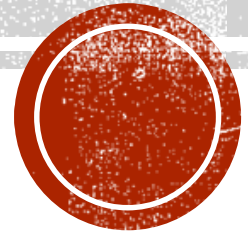
Universidade Tecnológica Federal do Paraná
Departamento Acadêmico de Informática – DAINF
Curso: Engenharia de Computação

Disciplina: Segurança Computacional

Artigo: Determinando o Risco de Fingerprinting em Página Web



AUTORES: ADRIANA R. SARAIVA, ADRIA M. DE OLIVEIRA, EDUARDO L. FEITOSA

DISCENTE: MAICON PAULO ASSMANN

DOCENTE: BRUNO CÉSAR RIBAS

Pato Branco
2017

Website Fingerprinting

- O que é *Fingerprinting*?
  - Impressão Digital.
- Mas, em termos computacionais?
  - Segundo [Cooper et al. 2013], o termo *fingerprint* "um conjunto de elementos de informação que caracteriza um dispositivo ou uma instância de uma aplicação" e *fingerprinting* como "o processo pelo qual um observador ou atacante identifica, de maneira única e com alta probabilidade, de um dispositivo ou instância de um aplicativo com base em um conjunto de múltiplas informações".
- Algumas destas informações podem ser:
  - Tamanho da tela do dispositivo;
  - Versão do sistema operacional;
  - Número de processador.

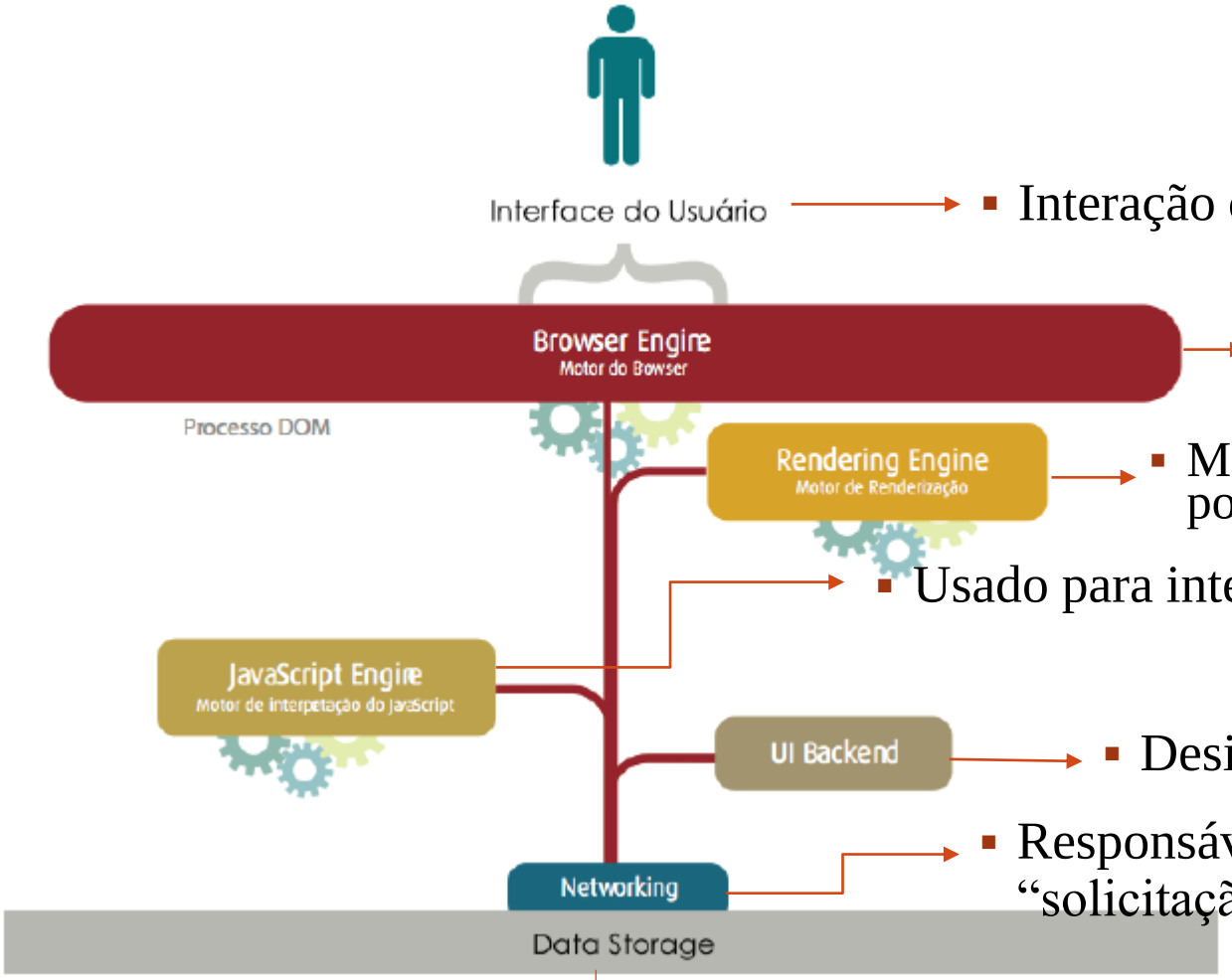
Classificações de técnicas de *Website Fingerprinting*

- Premissa básica, é de que todos deixamos rastros de no ato de navegar na internet como:
  - IP, *Cookies* e entre outros.
- Podem ser classificados em:
  - **Passiva:** É baseado nas características observáveis no conteúdo de solicitações Web.
    - Cabeçalhos HTTP;
    - Endereço IP.
  - **Ativa:** Levam em consideração técnicas onde o site é executado via JavaScript ou por outro código, no lado do cliente, para observar características adicionais sobre o navegador.
    - Tamanho da janela do navegador;
    - Enumerar fonte ou *plug-ins*.
  - ***Cookie-like*:** Os dispositivos também podem ser pré-identificados por um site que primeiro configura e depois recupera o estado armazenado pelo *user-agent* do navegador ou dispositivo.

Relação de *Website Fingerprinting*

- Serve para:
  - Medida de segurança;
  - Antifraude.
- O que isto acaba ocasionando?
  - Um impacto na privacidade do usuário.
- **Identificação do usuário:** Um *fingerprinting* tem o potencial de obter dados do usuário sem autorização prévia que o deixa exposto e vulnerável.
- **Correlacionar as atividades da navegação:** Os usuários podem se surpreender ao perceber que terceiros podem correlacionar suas várias visitas ao mesmo ou diferentes sites com a finalidade de elaborar um perfil.
- **Inferências sobre o usuário:** As informações coletadas podem revelar dados sobre os quais se pode tirar conclusões sobre o usuário.

Navegadores



- Interação com usuário
- Motor de navegação
- Motor de Renderização: Responsável por exibir o conteúdo na tela.
- Usado para interpretar código JavaScript.
- Design básico do navegador.
- Responsável por fazer a chamada de rede "solicitação HTTP"
- Responsável por onde o navegador vai guardar os dados locais.

Figura 1: Arquitetura simplificada de um navegador.

Usuário no navegador

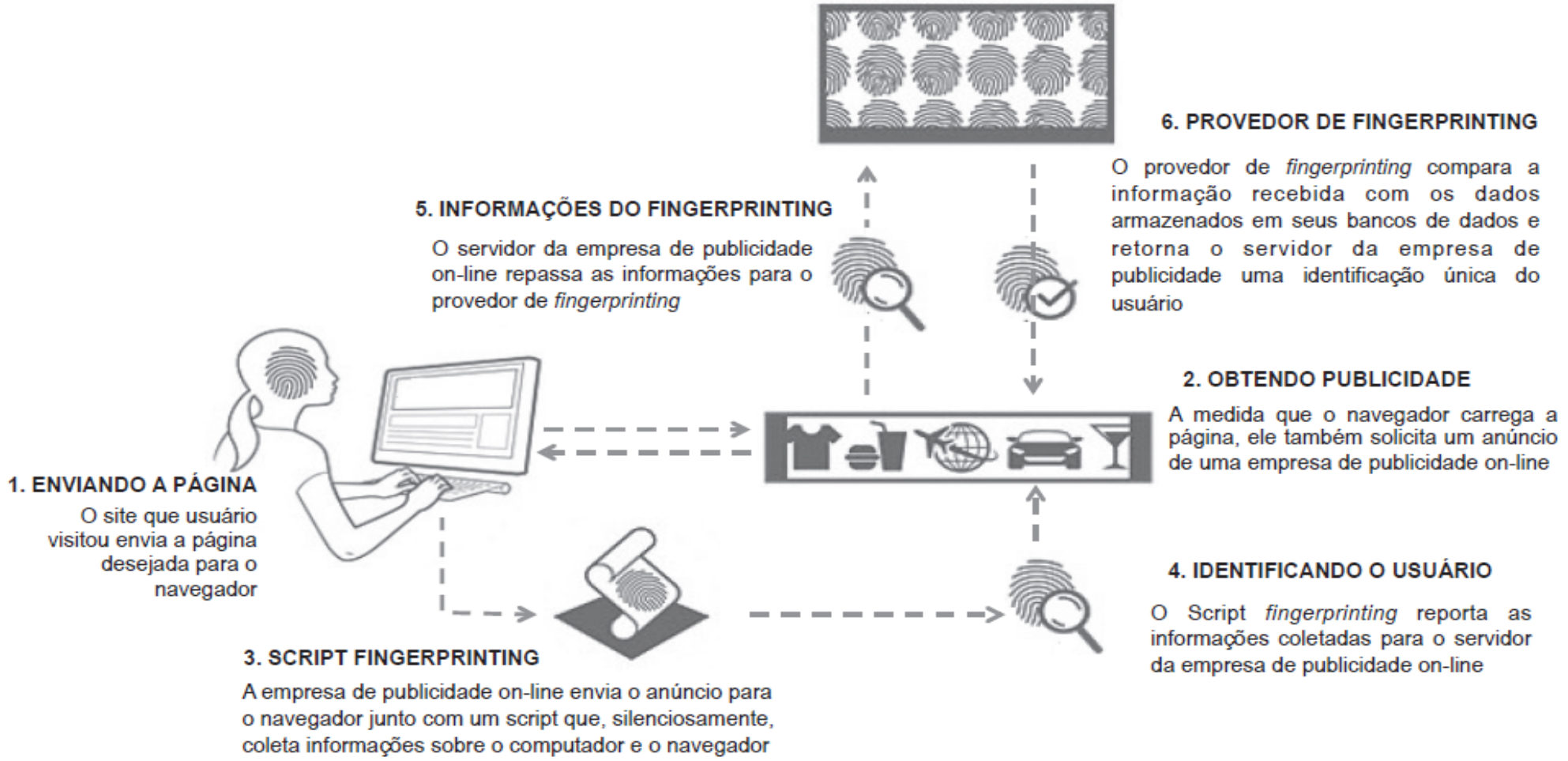


Figura 2: Exemplo do processo de fingerprinting.

Tecnologias que podem ser usadas para *Website Fingerprinting*

- HTML5 Canvas

- Tela do navegador;
- Dados pessoais do usuário.

- JavaScript

- Extrair informações do navegador;
- E dados do usuário.

- Abobe Flash

- Informações do navegador;
- Tipo de sistema operacional.

- WebGL

- Identificar o navegador;
- Realizar ataques de negação de serviço.

- Silverlight

- Versão do sistema operacional;
- Número de processadores;
- Fuso horário;
- Fontes instaladas;
- Sistema;
- Idioma do sistema operacional

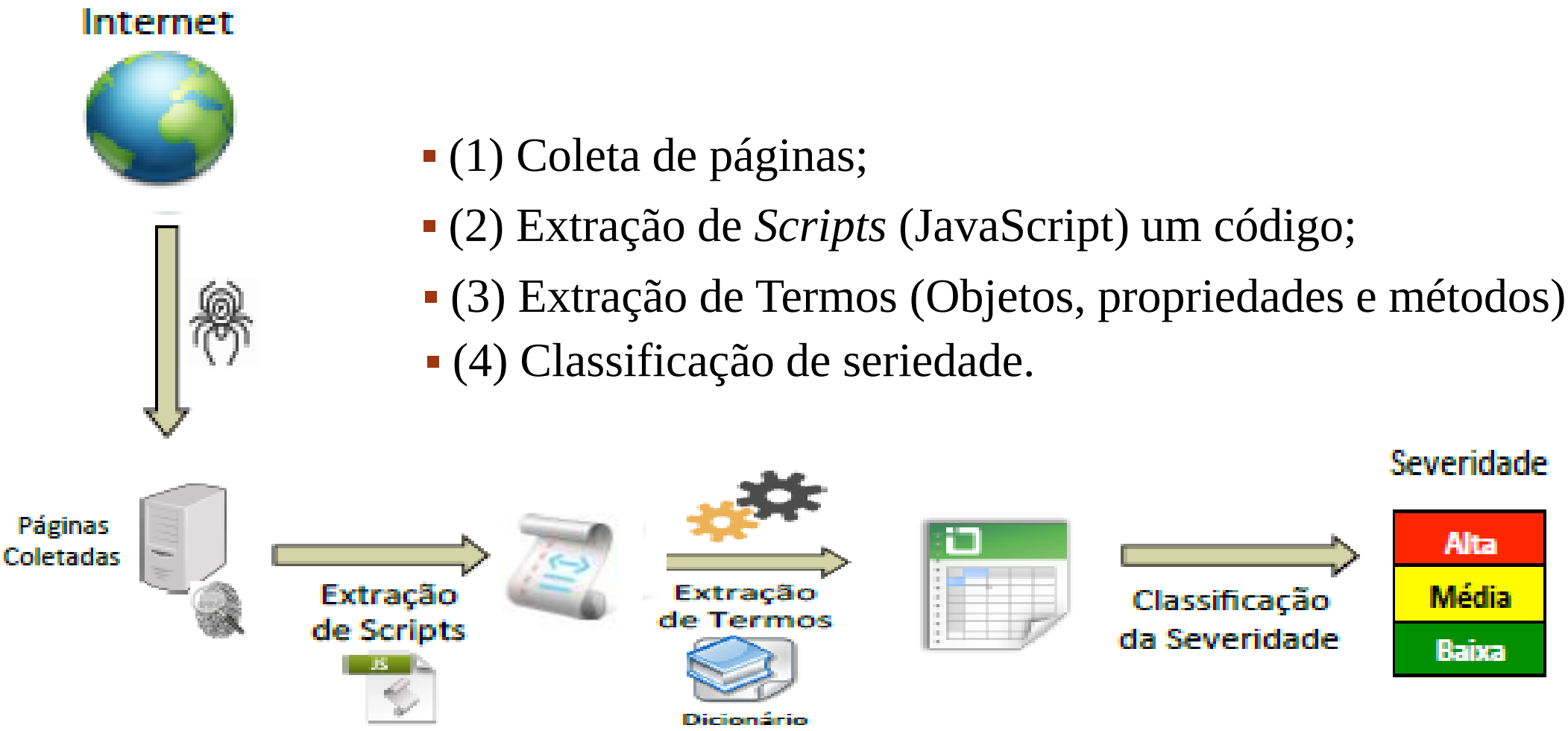
Trabalho que foi realizado no Artigo

- Duas máquinas/computadores:
  - Linux: Ubuntu 14.04
    - Intel Core i7, com 8 GB de memória RAM e 500 GB de disco.
  - Microsoft: Windows 8
    - Intel Core i5, com 8 GB de memória RAM e 1 TB de disco.
- Princípio do trabalho:
  - Realizar a identificação de páginas webs que contêm *Fingerprinting* e classificá-los de acordo com expressões regulares.
  - Sendo que retorno da expressão indica em qual tipo o atributo, objeto, propriedade e método se encaixa.

Expressão Regular	Níveis
$("N2", r'([nw]*n.[nw]*)')$	Essa expressão regular busca por objetos que utilizam em sua sintaxe de chamada até dois níveis de acesso a propriedade. Exemplo: *navigator.userAgent*
$("N3", r'([nw]*n.[nw]*n.[nw]*)')$	Essa expressão regular busca por objetos que utilizam em sua sintaxe de chamada até três níveis de acesso a propriedade. Exemplo: *window.navigator.plugin*
$("N4", r'([nw]*n.[nw]*n.[nw]*n.[nw]*)')$	Essa expressão regular busca por objetos que utilizam em sua sintaxe de chamada até quatro níveis de acesso a propriedade. Exemplo: *window.navigator.plugin.name*

Tabela 1: Expressões Regulares para Extração de Termos

Metodologia



- (1) Coleta de páginas;
- (2) Extração de *Scripts* (JavaScript) um código;
- (3) Extração de Termos (Objetos, propriedades e métodos);
- (4) Classificação de seriedade.

Classificação de severidade

- Classificação **Baixa**:
  - Quando encontram-se os elementos capazes de fornecer informações usadas apenas para adaptação de conteúdos (Tamanho da tela e a versão do navegador).
    - Ex: HTML Window
- Classificação **Média**:
  - Quando encontram-se os artefatos que listam os *plug-ins* instalados, os *mine-types* suportados e verificam a presença dos *plug-ins* SilverLight e WebGL.
    - Ex: Plungins Bancários
- Classificação **Alta**:
  - Quando encontram-se elementos como: Histórico, o que o usuário usa, tentativa de identificação do usuário, coleta de imagens do navegador, idioma do sistema operacional e tipo do sistema operacional.

Classificação de severidade

Níveis	Objetos
1	window.innerHeight; window.outerWidth; window.outerHeight; window.devicePixelRatio; document.domain; navigator.userAgent; navigator.appCodeName; navigator.appName; navigator.appVersion; navigator.appMinorVersion; navigator.buildID; navigator.browserLanguage; navigator.cpuClass; navigator.doNotTrack; navigator.language; navigator.onLine; navigator.oscpu; navigator.platform; navigator.userLanguage; navigator.product; navigator.productSub; navigator.securityPolicy; navigator.systemLanguage; navigator.vendor; navigator.vendorSub; navigator.geolocation; navigator.savePreferences; screen.availHeight; screen.availWidth; screen.availLeft; screen.availTop; screen.height; screen.width; screen.colorDepth; screen.pixelDepth; screen.bufferDepth; screen.deviceXDPI; screen.deviceYDPI; screen.logicalXDPI; screen.logicalYDPI; screen.systemXDPI; screen.systemYDPI; screen.updateInterval
2	document.referrer; document.cookie; Modernizr; Silverlight; WebGL; navigator.cookieEnabled; navigator.javaEnabled; mimeTypes; plugins; window.localStorage
3	navigator.getUserMedia; canvas.toDataURL; canvas.getImageData; window.history

Tabela 2: Níveis de Severidade

Base de controle

- O que é Base de controle?
  - É uma base de Websites que enconta-se nos artigos de Nikiforakis [Nikiforakis et al. 2015], Khademi [Khademi 2014] e Acar [Acar et al. 2013].
  - Tendo 250 páginas:
    - conibase.com
    - tumblr.com
- Foram encontrados mais 27.000 métodos no nível 1.
- Mais 996 nos níveis 2 e 3.
  - canvas.getImageData()-4
  - navigator.getUserMedia()-97
  - canvas.toDataURL()-60
  - window.history()-835

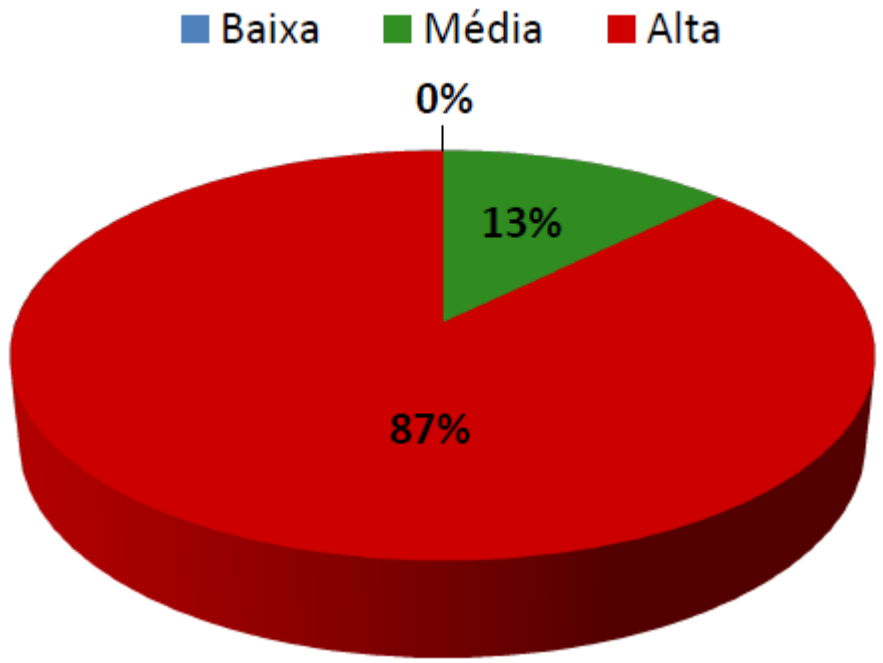


Gráfico 1: Classificação dos 250 Websites que compõem a Base de Controle.

Base de canvas

- O que é o Base canvas?
  - É uma base de Websites listados no trabalho de Englehardt [Englehardt and Narayanan 2016].

- Base do Canvas 2127 *Websites*:
- Níveis de seriedade foram menores;
- Existência de referências para outros níveis.

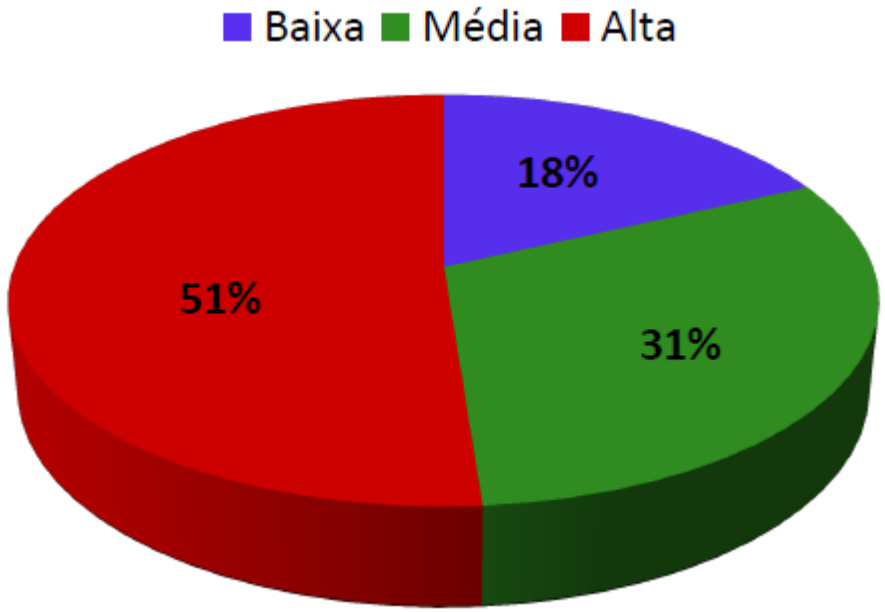


Gráfico 2: Classificação dos 2127 Websites que compõem a Base Canvas

Base DMOZ

- O que é a Base DMOZ?
  - É uma Base de Websites do diretório do [http://www.dmoz.org/].
- A Base DMOZ contém 1585 Websites.

Websites	Nível 1 - Baixa					Nível 2 - Média					Nível 3 - Alta			
	innerWidth	innerHeight	userAgent	screen.height	screen.width	referrer	cookie	mimeTypes	plugins	localStorage	getUserMedia	getImageData	toDataURL	history
robertkleingallery.com	0	0	0	2	2	0	0	0	0	0	0	0	0	0
pranapoweryoga.com	1	0	3	2	2	0	0	0	0	0	0	0	0	0
joegreenephoto.com	1	5	3	1	0	0	0	0	0	0	0	0	0	0
captel.com	0	0	0	1	1	0	0	0	0	0	0	0	0	0
dynamicpost.co.uk	5	5	1	0	0	0	0	0	0	0	0	0	0	0
olddogpaws.com	0	0	2	2	3	1	36	0	3	0	0	0	0	0
ilight-tech.com	9	4	8	1	1	1	2	4	8	0	0	0	0	0
bbc.com	10	7	11	32	32	52	118	6	9	7	0	0	0	0
gourmetsleuth.com	6	6	22	6	6	12	6	16	0	0	0	0	0	0
nordicacademy.com.au	4	4	1	0	0	0	3	3	5	0	0	0	0	0
nikonusa.com	29	30	37	10	12	11	49	14	58	8	0	0	0	19
flairstrips.com	14	14	17	0	3	0	12	13	3	0	0	6	1	2
robotshop.com	8	8	13	3	3	7	13	4	6	4	0	3	1	1
fibergarden.com	13	17	9	2	2	5	20	4	5	4	0	3	1	4
newscooters4less.com	5	3	0	0	0	0	0	4	0	0	0	1	1	9

Tabela 3: Ocorrencia de Termos encontrados nos sites da Base DMOZ

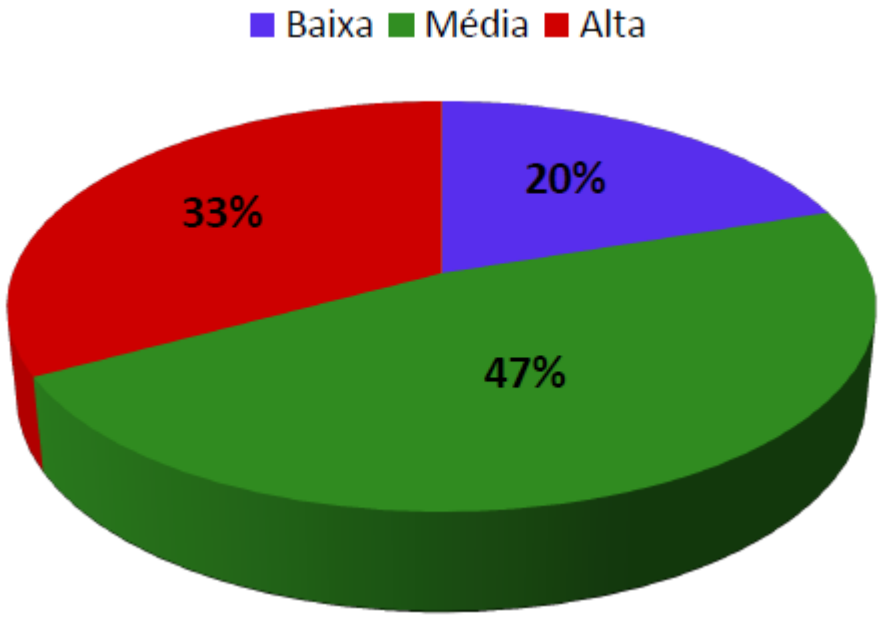


Gráfico 3: Classificação dos 1585 Websites que compõem a Base DMOZ

Possíveis soluções

- Existem soluções propostas academicamente, mas poucas são implantadas.

- Bloqueio de anúncios indesejados.
- Não ser rastreado.
- Ser um completo “fantasma”.
- Não deixa executar scripts, detectar rastreamento, plugs indesejado.

Referências

SARAIVA, A. R., DE Oliveira, A. M., FEITOSA, E. L. Determinando o Risco de *Fingerprinting* em Página Web. XVI Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais – SBSeg, 2016.

Saraiva, A. R., Elleres, P. A., Carneiro, G. B., Feitosa, E. (2014). Device *Fingerprinting*: Conceitos e técnicas, exemplos e contramedidas. No Livro de Minicursos do XIV Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais – SBSeg, 2014, Belo Horizonte, MG, Brasil. SBC.