FICE 7ª FEIRA DE INICIAÇÃO CIENTÍFICA E EXTENSÃO 05 e 06 de setembro

LEVANTAMENTO E MAPEAMENTO DA DISTRIBUIÇÃO DE *MIDDLEBOXES* PRESENTES NA *WORLD WIDE WEB*

Bruno Chagas[1]; Adenes Schwantz[2]

INTRODUÇÃO

A internet é baseada em roteadores e *hosts*. Conforme Sherry et al. (2012) evidenciam, "com o crescimento da rede, surgiram alguns problemas, tais como vírus, segurança de dados, falta de endereços IPv4, entre outros". Para contornar esses problemas, antivírus, *firewalls*, NATs (Tradudores de Endereço de Rede), etc., são usados. Entretanto, a rede mundial de computadores não foi projetada para lidar com tais dispositivos.

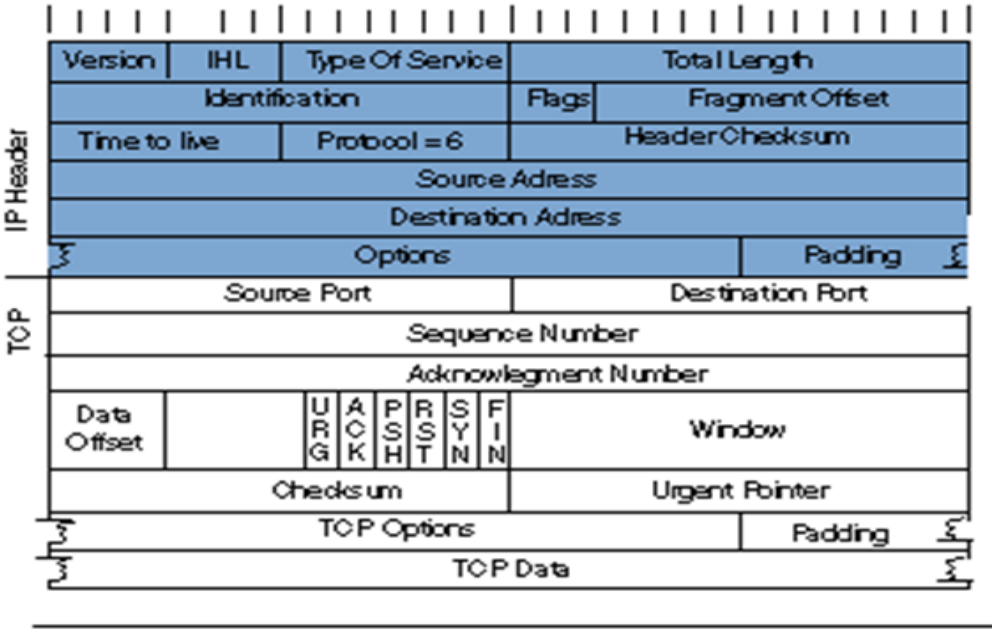
Outros dispositivos, que não roteadores e *hosts*, executando qualquer função são chamadas *middleboxes*, nome proposto por Lixia Zhang e formalizado na RFC 3234. De acordo com Medina et. al. (2007) "uma *middlebox* pode modificar a informação contida nos campos dos cabeçalhos do protocolo TCP/IP (Protocolo de controle de transmissão/Protocolo de internet)". Os autores ainda apontam que estes dispositivos atuam principalmente nas camadas de transporte e rede.

Os autores Detal et al. (2013) apresentam Tracebox, que é uma ferramenta que permite a detecção de *middleboxes* em qualquer caminho de rede, ou seja, entre uma fonte e um destino. Ainda, este *software* permite gerar *probes* de diferentes especificidades para então envia-las a qualquer endereço IP desejado. Uma *probe* funciona como uma sonda, analisando as modificações que esta sofre por parte das *middleboxes* a que é submetida. O Tracebox ainda compara as *probes* que chegaram

[1] Aluno do Instituto Federal Catarinense, Campus Videira. Curso superior de Engenharia Elétrica. E-mail: abschagas@hotmail.com
[2] Professor Orientador do Instituto Federal Catarinense, Campus Videira. Curso superior de Engenharia Elétrica. E-mail: adenes.schwantz@ifc.edu.br

ao destino com aquelas que foram enviadas, detectando assim as *middleboxes* presentes no caminho testado e as modificações sofridas. Todos os campos do datagrama TCP/IP, mostradas na figura 1, são passiveis de sofrerem modificações resultantes de uma operação efetuada por uma *middlebox*.

Figura 1. Datagrama TCP/IP



Fonte: FORD et. al. (2013).

Devemos atentar para o fato de que Tracebox é capaz de detectar a maioria dessas modificações, desde que não sejam desfeitas. Cabe salientar que algumas *middleboxes* efetuam modificações e depois inserem os dados originais nos campos modificados.

PROCEDIMENTOS METODOLÓGICOS

Através de um estudo concentrado, em diversos caminhos de rede, levantou-se os tipos mais comuns de modificações performadas por *middleboxes*. Esse estudo envolveu uma grande amostra de endereços IP da *world wide web* (geograficamente distribuídos), da ordem de 1152 endereços. A partir disso pode-se identificar as

middleboxes e as modificações que estas causaram. Os testes foram realizados de quatro maneiras, identificadas de 0 a 3, afim de tentar localizar *middleboxes* que atuam e reagem com determinados tipos específicos de pacotes de dados. Em nenhum momento foram contabilizadas as mudanças nos campos IP TTL (*Time to live*) e IP *CheckSum*, já que são modificações esperadas em todas as *probes*, bem como em todos os saltos de um caminho de rede qualquer que seja.

A *probe* 0 faz parte do grupo controle, nela será possível observar modificações que as *middleboxes* normalmente provocam. Já as *probes* de 1 a 3 possuem parâmetros para detectar modificações diferentes do grupo controle. As *probes* 1 e 2 testaram argumentos do cabeçalho TCP do pacote. Argumentos estes tais como *MSS*, que define o tamanho do *payload*, *WSCALE*, que é expansão para uma janela de 32 bits e adiciona um fator de escala na janela de 16 bits do cabeçalho TCP. Ainda se testou *MPCAPABLE*, que verifica a capacidade da rede de executar uma tarefa em múltiplos caminhos, como definido na RFC 6824. Ainda se averiguou *TS*, que auxilia na performance do protocolo TCP como exemplifica Silbersack (2005) em seu estudo. O parâmetro *NOP*, que é usado para preencher a lista de opções, bem como *SACKPermitted*, que altera o comportamento de confirmação do TCP, também foram exaustivamente testados. A *probe* 3 testou única e exclusivamente o argumento *MPCAPABLE*, afim de verificar a adaptabilidade da rede à esta recente opção adicional do protocolo.

Para a realização dos testes, foi composta uma lista com os endereços IP que serviram de alvo. Um *script* previamente preparado realizava a leitura dessa lista e executou os comandos, integrados ao Tracebox, no terminal do sistema Linux. Ao todo, foram testados 1152 endereços IP com quatro diferentes configurações de argumentos e em três modos de conexão, sendo estes Ethernet (802.3), Wi-Fi (802.11) e 3G (HSDPA). As figuras apresentadas na sequência final irão evidenciar as alterações sofridas pelos pacotes, em percentual, de todos os lançamentos realizados.

Os endereços foram selecionados por sua localização geográfica, logo, existirão endereços localizados em diferentes continentes afim de determinar possíveis interferências em regiões especificas. Como exemplo destaca-se casos encontrados em endereços localizados na Rússia. Assim gerando a possibilidade de teste também em um panorama global.

Desta forma, todas as *probes* são enviadas a todos os endereços IP disponíveis na lista. Os dados resultantes gerados pelo Tracebox são automaticamente armazenados em disco. Caso ocorresse algum problema associado ao *script* de execução, seria possível fazer a contagem de ocorrências de forma alternativa.

Após a execução de todos os testes, o *script* inicia a leitura dos arquivos resultante destes, contabilizando as mudanças e adicionando-as em uma tabela. Cada *probe* foi contabilizada em uma tabela diferente, para que assim seja possível saber quais são as diferenças entre as mesmas e a soma total é feita manualmente.

As comparações foram feitas entre *probes* 0 e 3, buscando diferenças nas *middleboxes* que são acionadas com a função *MPCAPABLE*. Isso também é feito entre as *probes* 1 e 2, identificando diferenças principalmente no protocolo TCP, visto

que as funções são, em sua maioria, relacionadas à este protocolo. Por fim, são apresentadas algumas figuras que denotam a localização (ampla) de alguns endereços IP que foram testados, bem como os caminhos entre fonte (Videira/SC) e os diversos destinos (alguns mostrados nos mapas abaixo). Os caminhos físicos e a localização geográfica das *middleboxes* pode ser observada, com maiores detalhes, no estudo feito por Schwantz (2016). Aqui o foco é dado apenas no endereço final e nas modificações ocorridas.

Figura 2. Alguns dos endereços IP localizados no continente Europeu

Fonte: Autores

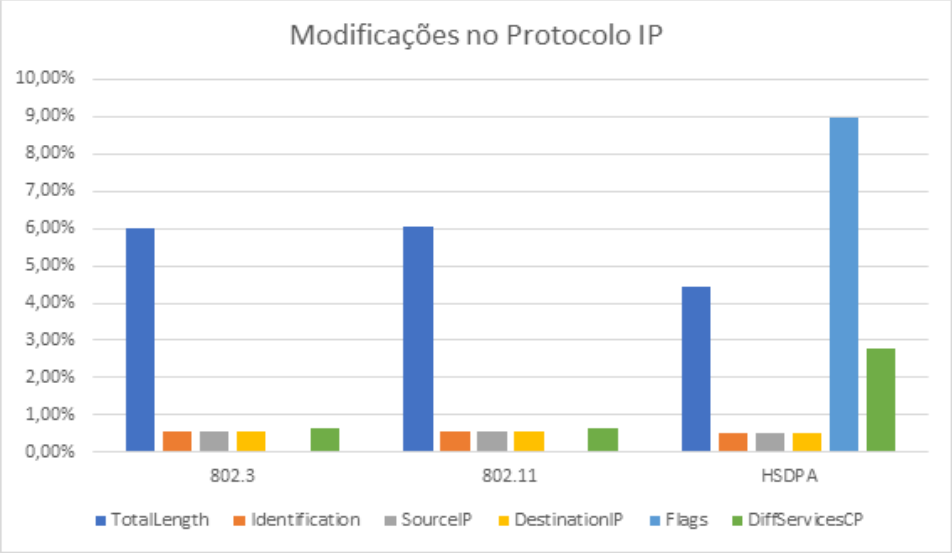
Figura 3. Alguns dos endereços IP localizados na América do Norte

Fonte: Autores

RESULTADOS E DISCUSSÕES

Os dados levantados pelos testes mostraram que o protocolo IP apresentou um comportamento anômalo em vista do padrão esperado. Em vista de uma porcentagem menor que 0,1% das modificações no campo *Flags*, se encontrou valores bem superiores em testes realizados via HSDPA. O campo *Flags* indicou 9,04% do total de modificações, sendo que 8,95% foram apenas com a conexão HSDPA, assim como o campo *DiffServicesCP*, que indicou um total de 4,05% das modificações, sendo 2,76% com a conexão HSDPA.

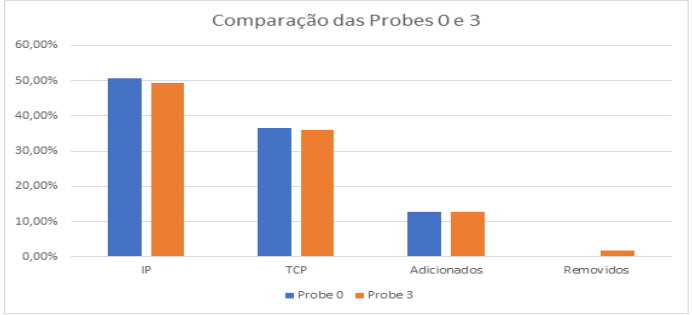
Figura 4. Modificações encontradas no protocolo IP



Fonte: Autores

O quantitativo de dados mostrou que os pacotes que foram enviados via protocolo 802.3 e 802.11 receberam uma quantidade menor que 0,2% de *flags*, enquanto na conexão HSDPA o mesmo campo recebeu aproximadamente 9%.

Figura 5. Comparação de modificações entre as *probes* 0 e 3

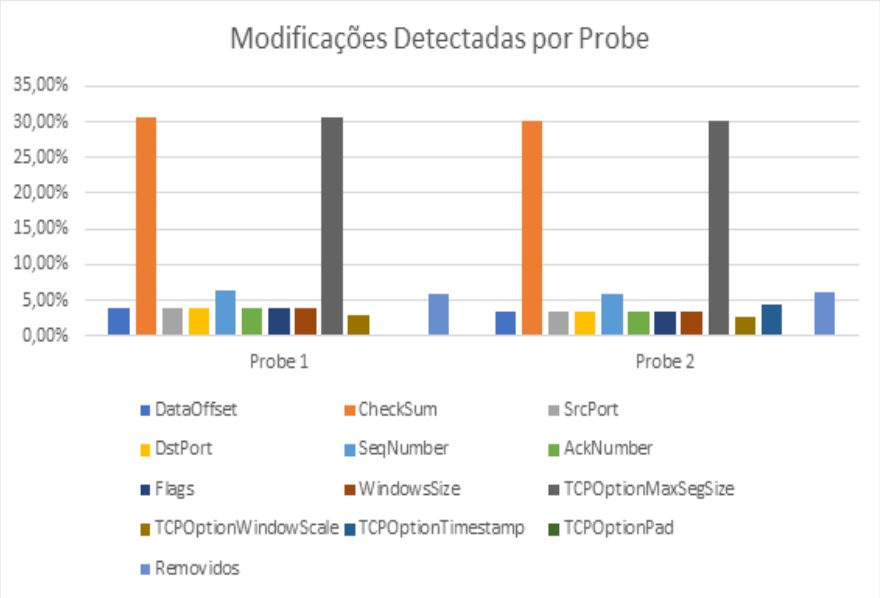


Fonte: Autores

Logo, analisando o percentual de modificação dos campos *flags* e *DiffServicesCP* nas três conexões e o tempo total de execução dos testes, entende-se que a conexão HSDPA obteve um pior desempenho em relação as conexões 802.3 e 802.11. Portanto outros pacotes estavam sendo priorizados. Com os dados obtidos com a pesquisa não é possível identificar claramente o motivo da prioridade dada por uma *middlebox* especifica, porém, com a análise geral dos dados foi possível determinar o desempenho de cada conexão, bem como uma possível prioridade dada pela rede a certos dados.

Desta maneira, as *probes* 0 e 3 identificaram o mesmo padrão na maioria dos campos, com a exceção da alteração “Removidos” que apresentou uma taxa de 1,70% na execução da *probe* 3. Isso aponta para alguns endereços IP que foram testados pela *probe* não possuem múltiplos caminhos a serem percorridos pelos pacotes.

Figura 6. Comparação do Protocolo TCP das *probes* 1 e 2



Fonte: Autores

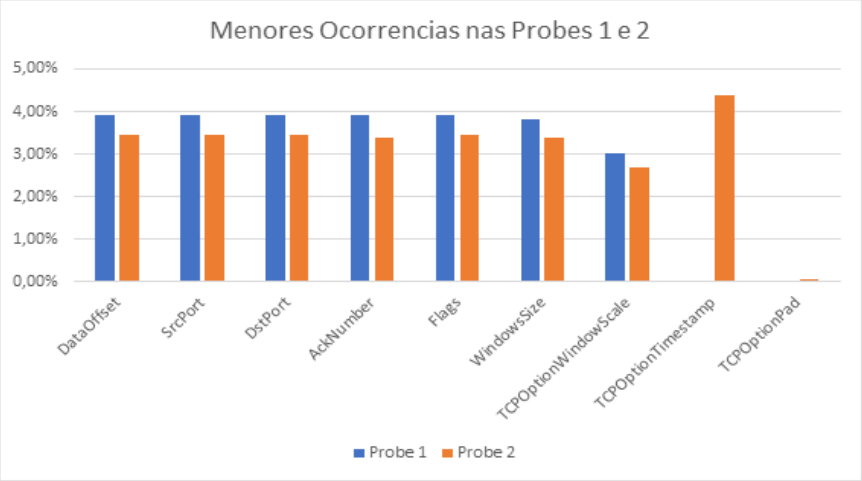
Na execução da *probe* 1, os argumentos possibilitaram a visualização de *middleboxes* atuando mais intensamente no protocolo TCP dos pacotes. Um dos campos com maior percentual foi o campo *TCPOptionMaxSegSize* que define o tamanho do campo *Payload* e *MSS*. Em alguns caminhos de rede, conforme aponta Cohen (2017) “a fragmentação do pacote não é permitida, também sempre que possível deve ser evitada”. A melhor forma de evitar a fragmentação é então ajustar o tamanho máximo de segmento. Isso faz com que a *middlebox* possivelmente presente modifique o pacote, seguindo sua configuração pré-estabelecida.

Nos testes envolvendo a *probe* 2, o mesmo padrão pode ser observado na maioria dos campos. Entretanto, esta *probe* possui argumentos inexistentes na *probe* anterior, logo, encontrou um padrão de atuação e campos modificados que não

haviam sido encontradas. O campo *TCPOptionTimeStamp* (*TS*) sofreu 4,38% das modificações detectadas. Logo, para alguns endereços IP, o pacote foi perdido e na tentativa de recuperar o pacote, a partir de um "número de sequência" acondicionado, sofreu uma modificação em um de seus campos.

O campo *TCPOptionPad* foi modificado apenas três vezes no decorrer dos experimentos, contabilizando 0,04% das modificações do protocolo TCP. Como essa modificação não possui um valor significativo, não dever-se-ia alterar a experiência para o usuário, tão pouco para a rede como um todo. Apesar de ser uma modificação com efeitos altamente negativos, ocorreu em uma quantidade muito pequena da amostra.

Figura 7. Caracterização das ocorrências em menor quantidade



Fonte: Autores

Em determinado teste, realizado utilizando a conexão 802.3, envolvendo a *probe* 3, foi encontrado um resultado inesperado e que, idealmente, nunca deveria ocorrer. O cabeçalho TCP do pacote foi removido por completo.

Neste caso, se sucedeu o mesmo padrão que foi apresentado nas *probes* 1 e 3, porém, neste caso específico, no salto de número 23, do caminho de rede, o tipo "removidos" apresentou a modificação "-TCP", ou seja, o protocolo TCP foi removido. A modificação ocorreu no IP 61.19.7.2, que está localizado em Bangkok, Tailândia (suas coordenadas geográficas são 13.7563º Norte, 100.502º Leste), próximo ao seu destino (14.9716º Norte, 102.0825º Leste), região que possui restrições para o tráfego de dados.

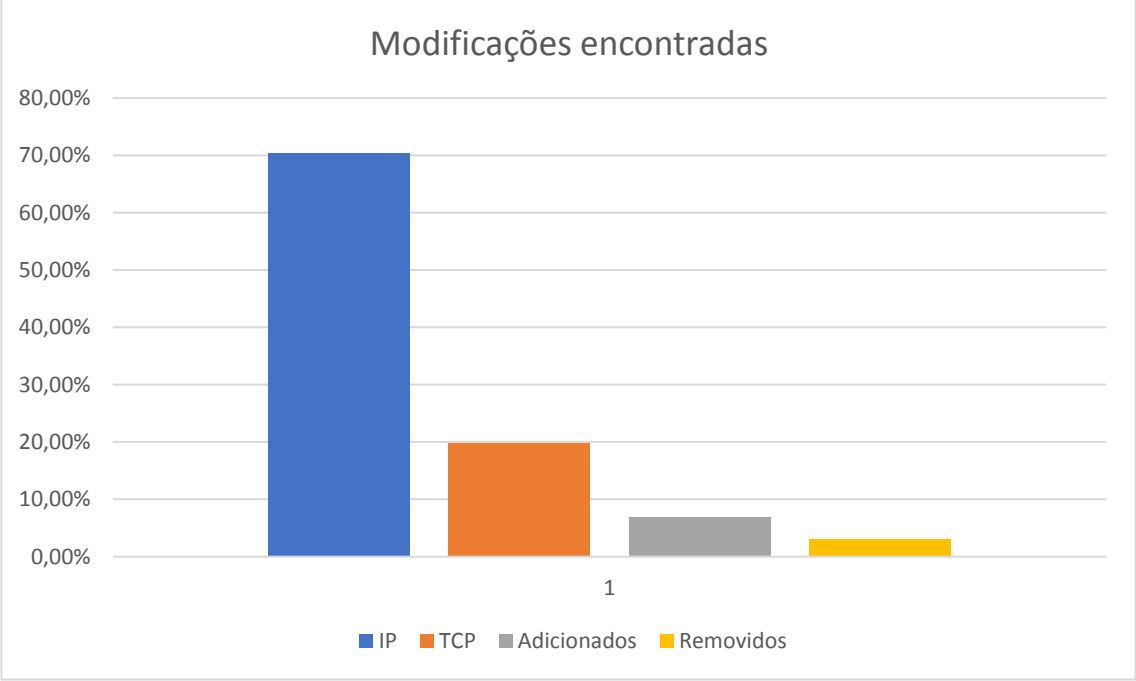
Como normalmente essas modificações não ocorrem, e neste caso ocorreu apenas uma única vez, não há padrão para ser observado, portanto, não é possível apresentar com certeza o real motivo da modificação. Porém este mesmo caso é passível de menção, já que mostra o poder de atuação e modificação das *middleboxes* atualmente presentes na rede.

Outra modificação ocorreu uma vez nas três conexões com a utilização da *probe* 2. Essa modificação ocorreu no IP 80.68.253.9, localizado em Moscou

(55.7522º Norte, 37.6156º Leste) que é o destino do endereço IP testado.

A modificação identificada foi um *TCPOptionPad Kind*, porém, esta é uma opção de TCP conhecida como *no operation*, que foi mencionado anteriormente, e é definido na RFC 793, também reforçando a conclusão apresentada no parágrafo anterior.

Figura 8. Panorama geral das modificações detectadas



Fonte: Autores

Por fim, na figura 8, é apresentado o panorama geral de todos os testes. Essa figura evidencia o resultado geral do estudo. Como pode-se notar, cerca de 70% do total das ações de *middleboxes* se concentrou em campos pertencentes ao cabeçalho IP dos pacotes enviados. A seguir, próximo de 20% do total, temos as modificações relativas ao cabeçalho TCP do pacote. Por último, com números bem menores, temos a adição e remoção completa de campos do pacote. Apesar de mais grave, o comportamento de adição e remoção, quando ocorre em redes privadas, o tráfego geral não deverá ser afetado de forma importante.

Já em relação as demais modificações, pode-se observar que o protocolo IP vem sendo o que mais sofre com a atuação das *middleboxes*. Sendo assim, estudos futuros deveriam focar nesse cabeçalho do pacote, afim de mitigar interferências mais sérias em relação ao tráfego geral. Além do mais, esse dado nos mostra que mesmo com a expansão da disponibilidade de endereçamento, através de endereços IPv6, as *middleboxes* já implantadas ignoram, ou não detectam, tal disponibilidade. Como resultado disso impõem suas restrições sem explorar alternativas. Essas continuam performando suas atividades, o que vemos na forma de modificações nos campos do

pacote.

CONSIDERAÇÕES FINAIS

Dadas as devidas proporções, bem como limitações, a execução da pesquisa atendeu plenamente seus objetivos. Foi possível identificar padrões e casos específicos que vem sendo executados pelas mais diversas *middleboxes* presentes na grande rede. A presença destas, ainda que conhecida, mostrou-se forte, bem como sua atuação, atendendo seus requisitos. Requisitos estes que obedecem às mais diversas especificações, determinadas por aqueles (provedores, empresas, usuários) que as instalaram.

A presença de *middleboxes* se mostrou incontestável, e supreendentemente as redes atuais dependem fortemente destas, em um nível superior ao pensado. Este estudo apontou diversas características das mesmas, além de revelar alguns padrões de atuação e suas prioridades de execução. Ainda, pode-se observar quais protocolos de rede vem sofrendo maiores e menores modificações de seus campos, quando submetidos a *middleboxes*.

REFERÊNCIAS

CARPENTER B.; BRIM S., **Middleboxes: Taxonomy and issues**. Internet Engineering Task Force, RFC 3234, 2002.

COHEN, M. **MTU and MSS**. Disponível em: https://www.incapsula.com/blog/mtu-mss-explained.html. Acesso em 09/04/2018

DETAL, G.; HESMANS, B.; BONAVENTURE, O.; VANAUBEL, Y.; DONNET, B. **Revealing Middlebox Interference with Tracebox**. In: Proc. ACM SIGCOMM Conference, v. 18 n. 85 p. 1-8, 2012.

FORD, A.; RAICIU, C.; HANDLEY, M.; BONAVENTURE, O. **TCP extensions for multipath operation with multiple addresses**. Internet Engineering Task Force, RFC 6824, 2013.

HONDA, M.; NISHIDA, Y.; RAICIU, C.; GREENHALGH, A.; HANDLEY, M.; TOKUDA, H. **Is It Still Possible to Extend TCP** In: Proc. ACM SIGCOMM Conference, v. 17 n. 13 p. 181-194, 2011.

MEDINA, A.; ALLMAN, M.; FLOYD, S. **Measuring Interactions Between Transport Protocols and Middleboxes**. ICSI Center for Internet Research. Proceedings of the ACM SIGCOMM 2004. 10-2004.

RFC 3234, **Middleboxes: Taxonomy and Issues** Disponível em: http://tools.ietf.org/html/rfc3234.txt Acesso em 04-04-2015.

SCHWANTZ, A. **Detecção, análise e geolocalização de middleboxes presentes na internet**. IF-SOPHIA Revista eletrônica de investigações filosóficas, científicas e tecnológicas, Ano II – Volume II – Número VIII. p. 10-27, 2016.

SHERRY, J.; HASAN, S.; SCOTT, C.; KRISHNAMURTHY, A.; RATNASAMY, S.; SEKAR, V. **Making Middleboxes Someone Else's Problem: Network Processing as a Cloud Service** University of Washington, UC Berkeley, Intel Labs. Proceedings of the ACM SIGCOMM 2012. 10-2012

SILBERSACK, M. J. **Improving TCP/IP security through randomization without sacrificing interoperability**. FreeBSB Project. 2005.