A Multi-Level Framework to Identify HTTPS Services

WAZEN M. SHBAIR , THIBAULT CHOLEZ , JEROME FRANCOIS , ISABELLE CHRISMENT

UNIVERSITY OF LORRAINE, LORIA

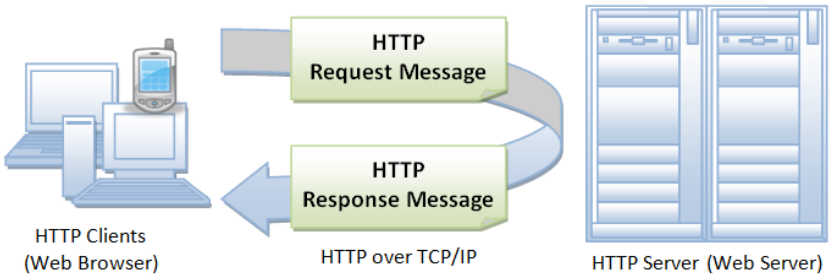
NOMS 2016

Roteiro

- Introdução
- Trabalhos Relacionados
- Framework
- Características e Metodologia
- Avaliação
- Conclusão
- Análise Crítica

Introdução

HTTP + TLS = HTTPS

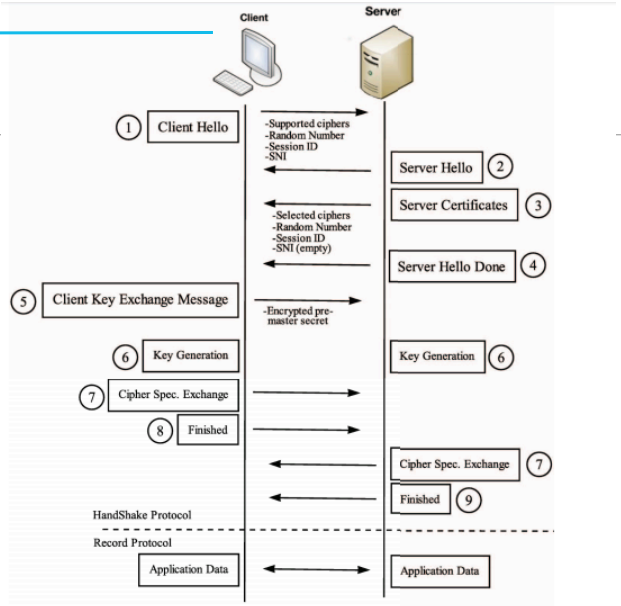


Introdução

HTTP + TLS = HTTPS

Extensão SNI

SEGURANÇA



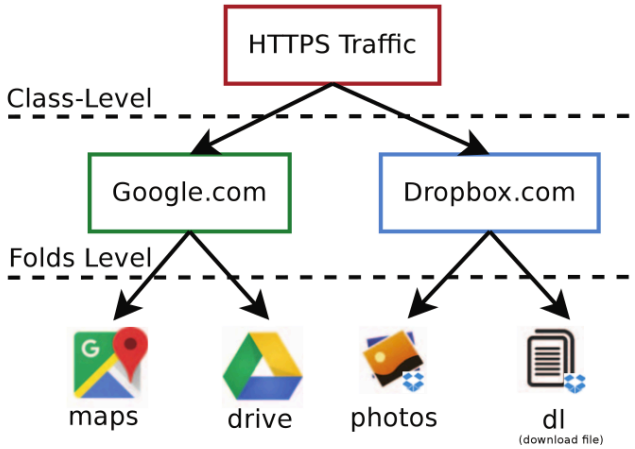
Introdução

HTTP + TLS = HTTPS

- Crescimento de websites usando HTTPS nos últimos anos - 50% do tráfego na Internet em 2015 contra 5% em 2012 (French ISPs)
- Funcionalidades oferecidas na nuvem acessadas pelo HTTPS em browsers e aplicações móveis a todo tempo

Proposta

Um framework completo para identificar os serviços HTTPS acessados em um “traffic dump”.



Trabalhos Relacionados

Identificação por tipo de Aplicação

L. Bernaille and R. Teixeira, 2007
Z. Cao, S. Cao, G. Xiong, and L. Guo, 2013
R. Alshammari *et al , 2009*
D. Schatzmann, W. M¨uhlbauer, T. Spyropoulos, and X. Dimitropoulos, 2010
T. T. Nguyen and G. Armitage, 2008
C. McCarthy *et al, 2011*
P. Velan, M. Cermak, P. Celeda, and M. Drasar, 2015

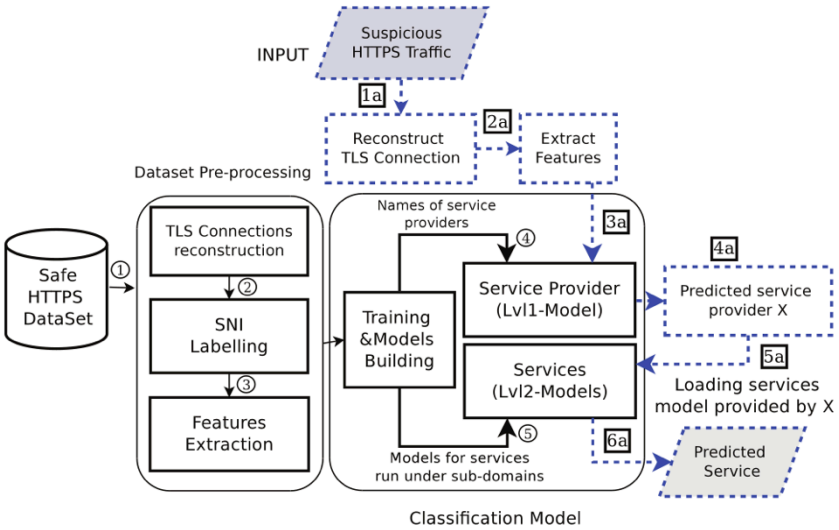
Website Fingerprinting (WF)

M. Liberatore and B. N. Levine, 2006
D. Herrmann, R. Wendolsky, and H. Federrath, 2009
A. Panchenko, L. Niessen, A. Zinnen, and T. Engel, 2011
B. Miller, L. Huang, A. D. Joseph, and J. D. Tygar, 2014
H. Cheng and R. Avnur, 1998

Identificação por nível de serviço

W. M. Shbair, T. Cholez, A. Goichot, and I. Chrisment, 2015
S.-M. Kim, Y.-H. Goo, M.-S. Kim, S.-G. Choi, and M.-J. Choi, 2015

Framework



Características

- 42 caracteristicas propostas (30 clássicas + 12 novas)

Client ↔ Server
Total number of packets, Packet size (Average, 25th,50th,75th percentile, Variance, Maximum), Inter Arrival Time (25th,50th,75th percentile)
Client → Server
Total number of packets, Packet size (Average, 25th,50th,75th percentile, Variance, Maximum), Inter Arrival Time (25th,50th,75th percentile)
Server → Client
Total number of packets, Packet size (Average, 25th,50th,75th percentile, Variance, Maximum), Inter Arrival Time (25th,50th,75th percentile)

Feature name	Directions
Average size	Client→Server, Server→Client
25th percentile size	Client→Server, Server→Client
50th percentile size	Client→Server, Server→Client
75th percentile size	Client→Server, Server→Client
Variance of size	Client→Server, Server→Client
Maximum size	Client→Server, Server→Client

- Aprendizado de máquina: C4.5 e RandomForest

Dataset

- Ambiente controlado, e seguro (sem SNI forjados)
- Pré-processamento de valores SNI

maps.google.com	mt0.google.com	mt1.google.com
khm.google.com	khm0.google.com	khm1.google.com
khmdb0.google.com	khmdb1.google.com	maps.gstatic.com

Avaliação

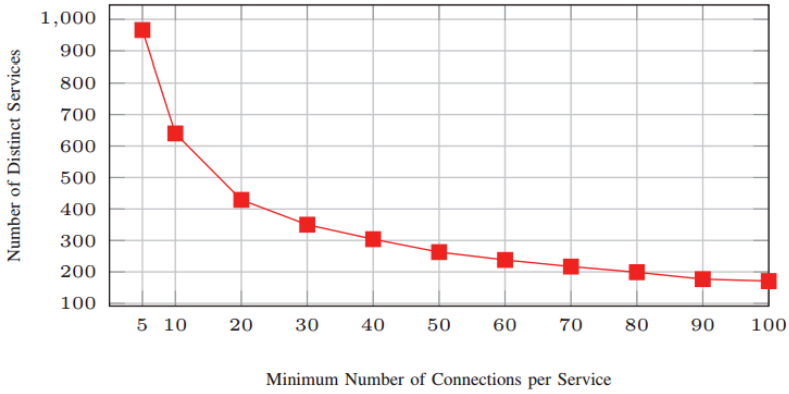
- K-fold
- Características de Operação de Receptor (ROC)
- F-Measure

$$F - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

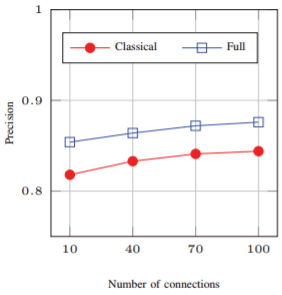
Avaliação

Visão geral

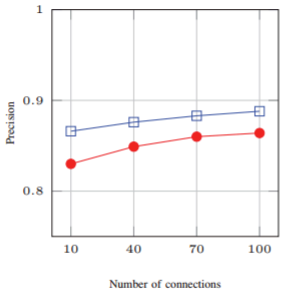
- 288.901 conexões HTTPS
- Conexões mínimas por serviço: 10, 40 e 100



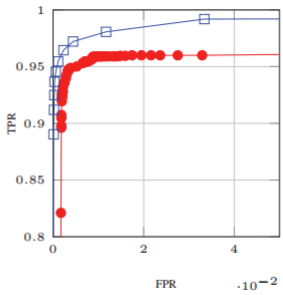
Avaliação Características



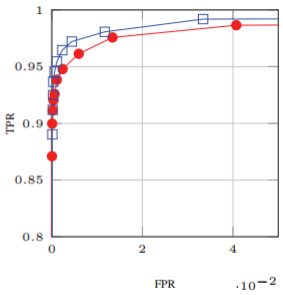
(a) Precision comparison between Classical and Full features with C4.5



(b) Precision comparison between Classical and Full features with RandomForest



(c) ROC analysis for classic and full features with C4.5 Classifier

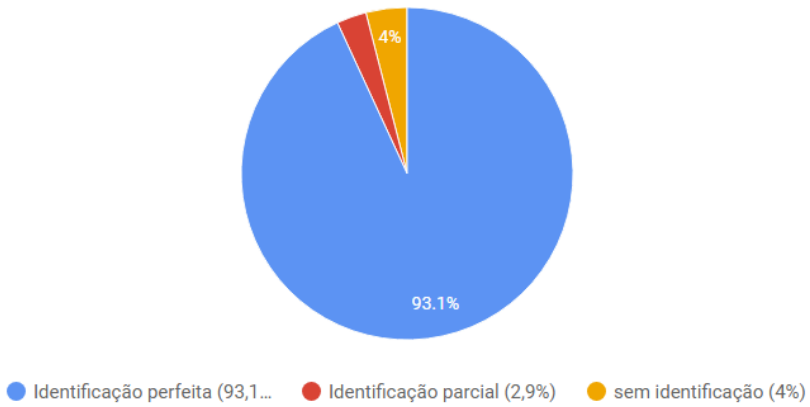


(d) ROC analysis for classic and full features with RandomForest Classifier

- 90,95% de identificação perfeita de serviços HTTPS e 4,5% de identificação parcial
- Abordagem de classificação de vários níveis foi adicionada à estrutura de identificação HTTPS.

Avaliação Framework

Avaliação com o RandomForest com 100 conexões mínimas por serviço.
A estrutura (Níveis 1 e 2) foi avaliada por validação cruzada de 10-Fold.



Conclusão

Os métodos como SNI para rotulação, novo conjunto de recursos extraídos da carga útil criptografada, a classificação multi-level obteve uma excelente taxa de identificação.
O projeto obteve sucesso e será adaptado para identificação em tempo real;

Análise Crítica

Pontos positivos:

- Identificação dos serviços sem descriptografia
- Identifica o provedor e o serviço

Ponto negativo:

- Testes apenas em ambiente controlado (sem SNI falso, e nomes SNI alterados em casos de solicitações simultâneas)

Dúvidas?