Proposta para Grupo de Trabalho

GT-CIRD: Caracterização e Identificação Remota de Dispositivos

Programa: Pesquisa e Desenvolvimento em Internet Avançada
Tema: Serviços/produtos de rede

João Paulo de Souza Medeiros (DCT/UFRN)
14 de março de 2017

1 Título

GT-CIRD: Caracterização e Identificação Remota de Dispositivos

2 Coordenador

João Paulo de Souza Medeiros (DCT/UFRN)
- E-mail: `jpsm@dct.ufrn.br`;
- Sítio: http://labepi.ufrn.br/~jpsm;
- Currículo: http://lattes.cnpq.br/8782777013152714.

3 Programa de P&D

Internet Avançada

- **Serviços/produtos de rede**: sistema de caracterização e identificação remota de dispositivos em rede com base no monitoramento e análise de tráfego das camadas de rede, transporte e aplicação.

4 Resumo

O processo de caracterização e identificação de computadores possui várias aplicações em segurança da informação. Em análise forense em redes de computadores, por exemplo, tal processo pode ser usado em conjunto com com sistemas de detecção de intrusão a fim de caracterizar máquinas utilizadas em ataques de rede (e.g., negação de serviço). A caracterização de dispositivos remotos é baseada na análise de dados de rede gerados pela máquina de origem e a abordagem clássica é a de explorar características peculiares das diferentes implementações dos protocolos em cada camada da pilha de protocolos (i.e. enlace, rede, transporte e aplicação). Em trabalhos recentes é demostrado que o uso de inteligência computacional pode melhorar o desempenho de identificação quando comparado a métodos e ferramentas clássicas. Este projeto tem como objetivo a criação de um sistema de caracterização e classificação de assinaturas digitais para identificação de dispositivos, algo ainda não encontrado na literatura disponível.

5 Abstract

The process of remote characterization and identification of computers has many applications in information security. On network forensics analisys, this process can be used together with intrusion detection systems to characterize machines of remote attackers (e.g., denial of service). The characterization of remote computers is based on the analysis of network data originated from the remote machine and the classical approach is to exploit peculiar characteristics of different implementations of network protocols at each layer of the protocol stack (i.e., link, network, transport and application layers). Recent works show that the use of computational intelligence techniques can improve the identification performance when compared to classical classification algorithms and tools. This project aims to build a system to perform remote identification of computers and evaluate its application, something not found in the available literature.

6 Parcerias

Este projeto será executado através de uma parceria interinstitucional envolvendo a Universidade Federal do Rio Grande do Norte (UFRN), a Universidade do Estado do Rio Grande do Norte (UERN) e a Universidade Federal de Minas Gerais (UFMG). Na UFRN, estão envolvidos o Departamento de Computação e Tecnologia (DCT), localizado no interior do estado, no município de Caicó/RN, e o Departamento de Engenharia de Computação e Automação (DCA), localizado no *campus* central, no município de Natal/RN. Na UFMG, é parceiro o Departamento de Ciência da Computação (DCC). Na UERN, é parceiro o Departamento de Informática (DI). Os participantes de cada uma das instituições são discriminados a seguir.

- DCT/UFRN:
  - Prof. João Paulo de Souza Medeiros;
  - Prof. João Batista Borges Neto.
- DCA/UFRN:
  - Prof. Agostinho Medeiros Brito Júnior.
  - Prof. Paulo Sérgio Motta Pires;
- DCC/UFMG:
  - Prof. Antônio Alfredo Ferreira Loureiro.
- DI/UERN:
  - Prof. Rommel Wladimir Lima.

Em seguida são apresentados os laboratórios de pesquisa onde serão desempenhadas as atividades relacionadas ao projeto e seus respectivos coordenadores.

- Laboratório de Elementos do Processamento da Informação (LabEPI/UFRN)
  - Coordenador: Prof. João Paulo de Souza Medeiros.
  - Sítio: http://labepi.ufrn.br/.
- Laboratório de Segurança da Informação (LabSIN/UFRN)
  - Coordenador: Prof. Agostinho Medeiros Brito Júnior.
- Laboratory of Wireless, Innovative, Sensing Embedded Systems & Models, Algorithms and Networking Protocols (WISEMAP/UFMG)
  - Coordenador: Prof. Antônio Alfredo Ferreira Loureiro.
  - Sítio: http://wisemap.dcc.ufmg.br/.
- Laboratório de Redes e Sistemas Distribuídos (LORDI/UERN)
  - Coordenador: Prof. Rommel Wladimir de Lima.
  - Sítio: http://lordi.uern.br/.

As tarefas e atividades de cada laboratório de pesquisa e de cada pesquisador são apresentadas de forma detalhada na Seção 7.

7 Duração do projeto e marcos

A execução do projeto tem previsão de duração de 12 (doze) meses. Os quatro laboratórios envolvidos terão infraestrutura de rede para realização de experimentos em rede local e através da Internet. Nesse último caso, pretende-se utilizar segmentos da Rede Ipê que interligam Natal/RN e Belo Horizonte/MG para estimar de forma mais precisa a influência das métricas de rede (coletadas pelo MonIPÊ) nos resultados da identificação. A infraestrutura que interliga UERN em Mossoró/RN e UFRN em Natal/RN e Caicó/RN também será utilizada nos experimentos para comparar com os resultados obtidos na Rede Ipê. Considerando essa infraestrutura e as definições e a discussão sobre o estado da

arte apresentadas na Seção 8, são listadas a seguir as tarefas que compõem os objetivos específicos desta proposta.

Tarefa 0: montar e/ou configurar uma infraestrutura local de máquinas e serviços em cada um dos quatro laboratórios de pesquisa utilizando *hardware* e *software* diversificado;

Tarefa 1: quantificar a singularidade das assinaturas obtidas com o método de caracterização proposto por Kohno et al. (2005) e Lanze et al. (2012);

Tarefa 2: quantificar a singularidade da associação do método anterior com outras técnicas de identificação de sistemas e serviços discutidas em (Medeiros et al., 2014a,b);

Tarefa 3: projetar e desenvolver um protótipo capaz de criar caracterizações de máquinas remotas utilizando dados das camadas de rede, transporte e aplicação;

Tarefa 4: projetar uma base de dados de caracterizações para armazenar caracterizações e confrontar evidências;

Tarefa 5: projetar, desenvolver e avaliar um sistema de classificação eficaz para caracterização de dispositivos e serviços;

Tarefa 6: realizar integração do protótipo para algum sistema de detecção/prevenção de intrusão baseado em *software* de código aberto;

Tarefa 7: elaboração do relatório final do projeto;

Tarefa 8: elaboração do artigos científicos para periódicos da área.

Cada uma dessas tarefas estão sincronizadas com os Relatórios de Planejamento (RP), Relatórios Técnicos (RT), Relatórios de Acompanhamento (RA) e o Relatório de participação no WRNP (RWRNP). Na Figura 1, é apresentada a distribuição dessas tarefas ao longo do período de execução do projeto.

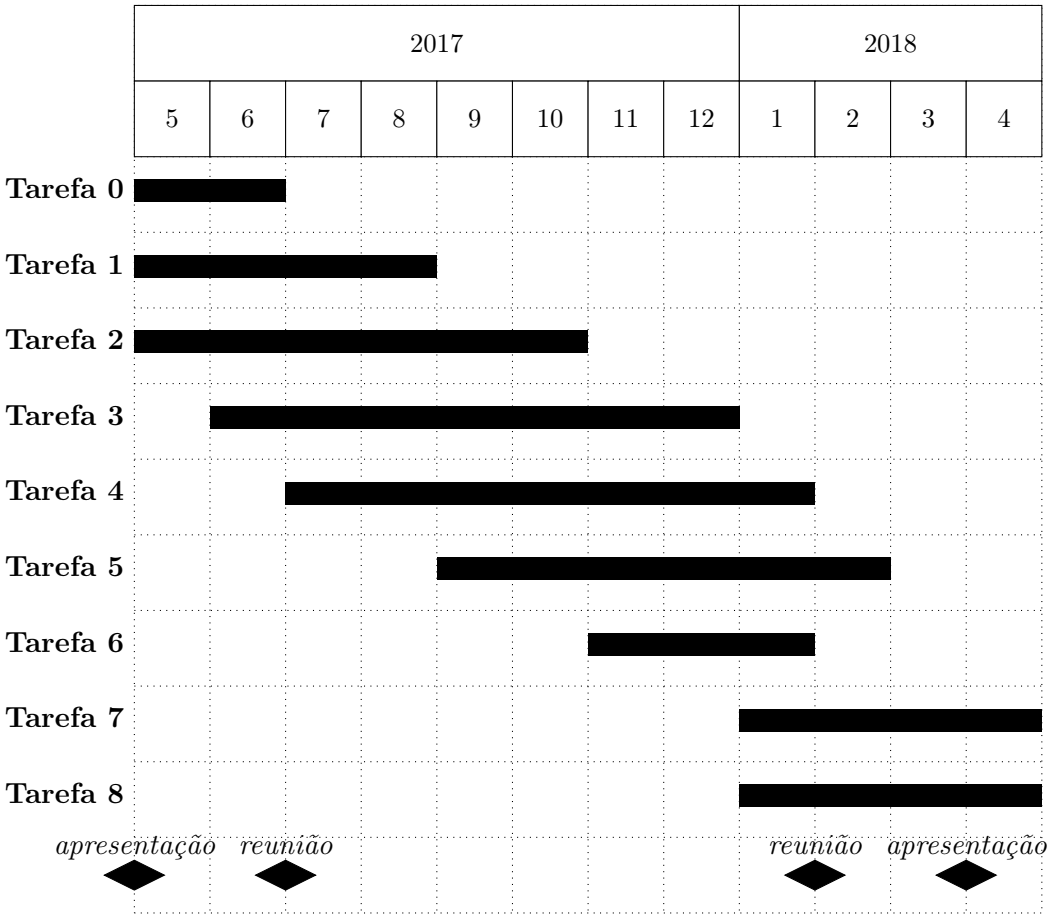

Figura 1: Cronograma de atividades.

O primeiro marco "apresentação" está associado à Apresentação em sessão técnica no

Workshop da RNP nos dias 15 e 16/05, em Belém/PA. O segundo ao Workshop de Apresentação Final para o comitê de avaliação e colaboradores da RNP. Os marcos "reunião" representam encontros previstos entre os coordenadores dos laboratórios de pesquisa que compõem o projeto. Na Tabela 1, são discriminadas as atribuições de cada laboratório de acordo com as tarefas apresentadas.

Tabela 1: Competências de cada laboratório.

Laboratório	**Tarefas**
LabEPI	0, 1, 2, 3, 5, 7, 8
LabSIN	0, 1, 2, 6, 7, 8
WISEMAP	0, 1, 2, 4, 7, 8
LORDI	0, 1, 2, 3, 4, 7, 8

As tarefas apresentadas na Figura 1 estão sob a responsabilidade de cada um dos laboratórios listados na Seção 6. As atribuições apresentadas estão de acordo com a competência e experiência profissional dos coordenadores e colaboradores de cada laboratório.

8 Sumário executivo

O processo de caracterização e identificação remota de dispositivos possui diversas aplicações nas áreas de segurança e práticas forenses em redes de computadores. Em análise forense em redes de computadores, esse processo pode ser usado juntamente com sistemas de detecção de intrusão, ou *Intrusion Detection Systems* (IDSs), a fim de caracterizar máquinas utilizadas em atividades suspeitas. O processo de caracterização tem como base a análise de dados de rede originados nas máquinas ou dispositivos remotos de interesse. A abordagem clássica é a de explorar características peculiares de diferentes implementações de protocolos de rede em diferentes camadas da pilha de protocolos. Em trabalhos recentes (e.g., Medeiros et al. (2014b)), é demostrado que o uso de inteligência computacional torna o processo de identificação mais eficaz e eficiente, quando comparado com algoritmos e ferramentas clássicas. O objetivo geral deste grupo de trabalho é criar um protótipo capaz de criar e comparar assinaturas de dispositivos remotos. Ainda visa quantificar a singularidade dessas assinaturas com base nos diferentes métodos encontrados na literatura. Um levantamento recente da literatura é apresentado por Medeiros et al. (2014a).

8.1 Fundamentos

A identificação de dispositivos que são usados para desempenhar atividades maliciosas ou ilegais é um importante passo para rastrear os indivíduos responsáveis. Essa identificação pode ser usada em investigações para criar evidências do uso de dispositivos em possíveis crimes cibernéticos. Nesse caso, o sistema de identificação deve criar uma impressão digital do dispositivo remoto usado pelo suspeito e, quando algum equipamento for apreendido, um especialista treinado poderá verificar se a impressão digital capturada durante a atividade criminosa é equivalente à do dispositivo capturado. Essa aplicação é exemplificada por Novotny et al. (2002, 2004) com o objetivo de capturar evidências digitais de atividade criminosa em páginas de bate-papo. Esse processo de identificação é denominado *Remote Computer Fingerprinting* (RCF).

Neste projeto, a aplicação de RCF em forense de redes de computadores está associado à criação automática de impressões digitais (assinaturas). Esse processo de criação de evidências pode ser iniciado a partir de um IDS que detecte alguma atividade suspeita. Esse cenário é ilustrado pela Figura 2.

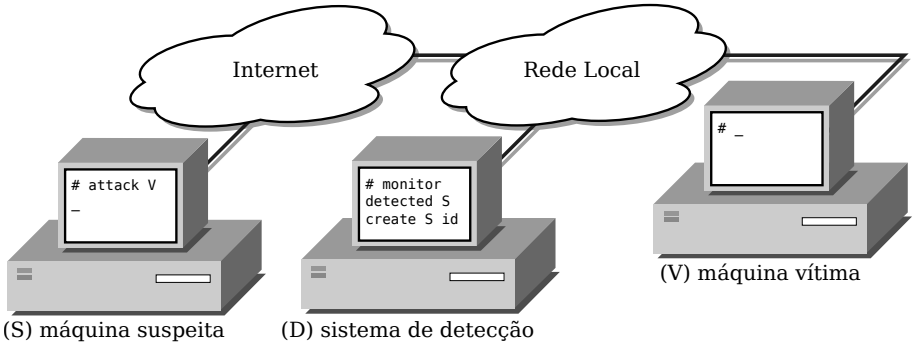

Figura 2: Representação de possível aplicação (Medeiros et al., 2014a).

Segundo Garfinkel (2010), sistemas autônomos que são capazes de detectar e apresentar *outliers*, assim como outros elementos suspeitos, são de fundamental importância para cobrir os desafios da forense digital. No cenário apresentado na Figura 2, o sistema de detecção (D) monitora a rede local com os seguintes objetivos: (i) identificar atividade suspeita a partir da análise do tráfego destinado as máquinas da rede local; e (ii) criar uma impressão digital da máquina remota do suspeito para ser usada como evidência em uma eventual investigação. Para atingir o primeiro objetivo, é possível utilizar IDSs, como por exemplo o *software* livre Snort[1]. Já para o segundo, não há solução conhecida, assim como persistem ainda desafios tecnológicos fundamentais para a concepção de tal sistema. Este projeto tem como finalidade a criação do primeiro protótipo desse tipo de sistema.

Impressões digitais de dispositivos remotos são constituídas de informações que caracterizam serviços, sistemas ou dispositivos da máquina. Dessa forma, RCF é uma solução natural para compor o sistema representado na Figura 2. O processo de caracterização e classificação de impressões digitais pode ser representado pelo diagrama conceitual apresentado na Figura 3.

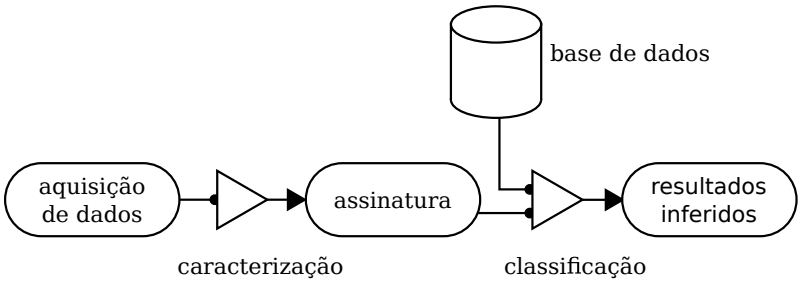

Figura 3: Representação conceitual do processo de RCF.

O grande desafio para tornar possível a identificação de impressões digitais de dispositivos em rede está em criar representações singulares por meio de um processo de caracterização eficaz. Estudos recentes (cf., Kohno et al. (2005), Lanze et al. (2012) e Polcak et al. (2014)), concluem que isso é possível em determinadas condições.

8.2 Proposta

Considerando o desafio de caracterizar de forma singular máquinas remotamente, Medeiros et al. (2014a) propõem a utilização de classificadores especializados para serviços, sistemas operacionais e dispositivos da máquina que possuem características semelhantes. Isso pode ser feito projetando um classificador para cada grupo de impressões digitais similares. Na Figura 4, esse conceito de agrupamento é exemplificado utilizando impressões

[1] https://www.snort.org/

digitais de diferentes sistemas operacionais.

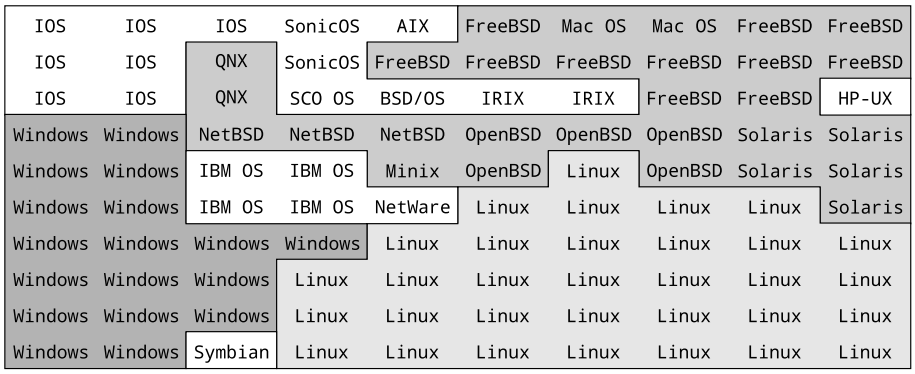

Figura 4: Agrupamento de impressões digitais de sistemas operacionais (Medeiros et al., 2010).

Com base em grupos definidos por alguma medida de similaridade, uma evidência pode ser classificada utilizando o algoritmo mais adequado. Esse processo de classificação guiada por agrupamento é ilustrado na Figura 5.

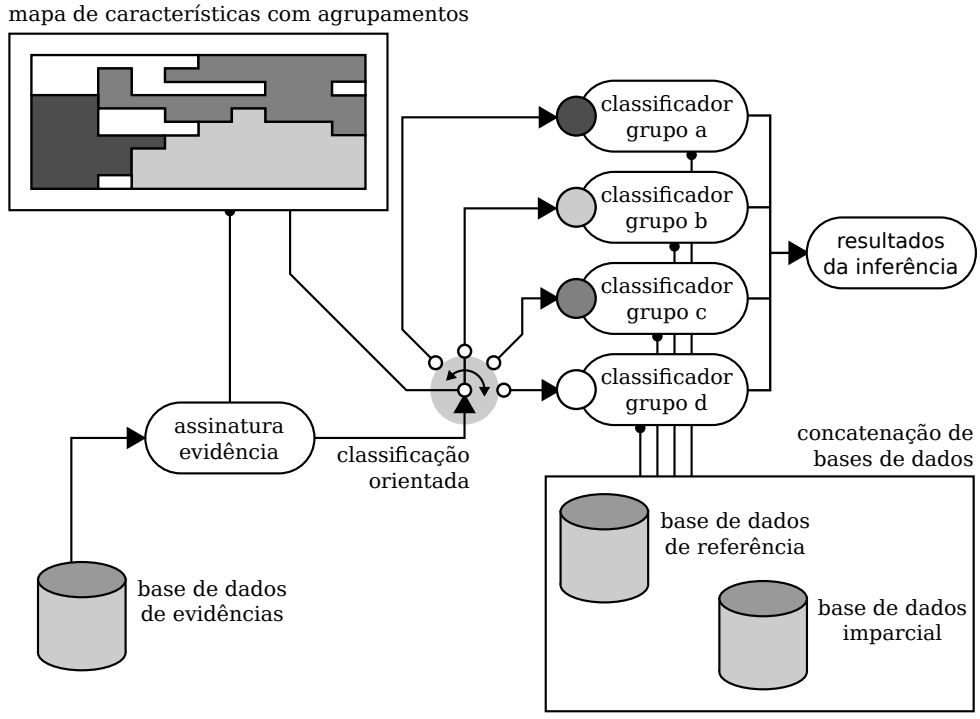

Figura 5: Sistema composto por vários classificadores especializados (Medeiros et al., 2014a).

A proposta então é a de utilizar os diferentes métodos de caracterização apontados na literatura, e criar classificadores especialistas a fim de maximizar a possibilidade de criação de impressões digitais singulares. Em adição, deve-se considerar ativos da rede que possam influenciar na validade dos dados capturados, e.g., *Network Address Translation* (NAT), *firewalls* e *protocol scrubbers* (Watson et al., 2004).

9 Recursos financeiros

Para execução desta proposta são solicitados equipamentos, gasto com pessoal, passagens e diárias. Todo *software* utilizado está disponível.

10 Ambiente para testes do protótipo

Para avaliar o protótipo desenvolvido será utilizado um ambiente similar ao apresentado na Figura 2. Porém, além de considerar a etapa de classificação de evidências, o ambiente em questão tem como extensão a criação de uma ambiente controlado para a criação da impressão digital de possíveis dispositivos apreendidos. A arquitetura desse ambiente de testes do protótipo é apresentada na Figura 6.

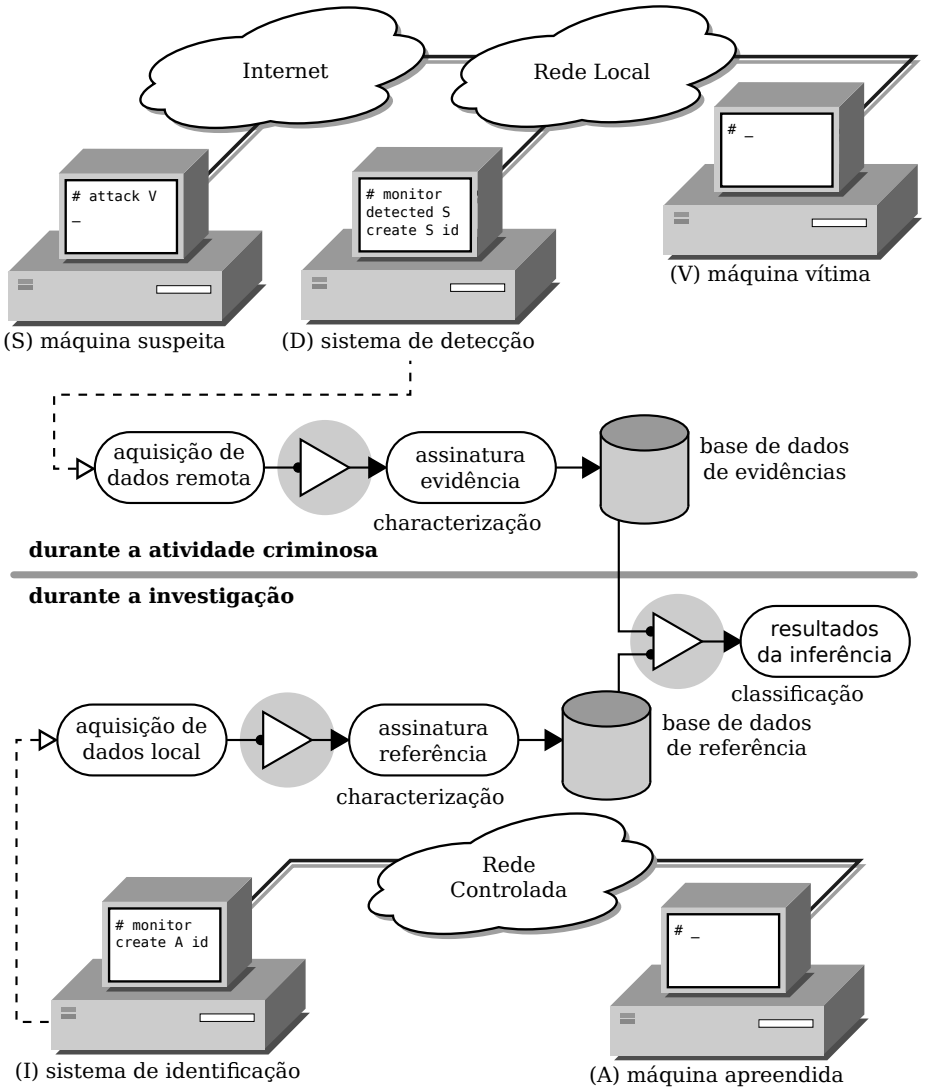

Figura 6: Ambiente de testes do protótipo (Medeiros et al., 2014a).

Arquiteturas similares à proposta na Figura 6 são encontradas na literatura da área de análise forense em redes. Por exemplo, Novotny et al. (2004) propõem a utilização de ferramentas como o Nmap[3] para auxiliar na investigação de crimes cibernéticos. De forma similar, Meehan et al. (2001) aplicam o mesmo conceito para automação do monitoramento de páginas de bate-papo. O mecanismo de caracterização e classificação de evidências deve ser avaliado quanto à confiabilidade. Esse aspecto é fundamental para que essas evidências sejam judicialmente admissíveis. Para cobrir esse aspecto, a arquitetura proposta está em acordo com os trabalhos mais recentes na literatura (Medeiros et al., 2014a).

[3] http://nmap.org/

Protótipo e serviço da RNP Demonstrada a viabilidade técnica do protótipo, planeja-se em uma segunda fase do projeto a criação de um serviço de caracterização, classificação e armazenamento de assinaturas da RNP.

Referências

T. Kohno, Andre Broido, and K. C. Claffy. Remote Physical Device Fingerprinting. *IEEE Transactions on Dependable and Secure Computing*, 2(2):93–108, February 2005. ISSN 0361-1434. doi: 10.1109/TDSC.2005.26.

F. Lanze, Andriy Panchenko, Benjamin Braatz, and Andreas Zinnen. Clock skew based remote device fingerprinting demystified. In *Proceedings of the 2012 IEEE Global Communications Conference (GLOBECOM)*, pages 813–819. IEEE, December 2012. ISBN 978-1-4673-0921-9. doi: 10.1109/GLOCOM.2012.6503213.

J. P. S. Medeiros, J. B. Borges Neto, A. M. Brito Júnior, and P. S. Motta Pires. Learning Remote Computer Fingerprinting. In Azah Kamilah Muda, Yun-Huoy Choo, Ajith Abraham, and Sargur N. Srihari, editors, *Computational Intelligence in Digital Forensics*, pages 253–283. Springer International Publishing, 2014a. ISBN 978-3-319-05884-9. doi: 10.1007/978-3-319-05885-6_12.

J. P. S. Medeiros, J. B. Borges Neto, G. S. D. Queiroz, and P. S. Motta Pires. Intelligent Remote Operating System Detection. In Biju Issac and Nauman Israr, editors, *Case Studies in Intelligent Computing*, pages 177–196. CRC Press (Auerbach Publications), September 2014b. ISBN 978-1-4822-0703-3. doi: 10.1201/b17333-10.

J. M. Novotny, A. Meehan, Dominic Schulte, Gavin W. Manes, and Sujeet Shenoi. Evidence acquisition tools for cyber sex crimes investigations. In *Proceedings of the SPIE, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Defense and Law Enforcement*, volume 4708, pages 53–60, 2002. doi: 10.1117/12.479292.

J. M. Novotny, Dominic Schulte, Gavin W. Manes, and Sujeet Shenoi. Remote computer fingerprinting for cyber crime investigations. In *Proceedings of the Seventeenth Annual Working Conference on Data and Applications Security*, Data and Applications Security XVII, pages 3–15, 2004. doi: 10.1007/1-4020-8070-0_1.

S. L. Garfinkel. Digital forensics research: The next 10 years. *Digital investigation*, 7: S64–S73, 2010. doi: 10.1016/j.diin.2010.05.009.

L. Polcak, Jakub Jirasek, and Petr Matousek. Comment on "Remote Physical Device Fingerprinting". *IEEE Transactions on Dependable and Secure Computing*, 11(5):494–496, September 2014. ISSN 1545-5971. doi: 10.1109/TDSC.2013.26.

J. P. S. Medeiros, A. M. Brito Júnior, and P. S. Motta Pires. Using intelligent techniques to extend the applicability of operating system fingerprint databases. *Journal of Information Assurance and Security*, 5(4):554–560, 2010.

D. Watson, M. Smart, G.R. Malan, and F. Jahanian. Protocol scrubbing: network security through transparent flow modification. *IEEE/ACM Transactions on Networking*, 12(2): 261–273, 2004. doi: 10.1109/TNET.2003.822645.

A. Meehan, G. Manes, L. Davis, J. Hale, and S. Shenoi. Packet sniffing for automated chat room monitoring and evidence preservation. In *Proceedings of the 2001 IEEE Workshop on Information Assurance and Security*, pages 285–288, 2001.