern Recognition**, v. 33, n. 6, p. 907-916, 2000.

Digital Rocks Portal: a repository for porous media images, 2015.

Empirical Analysis of Semantic Metadata Extraction from Video Lecture Subtitles

Marcos Vinícius M. Borges[1], **Julio Cesar dos Reis**[1], **Guilherme P. Gribeler**[1]

[1]Institute of Computing – University of Campinas (UNICAMP) – Campinas – SP – Brazil

m211847@dac.unicamp.br, jreis@ic.unicamp.br, guigribeler@gmail.com

Abstract. *Video lectures improve the learning experiences considering individual's needs and learning styles. However, the large amount of educational content and their availability turns difficult the tasks of accessing these resources. The extraction of semantic metadata from a video subtitle involves challenges in dealing with informal aspects of language and the detection of semantic classes from the text. In this paper, we conduct an empirical analysis of semantic annotation approaches supported by ontologies in the extraction of relevant metadata from textual transcriptions of video lectures in Computer Science. The obtained results indicate that existing tools can be useful and the semantic metadata extraction process is highly influenced by the underlying ontologies.*

1. Introduction

The growth of information dissemination enabled the easy access for multimedia content that helps in the learning process, resulting in a significant increase in the amount of educational resources available to students. In this context, efforts are required by students to select the appropriate resources in the learning process. Potentially, video lectures from other courses or teachers may be interesting to replace or complement the concepts of a lesson. The filtering and searching of education contents could benefit from techniques exploring the meaning of concepts appearing in the video lectures.

The key challenge in this research is to investigate techniques that allow semi-automatically annotation of text transcriptions from video lectures based on Semantic Web standards and knowledge bases. The required techniques are complex and can be influenced by factors such as the quality of video transcriptions, language, ambiguity and context. This investigation addresses the challenge in creating annotations in text as metadata that associate concepts represented in an ontology with a particular piece of text or multimedia resource.

In this paper, we analyze existing semantic annotation tools to enable extraction of relevant semantic metadata from video lectures. These metadata must be able to describe the video well so it could be used as input to automatic semantic-enhanced recommendation methods. Our contribution enables further analysis of semantically annotated videos using existent annotation tools associated with general-domain or specific ontologies.

2. Study Design

We conducted experiments to assess the quality of semantic annotations obtained from a set of real-world video lectures in Computer Science area available on *Youtube*. The semantic metadata extraction process automatically retrieved the subtitles from a video

lecture in a textual format. The procedure used these textual subtitles as input for existing semantic annotation tools.

Our investigation considered software tools for semantically annotating texts such as *AutoMeta*[1], *CSO-Classifier*[Salatino et al. 2018], *NCBO Annotator* and OntoText [Kiryakov et al. 2004]. As support for the semantic annotation task we used the following ontologies: *DBpedia* ontology, Computer Network Ontology[2], and two releases of Computer Science Ontology (CSO)[Salatino et al. 2018].

3. Results

Table 1 presents the obtained results for a total of seven setups considered to conduct the evaluation, with average results and confidence interval for the mean of 95%, representing an interval of plausible values for population mean to analyze the overall effectiveness for each setup.

Table 1. Overall results. Table presents the tool's name, ontology used, distinct relevant terms (DRV), distinct terms annotated (DTA), distinct relevant terms correctly annotated (DRA), precision (Pr), recall (Re) and f-score

Tool	Ontology	DRV	DTA	DRA	Pr	Re	F-Score
AutoMeta	DBPedia	49 [36; 61]	62 [44; 81]	20 [18; 30]	0.318	0.416	0.360
AutoMeta	Computer Science V_1	49 [36; 61]	50 [26; 64]	13 [7; 20]	0.278	0.290	0.283
AutoMeta	Computer Science V_2	49 [36; 61]	21 [13; 28]	11 [5; 17]	0.553	0.243	0.337
CSO-Classifier	Computer Science V_1	49 [36; 61]	36 [22; 49]	13 [8; 18]	0.383	0.282	0.324
CSO-Classifier	Computer Science V_2	49 [36; 61]	25 [14; 36]	15 [7; 23]	0.633	0.324	0.428
Ontotext	DBPedia	49 [36; 61]	32 [20; 44]	7 [4; 10]	0.193	0.169	0.180
NCBO	Computer Network	49 [36; 61]	10 [5; 15]	6 [2; 10]	0.838	0.324	0.467

We found that the ontology used by the annotation tools plays an important role in the task of annotating the terms. A higher coverage of concepts matching with relevant terms in the video leads to better results. The results obtained with domain-specific ontologies (CSO V_1, CSO V_2 and Computer Network) showed that the number of distinct terms annotated (DTA) and the number of distinct relevant terms (DRA) decreased in general. However, the overall results for precision and recall were higher using these ontologies

4. Conclusion

Our findings point out that obtained annotations considering an ontology related to the specific domain can achieve more precise results, even though less domain-specific ontologies like *DBpedia* can help in the process. Our experimental results were relevant to understand which parts of the whole metadata extraction process can influence the most on the quality of the extracted metadata. Future work involves the development of further techniques to enrich a computer science ontology from book resources.

References

Kiryakov, A., Popov, B., Terziev, I., Manov, D., and Ognyanoff, D. (2004). Semantic annotation, indexing, and retrieval. *Web Semantics*, 2(1):49–79.

Salatino, A. A., Thanapalasingam, T., Mannocci, A., Osborne, F., and Motta, E. (2018). The computer science ontology: a large-scale taxonomy of research areas. In *International Semantic Web Conference*, pages 187–205. Springer.

[1]https://github.com/celsowm/AutoMeta

[2]https://bioportal.bioontology.org/ontologies/CN

Busca Semântica de Dados Científicos

Gustavo Caetano Borges[1], **Claudia Maria Bauzer Medeiros**[1], **Julio Cesar dos Reis**[1]

[1]Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)

gustavo.borges@students.ic.unicamp.br,{cmbm,jreis}@ic.unicamo.br

Resumo. *Dados científicos são conjuntos de dados usados como entrada e ou resultados de pesquisas científicas, se distribuídos abertamente, contribuem para a reprodutibilidade total ou parcial de pesquisas. A distribuição aberta de dados apresenta diversos desafios, como sua publicação e heterogeneidade dos metadados. O uso de semântica se faz necessário por proporcionar à máquina formas de efetuar buscas com resultados mais relevantes. Portanto o objetivo de pesquisa é desenvolver um mecanismo de recuperação de dados distribuídos entre os diversos repositórios científicos, visto que cada repositório possui sua política de anotação dos dados, dificultando uma busca centralizada em todos os repositórios. A pesquisa irá conduzir uma investigação de como tais dados científicos são publicados e recuperados, visando definir formas de recuperação em repositórios científicos.*

Introdução

O avanço da tecnologia e sua aplicação em pesquisas de todas as áreas, faz com que grandes montantes de dados sejam gerados, contribuindo com o dilúvio de dados. Os dados de algumas pesquisas em determinadas áreas podem acabar se assemelhando pelo escopo de cada pesquisa. Isso levanta uma boa prática, o compartilhamento de dados científicos.

Desde que os dados da pesquisa sejam compartilhados e abertamente, eles contribuem para a reprodutibilidade total ou parcial de pesquisas a partir do reuso dos dados existentes. Com isso a prática de se compartilhar e distribuir os dados de pesquisas vem sendo cobrada por algumas instituições como a FAPESP.

A distribuição aberta de dados apresenta diversos desafios, visto que a heterogeneidade dos metadados e forma como eles são publicados dificultam sua recuperação de forma eficiente. A heterogeneidade dos metadados faz com que seja necessário encontrar pontos em comum nos repositórios, a fim de melhorar a forma de recuperação centralizada.

Visando centralizar a busca por tais dados, este trabalho objetiva desenvolver uma plataforma de busca e recuperação de dados científicos utilizando métodos semânticos, dado que com utilização de web semântica e ontologias é possível dar significado aos dados e à busca, retornando ao usuário não apenas aqueles conjuntos de dados que ele buscou, mas também aqueles conjuntos de dados que ele deseja mas não expressa.

Trabalhos Correlatos

Um trabalho que faz busca comum por palavras chave e semântica é o de [Diaz and Medeiros 2017]. O autor propõe e desenvolve um sistema de busca de workflows em repositórios da internet. Sua metodologia divide todo o sistema em quatro grandes componentes responsáveis por pré-processamento, processamento, recursos externos

e armazenamento. Com isso o autor consegue apresentar resultados por meio de consultas semânticas e de palavras chave.

O trabalho de [Keong and Anthony 2011] também trata sobre busca semântica, porém os autores trabalham em um meta buscador alimentado pela DBpedia. Em seu trabalho os autores possuem quatro estágios no processo de busca por metadados para que seja possível transformar os resultados de uma busca não semântica em resultados de uma busca semântica.

Metodologia

A metodologia possui dois passos principais, um onde é feito um mapeamento dos metadados dos repositórios utilizados e outro do buscador em si.

A metodologia utilizada para se desenvolver o buscador é baseada na utilizada por [Diaz and Medeiros 2017] em sua pesquisa. As etapas para se alcançar uma lista de dados são divididas em duas linhas de frente, onde em uma são vistas as atividades relacionadas ao usuário (b) e outra as atividades relacionadas ao sistema que roda o buscador (a). Podemos verificar como se dão esses passos na figura 1, onde são recuperados os metadados em repositórios externos, armazenados, anotados semanticamente, e assim gerar um índice para o usuário consultar.

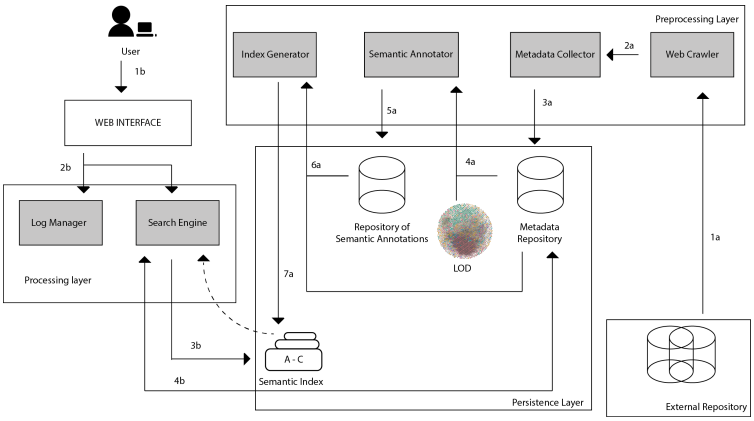


Figura 1. Metodologia baseada na de [Diaz and Medeiros 2017]

Benefícios Esperados

Espera-se que com a metodologia desenvolvida seja possível alcançar os seguintes resultados: Mapeamento de metadados de repositórios de dados científicos; Framework de recuperação de conjunto de dados científicos; Buscador centralizado de conjunto de dados científicos.

Referências

Diaz, J. S. B. and Medeiros, C. B. (2017). Workflow hunt: Combining keyword and semantic search in scientific workflow repositories. *Proceedings - 13th IEEE International Conference on eScience, eScience 2017*, pages 138–147.

Keong, B. V. and Anthony, P. (2011). Meta Search Engine Powered by DBpedia. *2011 International Conference on Semantic Technology and Information Retrieval*, (June):89–93.

Problemas de Empacotamento com Relação entre Itens

Vítor Gomes Chagas[1], **Flávio Keidi Miyazawa**[1]

[1]Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)

vitorvgc07@gmail.com, fkm@ic.unicamp.br

Abstract. *In this project, we introduce a bin packing problem variant denoted by Bin Packing Problem with Relation between Items, in which each pair of items has a dissimilarity value, and we desire to find a packing of the items so that similar items are assigned to the same bin while dissimilar items remain in different bins.*

Resumo. *Neste projeto, introduzimos uma variante do problema de empacotamento denominada Empacotamento em Recipientes com Relação entre Itens, em que cada par de itens possui um valor de dissimilaridade, e deseja-se empacotar os itens de forma que itens similares sejam atribuídos em um mesmo recipiente enquanto itens dissimilares se mantenham em recipientes diferentes.*

Problemas de empacotamento podem ser descritos informalmente da seguinte forma: dado um conjunto de itens de diferentes tamanhos e uma quantidade ilimitada de recipientes com certa capacidade, empacote todos os itens utilizando a menor quantidade possível de recipientes. Esses problemas são de grande relevância prática por representar uma questão fundamental em diversas situações, que é a de como utilizar um espaço limitado da forma mais eficiente possível. Por conta disso, essa classe de problemas possui uma grande quantidade de aplicações, como empilhamento de caixas de diferentes alturas, carregamento em caminhões e contêineres, e escalonamento de tarefas.

No trabalho que será realizado para o mestrado de Vítor Gomes Chagas, no Instituto de Computação da UNICAMP, pretendemos investigar o problema de empacotamento em recipientes com relação entre itens, que denotaremos por BPPRI. Nessa variante, cada recipiente possui um custo para ser utilizado e cada par de itens possui um valor de distância ou dissimilaridade, que pode ser tanto positivo como negativo, representando respectivamente uma penalidade ou um benefício ao serem empacotados juntos. Deseja-se encontrar um empacotamento dos itens que minimize a soma das distâncias entre os pares de itens de cada recipiente e os custos de utilização dos recipientes. Dessa forma, torna-se relevante tanto a quantidade de recipientes utilizados como a escolha dos itens que serão empacotados no mesmo recipiente.

Essa versão do problema de empacotamento possui várias aplicações em cenários do mundo real, como por exemplo no carregamento de uma grande variedade de produtos em caminhões em que se queira minimizar a quantidade de caminhões utilizados, porém considerando que não se deseja carregar itens muito diferentes juntos, como remédios e veneno, ou eletrônicos e alimentos. Outro exemplo consiste na alocação de produtos que devem aparecer em espaços de anúncios, e deseja-se que anúncios similares fiquem juntos.

Note que o BPPRI está relacionado com vários outros problemas. Se os itens tiverem dimensão 0 (pontos no espaço), tem-se o problema de *Correlation Clustering*,

primeiramente proposto por [Bansal et al. 2004]. Caso sejam removidos os valores de distâncias entre os itens, tem-se o problema de empacotamento em recipientes tradicional. Se todos os valores de distância entre os itens forem 0 ou $-\infty$, tem-se o problema de empacotamento em recipientes com conflitos, onde alguns pares de itens não podem ser empacotados juntos [Jansen and Öhring 1997, Epstein et al. 2008].

Temos como objetivo desenvolver algoritmos exatos e formulações para o BPPRI em sua versão unidimensional e bidimensional. A seguir, a versão unidimensional do problema é descrita mais formalmente.

Problema de empacotamento unidimensional com relação entre itens (1D-BPPRI): Dado conjunto $I = \{1, \ldots, n\}$ de n itens, função $w : I \to \mathbb{Q}_+$ que representa o peso de cada item, função $d : I \text{x} I \to \mathbb{Q}$ que representa a distância entre pares de itens, com $d_{ii} = 0, d_{ij} = d_{ji}\ \forall i, j \in I$, capacidade $W \in \mathbb{Q}_+$ e custo $c \in \mathbb{Q}_+$ de cada recipiente, encontre uma partição de I dada por $\mathcal{C} = \{C_1, \ldots, C_k\}$ tal que $ck + \sum_{C \in \mathcal{C}} \sum_{i,j \in C} d_{ij}$ é mínimo e para todo $C \in \mathcal{C}$, $\sum_{i \in C} w_i \leq W$.

Seja x_{ij}^k uma variável de decisão binária que indica se os itens i e j são empacotados no recipiente k (consequentemente x_{ii}^k indica se o item i é empacotado no recipiente k), e y_{ik} uma variável que indica se o recipiente k está sendo utilizado. Uma formulação para o 1D-BPPRI é fornecida a seguir:

$$\min \sum_{i=1}^{n} \sum_{j=i+1}^{n} \sum_{k=1}^{n} d_{ij} x_{ij}^k + \sum_{k=1}^{n} c y_k \tag{1}$$

$$\text{s.a } y_k \geq x_{ii}^k \qquad \forall\, i, k \in I \tag{2}$$

$$\sum_{k=1}^{n} x_{ii}^k = 1 \qquad \forall\, i \in I \tag{3}$$

$$x_{ij}^k \geq x_{ii}^k + x_{jj}^k - 1 \qquad \forall\, 1 \leq i < j \leq n, k \in I \tag{4}$$

$$x_{ij}^k \leq x_{ii}^k \qquad \forall\, 1 \leq i < j \leq n, k \in I \tag{5}$$

$$x_{ij}^k \leq x_{jj}^k \qquad \forall\, 1 \leq i < j \leq n, k \in I \tag{6}$$

$$\sum_{i=1}^{n} w_i x_{ii}^k \leq W \qquad \forall\, k \in I \tag{7}$$

$$x_{ij}^k \in \{0, 1\} \qquad \forall\, 1 \leq i \leq j \leq n, k \in I \tag{8}$$

$$y_k \in \{0, 1\} \qquad \forall\, k \in I \tag{9}$$

A restrição (2) impõe que um item é empacotado apenas em recipientes utilizados. A restrição (3) garante que cada item é empacotado em exatamente um recipiente. A restrição (4) indica que $x_{ij}^k = 1$ sempre que os itens i e j forem empacotados no recipiente k, enquanto as restrições (5) e (6) fazem com que $x_{ij}^k = 0$ se o item i ou item j não foram empacotados no recipiente k. A restrição (7) representa a restrição de capacidade de cada recipiente. Por fim, as restrições (8) e (9) são referentes às restrições de integralidade das variáveis.

Referências

Bansal, N., Blum, A., and Chawla, S. (2004). Correlation clustering. *Machine Learning*, 56(1):89–113.

Epstein, L., Levin, A., and van Stee, R. (2008). Two-dimensional packing with conflicts. *Acta Informatica*, 45(3):155–175.

Jansen, K. and Öhring, S. (1997). Approximation algorithms for time constrained scheduling. *Information and Computation*, 132(2):85 – 108.

Routing Protocol Using Deep Graph Networks Applied to Organic Growth Topologies

Caio V. Dadauto[1], **Nelson L. S. da Fonseca**[1], **Ricardo da S. Torres**[1]

[1]Instituto de Computação (IC) – Universidade Estadual de Campinas (Unicamp)
Campinas – SP – Brazil

{caio.dadauto, nfonseca, rtorres}@ic.unicamp.br

Abstract. *We propose a routing protocol based on deep learning, which processes the whole graph structure and keeps the inherent distributed environment of the Internet. We validate our model in realistic topologies provided by software BRITE, in which we compare our protocol with the Dijkstra's algorithm. Our model achieved the accuracy of 85% for the routing links, where, it increases in just 3 hops to the path for the worst case; moreover, we shows that our model can be generalize to a more lager topologies without retraining, keeping the accuracy greater than 80%.*

1. Introduction

The dramatically increasing of the complexity in the network environment and the amount of network traffic data make difficult the use of closed-form protocols; this leverage the development of data-driven solutions to achieve the goal of a self-driven network [Feamster and Rexford 2017]. On the other hand, the recent improvements in the deep learning techniques make feasible the investigation of robust and general solutions for problems that needs to process non-euclidean objects, like graphs. Therefore, the deep learning for graphs can naturally applicable in a self-driven network environment, especially in the control plane. The authors in [Geyer and Carle 2018] propose a routing protocol based in the graph deep learning techniques; however they are not clear in how to generalize the model for any topology and they evaluate their protocol only for non-realistic random topologies. Thus, we propose another routing protocol utilizing the Graph Networks abstraction [Battaglia et al. 2018] and evaluate it in realistic topologies created by BRITE software[1].

2. Model

Our model uses the Graph Networks abstraction, in which there is defined the elementary structure named by GN block. Basically this block can be composed by three functions, namely the update function for nodes and edges, and the aggregation function; the blocks uses those functions to extract the relation between nodes, which is imposed by graph topology. We implement three different GN blocks to architect our model, they named by encoder, core and routing block. The encoder and routing block use feed forward networks

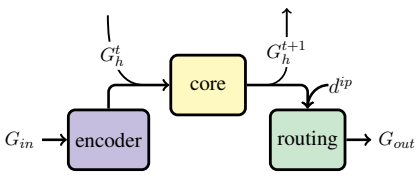


Figure 1. The model architecture.

[1]http://www.cs.bu.edu/brite/

for update functions, while the core uses recurrent networks for the update and a simple sum for the aggregation function; the architecture is presented in the figure 1. The encoder is responsible to map the vectors features from nodes and edges to a latent dimension. On the other hand, the core spreads the features among the graph through the massage-passing. Finally, the routing block decides, based on a destination IP (d^{ip}), which is the link that a node should route the packet.

3. Numerical Evaluation

We training the model using batches of topologies with size between 8 and 20 nodes, and they are created by BRITE software, which uses the Barabási-Albert algorithm that lies in the organic-growth paradigm, concept commonly observed in the real-world network. Each topology uses the vector features composed by the IP address and the euclidean distance, respectively, for nodes and edges. Moreover, as ground truth, we use the paths generated by Dijkstra's algorithm; each path is computed from all nodes to an only random node destination for each topology. After the training step, we evaluate the final model in two batches with 200 graphs, in which one batch is composed by graphs with the same size of training step (called by non-generalization) and other batch with graphs with size between 16 and 35 nodes (called by generalization). In this context, we measure the accuracy (the ratio between the number of edges that was predict according to Dijkstra's algorithm and the number of edges), the massage delivery ratio, *i.e.*, the ratio between the number of massages that was correctly delivery to the destination and the total number of the sent massages, and the difference of hops between paths created by our model and the ground truth; these measures was presented in figure 2

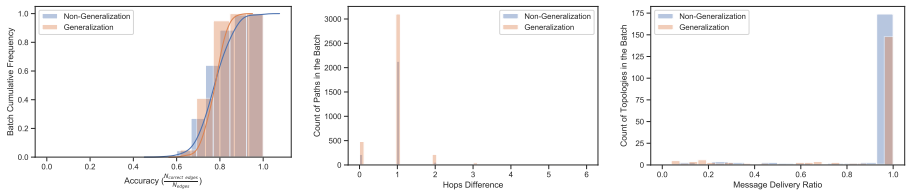


Figure 2. The measures from numerical evaluation.

4. Conclusion

Compared with Dijkstra's algorithm, our model achieved the accuracy of 85%. Also, it increases in just 3 hops to the Dijkstra's path, in the worst case, and our results show that the model can be generalize to larger topologies without retraining. Finally, our solution leverages the discussion of self-driven networks; once it uses the GN blocks that allows the use of any type of feature and not depends of the input topology.

References

Battaglia, P. W. and et al (2018). Relational inductive biases, deep learning, and graph networks.

Feamster, N. and Rexford, J. (2017). Why (and how) networks should run themselves.

Geyer, F. and Carle, G. (2018). Learning and generating distributed routing protocols using graph-based deep learning.

Using function expansion to increase Shadow Stack viability

Pedro Terra Delboni, Heitor Boschirolli, João Moreira, Sandro Rigo

[1]Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)

Abstract. *Refined protections against Control-Flow Hijack involve the use of a shadow stack. Unfortunately, their implementation induces an increase in execution time which in many cases is unacceptable. A proposed solution to this issue is to expand selected calls, but this solution hasn't been tested yet. In this paper, we'll explain our strategy to evaluate this proposition.*

Resumo. *Proteções refinadas contra Sequestro de Controle de Fluxo involvem o uso de uma Shadow Stack. Infelizmente suas implementações induzem um aumento no tempo de execução que em vários casos é inaceitável. Uma solução proposta para esse problema é expandir chamadas de funções em lugares chaves do programa. Nesse artigo vamos explicar a nossa estratégia para avaliar essa proposta.*

1. A brief story of Control-Flow Hijack

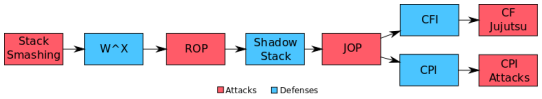


Figure 1. Iterations of attacks and defenses.

Figure 1 shows a brief sequence of attacks and defenses that illustrate the importance of a shadow stack. Stack Smashing [One 1996] presented the dangers of unprotected stacks by injecting code and replacing the stack's return address to a pointer to the injection. Write XOR Execute is a hardware defense that doesn't allow execution of instructions at writable memory. ROP [H. 2007] is an attack that instead of injecting code, manipulates return addresses to execute code already in the program in an unpredicted order which gives the attacker control of the machine. Shadow Stacks [Cowan et al. 1998] is a defense against stack corruption that involves creating a second, more protected, stack which will contain redundancy of return addresses so they can be validated before exiting a function. JOP [S. et al. 2010] is another attack that uses function pointers instead of return address. This lead to defenses like CFI [Abadi et al. 2005] and CPI [Kuznetsov et al. 2014], which if not implemented strictly can still be bypassed by other attacks [I. et al. 2015] [Evans et al. 2015], and a good implementation needs a shadow stack.

2. Expanding Functions (inline)

By expanding (inlining) a function, no call is issued, thus no return address can be corrupted. Return addresses are also the most frequently accessed code pointers inside our program, so by removing them we are removing most of the places which need to be protected, removing the cost to protect them.

2.1. Selecting functions to inline

Inlining every function is not an option, because it would make the program binary a lot bigger. Another issue is that by inlining functions, we may be removing the cost of protecting a call, but we may also be adding the cost of cache misses, since two different calls to the same function will now lead to two different parts of the binary.

In order to maximize the efficiency, we need to inline the calls which are used the most. This means that our choice will be based on a specific execution flow. In order to choose these calls, we modified the compiler to create one global variable associated with each call, and added instructions right before the call to increment this variable. At the end of our desired execution, we have a list of how many times each call was made. This will help us determine which calls to inline.

2.2. Inlining a call

Compilers already have the capacity to inline a call, but they can't inline **any** call. In order for a call to be expanded, the callee must either be at the same file that the caller or the expansion must be made at a link-time optimizer. Unfortunately, link-time optimizers aren't well adopted yet, and many projects can't be compiled with them.

We propose two solutions:

- **Cloning calls** The simplest solution to remove the call and return instructions was to instead of expanding the function we modify the compiler to create a clone of the callee and replacing the call and return instructions to jumps in and out of the clone.
- **Inlining calls** A more complex solution is to use the compiler to get every function that will be inlined and create a special module that will contain only a copy of the function with every symbol used by it marked as an external one. We'll fuse this module with the one with the call and point the call to the new function, so the compiler will now be able to inline it.

3. Evaluating the results

We are finishing the implementation of both solutions and once done we'll try to answer the following questions:

- Given a specific execution flow is it possible to inline a set of calls that will make the cost of shadow stacks negligible?
- Is there a set of functions that will make so that most execution flows won't be penalized by the cost of the shadow stack?
- Is the impact of cash misses on expanded functions greater than the impact of the shadow stack, and if not, is this impact acceptable?

References

Abadi, M., Budiu, M., Erlingsson, Ã., and Ligatti, J. (2005). Control-flow integrity: Principles, implementations, and applications. *ACM SIGSAC Conference on Computer and Communications Security (CSS)*.

Cowan, C., Pu, C., Maier, D., Hintony, H., Walpole, J., Bakke, P., Beattie, S., Grier, A., Wagle, P., , and Zhang, Q. (1998). Stackguard: Automatic adaptive detection and prevention of buffer-overflow attacks. *8st USENIX Security Symposium*.

Evans, I., Fingeret, S., Gonzalez, J., Otgonbaatar, U., Tang, T., Shrobe, H., Sidiroglou-Douskos, S., Rinard, M., and Okhravi, H. (2015). Missing the point(er): On the effectiveness of code pointer integrity. *IEEE Symposium on Security and Privacy*.

H., S. (2007). The geometry of innocent flesh on the bone: Return-into-libc without function calls (on the x86). *Proceeding CSS 07 Proceedings of the 14th ACM conference on Computer and communications security*, pages 552–561.

I., E., F., L., U., O., H., S., M., R., H., O., and S, S.-D. (2015). Control jujutsu. *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security - CCS '15*, pages 901–913.

Kuznetsov, V., Szekeres, L., Payer, M., Candea, G., Sekar, R., , and Song, D. (2014). Code-pointer integrity. *11th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*.

One, A. (1996). Smashing the stack for fun and profit. *Phrack Magazine 49(14)*.

S., C., L., D., A., D., A., S., H., S., and M., W. (2010). Return-oriented programming without returns. *7th ACM conference on Computer and communications security - CCS Í0*.

Implantação de Contêineres Docker em clusters HPC para execução de programas MPI

Rodrigo C. Freitas[1], **Hervé Yviquel**[1], **Marcio Pereira**[1], **Guido Araújo**[1]

[1]Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)

r176848@dac.unicamp.br, {herve.yviquel,mpereira,guido}@ic.unicamp.br

Abstract. *Containers allow applications to be packaged with their dependencies, making them portable. Despite this encapsulation, distributed applications may still require configuration time from system administrators. This paper investigates an automated way to instantiate containers cluster ready to run MPI applications on an arbitrary number of physical machines. This model has proven to be practical for testing distributed applications and allows the replacement and experimentation of different implementations of MPI specifications without changing physical machine software.*

Resumo. *Contêineres permitem o empacotamento de aplicações junto de suas dependências, tornado-as portáteis. Apesar deste encapsulamento, aplicações distribuídas ainda podem exigir tempo de administradores de sistema. Este trabalho investiga uma forma automatizada de se instanciar um cluster de contêineres prontos para executar aplicações que usam MPI em um número arbitrário de nós físicos. Esse modelo se mostrou prático para a execução de testes em aplicações distribuídas, além de permitir substituir e experimentar diferentes implementações das especificações de MPI sem alterar instalações nas máquinas físicas.*

1. Introdução

Contêineres como Docker têm se popularizado por fornecerem isolamento de ambiente, tornando aplicações mais portáteis e de comportamento reprodutível. Como o *runtime* do Docker cria um camada de tradução para acesso aos recursos do *host*, a penalidade de desempenho é muito inferior à de máquinas virtuais e é geralmente desprezível [Ruan et al. 2016]. Apesar dessa abstração e isolamento, o uso associado à aplicações de computação distribuída ainda pode exigir tempo razoável de configuração por parte dos administradores de sistema [Nguyen and Bein 2017].

2. Proposta e metodologia

Para o funcionamento de uma aplicação distribuída, a mesma implementação das especificações da biblioteca MPI (Message Passing Interface) precisa estar instalada em todos os ambientes que a executarão. Além disso, é necessário que a autenticação por SSH funcione sem a necessidade de inserção de senhas de forma interativa entre todos os nós participantes. Para endereçar este problema, projetamos um *script* que utiliza as ferramentas de orquestração disponíveis no Docker para instanciar uma rede de contêineres (Docker Swarm e Docker Secret). A Figura 1 mostra o fluxograma proposto. O *script* acessa

uma lista de *hostnames*, fornecida pelo usuário, correspondente às máquinas físicas que executarão os ambientes virtualizados. É utilizada uma rede sobreposta fornecida pelo Docker que permite aos contêineres pertencer à uma mesma rede, mesmo que estejam em nós físicos distintos, possibilitando assim a criação do *cluster*. A parte final do *script* faz com que todos os contêineres adotem uma mesma chave RSA privada gerada em tempo de *deploy*, para que a autenticação do SSH possa ser feita entre os participantes do *cluster*, condição necessária para o funcionamento de aplicações MPI.

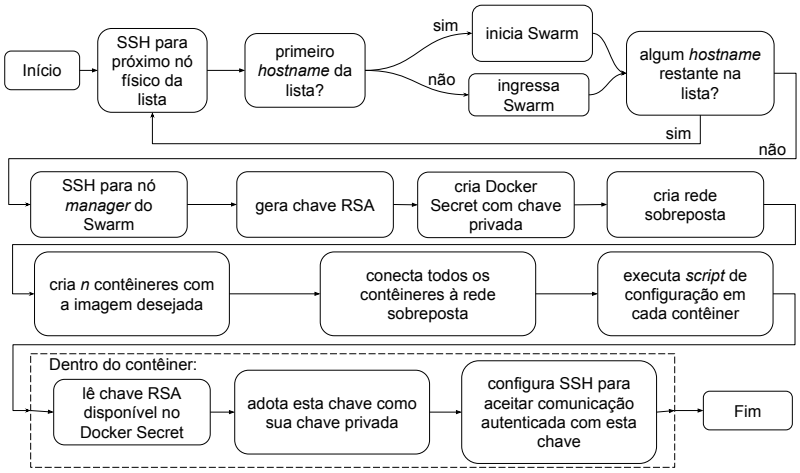


Figura 1. Fluxograma com os passos que o *script* de *deploy* proposto realiza

3. Resultados e conclusões

A solução proposta se mostrou prática para instanciar *clusters* de contêineres, permitindo, por exemplo, realizar diversos testes em aplicações distribuídas de forma local, em um único nó físico. Também permitiu o experimento, de forma rápida, de diferentes implementações de SSH e MPI, sem alterar nenhuma instalação na máquina *host*, permitindo, por exemplo, mudar de *OpenMPI* para *MPICH* apenas substituindo a instalação feita na imagem de *Docker* usada no *cluster*. Apesar da rede sobreposta ter se mostrado adequada nos testes realizados, ela adiciona um *overhead* que pode ser significativo dependendo da natureza da aplicação. Desse modo, é sugerido experimentar diferentes conexões e implementações de contêineres, MPI e SSH, que diferem no modo de tratar I/O para se encontrar a solução mais adequada para cada caso [Ermakov and Vasyukov 2017].

Referências

Ermakov, A. and Vasyukov, A. (2017). Testing Docker Performance for HPC Applications. *CoRR*, abs/1704.05592.

Nguyen, N. and Bein, D. (2017). Distributed MPI cluster with Docker Swarm mode. *2017 IEEE 7th Annual Computing and Communication Workshop and Conference, CCWC 2017*, pages 1–7.

Ruan, B., Huang, H., Wu, S., and Jin, H. (2016). A performance study of containers in cloud environment. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10065 LNCS:343–356.

Ferramenta de Geração Automática de Códigos Maliciosos Distribuídos

Victória Serra de Lima Moraes[1], **Paulo Lício de Geus**[1]

[1]Instituto de Computação (IC) – Universidade Estadual de Campinas (UNICAMP) – Campinas – SP – Brasil

victoriamoraes42@gmail.com, pgeus@unicamp.br

Resumo. *Códigos maliciosos tornam-se cada vez mais perigosos dia após dia, com novas arquiteturas ou melhores formas de processamento de dados. Surge, então, a necessidade de desenvolver ferramentas que impeçam a proliferação desses códigos. Com esse objetivo, é proposto o projeto e implementação de uma ferramenta de geração automática de códigos maliciosos distribuídos a fim de testar ferramentas de detecção de malware no contexto de computação de múltiplos núcleos.*

1. Introdução

Desenvolvedores de malware elaboram técnicas sofisticadas para burlar as ferramentas de detecção baseadas em assinatura existentes. Com a migração de CPUs de único núcleo para processadores multi-core, e a mudança de sistemas de 32 bits para sistemas de 64 bits, a realidade dos códigos maliciosos acompanha tais alterações. Assim, desenvolvedores de malware podem começar a usar essas informações para detectar se o código está sendo executado em uma solução *sandbox* ou em uma máquina real, pois a maioria das soluções de segurança ainda é de núcleo único. Ao verificar o número de núcleos disponíveis para determinada execução, o código malicioso pode evitar não apenas uma *sandbox*, mas também emuladores de antivírus, evitando assim a detecção.

Contudo, ferramentas de detecção de softwares nocivos não realizaram muitas atualizações que fossem capazes de detectar esse modo de ataque[Botacin et al. 2019]. Neste cenário, este trabalho propõe uma ferramenta de geração automática de malwares distribuídos a partir de amostras reais, para testar essas ferramentas e assegurar que os esforços feitos para reparar as falhas encontradas sejam definitivos, visto que não seria apenas um malware específico a ser detectado, mas inúmeros malwares gerados da mesma maneira.

2. Reescrita Binária

A reescrita binária pode ser dividida em quatro etapas[Wenzl et al. 2019] – análise sintática, análise, transformação, e geração de código. Executáveis consistem em dados administrativos e de *payload*. O foco da reescrita é obter e manipular os dados de *payload*. No entanto, essa informação geralmente se encontra dispersa ao longo do arquivo; instruções (em arquiteturas CISC) e variáveis não são bem delimitadas; binários não possuem informações de tipo de variável; e o tipo do endereço deve ser recuperado separadamente. Assim, o propósito da primeira etapa é obter o fluxo bruto de instruções e passá-lo ao *disassembler*.

A segunda etapa recupera a estrutura do código fonte do programa. Primeiramente, o fluxo bruto binário é avaliado a fim de que um grafo de controle de fluxo seja gerado. Então, algoritmos de recuperação de funções têm a tarefa de achar e agrupar séries de instruções conectadas por condições a blocos de funções, assim como determinar os pontos de entrada e saída da função. Após obtidas essas informações, o binário poderá ser alterado em pontos de instrumentação. Esses pontos são definidos como locais especificados pelo usuário onde o fluxo de controle muda; ou mudanças de instruções podem ser aplicadas. Finalmente, as mudanças planejadas são integradas ao binário de tal forma que ele se mantenha executável.

3. Metodologia

A princípio, construímos, estaticamente, um grafo de dependências dos binários, de modo a entender quais partes podem ser distribuídas. O próximo passo é realizar uma análise de dados do programa, colocando-o na forma de atribuição estática única (SSA)[Pradelle et al. 2011]. Isso fornece um número de versão exclusivo para cada definição de registro e fornece um link direto para cada uso de registro.

Partimos, então, para a reescrita do binário. A aplicação é carregada em um programa similar a um *debugger*, com o intuito de monitorar cada instrução executada e variável acessada. Isso pode ser realizado com a API `PTRACE` (Linux) ou a API de aplicação de *debug* (Windows). Durante a execução, o binário é "desmontado"(*disassembled*) ao longo dos caminhos cobertos por seus dados de entrada, obtendo, assim, as estruturas de interesse.

Com essas estruturas, obtemos um código com todos os acessos à memória executados no código binário, mas onde a semântica está oculta. Finalmente, com esse código e o grafo de dependências previamente obtido, podemos realizar a divisão do malware em múltiplos núcleos com o intuito de evitar detecção.

4. Conclusões

A ferramenta proposta estende o trabalho realizado por [Botacin et al. 2019], buscando um sistema de geração automática de códigos maliciosos distribuídos paralelizados de diferentes formas a fim de avaliar diferentes ferramentas em contextos variados. Ao fim, será gerado um possível alerta destinado aos fabricantes de softwares de detecção de malware para que haja uma atualização de suas ferramentas.

Referências

Botacin, M., de Geus, P. L., and Grégio, A. (2019). "VANILLA" malware: vanishing antiviruses by interleaving layers and layers of attacks. *Journal of Computer Virology and Hacking Techniques*.

Pradelle, B., Ketterlin, A., and Clauss, P. (2011). Transparent Parallelization of Binary Code. In *First International Workshop on Polyhedral Compilation Techniques, IMPACT 2011, in conjunction with CGO 2011*, Chamonix, France. Christophe Alias, Cédric Bastoul.

Wenzl, M., Merzdovnik, G., Ullrich, J., and Weippl, E. (2019). From hack to elaborate technique—a survey on binary rewriting. *ACM Comput. Surv.*, 52(3):49:1–49:37.

Link Maintenance in the Semantic Web

Andre Gomes Regino[1], **Julio Cesar dos Reis**[1]

[1]Institute of Computing – University of Campinas (Unicamp)
Campinas – SP – Brazil

{andre.regino,jreis}@ic.unicamp.br

Abstract. *Connections among data elements represent the core of Semantic Web. The connections are built with semi-automatic linking algorithms using a variety of similarity calculus. The data interconnected by these algorithms demands automatic methods and tools to maintain its consistency. Even though the constant update of RDF connections is considered an important process for the evolution of these structured datasets, such changing operations can influence the well-formed links, which turns difficult the consistency of the connections over time. In this work, we aim to investigate new methods responsible for fixing and updating links among ontologies in the Linked Open Data context.*

1. Context

Links between LOD datasets are at the heart of the Web of Data. Although the implementation of change operations in LOD datasets is essential to assure structured data evolution these operations can affect established links, which might turn them invalid or inconsistent. These links are maintained sporadically and manually [Bizer et al. 2009]. Also, ontologies, vocabularies and data schemas can change the definition and structure of RDF data. The manual maintenance remains hardly accomplishable due to the overwhelming number of links available.

2. Goal

We aim to investigate, formalize and implement semi automatic link maintenance actions in order to recognize affected links and turn them up-to-date. Figure 1 shows an evolution of a removal of a given triple and the absence of a link removal associated to that triple.

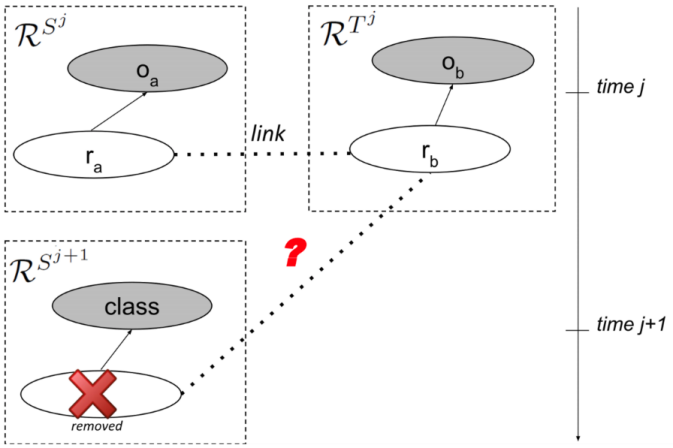


Figure 1. Problem Characterization

3. Methodology

In order to keep to links up-to-date, we are building a framework composed by three main steps, listed as:

- **Step A:** Detection of changes, given two versions of the same dataset as input, the framework maps every simple or complex change through these versions;
- **Step B:** Recognition of affected links, given the changes mapped at Step A, discover which of the links became structurally or semantically broken;
- **Step C:** Application of maintenance action, given the list of broken links as output of Step B, select based on a given number of actions which one is appropriate to make the links consistent.

4. Initial Results

Table 4 shows the results we collected in a study [Regino et al. 2019] that interrelates changes in triples (lines) with changes in links (columns) performed in a life sciences dataset named Agrovoc[1]. The first table shows that Agrovoc dataset applies the concept of Linked Data, linking 99% of their newly added triples to an external dataset. In second table, however, 96.15% of identified removed cases shows that if an internal triple is removed, the connected link remained untouched, generating cases of structurally broken links. The third table, regarding modification, shows that the fourth sub-case concerns the most frequent one, in which the modification of triples led to unchanged links. This case needs additional studies to further observe to which extend these unchanged links remained semantically inconsistent due to the modifications of the associated RDF triples.

Triples / Links	Add	No Add
Add	98.84%	0.41%

Triples / Links	Remove	No Remove
Remove	3.85%	96.15%

Triples / Links	Add	Remove	Modify	No Change
Modify	0%	0.04%	4.41%	95.55%

Table 1. Add, Remove and Modify Actions

5. Next Steps

We are now focusing on developing novel strategies to address the challenges on identifying broken links and maintaining them (Steps B and C of Section 3).

References

Bizer, C., Heath, T., and Berners-Lee, T. (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22.

Regino, A. G., dos Reis, J. C., Matsoui, J., Bonacin, R., Morshed, A., and Sellis, T. (2019). Understanding link changes in lod via the evolution of life science datasets. In *Workshop Semantic Web solutions for large-scale biomedical data analytics (SeWeBMeDA-2019) co-located at the 18th International Semantic Web Conference (ISWC'19), Auckland, New Zeland. (accepted for publication).*

[1]http://aims.fao.org/standards/agrovoc/linked-data

Detecting the Misuse of Cryptographic Methods with Machine Learning

Gustavo Eloi de P. Rodrigues[1], **Ricardo Dahab**[1], **Alexandre M. Braga**[1]

[1]Instituto de Computação – Universidade Estadual de Campinas (Unicamp)

g230218@dac.unicamp.br, rdahab@ic.unicamp.br, alexmbraga2007@gmail.com

Abstract. *Misuse of cryptographic libraries leads to vulnerabilities that, if exploited, can cause problems related to application data confidentiality, integrity, and availability, causing major harm to software developers. Therefore, this research project proposes the use of machine learning for detecting cryptographic library misuse in combination with source code representations that preserve syntactic, semantic, and data-flow structures. Additionally, this project aims to indicate in the source codes where the detected misuse occurs.*

Resumo. *Mau-usos de bibliotecas criptográficas levam a vulnerabilidades que, se exploradas, podem causar problemas relacionados à confidencialidade, integridade e disponibilidade de dados de aplicações, causando grandes prejuízos aos desenvolvedores de software. Sendo assim, este projeto de pesquisa propõe o uso de aprendizado de máquina para a detecção de maus-usos de bibliotecas criptográficas em combinação com representações de códigos-fonte que preservem estruturas sintáticas, semânticas e de fluxo de dados. Adicionalmente, este projeto visa indicar nos códigos-fonte o local em que ocorre os maus-usos detectados.*

1. Introduction

With the increasing use of technologies and applications that require assurance of requirements such as data integrity, confidentiality, authenticity and availability, encryption has been increasingly used for this purpose. However, most developers responsible for building these applications have limited knowledge of cryptography. In addition, the cryptographic libraries used are not easy to understand and have limited documentation, which results in the difficulty of their use and often leads to their misuse [Lazar et al. 2014], which causes cryptographic misuses.

As a result, many companies rely on tools to aid in the development of cryptographic applications. However, only 30% of these misuses are detected on average by these tools and the combination of two or more of them detects a maximum of 50%, leading to the introduction of software vulnerabilities in these applications [Braga et al. 2017]. Therefore, solutions that effectively support application developers to incorporate encryption simply and effectively into their applications are urgently needed [Nadi et al. 2016].

2. Goals

In order to improve the detection of such misuse, we propose an approach that uses machine learning to detect misuse rather than the matching of predefined patterns and rules employed by the available tools. Thus, the main expected goals of this project are:

- Develop a machine learning model capable of identifying misuse of cryptographic libraries through the use of source code representation structures.
- Classify detected misuse according to defined categories.
- Indicate in the source code analyzed where a detected misuse occurs.

3. Methodology

Our approach will be divided into four phases:

1. **Data Collect**: Collect source codes (data) from other works such as [Braga and Dahab 2016] and others.
2. **Feature Engineering**: Transform source code into Abstract Syntactic Trees and Program Dependency Graphs and extract features using the Bag of Graphs [Silva et al. 2014] method.
3. **Train and Validation of Classifiers**: Train and Test of various classifiers using specific metrics.
4. **Test and choice of Classifiers**: Test of trained classifiers and choice of best ones.

4. First Results, Conclusion and Next Steps

At the time of this poster, we have assembled a dataset of approximately 19,000 cryptographic misuse source codes that are yet to be categorized according to [Braga and Dahab 2016]. Collect data is difficult due the lack of datasources. However, this work is only at the beginning of its development, but with great prospects The next steps will be to implement the Bag of Graphs [Silva et al. 2014] methods for vectoring.

5. Acknowledgements

We thank CAPES (Coordination for the Improvement of Higher Education Personnel) and LASCA (Laboratory of Security and Cryptography) for all support.

References

Braga, A. and Dahab, R. (2016). Mining cryptography misuse in online forums. In *2016 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, pages 143–150. IEEE.

Braga, A., Dahab, R., Antunes, N., Laranjeiro, N., and Vieira, M. (2017). Practical evaluation of static analysis tools for cryptography: Benchmarking method and case study. In *2017 IEEE 28th International Symposium on Software Reliability Engineering (ISSRE)*, pages 170–181. IEEE.

Lazar, D., Chen, H., Wang, X., and Zeldovich, N. (2014). Why does cryptographic software fail?: a case study and open problems. In *Proceedings of 5th Asia-Pacific Workshop on Systems*, page 7. ACM.

Nadi, S., Krüger, S., Mezini, M., and Bodden, E. (2016). Jumping through hoops: Why do java developers struggle with cryptography apis? In *Proceedings of the 38th International Conference on Software Engineering*, pages 935–946. ACM.

Silva, F. B. et al. (2014). Bag of graphs = definition, implementation, and validation in classification tasks. *Repositório Unicamp. Universidade Estadual de Campinas, Instituto de Computação, Campinas, SP. Available in: http://www.repositorio.unicamp.br/handle/REPOSIP/275527. Access in: 25 aug. 2018.*

Acesso Universal em Sistemas Socioenativos*

Andressa Cristina dos Santos, Julio Cesar dos Reis

[1]Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
CEP 13083-852 – Campinas – SP – Brasil

andressacs.cc@gmail.com, jreis@ic.unicamp.br

Resumo. *O processo de design focado no acesso universal deve ser guiado por um conjunto de recomendações relevantes para melhorar design e avaliação de interação. Apresentamos uma análise de dois estudos de casos aplicados a um cenário de sistemas socioenativos. Buscamos a combinação de métodos de avaliação existentes visando a criação de um instrumento aplicado a esses novos sistemas.*

1. Introdução

Na computação ubíqua e pervasiva, assim como em outros ambientes computacionais contemporâneos, a interação passou a fazer uso de diversos dispositivos e sensores. A maneira como o *design* desses sistemas é conduzido afeta a interação e a facilidade de uso. Nesse contexto, Sistemas Socioenativos[1] exploram um conceito novo que relaciona a presença de novas tecnologias e novas formas de interação, aliada à onipresença da computação apresentando desafios que exigem a consideração de novos fatores no projeto de sistemas interativos [1].

Desta forma, os instrumentos de avaliação existentes não capturam todos os aspectos intrínsecos desses novos cenários, tais como a ampla gama de características, necessidades do usuário e a tecnologia envolvida. Desse modo, busca-se a criação e aplicação de um instrumento de avaliação que proporcione o acesso universal aplicados às tecnologias da informação atendendo a qualquer pessoa, em qualquer lugar e a qualquer momento [4]. Através de nosso instrumento, o processo de design focado no acesso universal será guiado por um conjunto de recomendações relevantes para melhorar o design e a avaliação da interação desses sistemas, permitindo que designers beneficiem ainda mais pessoas.

2. Metodologia

Dois estudos de caso foram realizados no Hospital - Sociedade Brasileira de Pesquisa e Assistência para Reabilitação Craniofacial - SOBRAPAR. As atividades realizadas foram aprovadas pelo Comitê de Ética em Pesquisa da Unicamp[2]. Um dos estudos de caso [3] ocorreu em uma oficina realizada em dezembro de 2018. Participaram do estudo 6 crianças com idades entre 7 e 11 anos, seus responsáveis, profissionais do hospital e 8 pesquisadores de Interação Humano Computador.

*Agradecimentos ao Laboratório de Interação Humano-Artefato Digital e ao Hospital SOBRAPAR. Este trabalho tem apoio financeiro da FAPESP (projeto Temático #2015/16528-0), da Pró-Reitoria de Pesquisa da UNICAMP (processo nº 2018/2132) e da CAPES - Código de Financiamento 001.

[1]Investigação conduzida no contexto de um Auxílio à Pesquisa - Linha de fomento Temático na FAPESP #2015/16528-0.

[2]CAAE 72413817.3.0000.5404

Como instrumento de análise foram utilizados os princípios de Design Universal (DU) [6] e as Heurísticas Naturais de Usuário (NUI) [5], cada um deles analisados por 2 pesquisadores através de uma escala de conformidade. O segundo estudo de caso [2] foi realizado através da análise dos vídeos desta mesma oficina. Contudo, foram analisados os Princípios de Afetibilidade (PAff) [7] a fim de verificar se a afetividade pode contribuir com a promoção do acesso universal em sistemas socioenativos. A análise foi conduzida com o objetivo de observar se tal sistema proporcionou afetividade na interação.

3. Resultados

Através do primeiro estudo de caso, apresentamos uma lista de verificação, que contém um total de 40 itens. Sugerimos usá-lo para a avaliação através de uma escala de conformidade onde pode ser marcado um valor de correspondência com o que está sendo observado. Já a análise dos Princípios de Afetibilidade revelou o potencial de sua aplicabilidade aos sistemas socioenativos. Além disso, geramos uma lista de recomendações a serem utilizados no design de sistemas voltados para cenários desses sistemas.

Trabalhos futuros envolvem a aplicação das recomendações e sua análise em outros cenários de sistemas socioenativos. Visamos combinar essas recomendações com outras adicionais que consideram o Design Universal para obter instrumentos que possam contribuir no projeto e avaliação de sistemas socioenativos universalmente acessíveis.

Referências

[1] Baranauskas M C C. Sistemas sócio-enativos: Investigando novas dimensões no design da interação mediada por tecnologias de informação e comunicação. In *FAPESP Thematic Project (2015/165280)*. FAPESP, 2015.

[2] Santos A C, Muriana L M, Pimenta J R O G, Silva J V da, Moreira E A, and Reis J C. Investigating aspects of affectibility for universal access in socioenactive system scenarios. In *Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems*, page 33. ACM, 2019.

[3] Santos A C; Maike V R M L; Mendoza Y L M; Silva J V; Bonancin R; Reis J C and Baranauskas M C C. Inquiring evaluation aspects of universal design and natural interaction in socioenactive scenarios. In *International Conference on Human-Computer Interaction*. Springer, 2019.

[4] Emiliani P L and Stephanidis C. Universal access to ambient intelligence environments: opportunities and challenges for people with disabilities. *IBM Systems Journal*, 44(3):605–619, 2005.

[5] Maike V R M L, Neto L S B, Goldenstein S K, and Baranauskas M C C. Heuristics for nui revisited and put into practice. In *International Conference on Human-Computer Interaction*, pages 317–328, Cham, 2015. Springer, Springer International Publishing.

[6] Connell B R, Jones M, Mace R, Mueller J, Mullick A, Ostroff E, Sanford J, Steinfeld E, Story M, and Gregg Vanderheiden. The principles of universal design. *http:www.ncsu.eduncsudesigncudabout_ududprinciplestext.htm*, 1997.

[7] Hayashi E C S and Baranauskas M C C. Designing for affectibility: Principles and guidelines. In Constantine Stephanidis, editor, *HCI International 2015 - Posters' Extended Abstracts*, pages 25–31, Cham, 2015. Springer International Publishing.

Modelo de *Machine Learning* para processamento de *Big Data* em *Fog Computing* aplicada à *Smart Cities*

Matteus Vargas[1], **Luiz Fernando Bittencourt**[1]

[1]Instituto de Computação – Universidade Estadual de Campinas (Unicamp)
CEP 13083-852 – Campinas – SP – Brazil

vargas.simao@gmail.com, bit@ic.unicamp.br

Abstract. *The massive spread of IoT devices, wich generate large amounts of data, has created many "smart environments". Although more convenient, Cloud can run into limited power, low bandwidth, or high latency issues. Including distributed learning algorithms in this case acting on devices or on network edge layers such as Fog Computing may be an alternative to address these issues. This concept can be extended to Smart Cities on a larger scale. This paper proposes a distributed learning model implemented in Fog Computing aimed to partition data of a mobile user, process, aggregate and return a response, with lower bandwidth and latency usage.*

Resumo. *A difusão maciça de dispositivos de IoT, geradores de grande quantidade de dados, tem criado muitos "ambientes inteligentes". Apesar de mais conveniente, a Cloud pode esbarrar nas questões de energia limitada, pouca largura de banda ou alta latência. A inclusão de algoritmos de aprendizado distribuído atuando nos dispositivos ou em camadas na borda da rede, como Fog Computing, pode ser uma alternativa para enfrentar essas questões. Esse conceito pode ser estendido para as Smart Cities, em uma escala maior. Este trabalho propõe um modelo de aprendizado distribuído implementado em Fog Computing que visa particionar os dados de um usuário móvel, processar, agregar e devolver uma resposta, com menor uso de largura e banda e menor latência.*

1. Introdução

A difusão de dispositivos de IoT tem criado muitos "ambientes inteligentes"[Vincentelli 2015]. Esses dispositivos são geradores de dados, assim enormes quantidades de dados serão geradas na borda da rede e conhecimento deve ser extraído disso. A *Cloud* pode parecer a solução mais conveniente para a análise em IoT, com alto volume, velocidade e heterogeneidade. Contudo, a transmissão de todos os dados para a *Cloud* esbarra em energia limitada [Valerio et al. 2017], largura de banda ou alta latência [Stolpe 2016].

Os aplicativos com restrição de comunicação requerem algoritmos de análise distribuídos que, em parte, trabalham diretamente nos dispositivos que geram os dados, como sensores e dispositivos incorporados ou encaminhando para camadas anteriores como a própria *Fog Computing*, na borda da rede [Garcia Lopez et al. 2015].

Cenários onde as análises calculadas sobre esses dados podem ser relevantes apenas por um curto período de tempo e em locais específicos, não é preciso grandes movimentações de dados, evitando desperdício de largura de banda. Essa abordagem visa

cenários limitados, mas pode vir a ser estendida para as *Smart Cities* [Valerio et al. 2017]. Este trabalho propõe um modelo para particionamento dos dados em *Smart Cities* com aplicação de aprendizado distribuído focado em usuários móveis.

2. Metodologia

Foi aplicada uma revisão bibliográfica que contempla a delineação do cenário e a problemática. O intuito foi procurar na trabalhos que explorem a aplicação de *Machine Learning* (ML) Distribuído para análise de *Big Data* em Ambientes de *Smart Cities*. Por fim, foram filtrados trabalhos que fossem de 2017 em diante, focado em artigos. A *string* ficou da seguinte forma:

- *"smart city"AND "distributed learning"*

As bases escolhidas foram a *Scopus*, IEEE, ACM, *Science Direct* e *Springer*. Foram feitas a checagem de duplicidades e a análise de título e *abstract*, fechando com 21 (vinte e um) artigos. Isso foi o bastante para identificar questões a serem trabalhadas.

3. Proposta

A análise dos artigos trouxeram questões em aberto [Stolpe 2016]. A principal é como manter a eficiência de comunicação, preservando ao mesmo tempo a precisão de seus equivalentes centralizados e consumindo menos largura de banda? Quanta informação deve ser no mínimo comunicada para aprender com êxito com os dados particionados? Como definir "hiper parâmetros"para que não haja conflitos ou repetições de dados?

A proposta aqui é utilizar a camada de *Fog Computing* para realizar o aprendizado distribuído, com mais recursos computacionais (*Cloudlets*), visando escalar para uma *Smart City*, com a premissa de que os dados podem ser quebrados e processados separadamente com ML distribuído na borda da rede [Valerio et al. 2017].

O destaque fica por conta dos usuários móveis. Enquanto eles se movem, as *Cloudlets* ao longo do trajeto capturam partes dos dados, processam, agregam e dão respostas. A princípio, as métricas possíveis de serem avaliadas aqui, do ponto de vista de rede, são o uso da largura de banda e a latência. A busca por métodos de ML Distribuídos mais adequados e a definição se a proposta será implementada ou simulada, bem como as ferramentas necessárias para ambos os casos, são passos futuros.

Referências

Garcia Lopez, P., Montresor, A., Epema, D., Datta, A., Higashino, T., Iamnitchi, A., Barcellos, M., Felber, P., and Riviere, E. (2015). Edge-centric computing: Vision and challenges. *ACM SIGCOMM Computer Communication Review*, 45(5):37–42.

Stolpe, M. (2016). The internet of things: Opportunities and challenges for distributed data analysis. *ACM SIGKDD Explorations Newsletter*, 18(1):15–34.

Valerio, L., Passarella, A., and Conti, M. (2017). A communication efficient distributed learning framework for smart environments. *Pervasive and Mobile Computing*, 41:46–68.

Vincentelli, A. S. (2015). Let's get physical: Adding physical dimensions to cyber systems. In *2015 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED)*, pages 1–2. IEEE.