Ambiente para Análise e Geração de Inteligência de Ameaças usando Fontes Abertas

José Valdy Campelo Júnior[1], **João José Costa Gondim**[1]

[1]Departamento de Ciência da Computação – Universidade de Brasília (UnB)
Brasília – DF – Brasil

{valdyjunior,gondim}@unb.br

Abstract. *Analysing attacks on computer networks may involve large volumes of data transmitted among several machines, even in small networks. Data volume is large and the time to process and analyze is short. The objective is to extract and analyse information on network attacks available from open sources in the Internet. Using a scalable, robust architecture that makes use of Hadoop processing techniques so that information is made available in a timely manner. With a proposed architecture implemented all the desired characteristics were the data file in near real time. The system provides the means for producing threat intelligence with agility and efficiency.*

Resumo. *A análise de ataques a redes de computadores é complexa dado o volume de dados trafegados e a quantidade elevada de máquinas, mesmo em pequenas redes. O volume de dados é grande e o tempo para processá-los e analisá-los é curto. O objetivo é extrair e analisar informações sobre ataques em rede, que foram obtidas de fontes abertas. Utilizando uma arquitetura robusta, elástica e escalável que faz uso de técnicas de processamento com o uso do Hadoop para que assim as informações sejam disponibilizadas em tempo hábil. Com a arquitetura proposta implementada todas as características desejadas foram obtidas permitindo o processamento dos dados em tempo próximo do real. O sistema provê informação de inteligência sobre ataques em grande escala com agilidade e eficiência.*

1. Introdução

O domínio da informação é ponto fundamental para a evolução da humanidade desde seu início. Conseguir analisar dados sobre ameaças aos sistemas é imprescindível para estruturar planos de proteção e recuperação. Entretanto garantir que seja possível obter essas informações completas e em tempo hábil é um desafio que se apresenta na atualidade. Processar e analisar esta quantidade de informações não era possível com uma arquitetura tradicional. *Big Data* se aplica as informações que não podem ser processadas ou analisadas utilizando processos ou ferramentas tradicionais onde o aumento considerável de problemas com esta natureza, com o foco dos sistemas na obtenção da maior quantidade de dados sobre as transações e seus usuários.

Segundo [IBM et al. 2011], três características são inerentes ao *Big Data*: volume, variedade e velocidade. O volume diz respeito ao que já foi abordado sobre a grande quantidade de dados que são e serão gerados no futuro, tornando sistemas centralizados

impossibilitados de processar todo conjunto de dados. A variedade diz respeito a diversidade de fontes das quais os dados são originados, registros de acesso de baixo nível, informações sobre localização, acelerômetros, padrões de escrita, entre outros formam a origem das informações. Por último, a velocidade para processar esses dados não pode ficar em segundo plano, num ambiente cada vez mais dinâmico alguns segundos podem significar a perda de valor da marca e consequentemente uma perda financeira.

Contudo, o processamento de grandes volumes de dados esbarra em alguns problemas que devem ser considerados, como por exemplo diminuir o tempo de acesso e gravação das informações, consultar todos os dados de uma grande base de dados e realizar a análise interativa de dados. A solução encontrada foi dividir e distribuir os dados ao longo de vários discos e paralelizar o acesso a esses, para reduzir o desperdício de espaço, o acesso a todos os discos seria compartilhado entre os processos e usuários garantindo que na média a utilização desse espaço adicional seja sempre otimizado.

A principal contribuição aqui apresentada é a construção de um ambiente de baixo custo para organização, compreensão e melhor visualização de ameaças em redes de computadores obtidas de fontes abertas. Como objetivo de longo prazo esta arquitetura busca a geração de inteligência de ameaças usando fontes abertas com capacidade de abstrair as dificuldades elencadas até aqui. Adicionalmente, demonstra-se como o ambiente pode ser usado para construção de perfis de atacantes e alvos.

Este trabalho esta organizado da seguinte forma: a Seção 2 discorre sobre alguns trabalhos presentes na literatura bem como as diferenças entre as abordagem entres eles e este trabalho. A Seção 3 descreve a arquitetura proposta e seus componentes, a Seção 4 versa sobre as fontes de dados utilizadas para desenvolvimento deste trabalho. Os resultados dos experimentos são detalhados na Seção 5, e por fim a Seção 6 apresenta as conclusões e algumas observações finais.

2. Trabalhos Relacionados

Na literatura, há diversos trabalhos que utilizam arquiteturas baseadas em Hadoop, tanto para extração quanto para análise de grandes massas de dados com objetivos relacionados à segurança da informação.

No trabalho de Bachupally et al. [Bachupally et al. 2016], por exemplo, uma arquitetura baseada em HDFS e Hive é utilizada para processamento de capturas de rede para identificação de ameaças e ataques de rede. A principal melhoria apontada pelo autor foi a possibilidade de se processar uma grande quantidade de dados em um curto período de tempo.

Janeja et al. [Janeja et al. 2014] propôs um sistema distribuído de detecção de intrusão, baseando seu desenvolvimento no *framework* HAMR [Wu et al. 2017] que opera com um fluxo de dados em tempo real com mecanismo de computação *in-memory*. A arquitetura busca detectar intrusões multifacetadas que podem estar distribuídas ao longo do tempo e também pela rede.

Nessa mesma linha Jia [Jia 2017] explana sobre uma arquitetura baseada em Big Data para detecção de Ameaças Persistentes Avançadas, decompondo seu exame camada por camada e utilizando diversas técnicas de segurança.

Assim como nesses trabalhos, o objetivo deste trabalho é propor uma arquitetura

para processamento de Big Data, apoiada sobre a suíte Hadoop para captura e análise de dados relacionados a segurança da informação. Entre as diferenças para os trabalhos relacionados esta o fator de simplificação da análise, permitindo que sem grandes conhecimentos sobre a arquitetura e processamento dos dados seja possível extrair informações.

3. Arquitetura

Para permitir a inserção e processamento da massa de dados obtida através das fontes abertas uma arquitetura robusta era necessária. Pensando nisso, foram definidas quatro características gerais que nortearam o desenvolvimento da desta abordagem, são eles:1. Escalabilidade 2. Elasticidade 3. Alta disponibilidade 4. Adaptabilidade

Seguindo esses conceitos, o sistema atenderá o crescente aumento no número de dados e permitirá diversos tipos de análise e uma extração de dados ágil frente ao aumento vertiginoso das informações sobre ataques geradas diariamente, permitindo desta forma a aplicação de várias técnicas de análise sobre os dados.

O artefato escolhido para compor a solução proposta neste trabalho é a suíte Hadoop que já conta com um sistema de arquivos distribuídos o **Hadoop Distributed File System** (HDFS) [Borthakur et al. 2008], uma ferramenta para processamento de grandes quantidades de dados o MapReduce [Dean and Ghemawat 2008] e o Yet Another Resource Negotiator(YARN) [Vavilapalli et al. 2013] que é a controla as tarefas executadas em um *cluster* Hadoop. Esta escolha teve por o objetivo de simplificar a gerência e instalação dos componentes que compõe o *cluster* Hadoop além de centralizar o controle e monitoramento de todos os serviços e equipamentos presentes. A partir disso, é possível adicionar e remover computadores do *cluster*, e também adicionar e remover serviços em todos os computadores pertencentes. Essas funcionalidades permitem que a execução de programas nesse ambiente sejam elásticos, escaláveis e adaptáveis. Para garantir a alta disponibilidade, o *cluster* foi configurado em uma nuvem pública que garante funcionamento ininterrupto do sistema.

A Figura 1 ilustra a configuração da arquitetura completa para implementação do sistema de captura. Os dados de ataques são capturados pelo *catcher* que age como um *proxy* simulando uma conexão real com o site que fornece as informação, como por exemplo o site da *NorseCorp* e da *LookingGlassCyber*. Esse programa recebe os dados das fontes abertas e cria uma requisição do tipo HTTP GET enviando em seguida esta mensagem ao programa *Flume* através de um *source* HTTP.

Este *source* está conectado e replicando os dados a dois *channels*, *Kafka Channel* e *File Channel*. Após os dados serem enviados a estes dois canais, por uma característica do *Flume*, ele só será retirado do *channel* após ter sido entregue em sua respectiva *sink*. Nesse caso existem duas *sinks*, *MorphlineSolr Sink* e *AsyncHBase Sink*. A primeira processa os dados para inserção desses no programa *Solr*, o segundo os processa para armazenamento do programa *HBase*.

Com todos os dados já inseridos e persistidos, a ferramenta HUE é responsável por exibir esses dados de forma clara, objetiva e simples. São inúmeras opções de exibição e filtragem dos dados, podendo o usuário em tempo real extrair informações e construir conhecimento prévio sobre uma série de aspectos dos dados de ataques capturados.

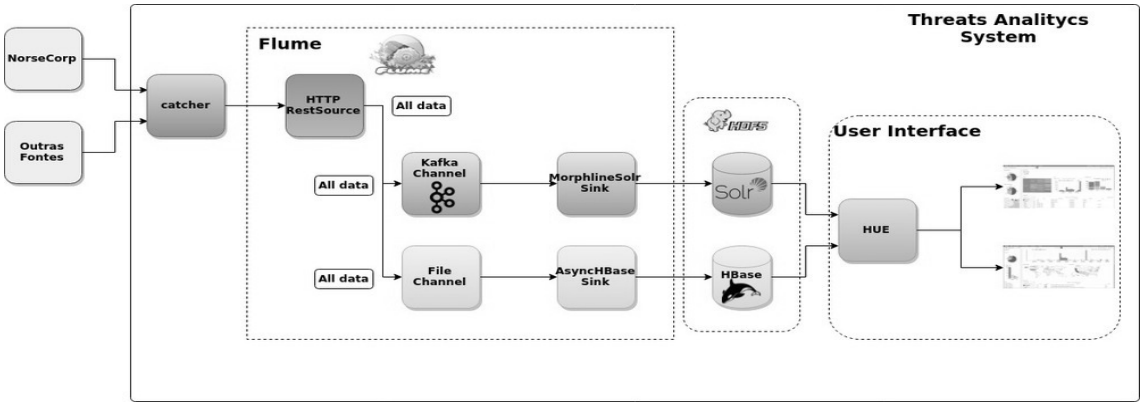


Figura 1: Arquitetura do sistema

4. Captura dos Dados

O protocolo para captura dos dados busca simular uma conexão legítima de um navegador para o site da fonte de dados. Utilizando a biblioteca *websocket-client*, o programa cria uma conexão do tipo *websocket* diretamente com o servidor da fonte. A utilização deste protocolo como base para o programa de captura se deve ao fato de grande parte das fontes de dados utilizar este protocolo para comunicação entre e aplicação cliente e o servidor.

Muitas outras fontes de dados sobre ataques podem ser encontradas na internet, como por exemplo: 1. *ThreatMap CheckPoint* - [CheckPoint 2018] 2. *Digital Attack Map* - [Arbor Networks 2018]. Uma das fontes de dados escolhida foi o site da empresa *NorseCorp*. Uma empresa de Cibersegurança especializada em entregar inteligencia contra ataques de rede, brechas de segurança e ameaças diversas. Ela oferece atualizações contínuas do estado de rede através de vários sensores espalhados pelo mundo.

Para divulgar a tecnologia utilizada, a *NorseCorp* disponibiliza parte destas informações através de uma ferramenta *online* em forma de um mapa interativo. O site pode ser acessado em `http://map.norsecorp.com/`, sendo possível visualizar em tempo real os ataques ocorrendo contra os clientes da empresa.

Os dados recebidos por essa fonte de dados são pré-processados pela empresa e limitados. A Tabela 1 trás um resumo dos dados que são armazenados no banco de dados acompanhados de exemplos de valores. Como não havia descrição dos dados disponível, restou a observação dos dados onde foi possível obter algumas informações iniciais como a existência de dados redundantes como por exemplo o *dport* e o *svc* que sempre possuem sempre o mesmo valor. O *vector_id* e *type* só tem significado quando utilizados dentro da página da empresa pois são utilizado durante as funções de desenho do mapa. Mesmo assim todos os dados são armazenados no banco por completude.

5. Resultados

Depois de implantar a arquitetura definida na Seção 3 a captura e ingestão dos dados se iniciou. Entre os meses de Agosto e Dezembro de 2017 cerca de 14 milhões de registros de ataque se encontravam internalizados.

Mesmo com essa grande quantidade de entradas as consultas na ferramenta não apresentavam variação em seu tempo de resposta, exibindo o resultado quase que imediatamente após a seleção dos filtros. Este comportamento esta relacionado ao fato de que a

Tabela 1: Descrição dos dados capturados da empresa NorseCorp

Nome Dado	Descrição	Exemplo
_id	Identificador único do objeto no banco de dados	ObjectId("58d1bbbfe2e558522a"),
city	Cidade de origem do ataque	Washington
city2	Cidade onde o alvo do ataque reside	De Kalb Junction
dport	Porta em que o ataque foi executado	25
countrycode	Código do país de origem do ataque	US
countrycode2	Código do país do alvo	US
country	Identificação do país de origem do ataque	US
country2	Identificação do país do alvo	US
latitude	Latitude aproximada da origem do ataque	38.95
longitude	Longitude aproximada da origem do ataque	-77.02
latitude2	Latitude aproximada do alvo	44.48
longitude2	Longitude aproximada do alvo	-75.3
svc	Serviço explorado pelo ataque	25
timestamp	Data e hora no momento da captura do dado	2017-03-21 23:48:15.161973
vector_id	Informação encaminhada pelo servidor para desenho em tela de animação	NumberLong("303241862584")
org	Organização que detêm o bloco de endereços de IP o qual originou o ataque	Microsoft Corporation
type	Informação sobre a ferramenta da empresa que capturou a informação do ataque	ipviking.honey
md5	Número de IP da máquina que enviou o ataque	65.55.169.250

indexação dos dados é realizado no momento da inserção desses. Assim o processamento exigido durante as consultas é reduzido consideravelmente.

A *Google Cloud Platform* foi a solução escolhida para hospedar o cluster. Neste ambiente duas máquinas foram criadas, uma chamada Master com 2 CPUs virtuais, 14GB de memória RAM e 100GB de memória secundária. E outra chamada Slave com 1 CPU virtual, 4GB de memória RAM e 100GB de memória secundária.

Depois de internalizar os dados e processá-los, foi observado que portas TCP com valores 50864 e 53413 eram frequentes. Assim, foram definidas duas opções de filtro para uma análise mais apurada: ataques que tenham a porta 50864 como alvo e ataques que utilizem a porta 53413 com o mesmo objetivo. A escolha das portas se baseou em todas as informações anteriores sobre vulnerabilidades nessas porta e por conta da concentração da origem desses. Após realizar essa seleção novamente todos os dados no painel de ferramentas exibem somente os dados selecionados.

A Figura 2 ilustra as regiões de onde partiram os ataques que exploravam possíveis vulnerabilidades nessas portas. Na Figura 2a somente a região leste da China é exibida onde os marcadores em verde representam o agrupamento das cidades de onde partiram os ataques, enquanto na Figura 4b além desta a região da Coreia do Sul também aparece como sendo uma das principais origens.

Porém, o mapa representado na Figura 4b no momento da captura da imagem estava centralizado em uma região específica, portanto não exibe completamente todos os países que utilizam essa porta. A Figura 2, que traz os países atacantes bem como suas respectivas cidades.

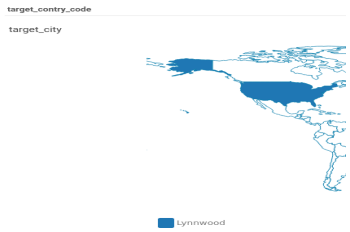
Prosseguindo, a Figura 3 ilustra os mapas que contêm a informações sobre as cidades alvo. No mapa à esquerda somente a cidade de *Lynnwood*(US) aparece como alvo de ataques que tem a porta 50864 como alvo, a direita entretanto mais cidades aparecem como alvo quando a porta 53413 é definida: a cidade de *Nama* (Japão), *Aix-En-provence* (França), *Dubai* (Emirados Árabes Unidos) e São Francisco (US).

As cidades atacantes estão representadas na Figura 4, onde os ataques se concentram em três nações: China, Coreia do Sul e Paquistão. Na China chama a atenção que

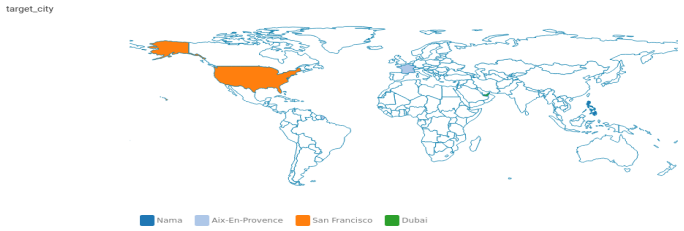
(a) Região de origem dos ataques porta 50864

(b) Região de origem dos ataques porta 53413

Figura 2: Regiões de origem dos ataques em portas altas



(a) País e cidade de destino dos ataques na porta 50864



(b) País e cidade de destino dos ataques na porta 53413

Figura 3: Países e cidades de destino dos ataques em portas altas

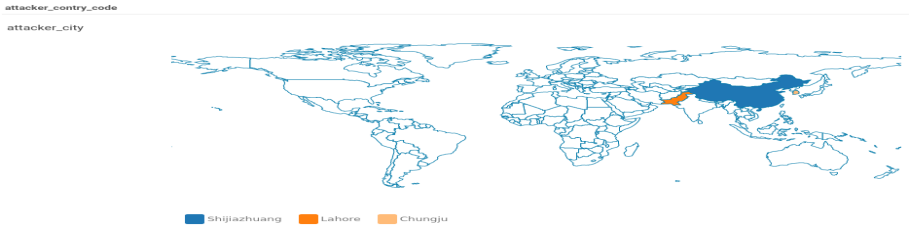
nas duas portas escolhidas para a análise há uma cidade diferente representada, na porta 50864 a cidade de *Guanzhou* e para a porta 53413 a cidade de *Shijiazhuang* que aparece como originária dos ataques.

Como complemento à informação de regiões e cidades atacantes e alvos, há também a alternativa de se estudar o número e origem dos ataques através das ferramentas de filtro: *attacker_country_code* e *target_country_code*. Por vezes é mais legível acessar as informações observando-se esses campos do que as outras opções de ferramentas.

Realizando uma busca rápida, vários fóruns e sites surgiram relatando o interesse de atacantes chineses em portas com valores acima de 40.000 com destaque para as portas 50.866, 50.864 e 54.413 [Gman_beeman , Jill Scharr 2014]. Outra informação que surgiu durante as pesquisas foi um possível *backdoor* em roteadores *Netis* de fabricação chinesa a partir da empresa *Netcore Group* [Eduard Kovacs 2014], em que usando a porta 54.413 [Jill Scharr 2014] os atacantes poderiam tomar total controle do dispositivo sem nenhum conhecimento de seu proprietário.

Essas referências corroboram o resultado obtido através da observação dos dados capturados, e além disso trazem novas informações sobre quais países estão utilizando dessas vulnerabilidades. Além dos filtros para países que aparecem como alvo de ataques,

(a) País e cidade de origem dos ataques na porta 50864



(b) País e cidade de origem dos ataques na porta 53413

Figura 4: Países e cidades de origem dos ataques em portas altas

também se tem a opção de filtrar os países de onde partem os ataques. essa opção tem como motivação demonstrar o papel de dois países distintos como origem de ataques. Para ilustrar bem a diferença foram escolhidos os dois maiores países atacantes registrados na base de dados, Estados Unidos(US) e China(CN).

Cada filtro foi selecionado independentemente, e os resultados das ferramentas de visualização podem ser conferidas nas Figuras 5 e 7. As primeiras duas ferramentas exibem as portas que receberam a maior quantidade de ataques, a Figura 5b e a Figura 5a dizem respeito a informações sobre as portas mais atacadas por máquinas situadas na China. Já a Figura 5d e a 5c ilustram as mesmas informações só que dos Estados Unidos.

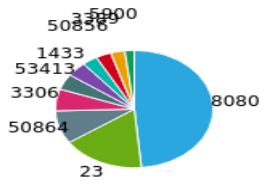
Enquanto, a ferramenta exibida nas Figuras 5a e 5c tem como propósito mostrar a proporção de ataques em cada porta através de um gráfico circular, a ferramenta exposta nas Figuras 5b e 5d tem um objetivo a mais. Elas podem ser utilizadas para aplicação de um novo filtro para as informações já exibidas desta forma restringindo ainda mais a aplicação do filtro gerando novas informações sub-sequentes.

A primeira vista fica evidente que existem várias diferenças entre os dois resultados, observando o gráfico circular é possível de imediato extrair informações claras sobre o ponto desejado. Enquanto, os ataques que partem da China tem múltiplas portas com quantidades significativas de ataques os Estados Unidos concentram os ataques em uma porta específica, sendo que as outras somente representam menos de 25% do total de ataques.

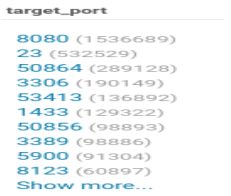
Observando a lista de portas nas Figuras 5b e 5d o usuário é capaz de perceber que a porta mais atacada pela China é a porta 8080 enquanto a partir dos Estados Unidos a porta mais utilizada é a 25.

A porta 23 figura em segundo lugar nos dois países, isso se deve ao fato de que essa porta pertence ao protocolo *Telnet* que além de altamente inseguro e alvo constante de ataques. Também chama a atenção que entre as portas mais atacadas pelos americanos

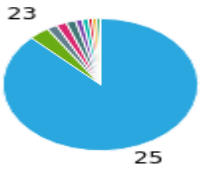
não há nenhuma que ultrapasse o valor de 40.000. Mas, observando as portas alvo de ataques vindos da China, há pelo menos três portas com valores superiores.



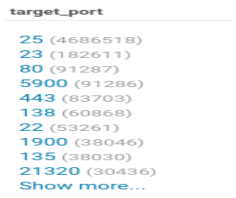
(a) Distribuição de portas atacadas a partir da China



(b) Tabela de portas atacadas a partir da China



(c) Distribuição de portas atacadas a partir dos Estados Unidos



(d) Tabela de portas atacadas a partir dos Estados Unidos

[1ex]

Figura 5: Informações sobre portas atacadas a partir da China e Estados Unidos

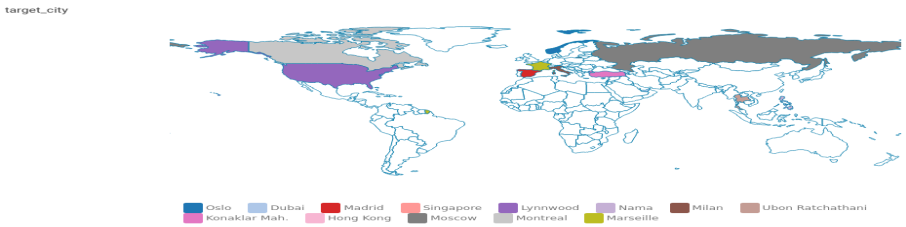
Continuando a percorrer o painel de ferramentas, duas opções de observação de mapas estáticos são mostradas, uma para representar os países e cidades que recebem mais ataques e outra para as cidades que são a origem destes ataques. Com o filtro devidamente selecionado a Figura 6 apresenta os mapas que contêm as cidade mais atacadas pela China e pelos Estados Unidos, respectivamente. A correspondência entre as cidades e os países é representada pelas cores dos retângulos.

Há várias similaridades entre o conjunto de países atacados a partir dos Estados Unidos e da China, como por exemplo: Noruega, Estados Unidos, Rússia, França, Itália. Com os Estados Unidos acontece um fato semelhante ao ocorrido com a França, o próprio país aparece como um de seus principais atacantes. No caso dos americanos a explicação parece mais simples já que a maioria dos *datacenters* que proveem serviços de *Cloud Computing* se encontram nesse país, fato esse que se confirma ainda mais quando observamos as organizações das quais partem o maior número de ataques.

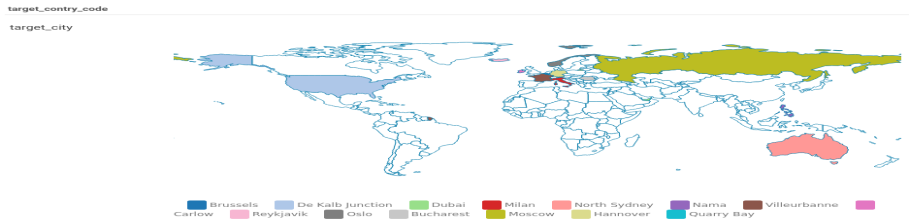
Na Figura 7b a organização *Microsoft Corporation* detêm quase que a totalidade dos atacantes hospedados no país. É importante frisar que, mesmo não sendo o foco da empresa fornecer serviços de acesso à Internet, a Microsoft possui sobre seu controle várias faixas de IPs, que são utilizadas dentro de seus serviços de *Cloud*. Por esse motivo, que ela figura na lista de organizações.

Por fim, o último gráfico apresentado para essa seleção de dados é representado na Figura 7. Ele exibe as dez organizações que são *Internet Service Providers(ISPs)* das máquinas que foram origem dos ataques capturados pela ferramenta. Os nomes de cada organização estão em ordem abaixo do gráfico e algumas vezes encontram-se deslocadas para direita.

A informação mais nítida ao se olhar para o gráfico é a da concentração dos ata-



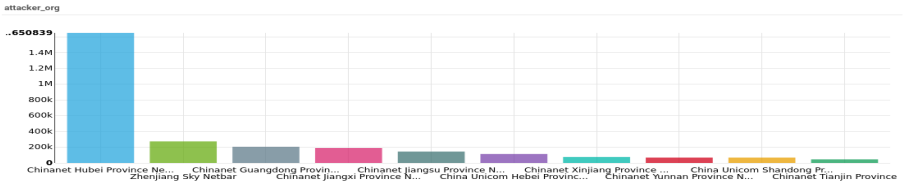
(a) Cidades mais atacadas a partir da China



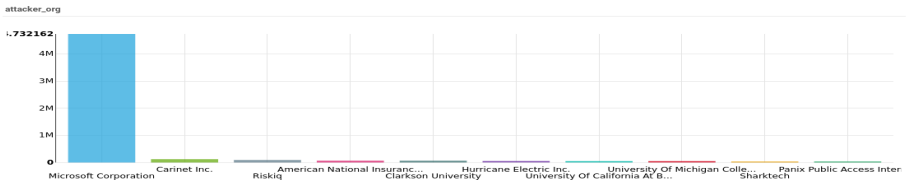
(b) Cidades mais atacadas a partir dos Estados Unidos

Figura 6: Cidades mais atacadas a partir da China e Estados Unidos

cantes em uma só organização para cada país, permanecendo as demais com baixíssima participação nesse ponto.



(a) Top 10 organizações responsáveis pelos IPs - China



(b) Top 10 organizações responsáveis pelos IPs - Estados Unidos

Figura 7: Top 10 organizações responsáveis pelos IPs da China e Estados Unidos

6. Conclusão

A abordagem adotada oferece um leque de possibilidades que não foram totalmente exploradas dado o tempo disponível voltado para o desenvolvimento da arquitetura e seu correto funcionamento, mas pode-se avaliar as perspectivas de sua utilização.

A apresentação das informações de forma visual com simplicidade e ilustrações de fácil compreensão permitem que um usuário, mesmo que desprovido de informações sobre como aqueles dados estão sendo processados, possa entender e interagir com esses. Isso permite que padrões sejam observados e que o usuário com seu conhecimento pregresso ou através de outras fontes possa enriquecer os dados gerando informações e as validando. Assim, a massa de dados contendo escassas informações pode ser convertida,

através de diversas técnicas de processamento, em informações que agregam diversos conceitos e que geram uma visão completamente nova do que se julgava ter pleno conhecimento.

No ambiente conseguiu-se representar as mudanças ao longo do tempo. Assim o comportamento geral das informações analisadas pode ser comparada para se definir uma tendência ou várias delas. Assim, o ambiente implementado permite a produção de inteligência de ameaças com a perfilização de ataques, atacantes e alvos. Uma possível melhoria ao ambiente seria a incorporação de facilidades de aprendizado de máquina

Referências

[Arbor Networks 2018] Arbor Networks, I. (2018 (acessado em 19 de Junho de 2018)). *Digital Attack Map*. `http://www.digitalattackmap.com/#anim=1&color=0&country=ALL&list=0&time=17487&view=map`.

[Bachupally et al. 2016] Bachupally, Y. R., Yuan, X., and Roy, K. (2016). Network security analysis using big data technology. In *SoutheastCon 2016*, pages 1–4.

[Borthakur et al. 2008] Borthakur, D. et al. (2008). Hdfs architecture guide. *Hadoop Apache Project*, 53.

[CheckPoint 2018] CheckPoint, S. (2018 (acessado em 19 de Junho de 2018)). *ThreatMap CheckPoint*. `https://threatmap.checkpoint.com/ThreatPortal/livemap.html`.

[Dean and Ghemawat 2008] Dean, J. and Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1):107–113.

[Eduard Kovacs 2014] Eduard Kovacs (2014). Easily Exploitable Vulnerability Found in Netis Routers.

[Gman_beeman] Gman_beeman. My Router is Showing a Bunch of IPs Attacking Me.

[IBM et al. 2011] IBM, Paul, Z., and Chris, E. (2011). *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. McGraw-Hill Osborne Media, 1 edition.

[Janeja et al. 2014] Janeja, V. P., Azari, A., Namayanja, J. M., and Heilig, B. (2014). B-dids: Mining anomalies in a big-distributed intrusion detection system. In *2014 IEEE International Conference on Big Data (Big Data)*, pages 32–34.

[Jia 2017] Jia, W. (2017). Study on network information security based on big data. In *2017 9th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, pages 408–409.

[Jill Scharr 2014] Jill Scharr (2014). Possible Backdoor Found in Chinese-Made Routers.

[Vavilapalli et al. 2013] Vavilapalli, V. K., Murthy, A. C., Douglas, C., Agarwal, S., Konar, M., Evans, R., Graves, T., Lowe, J., Shah, H., Seth, S., et al. (2013). Apache hadoop yarn: Yet another resource negotiator. In *Proceedings of the 4th annual Symposium on Cloud Computing*, page 5. ACM.

[Wu et al. 2017] Wu, Y., Zheng, L., Heilig, B., and Gao, G. R. (2017). Hamr: A dataflow-based real-time in-memory cluster computing engine. *The International Journal of High Performance Computing Applications*, 31(5):361–374.