Análise Experimental do Desempenho de Anonimização de Dados em Dispositivos IoT

Arthur C. Urbano[1], **Matheus M. Silveira**[1], **Rafael L. Gomes**[1]

[1]Universidade Estadual do Ceará (UECE)

{arthur.cordeiro,matheus.monteiro,rafaellgom}@larces.uece.br

Abstract. *The creation of Data Privacy Laws requires the use of data protection techniques such as anonymization algorithms in applications. However, few studies in the literature experimentally analyze the implementation of these algorithms in the Internet of Things (IoT) devices, which have different computational resource capabilities. Within this context, this work presents a performance analysis of anonymization techniques implemented for IoT devices, where heterogeneous hardware characteristics were evaluated. From the testbed performed, it was possible to identify the suitability of each algorithm for the contexts of IoT application, as well as to measure the execution capacity of each one in the different hardware used.*

Resumo. *A criação de Leis de Privacidade de Dados demandou a necessidade da aplicação de técnicas de proteção de dados, dentre elas os algoritmos de anonimização. Contudo, há poucos estudos na literatura que analisam de forma experimental a implementação desses algoritmos em dispositivos de Internet das Coisas (IoT), os quais possuem diferentes capacidades de recursos computacionais. Dentro deste contexto, este trabalho apresenta uma análise de desempenho das técnicas de anonimização implementadas para os dispositivos IoT, onde avaliou-se características de hardware heterogêneas. A partir do testbed realizado, foi possível identificar a adequação de cada algoritmo para os contextos de aplicação de IoT, bem como dimensionar a capacidade de execução de cada um nos diferentes hardwares utilizados.*

1. Introdução

A Internet das Coisas (*Internet of Things* - IoT) é uma tecnologia crucial para a implantação de novos serviços para a sociedade. Por consequência da inserção dos dispositivos IoT em diversos ambientes, tem-se um grande volume de dados que circulam diariamente destes dispositivos para a Internet, onde alguns desses dados possuem um teor sigiloso, o que pode gerar situações as quais as informações críticas usadas de forma maliciosa podem vir a comprometer o valor dos serviços prestados.

Esta realidade de dados sigilosos circulando pela Internet levaram a uma preocupação do impacto da exposição desses dados pessoais para entidades não desejadas, impulsionando a criação de leis de privacidade, como a Lei Geral de Proteção de Dados (LGPD - 13709/2018) no Brasil e *General Data Protection Regulation* (GDPR - 2016/679) na Europa [de Oliveira 2019].

Dentro deste contexto, este artigo apresenta uma análise de desempenho das técnicas de anonimização em dispositivos IoT, ou seja, executando dentro destes dispositi-

vos. Os experimentos foram realizados com os seguintes dispositivos (que possuem características de hardware heterogêneas) [Datasheet 2021, Datasheet 2015, Broadcom 2012]: ESP32, ESP8266/WEMOS, Raspberry Zero e Raspberry Pi. Com relação as técnicas de anonimização, foram avaliadas [Pang 2016]: Substituição, Generalização, IP por truncamento, *Black Marker*, Deslocamento e Supressão de Atributo. Os resultados obtidos possibilitam identificar o desempenho dos dispositivos IoT para cada técnica de anonimização no que se refere a tempo de execução, consumo de memória e tempo de resposta. Desta forma, este artigo possui as seguintes contribuições: (I) Análise dos algoritmos de anonimização em relação as leis de privacidade; e, (II) Identificação da viabilidade das técnicas de anonimização nos dispositivos IoT a partir de uma experimentação real.

O restante deste artigo está organizado da seguinte forma. A Seção 2 descreve a metodologia aplicada e as configurações dos experimentos, enquanto que a Seção 3 apresenta os resultados obtidos e conclui o artigo.

2. Metodologia e Técnicas Aplicadas

A proposta deste trabalho tem o objetivo de medir a aplicabilidade de algoritmos de anonimização de dados através da realização de medições da performance dos algoritmos em dispositivos IoT conectados entre si, tabelando então os contextos em que os dispositivos responderam de maneira esperada ou não.

Os seguintes dispositivos foram usados: ESP32, ESP8266/WEMOS, Raspberry Zero e Raspberry Pi. A placa WEMOS D1 R2 possui um módulo ESP8266EX (com clock de 80Mhz em modo regular a 160Mhz em modo burst), enquanto que o ESP32s é um módulo *WiFi* de alta performance (clock máximo de 240 MHz e uma memória RAM de 520Kbytes) com 4 *Megabytes* de memória flash. Por outro lado, a Raspberry Pi Zero possui um processador *Broadcom* BCM2835 *single-core* de 1GHz e memória de 512MB. Similarmente, a Raspberry Pi 3 tem um processador de 1,4 GHz Broadcom BCM2837B0 64bits ARM Cortex-A53 Quad-Core e 1GB de memória RAM.

A partir do exposto, percebe-se que os hardwares utilizados permitem uma avaliação ampla das técnicas de anonimização, visto que estes dispositivos possuem características e recursos computacionais diferentes, assim como é esperado nos diversos contextos de soluções IoT existentes. Baseado nisso, foram escolhidos seis algoritmos de anonimização utilizados em diversas soluções tradicionais de segurança e proteção de dados, mas que possuem abordagens e resultados distintos, permitindo a manutenção da privacidade e proteção dos dados de forma singular, e assim podendo ser aplicados de acordo com o contexto onde o dispositivo IoT está inserido.

Assim, foram analisados os seguintes algoritmos: (1) Black Marker, esta técnica marca os caracteres do campo anonimizável com valores constantes (por exemplo, ”*”, ”X”), provendo um mascaramento parcial (aplicado em apenas alguns caracteres do atributo) e sendo aplicada quando esconder parte do valor do campo é suficiente; (2) Deslocamento, consiste em somar ao valor a ser anonimizado um valor fixo, sendo assim é reversível, uma vez que a constante de incremento seja conhecida; (3) Anonimização de IP com Truncamento, visa eliminar os valores do endereço IP, que pode revelar informações sobre a estrutura de rede; (4) Supressão de Atributo, se refere à remoção de uma parte inteira dos dados de um certo conjunto, ocorrendo quando um atributo não é necessário

ou quando o atributo não pode ser anonimizados com outra técnica. (5) Generalização, é uma forma de redução de precisão deliberada, onde o intuito é separar as variáveis que compõem os campos que irão ser anonimizados em intervalos (definidos pelo autor da anonimização); e, (6) Substituição, visa trocar os valores que devem ser anonimizados por valores que já foram definidos, sendo reversível uma vez que tendo acesso à tabela de substituição é possível reverter a operação realizada.

É válido destacar que a aplicação de algoritmos de anonimização pode ser feita de forma agregada, ou seja, fazendo a anonimização final ser uma composição dos resultados de duas ou mais técnicas. Por exemplo, o algoritmo de substituição é bem utilizado quando os dados que estão sofrendo anonimização são alfanuméricos mesmo podendo ser utilizados em outras simbologias. Por ser um algoritmo simples, o algoritmo de substituição é acompanhado por outro algoritmo reversível também, para adicionar complexidade e tornar mais difícil de recuperar os dados.

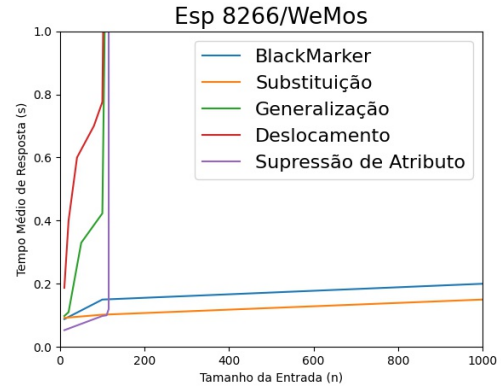
3. Experimentos e Resultados

O objetivo dos experimentos realizados foi analisar o limite de capacidade de cada dispositivo em relação aos algoritmos de anonimização descritos anteriormente, visto que identificar o nível de aplicabilidade de cada algoritmo nos dispositivos IoT poderá orientar o desenvolvimento de soluções de privacidade e proteção de dados. Portanto, durante os experimentos foram avaliadas duas métricas de desempenho: (1) Tempo médio de resposta (TMR), que se refere ao tempo que o dispositivo em questão leva para executar a anonimização dos dados de entrada; e, (2) Percentual de ocupação da memória RAM do dispositivo em questão, mostrando o impacto da anonimização sobre o dispositivo e sua concomitância de execução com outros serviços presentes. A análise do TMR permite identificar a viabilidade de cada algoritmo para os diversos contextos de IoT, os quais possuem requisitos singulares de atraso fim-a-fim para um QoS/QoE adequado.

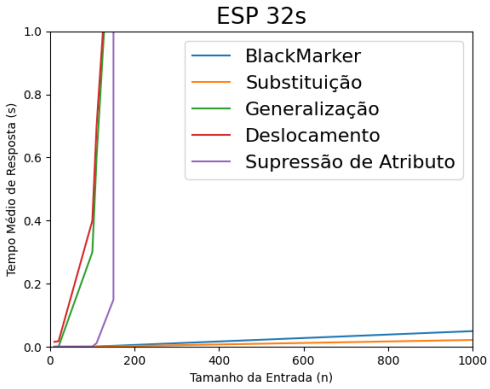
Com relação aos dados a serem anonimizados, foram escolhidos aleatoriamente dados sensíveis de acordo com as Leis de Privacidade, onde os parâmetros recebem um índice e são sorteados para compor uma tabela de dados para os testes [de Oliveira 2019]. Adicionalmente, variou-se o tamanho da entrada de dados de 10 a 10^4 bytes, a fim de possibilitar uma análise mais completa do comportamento dos algoritmos nos dispositivos IoT. É válido ressaltar que o processo de anonimização ocorre em anexo a outros serviços que executam nos dispositivos IoT. Portanto, a demanda de recursos da anonimização não pode inviabilizar e/ou impactar negativamente os demais serviços (comprometendo o QoS/QoE), como por exemplo monitoramento, automação, etc.

A Figura 1 apresenta os resultados obtidos nos experimentos, onde cada experimento foi repetido 50 vezes. Nos gráficos, é possível observar o comportamento assintótico dos algoritmos *Black Marker* e Substituição, onde a diferença do ESP32s para o WEMOS é o tamanho máximo de entrada suportada, uma vez que, a partir de $n > 10^5$ para a WEMOS e $n > 10^6$ para a ESP32s o tempo de resposta dos dispositivos é extremamente alto (tornando-se inviável). Este comportamento ocorre devido a alocação dinâmica de memória utilizada por estes algoritmos, a qual impacta de maneira mais suave no TMR (como pode ser observado nas Figuras 1(a) e 1(b)) e na disponibilidade de memória (apresentada nas Figuras 1(c) e 1(d)).

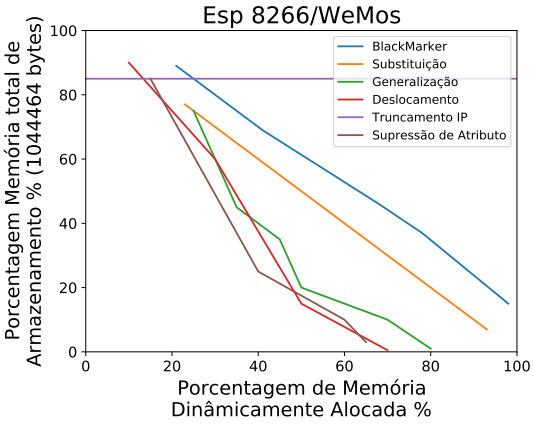
Os demais algoritmos (Generalização, Deslocamento e Supressão de Atributo)



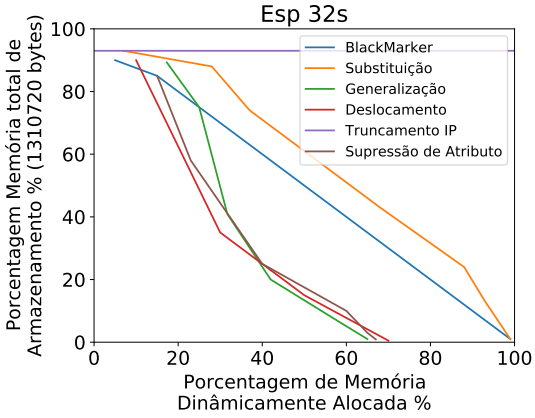
(a) ESP8266/WEMOS



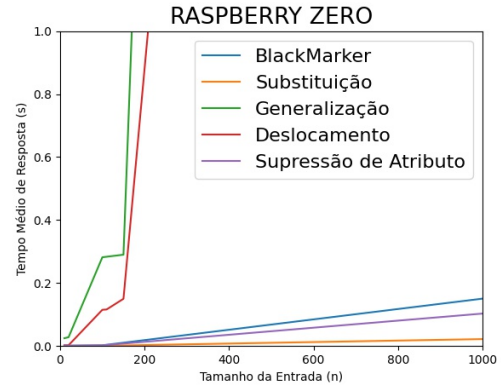
(b) ESP32s



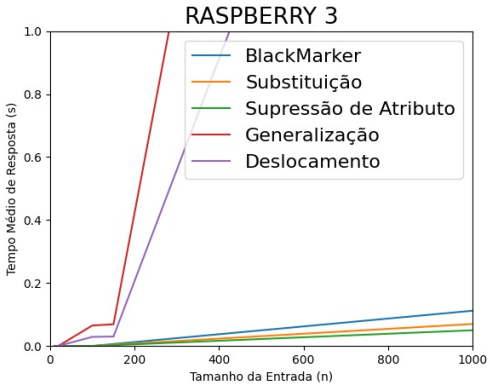
(c) ESP8266/WEMOS



(d) ESP32s



(e) Raspberry Pi Zero



(f) Raspberry Pi 3

Figura 1. Tempo Médio de Resposta (TMR) e ocupação de Memória.

possuem um comportamento diferente dos citados anteriormente, pois sua natureza de anonimização é menos dinâmica, resultando em um maior consumo de recursos computacionais e, consequentemente, restringindo o tamanho da entrada de dados. Os algoritmos de Substituição e Deslocamento seguem um comportamento quadrático, visto que possuem um padrão de ordem n^2. Por este motivo, quando a quantidade de dados de entrada supera o número de 10^3 bytes, os dispositivos operam com bastante dificuldade, inviabilizando a execução de outros serviços nos dispositivos IoT.

O algoritmo de Anonimização por Supressão de Atributo é um caso peculiar, apesar da função de processamento se comportar de acordo com uma função linear a alocação de memória do mesmo ocorre de forma quadrática, e os dispositivos ESPs, por serem bastante simples, possuem um espaço de memória utilizado para as operações bem limitado, de maneira mais objetiva, na ESP8266, instânciar uma matriz de números inteiros com dimensões 110 por 110 juntamente com um módulo de qualquer um desses algoritmos que foram implementados ocupa aproximadamente 100 % da memória do dispositivo, a ESP32s apesar de possuir um pouco a mais não está muito distante, podendo instanciar uma matriz de no máximo 150 linhas por 150 colunas.

Nos gráficos das Figuras 1(c) e 1(d) é possível notar que os algoritmos de Anonimização por substituição, Anonimização por Generalização e Anonimização por Supressão de atributo ao alocar memória dinamicamente seguem um padrão bem parecido com a de uma função quadrática e que os algoritmos *Black Marker* e Anonimização por substituição seguem o padrão linear. O caso peculiar deste experimento foi o Algoritmo de Truncamento do endereço IP, por sua entrada possuir tamanho constante o valor de memória consumida manteve-se constante em todos os casos de teste, por isso a representação de consumo deste algoritmo é dado por uma reta constante. Analisando de maneira mais precisa, é possível observar que as linhas das funções nunca começam com 100 % de memória disponível, isto se dá pelo fato de que é necessário uma parcela da memória para realizar o upload das instruções básicas do algoritmo, onde ocorre também a instância dinâmica das variáveis justamente por os dispositivos em questão serem desprovidos de recursos.

As Figuras 1(e) e 1(f) mostram o desempenho das raspberries. A principal diferença observada são os limites de entrada apresentados serem maiores, especialmente o *Supressão de Atributo* que apresentou um comportamento linear, enquanto que nos experimentos anteriores, assemelhou-se ao comportamento de uma função quadrática. Isto deve ao fato de que a memória de processamento das raspberrys são extremamente maiores do que as das ESPs.

3.1. Discussão Final e Conclusão

Com a realização dos experimentos foi possível observar de forma experimental a superioridade das Raspberry Pis em relação as ESPs: ESP8266/WEMOS consegue realizar em média 10^5 operações atômicas em 1 segundo, enquanto a Raspberry Pi 3 consegue realizar em torno 10^7 operações. O comportamento das funções de representação mantiveram-se semelhantes com exceção do Algoritmo de Anonimização por Supressão de Atributo: a alocação de memória nas Raspberry Pis não é um limitante, uma vez que a disponibilidade de memória tende a ser muito superior. Com isso, a restrição de memória do algoritmo de Supressão de Atributo que foi encontrada nos dispositivos ESPs, não supera a restrição de quantidades de operações realizadas pelos mesmos, diminuindo a quantidade de condições de gerar gargalo. Consequentemente, os limites de TMR do algoritmo de Anonimização por Supressão aumentaram de aproximadamente 10^2 nas ESPs para 10^7 nas Raspberries.

Como resultado final dos experimentos foi possível mapear o desempenho de cada algoritmo em cada dispositivo: o comportamento dos limites assintótico das funções que representam quantidade de operações executadas em detrimento do TMR em cada dispositivo, em conjunto com o experimento que aponta o consumo de memória dos algoritmos

em relação à quantidade de memória disponível, tornou possível o tabelamento dos limites de processamento de cada dispositivo no contexto o qual o algoritmo está implementado.

Tabela 1. Tabela com os Limites de Entrada

Entrada	Black Marker	Substituição	Generalização	Deslocamento	Supressão de Atributo	IP por Truncamento
Esp 8266 (WeMos)	10^5	10^5	10^2	10^2	10^2	10
Esp 32s	10^6	10^6	1,5x10^2	1,5x10^2	1,5x10^2	10
Raspberry Pi Zero	10^7	10^7	10^3	10^3	10^7	10
Raspberry Pi 3	10^8	10^8	10^3	10^3	10^8	10

De forma numérica, a Tabela 1 mapeia os resultados dos experimentos realizados nos dispositivos, apresentando os limites de entrada identificados. Assim, diante desses valores é possível identificar qual dispositivo IoT é adequado a solução em desenvolvimento, considerando a aplicação de técnicas de proteção de dados e evitando gargalos de processamento. Como trabalhos futuros, pretende-se desenvolver uma nova abordagem de anonimização que seja viável para dispositivos IoT e que atenda de maneira transversal os requisitos previstos nas leis de privacidade.

Referências

Broadcom, R. P. (2012). Bcm2835 arm peripherals. *Broadcom Europe Ltd.*, pages 1–202.

Datasheet, E. (2015). Esp8266ex datasheet. *Espr. Syst. Datasheet*, pages 1–31.

Datasheet, E. (2021). Esp32s datasheet. *Espr. Syst. Datasheet*, pages 1–60.

de Oliveira, N. S. (2019). Segurança da informação para internet das coisas (iot): uma abordagem sobre a lei geral de proteção de dados (lgpd).

Pang, R. (2016). The devil and packet trace anonymization. *Computer Communication Review*, 36(1):29–38.