Sexta Conferencia de Directores de Tecnología de Información, TICAL 2016
Gestión de las TICs para la Investigación y la Colaboración
Buenos Aires, 13 al 15 de septiembre de 2016

Catálogo de Fraudes e Catálogo de URLs Maliciosas: Identificação e Combate a Fraudes Eletrônicas na Rede Acadêmica Brasileira

Rogerio Bastos [a], Paula Tavares[a], Lucas Borges[a], Italo Brito[a],
Edilson Lima[b], Liliana V. Solha[b]

[a] Ponto de Presença RNP na Bahia, Superintendência de TI, Universidade Federal da Bahia,
Av. Adhemar de Barros s/n, 40170110 Salvador-Ba, Brasil
{rogerio.bastos,paula.tavares,lborges,italovalcy}@ufba.br

[b] Centro de Atendimento a Incidentes de Segurança, Rede Nacional de Ensino e Pesquisa,
Av. André Tosello 209, 13083886 Campinas-SP, Brasil
{edilson.lima,liliana.solha}@rnp.br

Resumo. Fraudes eletrônicas via e-mail configuram-se como um problema recorrente para todo usuário de serviços de Internet. A cada dia novas técnicas de ofuscação, evasão e engenharia social são adotadas pelos fraudadores a fim de enganar usuários menos preparados ou atentos, tornando-os vítimas desses ataques. Criado em 2008, o Catálogo de Fraudes da Rede Nacional de Ensino e Pesquisa consolida-se como um importante repositório de fraudes eletrônicas brasileiras disseminadas por e-mail. Ao longo de mais de oito anos de trabalho nesse projeto, alguns mecanismos de apoio foram criados para ajudar na proteção dos usuários, como base de conhecimento das fraudes, cartilhas de segurança e documentos de boas práticas. Em particular, recentemente foi iniciado um novo projeto, chamado Catálogo de URLs Maliciosas (CaUMa), cujo objetivo é manter um repositório de URLs maliciosas e auxiliar na identificação e combate destas fraudes. Este artigo apresenta o funcionamento do CaUMa, estatísticas das URLs catalogadas, suas características, tendências observadas e oportunidades para trabalhos futuros e colaborações visando melhorar a segurança dos usuários e instituições vítimas desses ataques.

Palavras Chave: phishing; URLs maliciosas; fraudes eletrônicas.

1 Introdução

Com o surgimento e posterior democratização da Internet, nas décadas de 1980 e 1990, surgiu a ferramenta de e-mail, ou correio eletrônico, uma nova maneira de comunicação entre as pessoas, tanto para fins profissionais como pessoais. Assim como o fenômeno tecnológico em que está contida, a utilização do e-mail cresceu rapidamente e estabeleceu-se como uma eficiente ferramenta de comunicação, sendo usada, por exemplo, na comunicação entre empresas e clientes, em campanhas publicitárias, nas relações pessoais e outros. Segundo pesquisa divulgada pelo grupo Radicati [10] cerca de 205,6 bilhões de mensagens foram enviadas por dia no ano de 2015. É importante ressaltar que o e-mail assumiu um teor de comunicação oficial das empresas, principalmente pelo fato de poder ser arquivado pelo destinatário, o que não ocorre em ligações telefônicas ou comunicação via páginas web.

Com o fato de o e-mail tornar-se um meio de comunicação tão importante, em pouco tempo ele foi explorado também para a prática de atividades ilícitas como a propagação de arquivos maliciosos, de conteúdos impróprios, de URLs de páginas contaminadas, conteúdos de falsidade ideológica, e outros. De acordo com um relatório da Kaspersky Lab [15], de junho de 2011 à junho de 2013, 37.3 milhões de pessoas reportaram terem sido vítimas de tais ataques, representando um aumento de 87% em relação ao ano anterior.

Nesse cenário, o Centro de Atendimento a Incidentes de Segurança (CAIS) área de segurança da informação da Rede Nacional de Ensino e Pesquisa (RNP) sentiu-se motivado a criar um serviço para a identificação e catalogação desses e-mails fraudulentos, um tipo de repositório que pudesse ser consultado pela comunidade acadêmica e a população em geral e servisse de apoio no combate a esse tipo de atividade fraudulenta. Assim, em 2008, surgiu o Catálogo de Fraudes da RNP.

O Catálogo de Fraudes da RNP é mantido pelo CAIS numa parceria com um grupo de pesquisadores e técnicos da Universidade Federal da Bahia (UFBA) e do Ponto de Presença da RNP na Bahia (PoP-BA/RNP) e, até onde se sabe é a primeira e maior fonte de informações pública online sobre fraudes eletrônicas do Brasil, sendo amplamente utilizado pela população em geral para a validação de e-mails suspeitos.

Ao analisar as fraudes que circulam via e-mail no Brasil e que são reportadas ao Catálogo de Fraudes da RNP, o grupo de pesquisadores e técnicos que atuam na manutenção do catálogo notou a presença frequente de URLs maliciosas nos e-mails e que as ferramentas de contenção dessas URLs apresentam baixa taxa de detecção no contexto de fraudes analisadas. Esses fatos culminaram na proposta de criação do Catálogo de URLs Maliciosas, CaUMa, cujo objetivo é prover à comunidade um serviço adicional para combate aos sites fraudulentos, a partir do seu bloqueio em navegadores web e clientes de e-mail.

Este artigo apresenta de forma geral a experiência de criação e manutenção dos catálogos de Fraudes e de URL Maliciosas, bem como os principais resultados identificados a partir do uso dessas ferramentas.

Este artigo está estruturado da seguinte maneira. Na Seção 2, apresentam-se trabalhos relacionados à detecção e tratamento de fraudes eletrônicas. Na Seção 3, discute-se o Catálogo de Fraudes da RNP, sua operação, benefícios e algumas estatísticas. A Seção 4, por sua vez, detalha o Catálogo de URLs Maliciosas, destacando sua arquitetura, funcionamento e possibilidades de utilização ou colaboração. A Seção 5 apresenta as estatísticas e tendências particularmente observadas no conjunto de dados analisados. Por fim, na Seção 6, conclui-se o trabalho e relacionam-se os trabalhos futuros.

2 Trabalhos relacionados

Existem diversos trabalhos recentes na área de fraudes eletrônicas (*phishing*) que tentam construir mecanismos de detecção automática ou analisar as técnicas utilizadas por atacantes.

Em [3] os autores propõem a criação de um sistema de detecção de *phishing* baseado em regras. Tais regras levam em consideração características frequentes de

URLs utilizadas para phishing, como presença de IP na URL, Uso ou não de certificado TLS/SSL, número de pontos na URL, tamanho do endereço e blacklist de palavras chave. Utilizando este conjunto de regras foi conseguida uma precisão de 99.14% na detecção automática de phishing. [4] propõe um detector de URLs fraudulentas utilizando apenas componentes léxicos e consegue uma precisão de até 97%. [5] também propõe um sistema de detecção utilizando regras e baseado em classificação associativa.

Aspectos comuns em ataques de *phishing* são examinados por [1], que levanta diversas características interessantes como registro e tempo de ativação dos domínios de *phishing*, máquinas utilizadas para hospedar tais sites e a anatomia das URLs e domínios fraudulentos. Ele ainda indica que tais resultados podem ser utilizados como heurísticas na filtragem de e-mails de *phishing* e na identificação de registros de domínios suspeitos. [6] analisa URLs utilizadas em ataques de *phishing* e identifica algumas técnicas utilizadas pelos fraudadores para enganar as vítimas, como mascarar o host com um endereço de IP, mascarar o host com outro domínio, criar domínios similares ao de organizações conhecidas e criar URLs muito grandes para confundir a vítima.

3 Catálogo de Fraudes da Rede Acadêmica Brasileira

Esta seção apresenta o funcionamento do Catálogo de Fraudes da Rede Nacional de Ensino e Pesquisa do Brasil (RNP), algumas estatísticas e tendências observadas durante mais de oito anos de operação do projeto.

3.1 Visão geral

Criado em 2008, o Catálogo de Fraudes da RNP alimenta uma base de dados de fraudes eletrônicas repassadas por diversos usuários da Internet no Brasil, principalmente, mas não restrito a, usuários da rede acadêmica brasileira. O processo de tratamento dessas fraudes passa pelas fases de coleta, triagem, categorização e publicação, cujas etapas e o fluxo de operação serão aprofundados nas seções subsequentes. Esse processo é executado por grupo de pesquisadores e técnicos da Universidade Federal da Bahia (UFBA), através de um acordo de cooperação firmado entre o Centro de Atendimento a Incidentes de Segurança da RNP (CAIS/RNP) e o Ponto de Presença da RNP na Bahia (PoP-BA/RNP).

3.2 Processo de recebimento e tratamento de Fraudes

O Catálogo de Fraudes da RNP recebe fraudes eletrônicas enviadas por e-mail por qualquer usuário na Internet, através do encaminhamento de mensagens para o contato phishing@cais.rnp.br. Através desse contato, os usuários podem encaminhar mensagens que já foram identificadas como fraudulentas ou que há dúvidas sobre a veracidade de seu conteúdo, sendo então analisadas e catalogadas para criação de uma base de conhecimento de fraudes identificadas. Essa base de conhecimento pode ser

consultada publicamente através do site do projeto, disponível em [8]. O objetivo do catálogo é, dessa forma, apoiar a comunidade brasileira na identificação e conscientização sobre os principais golpes eletrônicos que estão sendo veiculados na Internet.

Todos os e-mails encaminhados ao Catalogados de Fraudes são analisados e catalogados em uma ferramenta web, seguindo o processo de tratamento ilustrado na Fig. 1. Nessa figura, é possível observar as seguintes etapas do processo de tratamento das fraudes:

- Triagem – Nessa fase, mensagens são separadas em fraudes, spam e mensagens de língua estrangeira. Somente as mensagens de fraudes brasileiras são categorizadas e publicadas.
- Categorização – Após a triagem, as mensagens são categorizadas com base nas suas principais características, como Bancos, *e-Commerce*, Serviços de pagamento e outras.
- Publicação – Na última fase, são geradas imagens dessas mensagens, aplicados marcas d'água e criadas *tags* (marcações, através de palavras chaves) que facilitam a busca posterior. Após isso, as mensagens são disponibilizadas na página web do catálogo e liberadas para consultas pelo público em geral.

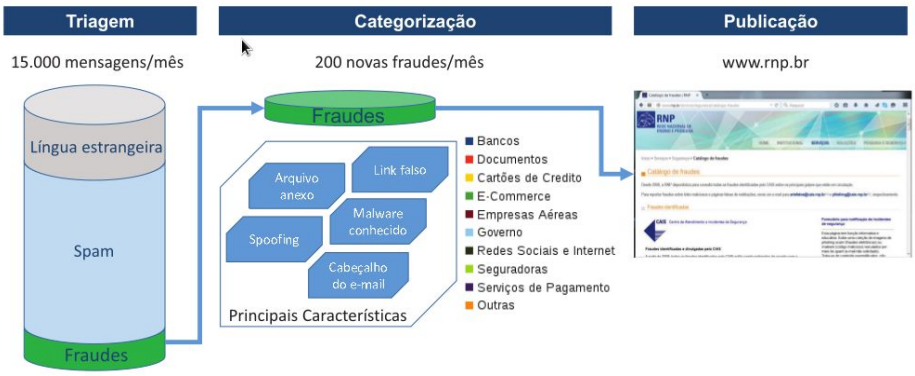

Fig. 1. Processo de catalogação de fraudes no Catálogo de Fraudes da RNP.

Atualmente, cerca de 15.000 mensagens são tratadas a cada mês, desse total são descartados os spams e as mensagens em língua estrangeira, depois descarta-se as mensagens repetidas. Com isso, são catalogadas uma média de 200 novas fraudes por mês. Após a mensagem ser classificada como fraude, inicia-se o processo de identificação das principais características do e-mail, tais como o uso de redirecionamento para sites falsos e/ou a presença de arquivos maliciosos em anexo ou disponíveis para download.

São registrados no Catálogo de Fraudes o corpo do e-mail na forma de texto e imagem através de captura de tela. As mensagens que direcionam o usuário para sites fraudulentos, também têm as páginas do site registradas como imagem, para isso é feito uma interação com esses sites, a fim de coletar o maior número de informações.

Quando há um *malware* anexado ou disponível para download, é utilizado a ferramenta VirusTotal [9] para análise do arquivo malicioso.

As informações como assunto da mensagem, tipo, classificação, nome do *malware*, hash md5 do malware são registradas no catálogo juntamente com as imagens. As principais tarefas relacionadas ao catálogo de fraudes são atualmente executadas manualmente, no entanto, objetivando aumentar a eficiência e os resultados obtidos, encontram-se em desenvolvimento novas ferramentas para a automatização de algumas etapas deste processo, tornando-o mais eficaz e possibilitando o aumento na quantidade de novas fraudes.

3.3 Benefícios do Catálogo de Fraudes

O Catálogo de fraudes é uma importante contribuição da RNP, tanto para a comunidade acadêmica, como também para a comunidade brasileira em geral. As fraudes catalogadas pelo projeto, por serem rigorosamente analisadas, são uma fonte confiável de informações sobre fraudes eletrônicas brasileiras. Como consequência do trabalho com fraudes eletrônicas foram desenvolvidos documentos de boas práticas, respostas para perguntas frequentes e cartilhas de segurança.

Os usuários em geral podem fazer uso dessas informações para consultar por período, identificar campanhas de fraudes, e também comparar o texto e imagem das mensagens recebidas com fraudes conhecidas.

Já para a comunidade de Segurança da Informação, esse catálogo pode ser usado para maior entendimento das fraudes direcionadas ao público latino-americano, especialmente para os brasileiros, uma vez que a maioria dos trabalhos anteriores apresenta dados de fraudes internacionais, essencialmente no idioma inglês. Nota-se, inclusive, a carência por repositórios dessa natureza aqui no Brasil, de forma que pudessem ser usados por ferramentas de segurança automatizadas para evitar que os usuários fossem vítimas das fraudes (e.g. filtros de conteúdo, *plugins* de navegadores web etc).

Nesse sentido, foi desenvolvido um trabalho relacionado ao Catálogo de Fraudes que visa incorporar o registro das URLs utilizadas nas fraudes, agregando um mecanismo de reputação às URLs para que possam ser usadas em ferramentas clientes. Este trabalho será apresentado na Seção 4.

3.4 Estatísticas de Fraudes Catalogadas

As fraudes brasileiras tendem a abordar diversos serviços eletrônicos, como vias de pagamento, serviços bancários e *e-commerce*, ou assuntos atuais que estejam sendo largamente discutidos na comunidade. O zika vírus, recentemente considerado um risco a saúde mundial, é um dos temas utilizados como fraude no ano de 2016. A imagem abaixo é de um phishing que aparenta disponibilizar mais uma informação sobre os testes de uma vacina contra o zika vírus.

Fig. 2. Exemplo de Fraude sobre o Zika Vírus.

Na fase de triagem grande parte dos e-mails, que são encaminhados por usuários ou recebidos diretamente, são classificados como e-mails de spam, uma parcela menor é dividida entre internacionais e os que realmente são fraudes. Dentre o total de fraudes catalogadas, a Fig. 3 apresenta um recorte dos últimos doze meses (entre Março/2015 e Junho/2016), onde é possível notar a proporção de e-mails recebidos, analisados e catalogados.

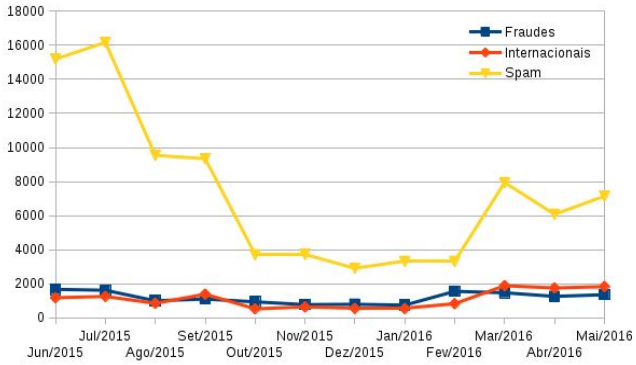

Fig. 3. Estatísticas de Fraudes Recebidas versus Catalogadas.

Os e-mails identificados como fraude são categorizados para facilitar a busca dos usuários no Catálogo de Fraudes. O gráfico da Fig. 4 mostra que a maior frequência de fraudes está relacionada às categorias de Bancos e Documentos.

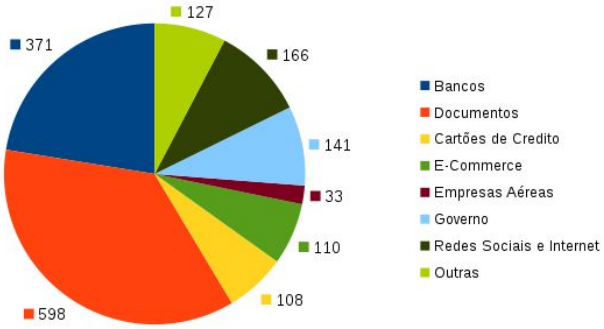

Fig. 4. Estatísticas sobre as categorias de fraudes.

4 CAUMA: Catálogo de URLs Maliciosas

Grande parte dos e-mails de *phishing* contêm URLs para sites falsos ou arquivos maliciosos [1], [2], levando o usuário a fornecer informações confidenciais nos sites fraudulentos ou até mesmo a infecção da máquina do usuário com vírus, *worms, spywares* ou *bots*. Em particular, cerca de 90% dos e-mails de fraude reportados ao Catálogo de Fraudes da RNP possuem URLs maliciosas.

Devido a essa característica, um dos mecanismos de proteção a esse tipo de ameaça é a exibição de alerta ou mesmo o bloqueio destas URLs nos navegadores web modernos. Esse bloqueio se dá através da consulta a bases de URLs maliciosas, também conhecidas como *blacklists*, ou sistemas de reputação de sites. O Google Safe Browsing [13] e o PhishTank [14] são os serviços de *blacklist* que mais se destacam nesse contexto devido à grande quantidade de URLs catalogadas e à possibilidade de consulta através de API, o que permite a integração com outras ferramentas. O Google Safe Browsing, por exemplo, é utilizado por browsers como o Chrome e o Firefox para alertar os usuários que tentam acessar URLs maliciosas.

Embora sejam amplamente utilizados e apresentem inúmeros benefícios, tais sistemas possuem algumas deficiências e oportunidades de melhorias principalmente quanto à sua aplicação no contexto brasileiro, conforme detalhado a seguir:

- **Inconsistência dos resultados das consultas através da API**: comparando o resultado das consultas através da API do Google Safe Browsing com a detecção do navegador Google Chrome é possível perceber que a taxa de detecção da API é inferior, ou seja, existem URLs maliciosas que são detectadas pelo browser, mas que não são detectadas através de consultas pela API.
- **Limite no número de consultas através da API**: tanto o Google Safe Browsing, quanto o PhishTank, limitam o número de consultas, seja por questões de interesse econômico, seja por limitações de recursos por parte de quem oferece o serviço.
- **Baixa taxa de detecção de URLs maliciosas brasileiras**: a partir das URLs extraídas das fraudes reportadas ao CAIS, verificamos que apenas uma

pequena quantidade era detectada pelo Google Safe Browsing ou pelo PhishTank, o que mostra que esses serviços possuem poucas URLs destinadas ao público brasileiro em suas respectivas bases. Os gráficos da Fig. 5 e Fig. 6 mostram a quantidade de URLs analisadas e a quantidade de URLs detectadas pelo Google Safe Browsing e PhishTank.

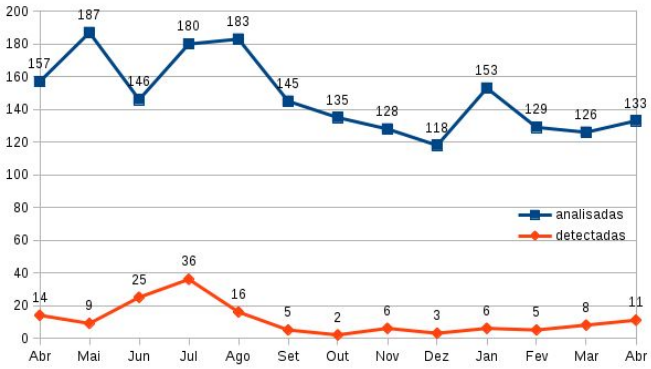

Fig. 5. Análise de URLs no Google Safe Browsing.

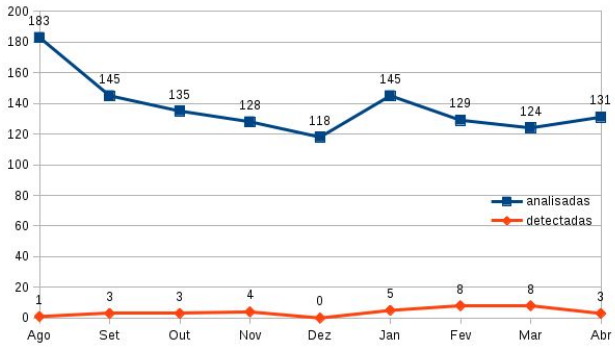

Fig. 6. Análise de URLs no PhishTank.

Buscando preencher essa lacuna, o PoP-BA/RNP, em parceria com o CAIS, iniciou o desenvolvimento do Catálogo de URLs Maliciosas (CaUMa) [12], um serviço de blacklist de URLs voltado para a comunidade brasileira.

A Fig. 7 ilustra a arquitetura do CaUMa e a relação entre os componentes e os agentes que interagem com o sistema. Nessa figura, os principais componentes do sistema CaUMa são: uma interface web, uma API HTTP, um banco de dados e ferramentas de apoio. Esses componentes serão objeto de estudo aprofundado nos parágrafos a seguir.

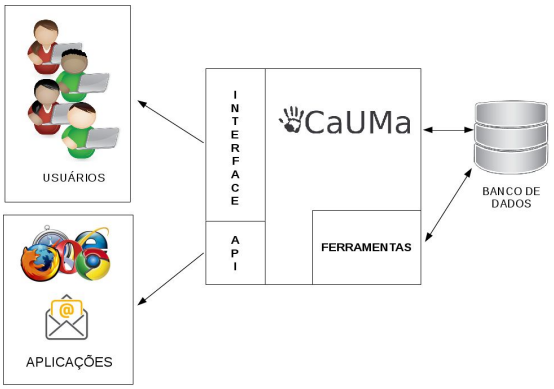

Fig. 7. Arquitetura do Catálogo de URLs Maliciosas (CaUMa).

A interface web possui uma página com formulário de consulta para as URLs cadastradas na base. Esta consulta está disponível para qualquer usuário e exibe informações adicionais, como o resultado da consulta em outras blacklists de URLs. A interface web também dá acesso à interface administrativa, que é restrita a usuário autenticados. Através da interface administrativa é possível cadastrar, classificar e remover as URLs da base.

Para facilitar a integração de outras ferramentas com o CaUMa, a API HTTP disponibiliza uma interface padronizada que permite realizar consultas à base de URLs do CaUMa. Através dessa API aplicações como browsers, clientes de e-mail e sistemas de detecção de phishing podem interagir com o CaUMa e fazer uso da base de dados para impedir o acesso às URLs maliciosas. O resultado de cada consulta é retornado em formato JSON, por ser um formato amplamente suportado e de fácil utilização. Abaixo um exemplo de consulta utilizando a API HTTP:

```
curl https://cauma.pop-
ba.rnp.br/api/v1.0/diagnostic/site=1.2.3.4/vision.php
{
  "status": 200,
  "message": "CaUMa detectou a url como maliciosa",
  "data": {
    "category": "Documentos",
    "in_database": true,
    "submitted_at": "22/06/2016 20:35",
    "url": "1.2.3.4/vision.php",
    "type": "Malware",
    "id": 1534
  }
}
```

O banco de dados armazena as URLs maliciosas e dados adicionais como: a data em que a URL foi cadastrada, o período em que ela permaneceu online, o tipo e a categoria. As URLs que levam ao download de arquivos maliciosos são classificadas como sendo do tipo malware, enquanto que URLs que levam a sites falsos são classificadas como do tipo phishing.

As URLs também são categorizadas de acordo com o tema do site falso ou da mensagem que divulga a URL maliciosa. A tabela a seguir mostra as categorias utilizadas no CaUMa:

Tabela 1. Categorias utilizadas pelo sistema CAUMA

Categorias	Descrição
Bancos e Financeiras	Serviços bancários como atualizações de segurança, acesso à internet banking e transações financeiras.
Documentos	Arquivos relacionados a pagamentos como boleto, nota fiscal, comprovantes, etc.
Cartões de Crédito	Benefícios relacionados a cartões de credito, solicitações de atualização cadastral.
E-Commerce	Promoções e descontos relacionados a lojas online.
Empresas Aéreas	Informações e alterações de passagens aéreas, promoções relacionadas a sistemas de pontos e milhas.
Governo	Informações relacionadas a serviços de instituições governamentais como intimações judicias, cobranças de impostos e outros.
Redes Sociais e Internet	Mensagens em redes sociais, solicitações de atualizações cadastrais de contas de e-mails
Seguradoras	Fraudes relacionadas a venda e cadastro de serviços de seguros
Serviços de Pagamento	Status de compras e atualizações cadastrais feitas através de serviços de pagamento online.
Outras	Temas sazonais e mensagens sem relações a marcas ou serviços.

Por fim, algumas ferramentas de apoio realizam tarefas complementares e de manutenção da base de dados. Dentre elas destacam-se: a ferramenta de monitoramento do período de disponibilidade das URLs, que monitora quanto tempo cada URL maliciosa permanece online; e a ferramenta de geração de estatísticas, que auxilia na análise e interpretação dos dados.

5 Estatísticas

Nessa seção serão apresentados os resultados das análises feitas a partir das URLs armazenadas no CaUMa. Esses resultados foram obtidos a partir de mais de 1500 URLs únicas extraídas de e-mails fraudulentos reportados por usuários brasileiros no período de março de 2015 a junho de 2016.

5.1 Domínios mais comuns nas URLs

Analisando as URLs maliciosas foram identificados os domínios mais utilizados nas fraudes brasileiras. Foram avaliados os domínios de um nível, também conhecidos como domínios de topo (TLD, do inglês *top-level domain*), e os domínios com dois e com três níveis. As Tabelas 2, 3 e 4 abaixo mostram que grande parte dos domínios estão registrados sob o domínio `.com`, apesar das URLs hospedarem *phishing* destinados ao público brasileiro, seguindo uma tendência de outros relatórios do Anti-Phishing Working Group [11].

Também foi observado que muitos dos arquivos maliciosos são hospedados em serviços de armazenamento em nuvem, como o Google Drive e o DropBox. O uso deste tipo de serviço possui duas vantagens para os phishers: é gratuito e passa credibilidade por serem amplamente utilizadas.

Tabela 2. Domínios fraudulentos de um nível.

Domínio	Quantidade
com	727
br	183
net	45
org	40
pl	21
ru	20
co	18
info	12
me	12
lt	9

Tabela 3. Domínios fraudulentos de dois níveis.

Domínio	Quantidade
com.br	170
googledrive.com	78
bitnamiapp.com	42
google.com	26
dropboxusercontent.com	21
tripod.com	18
formlogix.com	17
sugarsync.com	17
weebly.com	16
amazonaws.com	15

Tabela 4. Domínios fraudulentos de três níveis.

Domínio	Quantidade
googledrive.com	57
dl.dropboxusercontent.com	21
www.sugarsync.com	17
www.formlogix.com	17
s3.amazonaws.com	15
www.123contactform.com	14
docs.google.com	10
drive.google.com	10
www.dropbox.com	8
www.tungbachnhat.com	7

5.2 Nome de marcas e serviços na URL

Uma URL maliciosa pode ludibriar a vítima com maior facilidade quando são visualmente semelhantes às reais. Um dos artifícios utilizados pelos phishers é colocar elementos que façam a vítima confiar que aquele é um domínio verdadeiro, como incluir o nome de marcas ou serviços legítimos no subdomínio ou no path da URL falsa. [1] menciona que de 53% a 77% dos phishings possuem o nome da marca alvo em algum lugar da URL. [2] utiliza a presença de marcas na URL como característica importante para detecção automática de Phishing.

Conforme mostrado no gráfico da Fig. 8, foi constatado que de 20% a 50% das URLs apresentam nomes de marcas legítimas relacionadas ao tema da fraude, como nomes de instituições bancárias, empresas de cartão de crédito e comércio eletrônico.

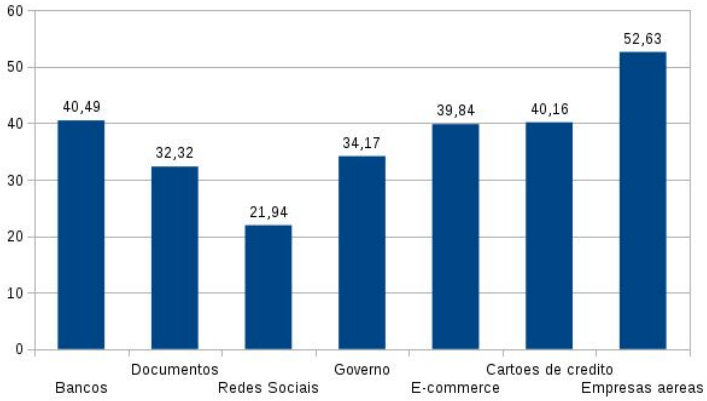

Fig. 8. Porcentagem das URLs que apresentam nome de marcas.

5.3 Tamanho da URL

Como consequência da técnica de ofuscação descrita na subseção anterior, o tamanho médio das URLs de phishing tende a ser maior do que URLs legítimas, com exceção das URLs encurtadas. A técnica de encurtadores de URL também é comumente utilizada pelos fraudadores para evasão de mecanismos de bloqueio ou para esconder do usuário o domínio fraudulento. Segundo McGrath et al. [1], URLs de domínios usados para praticar phishing têm, em média, de 67 a 107 caracteres, enquanto que endereços legítimos possuem 22 caracteres. A análise das URLs catalogadas no CaUMa mostram que a maior parte das URLs possuem entre 50 e 100 caracteres, como pode ser observado no gráfico da Fig. 9.

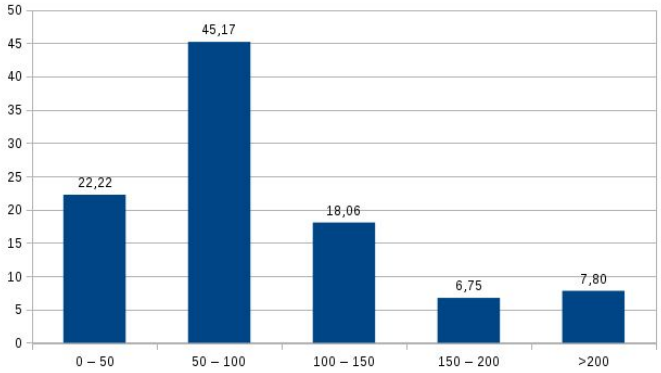

Fig. 9. Tamanho das URLs de acordo com o número de caracteres (em porcentagem).

5.4 Tempo de vida do *phishing*

Uma das principais características dos sites que hospedam fraudes eletrônicas é curta duração em que permanecem disponíveis nos provedores de conteúdo na Internet. Sheng et al. [7] aponta que cerca de 70% dos sites de *phishing* permanecem menos de 48 horas online. No entanto, a análise das URLs armazenadas no CaUMa apresentou comportamento ligeiramente diferente dos relatórios divulgados na literatura.

De acordo com a Fig. 10, mais de 50% das URLs contidas em fraudes eletrônicas brasileiras permaneceram online por um período igual ou superior a 5 (cinco) dias. A análise das mensagens de e-mail de *phishing* mostrou que a mesma URL pode ser detectada em mensagens diferentes, geralmente aquelas que fazem parte de uma mesma campanha de *phishing*.

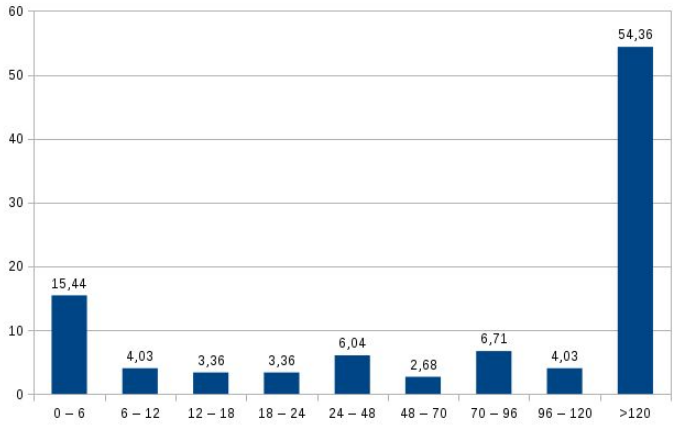

Fig. 10. URLs agrupadas por tempo de vida em horas (em porcentagem).

Os resultados obtidos da análise das URLs mostra que algumas características das campanhas de *phishing* realizadas no Brasil são diferentes dos resultados apresentados a partir da análise dos *phishing* que circulam em outros países.

6 Conclusões

O crescimento na disseminação de fraudes eletrônicas, via e-mail, redes sociais e outras mídas eletrônicas, associado com a falta de conhecimento e até discernimento de muitos usuários na Internet, evidenciam a necessidade e a importância de uma base de conhecimento das fraudes eletrônicas, que possa ser usada não apenas como uma fonte de consulta e validação de mensagens desconhecidas, mas também que possa ser usada pelas organizações na implantação de filtros e outros mecanismos de segurança a fim de evitar que os usuários sejam vítimas desse tipo de ataque.

Este artigo apresentou o processo de análise, triagem e registro do Catálogo de Fraudes da RNP, bem como algumas estatísticas e tendências observadas nesse processo. É notório observar o crescimento na quantidade das fraudes e também o nível de sofisticação nas páginas de captura de informações dos usuários. Apresentou-se também uma análise de ambientes de filtros de URL, a saber o projeto *Google Safe browsing* e *Phishtank*, onde em média menos de 10% das URLs maliciosas encontradas em fraudes brasileiras são identificadas pelas bases supracitados, evidenciando-se a necessidade por uma base de dados voltada para a realidade brasileira de fraudes eletrônicas.

Propõe-se, portanto, o Catálogo de URLs Maliciosas, CaUMa, um serviço que disponibiliza uma base de links maliciosos específica para o publico alvo brasileiro, mantido de forma colaborativa a partir das mensagens encaminhadas pelos usuários e fornecendo uma interface web simples e uma API aberta para consulta. Essa base já contém mais de 1500 URLs únicas, catalogadas durante seu curto período de funcionamento. A análise dessas URLs apresenta características interessantes das

fraudes brasileiras, como, por exemplo, o tempo de permanência da fraude online, cuja média apresentada em outros trabalhos é 48h, ao passo que os sites fraudulentos direcionados ao público brasileiro ficam online por mais de cinco dias.

Como trabalhos futuros, espera-se: i) desenvolver plug-ins e ferramentas a serem integradas aos navegados web e clientes de e-mail, a fim de mitigar as fraudes antes que elas atinjam os usuários; ii) aumentar as fontes de inserção de URLs, seja através de sensores de monitoramento de atividade maliciosa ou de novas parcerias de colaboração para encaminhamento de fraudes; iii) investigar melhor o conteúdo dos e-mails fraudulentos, principalmente os anexos e códigos maliciosos; iv) fomentar a utilização da base de fraudes e URLs para pesquisas acadêmicas, tecnológicas e comportamentais.

Referências

1. McGrath, D. K., & Gupta, M. Behind Phishing: An Examination of Phisher Modi Operandi. LEET, 8, 4 (2008)
2. Marchal, S., François, J., State, R., & Engel, T. PhishScore: Hacking phishers' minds. In 10th IEEE International Conference on Network and Service Management (CNSM), pp. 46-54 (2014)
3. Moghimi, M., & Varjani, A. Y. New rule-based phishing detection method. Expert systems with applications, 53, 231-242 (2016)
4. Le, A., Markopoulou, A., & Faloutsos, M. Phishdef: URL names say it all. In IEEE INFOCOM, pp. 191-195 (2011)
5. Abdelhamid, N., Ayesh, A., & Thabtah, F. Phishing detection based associative classification data mining. Expert Systems with Applications, 41, 5948-5959 (2014)
6. Garera, S., Provos, N., Chew, M., & Rubin, A. D. A framework for detection and measurement of phishing attacks. In Proceedings of the ACM workshop on Recurring malcode, pp 1-8 (2007)
7. Sheng, S., Wardman, B., Warner, G., Cranor, L. F., Hong, J., e Zhang, C. An Empirical Analysis of Phishing Blacklists. Em Conference on Email and Anti-Spam (2009)
8. Serviço Catálogo de Fraudes da RNP, <https://www.rnp.br/servicos/seguranca/catalogo-fraudes>, último acesso em 20/06/2016
9. Serviço Vírus Total, <https://www.virustotal.com/>, último acesso em 20/06/2016
10. The Radicati Group. Email Statistics Report, 2015-2019. Disponível em: <http://www.radicati.com/>, Acesso em: 30/06/2016.
11. APWG Reports. Global Phishing Survey: Trends and Domain Name Use in 2H2014. Disponível em: <http://docs.apwg.org/reports/APWG_Global_Phishing_Report_2H_2014.pdf>, Acesso em: 30/06/2016.
12. Serviço CaUMa, <https://cauma.pop-ba.rnp.br>, último acesso em 20/06/2016
13. Serviço Google Safe Browsing, <https://developers.google.com/safe-browsing/>, último acesso em 20/06/2016
14. Serviço PhishTank, <https://www.phishtank.com/>, último acesso em 20/06/2016
15. Kaspersky Lab's. The evolution of phishing attacks 2011-2013. Disponível em <http://media.kaspersky.com/pdf/Kaspersky_Lab_KSN_report_The_Evolution_of_Phishing_Attacks_2011-2013.pdf>, último acesso em 30/06/2016