Análise de mensagens associadas à cibersegurança em redes IRC

Rodrigo Campiolo[1,2], **Daniel Macêdo Batista**[1]

[1]Instituto de Matemática e Estatística (IME)
Universidade de São Paulo (USP)
São Paulo – SP – Brasil

[2]Departamento de Ciência da Computação
Universidade Tecnológica Federal do Paraná
Campo Mourão – PR – Brasil

rcampiolo@utfpr.edu.br, batista@ime.usp.br

Abstract. *This paper presents the analysis of information related to computer security in chat rooms of Internet Relay Chat (IRC) networks. We monitored and analyzed channels about computer security and hackers activities. So, we proposed an approach for extracting warnings of security and malicious activities by extending simple search mechanisms using natural language processing and information retrieval. As a result, we identified messages related to network and computer security and proposed and developed mechanisms for extracting relevant notifications as threat alerts and security breaches. Our mechanisms can be used as sources of security alerts in existing early warning systems, improving their efficacy.*

Resumo. *Este artigo apresenta a análise e extração de informações de conteúdos discutidos em salas de Internet Relay Chat (IRC) relacionados à segurança de computadores. Foram monitorados e analisados canais que abordam segurança de sistemas computacionais e atividades suspeitas. A partir da análise, é proposta uma abordagem para extração de alertas de segurança, que estende mecanismos de monitoramento usando palavras-chave com métodos compostos por heurísticas, processamento de linguagem natural e recuperação de informação. Como resultados, obteve-se a identificação de mensagens relacionadas à segurança de redes de computadores e mecanismos para a extração de notificações relevantes, como alertas de ameaças e falhas de segurança.*

1. Introdução

A rede *Internet Relay Chat* (IRC) [Oikarinen and Reed 1993] foi uma rede social usada intensivamente por inúmeros usuários para a troca de mensagens instantâneas até meados de 2000. Devido ao surgimento de outras redes, o número de usuários diminuiu consideravelmente. No entanto, por possuir uma estrutura de rede descentralizada, a rede IRC acaba por ser usada para a realização de atividades ilegais, já que a arquitetura da rede dificulta o monitoramento do conteúdo por especialistas em segurança. Há diversos casos em que a rede é usada para controle de *botnets* e troca de conteúdos maliciosos (códigos de exploração, vulnerabilidades e organização de ataques) [Michels 2012]. Logo, monitorar e identificar essas ameaças é importante para a geração de alertas antecipados e para a mitigação de ações maliciosas antecipadamente, já que mesmo antes de um ataque começar, ele pode ser discutido inicialmente no IRC.

Entretanto, o monitoramento de fontes de informações não estruturadas, como é o caso da rede IRC, enfrenta problemas por conta da sintaxe e semântica das mensagens, bloqueio de monitoramento, confiabilidade da informação compartilhada, ocultação de grandes operações maliciosas por meio de canais privados e/ou protegidos, entre outros. Por isso, para conseguir extrair informações relevantes para identificação de ameaças que ocorrem ou que ocorreram recentemente, é necessário analisar e caracterizar o conteúdo discutido nessa mídia e, dessa forma, elaborar mecanismos que possibilitem destacar e priorizar informações relevantes.

Este trabalho analisa as mensagens postadas em canais abertos da rede IRC que estão associados à proteção e a ataques de sistemas e redes de computadores, identifica padrões para a especificação de heurísticas que viabilizam a extração de conteúdo relevante para a geração de alertas antecipados, e também identifica padrões que possibilitem caracterizar os tipos de atividades e usuários que podem estar associados a atividades suspeitas. Como principal contribuição, é proposto um arcabouço para monitorar, coletar, analisar e priorizar as informações relevantes como notificações de segurança.

2. Trabalhos Relacionados

Muitos trabalhos relacionados a redes IRC têm investigado a organização e detecção de *botnets*, redes de programas controladas remotamente e usadas geralmente para fins maliciosos [Mazzariello 2008, Lu and Ghorbani 2008, Wang et al. 2009, Ma et al. 2010, Houmansadr and Borisov 2013, Carpine et al. 2013]. Além disso, há também o monitoramento e investigação de outros tipos de atividades, como por exemplo, estruturação de ataques a organizações e atividades criminosas [Décary-Hétu et al. 2014]. Nosso trabalho segue uma linha próxima, mas se destacando pela análise de uma base de mensagens resultante de monitoramento de diferentes canais e pela proposta de um arcabouço para a identificação de notificações e conteúdos associados à segurança de sistemas e redes de computadores.

[Brown 2007] propõe a arquitetura de uma ferramenta automatizada para investigações de roubo de identidade na rede IRC. O trabalho estrutura a arquitetura em cinco módulos (coleta, armazenamento, análise, alerta e localizador) e discute princípios para a implementação de cada módulo. No módulo de análise é proposto o uso de algoritmos de mineração de dados e análise de palavras-chave, frases-chave e expressões regulares. Como apenas modela a ferramenta, apesar de destacar as dificuldades e detalhes de cada fase, o autor não apresenta prova de conceito da arquitetura. Em nosso trabalho, propomos e realizamos uma avaliação de um arcabouço por meio de um protótipo e uma base de mensagens de diferentes canais.

[Michels 2012] implementa e analisa uma ferramenta automatizada para auxiliar investigadores na análise de mensagens em tempo real no IRC. A coleta de informações é realizada durante 1 minuto em cada canal considerado suspeito. A análise consiste em identificar as palavras-chave, encontrar tópicos de interesse e realizar a análise de categorias. O investigador é alertado pela ferramenta se for encontrado conteúdo suspeito. Segundo o autor, a parte mais complexa foi identificar canais abertos que estejam abordando atividades suspeitas. Diferente de Michels, monitoramos canais abertos e com potenciais mensagens de segurança. Além disso, utilizamos identificação de entidades e um esquema próprio de etiquetamento de mensagens que foi elaborado a partir da análise

de canais com tópicos associados à cibersegurança.

Os autores de [Gainaru et al. 2010] usam processamento de linguagem natural, clusterização e análise de conhecimento para analisar sessões de conversação. O trabalho aborda a detecção de tópicos, identificação de cadeias léxicas e resolução de correferência. Concluem que os resultados podem ser melhorados por uso de heurísticas, em especial, a detecção de tópicos. Em contraste a Gainaru et al., analisamos vários canais e exploramos o uso de heurísticas elaboradas a partir de nossa análise.

Os autores de [Iqbal et al. 2012] realizam a extração de associações e identificação de tópicos a partir da análise dos registros de sessões de chat obtidas a partir de máquinas apreendidas para investigação criminal. Para tal, realizam a divisão dos registros em períodos temporais para identificação de associações entre os participantes e, em seguida, a identificação de tópicos nesses períodos. Os autores destacam a dificuldade na análise das sessões devido ao tamanho e informalidade das mensagens, além dos erros de grafia. Em nosso trabalho realizamos a extração considerando uma base resultante de monitoramento e, por termos caracterizado as mensagens dos canais de segurança, usamos gírias, acrônimos e ofensas para auxiliar no processo de filtro e classificação de relevância.

Os autores de [Décary-Hétu and Dupont 2012] identificam potenciais suspeitos de atividades hackers pela análise de sessões IRC obtidas a partir de computadores apreendidos. Apesar de analisarem comunicações privadas, a análise de redes sociais possibilita eleger os grupos que merecem mais atenção no monitoramento. Em nosso trabalho, identificamos usuários suspeitos ou com intenções de executar ataques, mas consideramos o monitoramento de canais.

Em [Benjamin and Chen 2014], os autores direcionam seus esforços em uma abordagem proativa e na proposta de novas metodologias para compreender a ação dos hackers e ameaças emergentes. Reforçam a ideia de que hackers visitam diversas comunidades para melhorar suas habilidades simplesmente consumindo os recursos compartilhados nessas comunidades. Os canais IRC e fóruns são os principais locais usados por comunidades hackers para divulgar suas ações e recursos. Eles analisaram os usuários do canal *anonops* durante seis meses e constataram que as atividades ilegais são discutidas em canais privados enquanto a maior parte das mensagens são relacionadas a tecnologia em geral. Em contrapartida, por analisarmos diferentes canais, nossos resultados mostraram que há informações relevantes para identificação de ameaças emergentes, como por exemplo, alvos de ataques.

Nossa proposta inova ao investigar características para a identificação de informações que podem ser usadas como alertas, preferencialmente antecipados, nas mensagens postadas em canais de IRC. Difere-se das demais propostas similares pelo monitoramento constante de diferentes canais e pela proposta de um arcabouço de extração de informações associadas à cibersegurança, que foi desenvolvida considerando a combinação de técnicas de processamento de linguagem natural e recuperação da informação adaptadas segundo os resultados da análise dos canais.

3. Metodologia

A metodologia consiste em monitorar canais de IRC sobre segurança e sobre atividades suspeitas e, em seguida, analisar os dados para identificar mensagens que possam ser

usadas como alertas para administradores. Após a análise, são propostas heurísticas para extrair as potenciais mensagens de interesse. Ao final, as heurísticas são avaliadas em um conjunto de dados mais amplo.

Para investigar as mensagens coletadas no IRC, foram estabelecidas as seguintes questões de pesquisa:

Q1 Há compartilhamento de informações que podem ser usadas como alertas por administradores de redes?
Q2 É possível identificar alvos de ataques nos canais de atividades suspeitas?
Q3 Como automatizar o processo de identificação de alertas nas redes IRC?

A Figura 1 apresenta o processo para a investigação das questões de pesquisa. Em seguida, cada um dos passos desse processo são explicados.

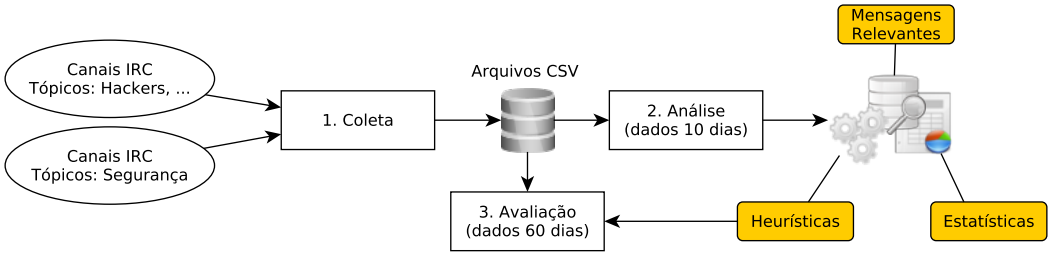

Figura 1. Método de pesquisa

Coleta de dados (1)

A fase de coleta de dados é responsável por capturar as mensagens postadas nos canais monitorados e armazenar em um formato padrão (CSV) na base de dados. Os canais monitorados foram selecionados a partir da popularidade do servidor IRC e pela quantidade de usuários. A seleção ocorreu por pesquisa Web e observações preliminares em servidores de IRC. Os canais com quantidade pequena de usuários foram descartados devido à facilidade na identificação do software de coleta, o que provavelmente levaria a um banimento e à falha no monitoramento. Os canais monitorados foram de dois servidores de IRC: freenode (irc.freenode.org) e anonops (irc.anonops.com). No freenode, foram monitorados os canais com tópicos relacionados à segurança: #security, #networking, #owasp e #oss-security. No anonops, foram monitorados os canais relacionados a atividades suspeitas: #anonops, #hackers, #ddos, #opnewblood, #defacement e #opferguson.

O monitoramento foi realizado a partir de duas redes distintas para dificultar a identificação do software de coleta. Os resultados de cada canal foram armazenados em arquivo CSV contendo as informações: horário, operação, identificação do usuário e mensagem. As operações monitoradas foram JOIN (usuário entra no canal), LEAVE (usuário deixa o canal) e MESSAGE (mensagem postada publicamente no canal). As mensagens privadas não são possíveis de monitorar, exceto as enviadas diretamente ao software de coleta.

A coleta foi realizada em diferentes períodos. No período de novembro/2014 a dezembro/2014, foram coletados os dados para a análise durante duas semanas. Nessas

duas semanas foram selecionados de 8 a 12 dias com dados completos para a análise. No período de março/2015 a maio/2015, foram coletados os dados para a avaliação dos mecanismos para extração de mensagens relevantes como alertas para administradores de rede. Nesses meses foram selecionados 60 dias de coleta que se apresentava completa. Durante o período mais longo, o monitoramento foi suspenso temporariamente (10 dias) devido à identificação do software de coleta e, por consequência, ao bloqueio de acesso ao servidor.

Análise de dados (2)

A análise de dados ocorreu essencialmente sobre os usuários e mensagens coletadas em 10 dias. O objetivo da análise foi caracterizar os resultados para possibilitar a criação de técnicas para extrair conteúdos associados à cibersegurança.

As etapas de análise foram:

- normalização: normalizar e filtrar os dados para isolar usuários e mensagens.
- estatísticas: gerar os dados estatísticos e gráficos sobre participação dos usuários e números de mensagens.
- identificação de mensagens relevantes: realizar a busca por potenciais mensagens que caracterizam informações de interesse à segurança de computadores.

Códigos foram escritos para normalizar os dados e gerar as estatísticas. A identificação de mensagens relevantes consistiu na leitura e etiquetamento manual das mensagens coletadas. Algumas técnicas foram usadas para excluir mensagens irrelevantes, como por exemplo, desconsiderar mensagens curtas, palavras irrelevantes (*stop words*) e eliminação de mensagens postadas por robôs.

Durante a análise de dados, também foram elaboradas heurísticas para caracterizar as potenciais mensagens relevantes. As heurísticas, estatísticas e algumas mensagens relevantes são apresentadas na Seção 4. Os resultados dessa fase auxiliam na investigação das questões de pesquisa Q1 e Q2.

Avaliação dos mecanismos (3)

Os mecanismos foram avaliados considerando o resultado produzido pelas implementações das heurísticas, técnicas de recuperação de informação e processamento de linguagem natural. Também é avaliada a proposta de um arcabouço para a extração de mensagens associadas à cibersegurança. Os resultados produzidos pelo arcabouço auxiliam diretamente na investigação da questão de pesquisa Q3.

4. Resultados e Discussão

Esta seção apresenta os resultados da coleta, processamento e avaliação dos dados coletados nos canais IRC selecionados para nosso estudo. Primeiramente, apresenta-se a caracterização da coleta, ou seja, a descrição e investigação dos dados coletados. Em seguida, é proposto um arcabouço para extração de mensagens associadas à cibersegurança em redes IRC. E, por fim, são avaliadas as questões de pesquisas.

4.1. Caracterização da Coleta

A caracterização da coleta consistiu em realizar diferentes análises visando extrair e priorizar informações de interesse como alertas em redes IRC. As mensagens foram investigadas por análise estatística, frequência de palavras, relações entre termos, conteúdo das URLs, identificação de entidades e correlação de padrões com outras fontes.

4.1.1. Análise estatística

A Tabela 1 apresenta a estatística descritiva dos dados coletados para a análise, isto é, os resultados da média de 10 dias de coleta. Os canais *oss-security* e *owasp* não foram considerados por contabilizarem apenas 17 mensagens no período e os canais *defacement* e *hackers* foram monitorados somente para a fase de Avaliação.

Tabela 1. Caracterização da coleta de dados no IRC (fase Análise)

Canais	Período (dias)	Total de mensagens	Média diária de mensagens	Desvio padrão (mensagens)	Média diária de usuários ativos	Desvio padrão (usuários)
anonops	12	39921	3326,75	740,44	162,25	17,09
ddos	12	5890	490,83	230,67	50,33	12,09
opferguson	8	28225	3525,13	2295,01	143,13	82,24
opnewblood	12	14578	1224,83	409,54	100,42	14,18
security	8	11352	1419,00	528,86	95,25	13,73
networking	8	22452	2806,50	495,82	112,00	10,90

Na Tabela 1, observa-se que o maior desvio padrão proporcional ao número médio de mensagens e usuários está associado ao canal *opferguson*. Isso acontece pelo fato do canal estar associado a um conjunto de ações hackers em protesto ao assassinato de um jovem negro desarmado por policial em Ferguson (EUA)[1]. Logo, um assunto não apenas de interesse de um grupo menor, mas relacionado à sociedade. Em canais do servidor anonops, é comum o uso do prefixo "op" para indicar uma orquestração de ataque visando um alvo ou em favor de causas políticas, sociais e/ou econômicas. Logo, é comum no IRC encontrar canais que começam com esse prefixo, como é o caso do *opnewblood* voltado à orientação de iniciantes, embora este último seja um caso a parte já que é um canal que permanece ativo no servidor.

Na maioria dos canais monitorados, foi observado um número elevado de mensagens diárias que, em sua grande maioria, são sobre tecnologia ou assuntos do cotidiano dos participantes. No entanto, é interessante notar que mesmo nestes canais, é possível acessar tutoriais e ferramentas para execução de ataques através de um simples comando. Neste caso, destacam-se os canais *ddos* e *opnewblood* que possuem usuários automatizados para indicar ferramentas e tutoriais didáticos para ataques. Já nos canais de segurança/rede, os usuários acabam comentando sobre assuntos relacionados a vulnerabilidades e formas de exploração.

A Tabela 1 não apresenta o número total de usuários diários devido à alta flutuação de entrada e saída de usuários nas salas. Em geral, a maior parte dos usuários não interage publicamente, especialmente nos canais de segurança. No canal *security*, o número de usuários ultrapassa 1000, mas a interação ocorre em média entre um décimo dos participantes.

A Tabela 2 apresenta a estatística descritiva dos dados coletados para a fase Avaliação, isto é, os resultados da média de 60 dias de coleta. Os canais *hackers* e *defacement* começaram a ser monitorados 10 dias após os outros, logo apresentam um número reduzido de dias de coleta. O monitoramento do canal *opferguson* foi descontinuado devido ao número pequeno de usuários e mensagens.

[1] http://rt.com/usa/179532-anonymous-op-ferguson-missouri/

Tabela 2. Caracterização da coleta de dados no IRC (fase Avaliação)

Canais	Período (dias)	Total de mensagens	Média diária de mensagens	Desvio padrão (mensagens)	Média diária de usuários ativos	Desvio padrão (usuários)
anonops	58	253522	4405,5	1202,22	162,05	22,47
ddos	58	22695	391,29	231,7	32,74	10,51
defacement	49	5350	109,18	58,51	8,35	3,51
hackers	50	19911	398,22	314,2	30,64	10,58
opnewblood	58	48369	833,95	322,44	73,57	15,81
security	68	118360	1740,58	592,74	90,04	15,17
networking	58	131310	2263,97	579,23	108,40	14,75

Na Tabela 2, o desvio padrão dos canais *ddos*, *defacement* e *hackers* é proporcionalmente maior devido à existência de usuários automatizados para fornecer ajuda por meio de tutoriais e indicação de ferramentas de ataques. O número de usuários ativos indica o interesse nos canais suspeitos de atividades maliciosas, mas também nos canais de segurança. Por meio de um histograma de frequências relativas, na prática, verificou-se que nesses canais, 20% dos usuários são responsáveis por mais de 70% das mensagens.

Comparando os dados das tabelas 1 e 2, foi observado que as relações médias entre mensagens e usuários dos canais comuns mantiveram-se uniformes, mesmo considerando um período de coleta maior.

4.1.2. Análise de frequência de palavras

A análise de frequência de palavras teve por objetivo caracterizar o vocabulário usado nos canais. Foram identificados termos para a remoção de mensagens irrelevantes e priorização das mensagens relevantes como alertas.

Cloud tags foram geradas para confirmar a suspeita do uso de termos mais informais nos canais suspeitos e de uma linguagem mais formal nos canais de segurança. Nos canais de segurança receberam destaque termos tecnológicos e assuntos relacionados à segurança de redes e computadores. Nos canais de atividades suspeitas receberam destaque ofensas, gírias e acrônimos. Os resultados também auxiliaram na definição de categorias de interesse para os termos. Foram definidas as seguintes categorias:

- acrônimos/gírias: acrônimos e gírias usados na Internet.
- segurança: termos associados à segurança de sistemas.
- atividades maliciosas/ameaças: termos associados às atividades suspeitas e ameaças à segurança de sistemas.
- termos frequentes (hackers): termos frequentes não comuns relacionados aos tópicos nos canais suspeitos.
- termos frequentes (security): termos frequentes não comuns relacionados aos tópicos nos canais de segurança.
- ofensas: termos com caráter ofensivo.

A categorização foi realizada manualmente e considerou os primeiros 3000 termos mais frequentes em cada canal. A Tabela 3 apresenta uma visão geral dos termos em cada categoria: O acrônimo “pm” significa “private message” e ocorreu algumas vezes em intenções de atividades suspeitas. Os termos “patch” e “cve” ocorreram em discussões de vulnerabilidades recentes. Os termos “ddos”, “down” e “target” ocorreram em

Tabela 3. Classificação dos termos em categorias

Categoria	Total	Amostra de termos
acrônimos/gírias	48	lol, xd, lmao, ur, pm, wtf, idk, ...
segurança	32	security, patch, cve, firewall, ssl, pentesting,...
atividades maliciosas / ameaças	124	ddos, dos, down, bonet, attack, target, injection, ...
termos frequentes (hackers)	401	off, police, top, site, windows, linux, server, government, ...
termos frequentes (segurança)	436	new, over, windows, server, linux, ip, problem, nsa, ...
ofensas	35	shit, idiot, bitch ...

situações de orquestração de ataques e notificação de sucesso de ataque. O termo "ddos" também ocorreu em mensagens irrelevantes, como em solicitações de informações sobre ferramentas. Os termos "police" e "government" estão associados a causas que o grupo *Anonymous* estava atuando na época da coleta. O termo "new" esteve associado a mensagens de novos códigos de exploração ou novas vulnerabilidades. Os termos de ofensas, em geral, estavam associados a mensagens irrelevantes como alertas.

4.1.3. Mineração de associação de palavras

A mineração de associação de palavras possibilita identificar relações paradigmáticas, sintagmáticas e *collocations* entre palavras dentro de um contexto [Manning and Schütze 1999]. São interessantes para prover variações em consultas em recuperação de texto e para auxiliar na identificação de tópicos e entidades. As relações paradigmáticas identificam palavras que podem ser substituídas por outras de uma mesma classe e mantêm o significado da sentença. As relações sintagmáticas relacionam palavras semanticamente. *Collocation* é uma expressão composta por dois ou mais termos comumente usados para expressar algo.

A identificação de associações de palavras ocorreu usando os métodos de extração de bigramas e trigramas. Foram selecionados os 200 melhores escores para três métricas de pontuação: Frequência, *Pointwise Mutual Information* (PMI) e Razão de Verossimilhança (*Likelihood Ratio*). Cada uma das associações foi analisada manualmente por meio de consulta às mensagens que as continham e próximas, no caso, as N anteriores e N posteriores (na pesquisa foi usado N=2). A Tabela 4 apresenta algumas das associações de interesse para o monitoramento.

Tabela 4. Amostra de associações relevantes para monitoramento

	Expressão	Exemplo de mensagem
1	take down	lets **take down** www.gadgetwide.com
2	taking down	I need help **taking down** this site but I think there is not enough people here
3	get involved	anyone want to **get involved**?
4	tango down	**tango down** target www.infotec.be
5	new target	send a **new target**
6	.check <URL>	**.check** *www.slmpd.org* / **.check** *www.micheldestot.fr*
7	.dns <URL>	**.dns** *thegrapevine.cc*
8	0day ... <SOFTWARE>	... play around with this new super-**0day** for *Windows*?"

As linhas de 1 a 5 da Tabela 4 apresentam *collocations* que, ao serem pesquisadas, devolveram informações relevantes como alertas nas próprias mensagens ou próximas. Outros bigramas (linhas 6 e 7) estavam sempre relacionando os termos *.check* e *.dns* com

diferentes URLs. Logo, caracterizaram uma relação sintagmática entre o termo e uma URL. No entanto, como a URL pode ser substituída por qualquer outra da mesma classe, também caracterizou uma relação paradigmática. A linha 8 foi obtida a partir de um trigrama, mas tratada como um bigrama, pois o "for" pode ser substituído por outros termos ou vazio.

É importante ressaltar que nem sempre uma busca por uma associação devolve bons resultados. Por exemplo, ao buscar por *"take down"* são retornadas mensagens fora do contexto como *"... if you wanna take down a country"* ou *"Anything going on about TPB being take down ..."*. Logo, não basta apenas verificar a presença da associação para filtrar efetivamente mensagens de interesse como alertas. No entanto, são de grande auxílio para diminuir o escopo de análise para outros mecanismos ou para um analista.

4.1.4. Análise de URLs

A análise de URLs é usada para identificar potenciais alvos, orquestrações de ataques e compartilhamento de informações nos canais de IRC. Observou-se que usuários postam URLs para indicar alvos de ataque, para compartilhar novas técnicas e códigos, e também para apontar a existência de vulnerabilidades documentadas. Por outro lado, também são usadas para propagar informações sem relevância, como por exemplo, notícias, vídeos, imagens e ofensas. Logo, ao filtrar essas URLs é reduzido o escopo de busca por potenciais ameaças ou alertas.

O processamento das URLs envolveu os seguintes passos: (1) extração e normalização das URLs; (2) agrupamento de URLs para identificar domínios e recursos relevantes e irrelevantes; (3) remoção das URLs irrelevantes; (4) uso de outros mecanismos para evidência ou exclusão de URLs; (5) inspeção por especialista da lista final de potenciais URLs com informações relevantes.

No passo 1, a extração e normalização das URLs foi realizada com o auxílio da API twitter-text[2]. No passo 2, sítios Web de notícias, entretenimento, pornografia e outros foram destacados e incluídos em uma lista de URLs para remoção. Sítios Web de compartilhamento, em especial de texto puro, foram destacados e incluídos em uma lista de URLs relevantes. Nesse passo também foram definidas heurísticas para a remoção de URLs, como por exemplo, URLs com data ou termos "news", "article", "story" e nomes longos separados por hífen, pois geralmente não caracterizam alertas ou ameaças. No passo 3, foram removidas as URLs irrelevantes e armazenadas em uma base para inspeção e análise de falsos negativos. No passo 4, foram aplicadas as seguintes heurísticas: acesso ao título do sítio e verificação por termos associados à segurança ou atividades suspeitas; uso de serviço de categorização[3] de sítios Web para identificar sítios irrelevantes; e análise do contexto das mensagens que contém a URL e das mensagens próximas. Esse passo não foi explorado intensivamente, apenas foi realizado como prova de conceito para futuras implementações. O passo 5 consistiu em averiguar os resultados (Tabela 5).

Nos resultados da aplicação dos passos 1 a 3, observamos reduções significativas no número de URLs em alguns canais. Por exemplo, no canal *anonops* a redução foi de

[2]https://github.com/twitter/twitter-text

[3]https://developer.similarweb.com/website_categorization_API

Tabela 5. Resumo do processamento de URLs (Passos 1 a 3)

Canal	Total de URLs	URLs - Inspeção	URLs - Relevantes
anonops	1229	233	11
ddos	219	155	8
opferguson	856	311	15
opnewblood	156	66	2
security	404	247	12
networking	667	332	15

aproximadamente 81%. No entanto, considerando a média de 10 dias de coleta, se totalizarmos o número de URLs para a inspeção, o valor resultante diário de URLs (134) pode ser considerado uma carga de trabalho alta para um especialista. O cenário pioraria se considerássemos mais canais a serem monitorados. Devido a experimentos preliminares seguindo o passo 4, acreditamos que é viável reduzir ainda mais o número de URLs para a inspeção.

4.1.5. Análise de extração de tópicos e entidades

A extração de tópicos e entidades possibilita identificar assuntos e entidades em documentos textuais. Nesta pesquisa, a extração de tópicos e entidades foi utilizada para identificar mensagens associadas à segurança de redes e para identificar potenciais alvos. A Tabela 6 apresenta uma síntese dos resultados da análise de extração de tópicos e entidades nos canais *security* e *ddos*.

Tabela 6. Extração de tópicos e entidades

Canal	Mensagens analisadas	Mensagens c/ tópicos	Mensagens c/ entidades	Tópicos frequentes	Exemplos de entidades
security	10401	1019	1696	Technology Internet, Human Interest, Entertainment Culture, Business Finance	(Windows XP, Operating System), (Java, Technology), (Austria, Country)
ddos	2677	110	534	Technology Internet, Human Interest, Entertainment Culture, Politics	(JSON, Technology), (Google, Company), (United Nations, Organization)

Como pode ser observado na Tabela 6, o número de mensagens com tópicos e entidades em relação ao total de mensagens é pequeno. Esse resultado é devido à linguagem informal usada na rede IRC e ao tamanho das mensagens. Após uma análise por inspeção manual dos tópicos e entidades, verificou-se que apenas essas informações são insuficientes para dizer se uma mensagem é relevante ou não como alerta. No entanto, podem ser usadas como uma característica para auxiliar na priorização ou descarte de mensagens. Por exemplo, pode-se criar regras para priorizar mensagens com o termo “exploit” e a entidade “Operating System”.

4.1.6. Análise de correlação com outras fontes

A correlação de informações de diferentes fontes possibilitou identificar informações relevantes como alertas. No nosso estudo, foi realizada a correlação com dados dos boletins da Microsoft e CVE. A Tabela 7 destaca algumas das mensagens relevantes como alertas.

Observando as mensagens é possível identificar o interesse em obter códigos de exploração para novas vulnerabilidades. Em um dos casos analisados, foi divulgada uma URL para um código de exploração, mas infelizmente ela não estava mais disponível ao

Tabela 7. Correlação com outras fontes

	Mensagens
1	Anyone have an exploit yet for **MS14-066**?
2	Anyone know what the exploit behind **MS14-068** is?
3	anyone experiment with the recent **CVE-2014-6352** exploit?
4	So, has there been a public exploit for **CVE-2014-6321** developed yet?

processar os registros. Foi observado que em canais de segurança é comum especialistas postarem notícias sobre novas vulnerabilidades. Logo, essa informação pode ser útil para a prevenção antecipada. A identificação de usuários interessados em códigos maliciosos para novas vulnerabilidades auxilia na escolha de usuários a serem supervisionados.

4.2. Extração de Mensagens Associadas à Cibersegurança

Esta seção apresenta um arcabouço para a extração de mensagens associadas à cibersegurança (Figura 2) baseado nas observações da análise de mensagens IRC e seis meses de observações dos processos de coleta e mensagens trocadas em salas de IRC.

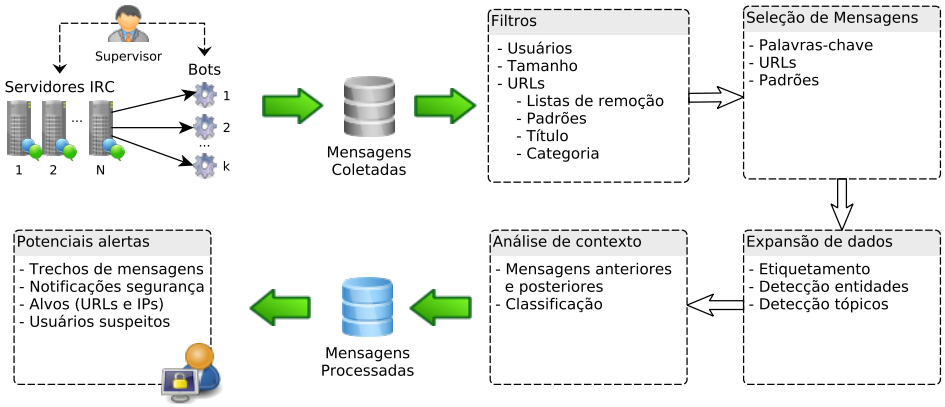

Figura 2. Arcabouço para a extração de mensagens associadas à cibersegurança

A coleta deve ser realizada com a supervisão humana, pois é comum os software de coleta automatizados (*bots*) serem identificados nos canais e banidos. É interessante implementar mecanismos para que os *bots* não fiquem muito tempo inativos e possam enviar e receber mensagens, mesmo que padronizadas e aleatórias. Além disso, a participação de um especialista possibilita obter registros de canais em que há a necessidade de senhas ou permissões para participação. O usuário não precisa ficar monitorando constantemente, mas deve entrar e deixar as salas. A mesma regra é válida para os *bots*, que devem ter origens diferentes e, aleatoriamente, deixar e retornar a rede IRC, além de trocar mensagens entre si. O monitoramento apenas por palavras-chave não é produtivo, pois não possibilita identificar o contexto das conversas, logo deve-se monitorar e armazenar todo o conteúdo discutido nos canais.

O processo de filtros deve remover as mensagens irrelevantes, mas mantê-las na base para a análise de contexto. Conforme apresentado nas seções anteriores, o número de mensagens descartadas desonera significativamente as próximas fases e o analista. Propomos no arcabouço filtros que consideram tamanho, usuários e URLs das mensagens.

A seleção de mensagens identifica potenciais alertas nas mensagens por meio do uso de regras baseadas em URLs suspeitas, palavras-chave associadas às atividades suspeitas ou notificações de segurança, e padrões descobertos nos canais.

A expansão de dados possibilita enriquecer o conjunto de dados, etiquetando as informações e usando algoritmos probabilísticos para detecção de tópicos e de entidades. Dessa forma, pode-se facilitar a priorização e descarte de mensagens via algoritmos que explorem as informações acrescentadas.

A análise de contexto envolve identificar se as mensagens próximas também possuem alguma associação com cibersegurança e, tentar classificar ou priorizar como alerta, usando as características obtidas nas outras fases. Com o conjunto de informações extraído, poderíamos optar por métodos de classificação supervisionados, como por exemplo Máquinas de Vetores de Suporte.

Os potenciais alertas devem ser inspecionados por um especialista em segurança e, podem conter informações que precisem de mais investigações ou devem ser descartadas. URLs e IPs podem ser monitorados na rede local do administrador para verificar se não há máquinas acessando endereços alvos. Notificações de vulnerabilidades e interesse por códigos de exploração podem indicar ao especialista a necessidade de atualização e atenção a tráfego suspeito na rede.

Como prova de conceito foi implementado um protótipo com parte dos métodos propostos na fase de filtros, seleção, expansão e classificação. O protótipo foi desenvolvido em Python 3.4 e foi avaliado processando os dados descritos na Tabela 2. A Tabela 8 apresenta a síntese dos resultados.

Tabela 8. Resultados do uso do arcabouço em diferentes canais

Canais	Período (dias)	Total de mensagens	Mensagens destacadas	Mensagens priorizadas	Precisão (1)	Precisão (2)
anonops	58	253522	7281	626	0,50	0,37
ddos	58	22695	3100	892	0,84	0,90
defacement	49	5350	96	96	0,76	0,77
hackers	50	19911	1681	203	0,50	0,64
opnewblood	58	48369	3266	217	0,47	0,52
security	68	118360	14885	434	0,31	0,36
networking	58	131310	8913	177	0,18	0,18

(1) amostra de 200 primeiras mensagens priorizadas (2) amostra aleatória de 50 mensagens priorizadas

Observa-se na Tabela 8 que o número de mensagens destacadas é bem menor que o número total de mensagens. Isso ocorreu devido aos processos de filtros, seleção e classificação. Esse resultado também deve-se à remoção de mensagens que não foram marcadas na fase de expansão de dados, logo não apresentam entidades, tópicos ou palavras-chave de interesse como alertas.

Na análise de precisão realizada em amostras, verifica-se que os canais *ddos* e *defacement* apresentam resultados melhores que os outros. Isso ocorreu devido a facilidade de identificar padrões e entidades nesses canais, que estão associadas a URLs, IPs, questionamentos sobre novos ataques e alvos. Nos outros canais, as mesmas expressões produzem uma quantidade maior de falsos positivos. Além disso, os canais *anonops* e *opnewblood* têm a política de não permitir a publicação de alvos, apesar de alguns usuários mesmo assim publicarem. Os canais *security* e *networking*, apesar de muitas mensagens associadas à cibersegurança, apresentaram baixa precisão pois essas mensagens eram associadas a pedidos de auxílio ou discussões de segurança.

A classificação de entidades possibilitou identificar alvos que não seriam detectados por IP ou URL, ou mesmo priorizados só por palavras-chave, como foi o caso da mensagem *"Anyone interested in a DDoS attack on Dolce & Gabbana?"* e notícias

associadas a um sítio web do governo *"brasil website hard to knock out using torshammer"*. No entanto, os melhores resultados com relação a alertas antecipados, são devido ao esquema de etiquetamento baseado nos grupos extraídos pela análise de frequência de palavras e análise de bigramas. Com esses grupos foi possível construir expressões regulares para identificar usuários interessados em realizar atividades maliciosas, potenciais alvos e ataques sendo realizados no momento.

Mesmo conseguindo classificar as informações relevantes, um especialista humano e o uso de filtros adaptativos são indispensáveis para minimizar os falsos positivos e, dessa forma, aumentar a precisão dos resultados. Verificamos que, somente a partir da análise e caracterização do canal a ser monitorado, é possível conduzir a extração de mensagens associadas à cibersegurança. Logo, a análise de diferentes canais é um diferencial desse trabalho em relação a outros associados ao monitoramento de IRC que visam a segurança de sistemas e alertas antecipados.

4.3. Avaliação das Questões de Pesquisa

Q1 *Há compartilhamento de informações que podem ser usadas como alertas por administradores de redes?* Sim. Como pode ser observado nas Tabelas 4 e 7, há várias mensagens identificando ações maliciosas, como definição de alvos para ataque e a procura por códigos de exploração. Também foram obtidas URLs para recursos suspeitos, como arquivos e descrição de alvos de ataques.

Q2 *É possível identificar alvos de ataques nos canais de atividades suspeitas?* Sim. Muitos alvos ou rumores de alvos são compartilhados abertamente nos canais, como por exemplo, pedidos de auxílio para executar um ataque. Apesar de que há alvos discutidos em seções privadas, ainda é interessante monitorar as ações e ferramentas usadas para a execução de ataques. Nos canais do anonops, uma operação a alvos é facilmente identificada pelo prefixo "op".

Q3 *Como automatizar o processo de identificação de alertas nas redes IRC?* O processo de identificação de alertas pode ser automatizado segundo o arcabouço proposto na Seção 4.2. No entanto, ainda é necessária a supervisão humana para evitar problemas no monitoramento e, se possível, garantir acesso a salas fechadas. Em nossa prova de conceito, como fizemos uma implementação simples, ainda há muitas mensagens irrelevantes para um analista fazer a conferência manual. No entanto, os métodos propostos, se refinados, podem conduzir a uma diminuição nos falsos positivos.

5. Conclusões

A análise das redes IRC confirmaram que ainda há informações relevantes como alertas nessa rede e que podem ser usadas como alertas antecipados, mesmo considerando o fato de que o IRC teve uma queda na sua utilização nos últimos anos. Nos canais de segurança são discutidas informações importantes sobre novos códigos de exploração e vulnerabilidades. Nos canais de atividades suspeitas são discutidas ferramentas e alvos de ataques. A análise e o arcabouço proposto para a extração de termos associados à cibersegurança possibilitam obter essas informações. Ainda há várias dificuldades com o monitoramento e processamento da relevância das mensagens retornadas como alertas, mas um lado positivo foi a diminuição de mensagens e evidência de informações importantes devido às técnicas e características extraídas da fase de análise. A continuação deste trabalho prevê

a otimização e extensão do protótipo baseado no arcabouço para aprimorar os resultados, em especial, na classificação de mensagens.

6. Agradecimentos

Agradecemos à Fundação Araucária, Secretaria de Estado da Ciência, Tecnologia e Ensino Superior (SETI-PR) e ao Governo do Estado do Paraná, pelo apoio financeiro. Agradecemos a Daniel Costa Bucher, pelo desenvolvimento do software de coleta.

Referências

Benjamin, V. and Chen, H. (2014). Time-to-event modeling for predicting hacker irc community participant trajectory. In *IEEE Joint JISIC, 2014*, pages 25–32.

Brown, D. A. (2007). Architecture for an automated irc investigation tool. Master's thesis, West Virginia University.

Carpine, F., Mazzariello, C., and Sansone, C. (2013). Online irc botnet detection using a soinn classifier. In *IEEE ICC 2013*, pages 1351–1356.

Décary-Hétu, D. and Dupont, B. (2012). The social network of hackers. *Global Crime*, 13(3):160–175.

Décary-Hétu, D., Dupont, B., and Fortin, F. (2014). Policing the hackers by hacking them: Studying online deviants in irc chat rooms. In *Networks and Network Analysis for Defence and Security*, pages 63–82. Springer International Publishing.

Gainaru, A., Dumitrescu, S. D., and Trausan-Matu, S. (2010). Toolkit for automatic analysis of chat conversations. In *8th COMM 2010 (IEEE)*, pages 99–102.

Houmansadr, A. and Borisov, N. (2013). Botmosaic: Collaborative network watermark for the detection of irc-based botnets. *Journal of Systems and Software*, 86(3):707 – 715.

Iqbal, F., Fung, B. C. M., and Debbabi, M. (2012). Mining criminal networks from chat log. In *IEEE/WIC/ACM*.

Lu, W. and Ghorbani, A. (2008). Botnets detection based on irc-community. In *IEEE GLOBECOM 2008. IEEE*, pages 1–5.

Ma, X., Guan, X., Tao, J., Zheng, Q., Guo, Y., Liu, L., and Zhao, S. (2010). A novel irc botnet detection method based on packet size sequence. In *IEEE ICC 2010*, pages 1–5.

Manning, C. D. and Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT press.

Mazzariello, C. (2008). Irc traffic analysis for botnet detection. In *Information Assurance and Security, 2008. ISIAS '08. Fourth International Conference on*, pages 318–323.

Michels, M. O. (2012). Real time text analysis on internet relay chat conversations. Master's thesis, Purdue University.

Oikarinen, J. and Reed, D. (1993). Internet Relay Chat Protocol. RFC 1459 (Experimental). Updated by RFCs 2810, 2811, 2812, 2813.

Wang, W., Fang, B., Zhang, Z., and Li, C. (2009). A novel approach to detect irc-based botnets. In *IEEE NSWCTC 2009*, volume 1, pages 408–411.