SMTD: Uma aplicação de Redes Definidas por Software no contexto de Redes Domésticas

Rodrigo C. O. Rocha[1], **Humberto T. Marques Neto**[2], **Dorgival Guedes**[1]

[1] Departamento de Ciência da Computação
Universidade Federal de Minas Gerais – Belo Horizonte, MG

[2]Instituto de Ciências Exatas e Informática
Pontifícia Universidade Católica de Minas Gerais – Belo Horizonte, MG

{rcor,dorgival}@dcc.ufmg.br, humberto@pucminas.br

Abstract. *Despite their popularity, the configuration and management complexity of home networks is recognized as a limiting factor for their deployment. This work presents our efforts in the development of SMTD, a home traffic monitoring system, that could create a wide understanding about the traffic patterns observed in each home network and serve as a basis for a management system. Using the Software Defined Network paradigm, SMTD can identify events of interest and act directly on the network to handle them. The system is under construction and our results confirm its power to identify complex behavior.*

Resumo. *Apesar de sua popularidade, a complexidade de configuração e administração de redes domésticas é apontada como um fator limitante em sua implantação. Este trabalho apresenta nosso trabalho no Sistema de Monitoramento de Tráfego Doméstico (SMTD) que pode desenvolver um entendimento abrangente dos padrões de tráfego observados em cada rede doméstica e servir de base para um sistema de gerência. Utilizando o paradigma de Redes Definidas por Software, SMTD pode identificar eventos de interesse e atuar diretamente sobre os comutadores da rede. O sistema está em desenvolvimento e os resultados confirmam a viabilidade da identificação dos padrões.*

1. Introdução

Nos últimos anos, as redes domésticas têm se tornado mais comuns, em parte pelo aumento da capilaridade das redes de acesso e em parte pela redução de preços para os dispositivos de rede necessários à sua instalação. Com a proliferação dos dispositivos com acesso à rede, como telefones celulares, vídeo-games e TVs, a demanda por redes domésticas vem aumentando. Entretanto, esse crescimento não ocorre sem problemas e limitações. A complexidade de configuração e administração dessas redes é apontada como um fator limitante em sua implantação, sendo que roteadores domésticos estão entre os dispositivos eletrônicos mais devolvidos às lojas logo após a compra, pela frustação dos usuários ao tentar fazê-los funcionar [Grinter et al. 2009]. Por esse motivo, já se reconhece que um desafio nessa linha é a criação de soluções de configuração e gerência de redes domésticas que sejam eficientes e de fácil utilização [Dixon et al. 2010].

O paradigma de Redes Definida por Software (SDN) oferece uma forma eficiente de implementar tais soluções. SDNs permitem que se crie uma visão da rede logica-

mente centralizada, sobre a qual aplicações podem atuar e determinar como os dispositivos de rede devem tratar cada fluxo que passe por eles [Casado et al. 2010]. Dessa forma, aplicações SDN podem ser implementadas usando essa visão de rede para ganhar um entendimento mais abrangente sobre as operações da mesma.

A partir desse entendimento, seria possível desenvolver um modelo que representasse os padrões de tráfego "usuais" para cada domicílio, identificando os endereços na rede externa que são mais contactados, os protocolos que são utilizados na rede local e os padrões de tráfego normalmente observados no domicílio. Uma vez de posse dessas informações, um controlador de rede doméstica poderia ser capaz de: (i) identificar que um novo dispositivo tenta se conectar à rede e auxiliar o usuário na sua configuração; (ii) identificar padrões de acesso permitidos ou não em função da sequência de protocolos observada; (iii) diferenciar padrões de acesso usuais entre dispositivos domésticos de um padrão de ataque de um *malware* que acaba de se instalar.

Com a possibilidade de identificar tais eventos na rede local, um sistema de gerência pode ser construído de modo que integre recursos de aprendizado de máquina, para assimilar os padrões aceitáveis em cada rede e detectar outros que possam ser considerados nocivos. Este trabalho apresenta os nossos esforços no desenvolvimento do Sistema de Monitoramento de Tráfego Doméstico (SMTD) que poderia ser integrado a um controlador SDN para desenvolver esse entendimento abrangente dos padrões de tráfego observados em cada rede doméstica.

2. Solução proposta

O Sistema de Monitoramento do Tráfego Doméstico se baseia em um modelo que identifique e armazene os padrões que representam o uso legítimo de um dado conjunto de protocolos, padrão esse que seria causado pela execução de uma aplicação conhecida no contexto da rede doméstica. O principal desafio do sistema é ser capaz de estabelecer relações de causa e efeito ou de co-ocorrência de certos padrões de tráfego a partir da observação dos fluxos de dados identificados pelo controlador da Rede Definida por Software. Para esse fim, utilizamos a técnica de identificação de correlações de tráfego proposta pelo projeto Magpie [Barham et al. 2004]. Naquele sistema, instrumentação era utilizada para criar identificadores que eram associados às principais operações registradas por um sistema e a passagem desses identificadores entre módulos (observadas através dos registros de operação nos arquivos de *logs*) permitia identicar quando dois eventos são relacionados. Com base na identificação desses marcadores, um algoritmo de *join* temporal era utilizado para identificar os padrões de ações correlacionadas.

Um desafio no caso do SMTD é que não é possível alterar os protocolos usados pelos dispositivos domésticos a fim de instrumentá-los, como no caso do Magpie. Entretanto, nossas observações têm mostrado que, no caso particular das redes domésticas, os protocolos observados podem já conter os elementos de marcação necessários para o acompanhamento de relações de causa e efeito entre pacotes nos fluxos de rede. Um fator favorável nesse caso é que, como o SMTD opera dentro da rede do usuário doméstico, sob seu controle e sem enviar qualquer informação para terceiros fora da rede, podemos contar com a inspeção do conteúdo dos fluxos sem violação da privacidade da rede doméstica, o que facilita a identificação de marcadores adequados.

A Figura 1 mostra como exemplo o comportamento legítimo de uma aplicação

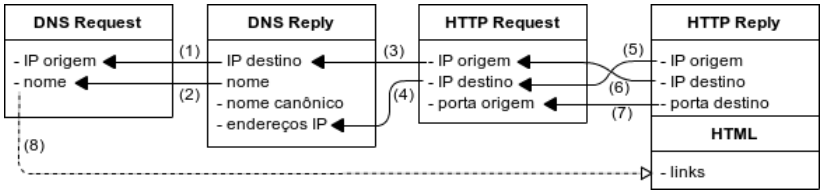

Figura 1. Modelo de uso legítimo de um fluxo HTTP.

HTTP. Um fluxo HTTP normal é iniciado por uma requisição DNS pelo endereço IP para um servidor HTTP identificado por um nome previamente obtido, seguida de uma resposta DNS. O pacote DNS Reply se relaciona ao DNS Request pelos atributos de IP de destino e IP de origem, respectivamente, e também pelo nome do domínio requisitado. Ambas as relações estão ilustradas pelas ligações (1) e (2) na Figura 1.

Um pacote HTTP Request se relaciona àquele DNS Reply pelo atributo de IP de destino da conexão HTTP e pelo IP contido em um dos endereços IP de resposta do DNS Reply. O IP de origem da conexão HTTP deve poder ser relacionado ao originador da consulta DNS anterior — itens (3) e (4) na figura. Um fluxo HTTP legítimo pode começar também por um pacote HTTP Request onde o endereço do domínio provavelmente já tenha sido resolvido e esteja armazenado em cache; nesse caso, entretanto, um DNS Reply já teria sido visto anteriormente pelo SMTD. Pacotes do fluxo da resposta HTTP também podem ser correlacionados ao mesmo padrão pelos endereços e portas de origem e destino dos fluxos de requisição e resposta — elementos (5–7) da figura.

Em certos casos, um DNS Request pode estar associado a uma URL no conteúdo HTML de um objeto obtido anteriormente, como ilustrado pela ligação (8) na Figura 1. Nesse caso, o DNS Request e os HTTP Requests posteriores serão agregados ao padrão de acesso já existente. Isto é, o nome de domínio do DNS Request e a URL do conteúdo HTML de um HTTP Reply são atributos que o SMTD considera estarem relacionados, se ocorrem um depois do outro, dentro de um intervalo de tempo limitado.

Uma vez identificados no tráfego padrões significativos, pretendemos usar uma técnica de agrupamento comportamental (*behavioral clustering*) semelhante à usada pelo Magpie para identificar padrões semelhantes — ou anômalos.

3. Implementação do Sistema

A Figura 2 apresenta o diagrama de componentes do SMTD, incluindo o controlador SDN utilizado, POX[1], e os módulos específicos do sistema. O controlador SDN recebe o primeiro pacote de cada novo fluxo e o entrega ao identificador de pacotes, que identifica os protocolos envolvidos e realiza um primeiro filtro para identificar o que fazer com o pacote. Alguns pacotes podem receber tratamento simples nesse ponto, como serem descartados por alguma regra de firewall. Aqueles pacotes que são identificados como de interesse prosseguem para o analisador de eventos, que contém o parser completo dos protocolos considerados relevantes.

O Analisador de Eventos de Rede é responsável por identificar sequências de pacotes que representem eventos conhecidos. Alguns dos principais eventos já implementados são os que compõem o padrão de uso legítimo de um fluxo HTTP: consultas DNS,

[1] www.noxrepo.org/pox/about-pox

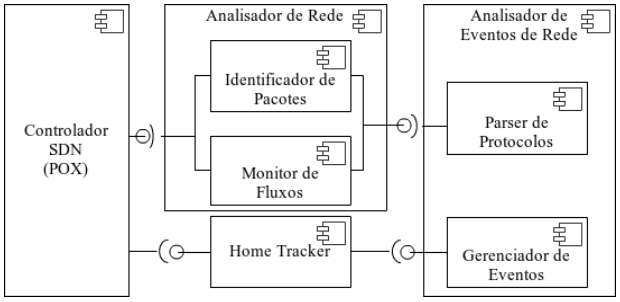

Figura 2. Diagrama do Sistema de Monitoramento do Tráfego Doméstico.

requisições e respostas HTTP e o processamento de um documento HTML. Esse módulo identifica marcadores relevantes que possam indicar a relação do novo fluxo com outros fluxos existentes e gera eventos para o *home tracker*. Esse módulo é responsável pela identificação de padrões de uso mais complexos, através da associação de identificadores entre fluxos diferentes. Em função dos padrões observados, esse módulo gera os comandos de programação dos elementos de rede através do POX, incluindo a inserção de regras para o espelhamento dos fluxos de interesse.

Tráfegos HTTPs que se enquadrarem adequadamente ao modelo de uso legítimo de um fluxo HTTP, serão considerados fluxos normais de HTTP. A partir desse momento, serão espelhados e acompanhados pelo monitor de fluxos para extrair do conteúdo recebido URLs que podem causar o acesso a outras páginas. Dessa forma, consultas DNS e requisições HTTP futuras podem ser relacionadas a uma sessão e usuário. Caso uma determinada sequência de pacotes HTTPs não se enquadre nas sessões existentes, esta será considerada uma sequência com potencial para ser maliciosa.

4. Avaliação

Nossos primeiros experimentos foram voltados para mostrar a viabilidade da identificação dos padrões de tráfego complexos descritos por diferentes relações entre parâmetros de protocolos, tanto do ponto de vista semântico quanto do ponto de vista de desempenho. Para análise dos resultados foi utilizado um roteador Netgear WNDR3700v2 configurado com a distribuição Linux OpenWRT[2]. Esse dispositivo é dotado de uma CPU Atheros AR7161 680 MHz com 64 MB de RAM, sendo considerado de desempenho intermediário no mercado atual. Todas as ferramentas necessárias foram compiladas e instaladas no roteador doméstico, para considerar o dispositivo que seria enfim responsável pelo processamento em uma instalação completa.

O experimento consistiu em capturar, usando tcpdump, dois *traces* de tráfego que incluiam sessões de navegação na Web. No primeiro, 16% dos pacotes eram devidos a conteúdo HTML, enquanto no segundo esse volume passou para 65%. Essa variação tinha por objetivo avaliar o impacto do monitor de fluxos no processo de remontagem do conteúdo das sessões (recuperação dos objetos HTML e extração de URLs dos mesmos). Os arquivos de captura foram então processados repetidamente pelo Sistema de Monitoramento do Tráfego Doméstico para medir e poder estimar a capacidade de processamento do sistema. Para um intervalo de confiança de 99%, todos os valores indicados apresentam erro inferior a 1% da média apresentada. A tabela 1 apresenta a caracterização da

[2]https://openwrt.org/

carga e os resultados de desempenho observados. Os resultados confirmam o custo extra devido ao processamento do conteúdo HTML, entretanto esse custo não é diretamente proporcional ao volume desse tipo de conteúdo: ao retirarmos o módulo de monitoração de HTML do processamento, a diferença de desempenho é menos significativa que o impacto do número de requisições HTTP do segundo caso. Isso sugere que o custo por requisição é mais significativo, algo que devemos ainda investigar.

Tabela 1. Caracterização da carga e taxas de processamento de pacotes.

Carga	Pacotes	DNS (Req/Rep)	HTTP (Req/Rep)	Proc. completo		Proc. sem HTML	
				pcts/s	t/pct (us)	pcts/s	t/pct (us)
16% HTML	3.786	27 / 25	266 / 335	11.518	86,8	14.929	67,0
65% HTML	3.243	38 / 51	1.195 / 1.615	6.561	152,4	9.298	107,6

A menor taxa de processamento observada é suficiente para processar tráfego até uma taxa de 60 Mbps, mais que satisfatória para processar o tráfego do canal de acesso banda larga atualmente disponível para usuários domiciliares, mas abaixo do desejável para analisar o tráfego interno à rede domiciliar. Trabalhar sobre o código e a arquitetura para melhorar esse desempenho é um dos objetivos atuais.

A saída do SMTD ao rotular cada fluxo com base nas regras definidas (não apresentada por limitações de espaço) confirmou que as operações de diversas sessões HTTP concorrentes (mas não correlacionadas) foram corretamente identificadas e agrupadas em função das sessões dos usuários que as geraram. Essas sessões incluíram acessos a sites de conteúdo dinâmico, com material delegado a terceiros (p.ex., anúncios e certificados) e não houve erros na identificação.

5. Trabalhos relacionados

Os problemas de redes domésticas e sua importância no contexto atual têm sido apontados por diversos pesquisadores. Os grupos de Calvert e Grinter têm trabalhado separadamente e em conjunto para identificar as demandas desse tipo de sistema do ponto de vista do usuário [Calvert et al. 2007, Grinter et al. 2009].

Alguns têm aplicado o conceito de Redes Definidas por Software no contexto de redes domésticas, entretanto com uma visão mais voltada apenas para o enlace de saída dessas redes e sua interação com o provedor banda larga. Esse enfoque oferece oportunidades interessantes para diagnóstico de problemas no link externo, mas coloca questões complexas do ponto de vista da privacidade do usuário, ao depender do envio de informações para o provedor de acesso [Feamster 2010, Calvert et al. 2011].

Do ponto de vista operacional, nosso trabalho tem semelhanças com o trabalho de Aggarwal et al., que oferece uma solução para diagnósticos de redes domésticas, com foco na informação de configuração [Aggarwal et al. 2009]. O princípio de análise e reconstrução de fluxos é semelhante ao ambiente `POX@home` [Mehdi et al. 2011], que pode até vir a ser usado futuramente. Esse sistema, entretanto, não tem os recursos de identificação e correlação de recursos do SMTD.

6. Conclusão

O crescimento das redes domésticas cria novos desafios na área, já que um grande número de usuários leigos é colocado frente a tarefas de configuração e gerência complexas. Uma

forma de resolver esse problema é a aplicação do paradigma de Redes Definidas por Software, utilizando a visão de rede oferecida por ele como ponto de partida para a integração de elementos de controle autônomo. Este trabalho apresentou o Sistema de Monitoração de Tráfego Doméstico, SMTD, que utiliza esse enfoque.

SMTD usa um analisador de protocolos para identificar elementos em cada fluxo observado na rede doméstica para tentar relacioná-lo com outros eventos observados na rede. Dessa forma, eventos podem ser interrelacionados em padrões mais complexos que podem indicar eventos como o surgimento de novos dispositivos, falhas de configuração ou tentativas de ataques, externos ou internos. Nosso protótipo identifica os principais elementos de implementação do sistema e nos permite realizar os primeiros testes operacionais, confirmando a viabilidade do sistema, apesar de ainda haver limitações de desempenho a serem tratadas.

Agradecimentos

Este trabalho foi parcialmente financiado pelo UOL (www.uol.com.br), através do programa UOL Bolsa Pesquisa, Fapemig, CNPq e Instituto Nacional de Ciência e Tecnologia da Web, InWeb (MCT/CNPq 573871/2008-6).

Referências

Aggarwal, B., Bhagwan, R., Das, T., Eswaran, S., Padmanabhan, V. N., and Voelker, G. M. (2009). Netprints: diagnosing home network misconfigurations using shared knowledge. In *Proceedings of the 6th USENIX symposium on Networked systems design and implementation*, pages 349–364, Berkeley, CA, USA. USENIX Association.

Barham, P., Donnelly, A., Isaacs, R., and Mortier, R. (2004). Using magpie for request extraction and workload modelling. In *Proceedings of the 6th conference on Symposium on Opearting Systems Design & Implementation*, volume 6, pages 18–18.

Calvert, K., Edwards, W., and Grinter, R. (2007). Moving toward the middle: the case against the end-to-end argument in home networking. In *Proceedings of the 6th ACM Conference on Hot Topics in Networks (HotNets-VI)*, pages 1–6, Atlanta, GA. ACM.

Calvert, K. L., Edwards, W. K., Feamster, N., Grinter, R. E., Deng, Y., and Zhou, X. (2011). Instrumenting home networks. *SIGCOMM Comput. Commun. Rev.*, 41(1).

Casado, M., Koponen, T., Ramanathan, R., and Shenker, S. (2010). Virtualizing the network forwarding plane. In *Proceedings of the Workshop on Programmable Routers for Extensible Services of Tomorrow*, PRESTO '10, pages 8:1–8:6, New York. ACM.

Dixon, C. et al. (2010). The home needs an operating system (and an app store). In *Proceedings of the Ninth ACM SIGCOMM Workshop on Hot Topics in Networks*, Hotnets '10, pages 18:1–18:6, New York, NY, USA. ACM.

Feamster, N. (2010). Outsourcing home network security. In *Proceedings of the 2010 ACM SIGCOMM workshop on Home networks*, HomeNets '10, pages 37–42, New York, NY, USA. ACM.

Grinter, R. E. et al. (2009). The ins and outs of home networking: The case for useful and usable domestic networking. *ACM Trans. Comput.-Hum. Interact.*, 16:8:1–8:28.

Mehdi, S. A., Khalid, J., McCauley, M., and Shenker, S. (2011). Pox-at-home. Poster apresentado no encontro Clean Slate/ONRC CTO Summit.