Preservando a Privacidade na Internet das Coisas com Pseudônimos Usando SDN

Antonio J. Pinheiro[1], **Caio A. P. Burgardt**[1], **Divanilson R. Campelo**[1]

[1]Universidade Federal de Pernambuco - Centro de Informática (CIn)

{ajp,capb,dcampelo}@cin.ufpe.br

Abstract. *Internet of Things devices usually capture information about their users, generating several privacy risks. Observers can infer information about their victims from the addresses of IoT devices. In this paper, we present a pseudo-anonymization solution based on Software-Defined Networking that hides IoT addresses. We evaluate this solution with real IoT traffic from a variety of devices. The results show that the proposed solution contributes to improving the privacy of IoT users. Hypothesis tests based on the Wilcoxon Signed-Rank Test attest that the impact on the communication performance generated by the solution is negligible.*

Resumo. *Dispositivos da Internet das coisas (Internet of Things, IoT) geralmente capturam informações sobre os seus usuários, gerando riscos à privacidade. Observadores podem inferir informações sobre suas vítimas a partir dos endereços de dispositivos IoT. Neste artigo, apresentamos uma solução de pseudo-anonimização baseada em redes definidas por software que oculta os endereços. Avaliamos essa solução com tráfego IoT real proveniente de diversos dispositivos. Os resultados mostram que a solução proposta contribui para aprimorar a privacidade dos usuários IoT. Testes de hipóteses baseados no Wilcoxon Signed-Rank Test atestam que o impacto ao desempenho da comunicação gerado pela solução é irrisório.*

1. Introdução

A *Internet* das coisas (*Internet of Things*, IoT) é composta por uma variedade de dispositivos com capacidades de sensoriamento e comunicação em rede, tais como câmeras, sensores de movimento, lâmpadas, etc. Dispositivos IoT estão presentes em diversos domínios de nossas vidas diárias, como casas, cidades, carros conectados, hospitais e indústria [Atzori et al. 2010]. A presença pervasiva desses dispositivos introduz inúmeros riscos à privacidade de seus usuários [Liu et al. 2018].

A criptografia é insuficiente para proteger a privacidade da comunicação em rede, e os endereços dos dispositivos podem ser mais relevantes em relação à privacidade que o conteúdo do tráfego [Cabaj et al. 2018]. A partir dos endereços *Internet Protocol* (IP) e *Media Access Control* (MAC), um observador é capaz de inferir, por exemplo, quantos e quais dispositivos estão ativos, com quem eles estão se comunicando, hábitos dos usuários e doenças ou problemas de saúde dos residentes de uma *smart home* [Liu et al. 2018]. É possível identificar o fornecedor e inferir o modelo de um dispositivo IoT a partir do seu endereço MAC. As portas da camada de transporte revelam as aplicações e serviços em execução nos dispositivos IoT [Sivanathan et al. 2017]. Esses riscos à privacidade

motivaram o desenvolvimento de soluções para ocultar os endereços dos dispositivos IoT [Davoli et al. 2017, Arana et al. 2018].

Ocultar a identidade dos participantes da comunicação é um grande desafio, e a pseudo-anonimização é uma técnica usada para alcançar esse objetivo. Em redes, pseudônimo é um identificador fictício usado para ocultar o endereço real de um sistema final [Wagner and Eckhoff 2018]. Soluções de pseudo-anonimização foram propostas para IoT, mas negligenciam as restrições técnicas destes dispositivos [Zeitz et al. 2018, Ullah et al. 2018, Davoli et al. 2017]. Essas soluções elevam o consumo de energia e o processamento dos dispositivos. Mecanismos de pseudo-anonimização usando Redes Definidas por *Software* (do inglês, *Software-Defined Networking*, SDN) também foram apresentados [Zhu et al. 2017, Sharma et al. 2018]. Tipicamente, esses mecanismos elevam a carga de trabalho do controlador, o que pode resultar em problemas no desempenho da rede.

Neste artigo, apresentamos uma solução de pseudo-anonimização que utiliza SDN para ocultar os endereços e portas de transporte de origem e destino dos dispositivos IoT. Essa solução configura os *switches* na borda da rede para substituírem os endereços dos dispositivos por pseudônimos. Uma aplicação externa ao controlador cria aleatoriamente esses pseudônimos. O controlador SDN obtém esses pseudônimos através de uma interface *REpresentational State Transfer* (REST) proposta neste trabalho. O controlador SDN será responsável apenas por configurar os *switches*, o que reduz a sua carga de trabalho. A solução proposta é transparente aos sistemas finais – dispositivos IoT e servidores.

A solução proposta garante que os pseudônimos de endereços MAC ocultem o fabricante do dispositivo. Ofuscamos o relacionamento entre dispositivos IoT e servidores remotos. Ao modificar as portas, ocultamos os serviços usados pelos dispositivos. Além disso, um observador será incapaz de inferir quais dispositivos estão ativos. Neste trabalho, focamos no cenário IoT; outros domínios de rede serão explorados em trabalhos futuros.

2. Trabalhos relacionados

Inúmeros mecanismos de pseudo-anonimização foram propostos para IoT [Zeitz et al. 2018, Haas and Yousefpour 2018, Davoli et al. 2017, Arana et al. 2018, Ullah et al. 2018]. Essas soluções exigem que os dispositivos IoT sejam modificados, o que resulta em maior consumo de energia e recursos computacionais. Além disso, tais soluções têm problemas de escalabilidade, visto que exigem a modificação de milhões de dispositivos IoT.

Mecanismos baseados em SDN foram propostos para aprimorar a privacidade da comunicação com pseudônimos [Zhu et al. 2017, Sharma et al. 2018]. Contudo, essas soluções usam o controlador para gerar os pseudônimos, o que eleva a sua carga de trabalho e pode resultar em problemas para a rede. Ademais, algumas soluções exigem modificações nos sistemas finais [Zhu et al. 2017].

Não foram encontrados trabalhos sobre pseudo-anonimização de endereços IoT usando SDN. A solução proposta neste trabalho torna desnecessárias modificações nos dispositivos IoT, dispensa adaptações nos protocolos de roteamento e é transparente para os sistemas finais. Os pseudônimos são gerados por um mecanismo externo ao contro-

lador SDN e são alterados endereços MAC, IPv4, IPv6 e portas *User Datagram Protocol*(UDP)/*Transmission Control Protocol* (TCP).

3. Solução de pseudo-anonimização proposta

A solução proposta tem como objetivo ocultar os endereços dos dispositivos IoT e substituí-los por pseudônimos criados aleatoriamente. Utilizamos as características de programabilidade dos dispositivos de rede e controle centralizado da arquitetura SDN. Utilizamos os *switches* SDN para modificar os endereços, o que evita restrições intrínsecas aos dispositivos IoT, como processamento e consumo de energia [Miettinen et al. 2017]. Assumimos que a rede local é segura, por isso é desnecessário realizar a modificação dos endereços nos dispositivos IoT. Os endereços são anonimizados apenas na rede gerenciada pelo controlador, mas essa limitação pode ser eliminada através da comunicação entre controladores de redes distintas. A Figura 1 apresenta uma visão geral da solução proposta.

Os endereços são modificados pelos dispositivos na borda da rede, como *home gateways* e roteadores. O tráfego no núcleo da rede é encaminhado com base nos pseudônimos. Modificações nos dispositivos IoT ou servidores são desnecessárias. Os *switches* conectados aos dispositivos IoT substituem os endereços originais por pseudônimos atribuídos pelo controlador, enquanto os *switches* conectados aos servidores modificam os pseudônimos para os identificadores originais e encaminham o tráfego para o seu destino. O servidor precisa do endereço original para enviar as respostas para requisições dos dispositivos IoT.

Implementamos um mecanismo responsável por gerar e armazenar os pseudônimos dos endereços MAC, IPv4, IPv6 e portas de transporte. Esse mecanismo segue as seguintes regras na criação dos pseudônimos: os três primeiros *bytes* do MAC são fixos, apenas os *bytes* restantes são aleatórios; são permitidas apenas portas superiores a 1024. Usamos uma faixa de endereços reservados nos 3 *bytes* iniciais do MAC para ocultar o fabricante do dispositivo. Portas inferiores a 1024 são reservadas para serviços conhecidos; ao evitá-las, impossibilitamos a identificação desses serviços. Não restringimos o número de endereços IPv4 e IPv6 que podem ser criados, o que aumenta a robustez da solução proposta [Wagner and Eckhoff 2018]. Contudo, a quantidade de endereços usados depende do cenário em que a solução proposta for aplicada. Ademais, o mecanismo pode fornecer endereços para múltiplos controladores.

O controlador SDN solicita os pseudônimos através de uma interface REST. O controlador obtém os endereços e portas usados pelos dispositivos IoT a partir de requisições *packet-in* e evita que esses identificadores sejam utilizados como pseudônimos. Ao receber uma mensagem *packet-in*, o controlador encontra o caminho fim-a-fim entre origem e destino do tráfego. Os *switches* que compõem esse caminho são configurados da seguinte forma: o primeiro *switch* substitui os endereços originais por pseudônimos; *switches* no centro do caminho encaminham o tráfego com base nos pseudônimos; por fim, o último *switch* – conectado a destinatário– modifica os pseudônimos para os endereços originais. Dessa forma, a solução proposta é transparente aos dispositivos IoT e servidores.

Ocultar os endereços nos switches permite escalar a solução de forma mais simples, pois seria complexo modificar centenas ou milhares de dispositivos IoT. Um *switch*

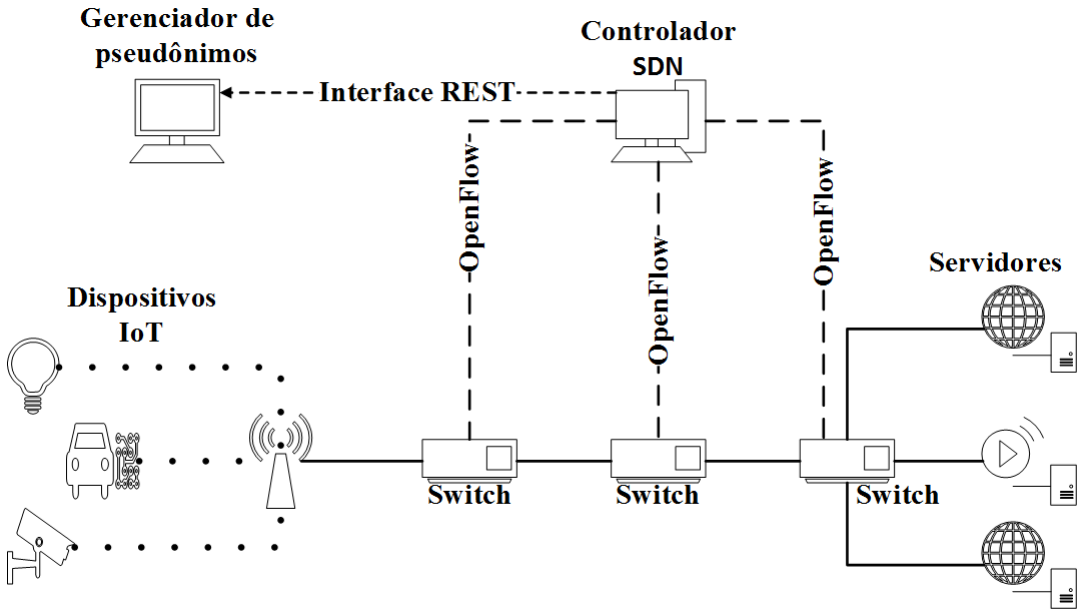

Figura 1. Visão geral da solução de pseudo-anonimização proposta.

é capaz de proteger inúmeros dispositivos. Os fluxos são identificados por: endereços MAC de origem e destino; endereços IP de origem e destino; portas de transporte de origem e destino. Para reduzir o número de entradas de fluxos nos *switches*, definimos o *idle timeout*[1] para 5 segundos. Dessa forma, as entradas de fluxos são removidas pelos *switches* após permanecerem inativas por mais que 5 segundos. Os autores de [Sivanathan et al. 2017] identificaram que 95% das conexões IoT analisadas por eles têm uma duração de 5 segundos.

4. Protótipo da solução proposta

Implementamos um protótipo com o *Mininet* 2.2.2 em uma máquina virtual (*Virtual Machine*, VM). Essa VM foi configurada com: 4 GB de *Random Access Memory* (RAM) *Double Data Range (DDR) 3*; duas CPUs virtuais do processador *Intel® Core™* i7 2.40GHz 64-bit; e o sistema operacional *Ubuntu* 14.04 *Long Term Support* (LTS) versão para servidor. O *software switch Open vSwitch* 2.9.2 foi usado para criar os *switches* da rede emulada. O gerenciador de pseudônimos foi implementado em *Python*. A interface REST para comunicação com o controlador foi implementada com o *framework Flask*. Os dados foram transferidos para o controlador no formato de texto *JavaScript Object Notation* (JSON).

A aplicação responsável por encontrar e configurar o caminho fim-a-fim foi implementada no controlador *Ryu*. A biblioteca *networkx* foi usada para criar um grafo a partir da topologia da rede e encontrar o caminho mais curto – em número de *switches* – com o algoritmo de *Dijkstra*. O *OpenFlow* 1.3 foi utilizado para configurar as regras de fluxos nos *switches*. A biblioteca ***requests*** foi usada para gerar solicitações ao gerenciador de pseudônimos.

[1]Período de permanência de uma entrada nas tabelas dos *switches* sem atividade no fluxo correspondente.

5. Avaliação do protótipo

Usamos o *Mininet* para emular uma rede composta por dois *hosts* (IoT1 e servidor1) e 50 *switches*[2]. Os *switches* foram posicionados em série para garantir que todo o tráfego entre esses *hosts* atravesse os 50 comutadores. A Figura 2 ilustra a topologia usada na avaliação da solução proposta. Usamos tráfego IoT real de 52 dispositivos para quantificar a melhoria na privacidade. Por isso, avaliamos as seguintes métricas para privacidade: conjunto de anonimização – quantidade de pseudônimos disponíveis para uso; quantidade de endereços ativos; número de portas observadas; relacionamentos entre sistemas finais [Wagner and Eckhoff 2018]. Além disso, usamos as ferramentas *Ping* e *iPerf* para mensurar o impacto da solução proposta ao desempenho da rede. Como métricas, utilizamos o atraso, vazão, *jitter* e perda de pacotes.

Utilizamos o seguinte modelo de ataque: um observador passivo captura tráfego em qualquer ponto na rede do *Internet Service Provider* (ISP). Usamos o *tcpreplay* [Turner 2011] para reproduzir o tráfego IoT do *host* **IoT1** para o **servidor1**[3], passando pelos 50 *switches*. Esse tráfego foi disponibilizado publicamente pelos autores de [Miettinen et al. 2017, Sivanathan et al. 2017]. Os autores de [Miettinen et al. 2017] capturaram o tráfego gerado durante a configuração inicial (*setup*) de 31 dispositivos IoT; os dados disponibilizados pelos autores de [Sivanathan et al. 2017] foram capturados de 21 *devices* ao longo de 24 horas. Esse experimento tem uma duração de 24 horas. Usamos o *TCPdump* para capturar o tráfego anonimizado[4] no *switch* 25 – contundo, poderia ser qualquer *switch*. O *Wireshark* foi usado para computar a quantidade de endereços observados e o relacionamentos entre sistemas finais.

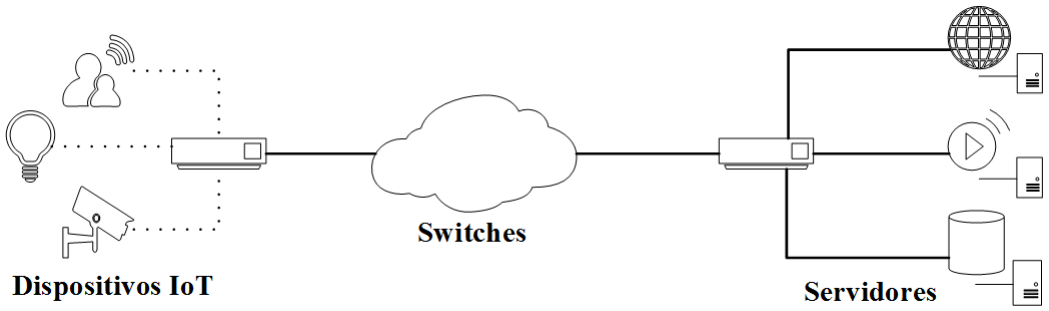

Figura 2. Topologia usada para avaliar a solução de pseudo-anonimização proposta.

Em outro experimento, enquanto o *tcpreplay* reproduziu o tráfego IoT disponibilizado por [Miettinen et al. 2017], o *Ping* foi usado para mensurar o atraso e o *iPerf* para quantificar a vazão, *jitter* e perda de pacotes. O *iPerf* foi configurado para gerar 20Mbps[5] de tráfego. Executamos essas duas ferramentas em paralelo ao longo de 12 horas para obter o desempenho na mesma carga de trabalho. Com uma amostragem a cada segundo, obtivemos 43.200 amostras. Calculamos os valores médios para as métricas avaliadas.

[2]O dobro do número de *switches* usados em [Zhu et al. 2017].
[3]A solução proposta lida com tráfego cliente → servidor e vice-versa.
[4]Tráfego cujo endereços originais foram substituídos por pseudônimos.
[5]Valor 20 vezes superior ao pico no tráfego disponibilizado pelos autores de [Sivanathan et al. 2017].

6. Resultados

Nesta seção, apresentamos os resultados obtidos na avaliação do protótipo de pseudo-anonimização.

6.1. Privacidade e anonimização

O conjunto de anonimização é uma das métricas mais relevantes na avaliação de uma solução de pseudo-anonimização [Wagner and Eckhoff 2018]. Essa métrica determina a quantidade de pseudônimos que a solução pode usar. Quanto maior é o conjunto de anonimização, mais robusto é o mecanismo proposto [Wagner and Eckhoff 2018]. Para identificadores IPv4 e IPv6, utilizamos todos os endereços possíveis. Para os três primeiros *bytes* do MAC, usamos uma faixa reservada pela *Internet Assigned Numbers Authority* (IANA) para *multicast*: 01-00-5E. Contudo, a solução proposta pode utilizar qualquer outra faixa de endereços reservados. A solução gera valores aleatórios para os 3 últimos *bytes* do MAC. A seguir são apresentados os conjuntos de anonimização para os endereços protegidos pela solução: MAC: 2^{24}; IPv4: 2^{32}; IPv6: 2^{128}; portas de transporte: $2^{16} - 1024$ ou 64.512.

Nem todos os dispositivos analisados possuem endereços IPv6. O tráfego foi capturado na rede local, por isso, foram observados somente os endereços MAC dos dispositivos IoT. A Tabela 1 apresenta os resultados da substituição dos endereços originais por pseudônimos. Nessa tabela, são apresentadas as quantidades de endereços observados dos dispositivos IoT e servidores. O número de dispositivos é igual para os dois cenários: sem e com o protótipo. Contudo, quando o protótipo está em execução, o número de endereços visíveis aos observadores é drasticamente elevado. Ao longo do tempo, cada dispositivo é identificado por diversos pseudônimos. Assim, um observador será incapaz de determinar quantos e quais dispositivos estão ativos em uma rede. Além disso, é impossível identificar os relacionamentos entre participantes da comunicação através dos endereços.

	Sem o protótipo	Protótipo
Endereços IPv4 (quantidade)	302	10602
Endereços IPv6 (quantidade)	49	6898
Endereços MAC (quantidade)	52	17743
Portas de destino (quantidade)	7911	37959
Relacionamentos IPv4 (quantidade)	338	5301

Tabela 1. Quantidade de endereços observados por um observador: com e sem o protótipo protegendo a rede composta por 52 dispositivos IoT.

O protótipo oculta os serviços usados – identificados pelas portas de transporte TCP e UDP. O protótipo torna observadores incapazes de determinar com quem os dispositivos se comunicam. Como cada conexão é representada por pseudônimos distintos, um observador verá múltiplos sistemas finais trocando informações, sendo o mesmo par de dispositivos. Dessa forma, a solução proposta reduz a validade das informações obtidas pelos observadores [Sharma et al. 2018].

6.2. Desempenho da comunicação

O desempenho da comunicação é apresentado na Tabela 2. Os resultados desta tabela demostram que o mecanismo de pseudo-anonimização praticamente não interferiu no de-

sempenho da rede. O impacto gerado pela modificação dos endereços nos *switches* é negligenciável [Zhu et al. 2017]. Além disso, como o controlador é responsável apenas por configurar as regras de fluxos – os pseudônimos são criados externamente – as requisições dos *switches* são atendidas rapidamente.

	Sem o protótipo	Protótipo
Atraso médio (ms)	0,172	0,168
Jitter médio (ms)	0,028	0,033
Vazão média (Mbps)	19,83	19,84
Perda de pacotes média (%)	0,082	0,084

Tabela 2. Desempenho da comunicação em carga de trabalho de 20Mbps.

Aplicamos testes de hipóteses para validar os resultados sobre o desempenho da comunicação. Usamos o teste de aderência *Kolmogorov-Smirnov* para identificar quais testes de hipóteses devem ser aplicados: paramétricos ou não-paramétricos. Os dados analisados são pareados[6] e o *Kolmogorov-Smirnov* indicou que devem ser aplicados testes não-paramétricos. Por isso, utilizamos o teste não-paramétrico *Wilcoxon Signed-Rank Test* (WSRT) com um nível de confiança de 95% para comparar os resultados dos cenários com e sem o protótipo. Formulamos as seguintes hipóteses para este teste: h_0 : o protótipo não interfere no desempenho da rede, e h_1: o protótipo interfere no desempenho da rede. O teste WSRT apontou que a hipótese h_0 não deve ser rejeitada com base nos dados analisados. Por isso, estatisticamente, o protótipo não interfere no atraso, vazão e perda de pacotes. Há apenas uma elevação do *jitter* da ordem de 0,005 ms, que não impacta a solução proposta.

7. Conclusão

Evidenciamos que a arquitetura SDN é relevante ao desenvolvimento de soluções para pseudo-anonimização de endereços IoT. Esses dispositivos introduzem inúmeros riscos à privacidade dos indivíduos. Observadores são capazes de inferir informações sobre suas vítimas a partir de endereços dos dispositivos. Por isso, mecanismos para pseudo-anonimização IoT são imprescindíveis.

Apresentamos uma solução para ocultar os endereços IoT de observadores. Os endereços originais são substituídos por pseudônimos nos *switches* localizados na borda da rede. Apenas a origem e o destino do tráfego conhecem os endereços originais de suas contrapartes. A solução proposta é transparente aos dispositivos IoT e servidores.

Avaliamos a solução proposta com tráfego IoT gerado por 52 dispositivos. Avaliamos a solução na preservação da privacidade e o impacto gerado ao desempenho da comunicação. Os resultados demonstram que a solução oculta o número de dispositivos ativos, serviços usados, fabricantes e o relacionamento entre sistemas finais. Testes de hipóteses baseados em *Wilcoxon Signed-Rank Test* confirmam que o impacto ao desempenho da comunicação é irrisório.

Agradecimentos

Este trabalho foi parcialmente apoiado pela FACEPE e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

[6]Por exemplo, resultados obtidos com e sem um determinado tratamento.

Referências

Arana, O., Garcia, F., Gomez, J., and Rangel, V. (2018). MSP: Providing location privacy in wlan networks with a mac swapping protocol. *Computer Networks*, 138:136 – 148.

Atzori, L., Iera, A., and Morabito, G. (2010). The internet of things: A survey. *Computer Networks*, 54(15):2787 – 2805.

Cabaj, K., Caviglione, L., Mazurczyk, W., Wendzel, S., Woodward, A., and Zander, S. (2018). The new threats of information hiding: the road ahead. *CoRR*, abs/1801.00694.

Davoli, L., Protskaya, Y., and Veltri, L. (2017). An anonymization protocol for the internet of things. In *2017 International Symposium on Wireless Communication Systems (ISWCS)*, pages 459–464.

Haas, Z. J. and Yousefpour, A. (2018). A privacy scheme for monitoring devices in the internet of things. *CoRR*, abs/1803.04453.

Liu, J., Zhang, C., and Fang, Y. (2018). EPIC: A differential privacy framework to defend smart homes against internet traffic analysis. *IEEE Internet of Things Journal*, 5(2):1206–1217.

Miettinen, M., Marchal, S., Hafeez, I., Asokan, N., Sadeghi, A. R., and Tarkoma, S. (2017). IoT SENTINEL: Automated device-type identification for security enforcement in iot. In *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, pages 2177–2184.

Sharma, D. P., Kim, D. S., Yoon, S., Lim, H., Cho, J.-H., and Moore, T. J. (2018). Frvm: Flexible random virtual ip multiplexing in software-defined networks. In *International Conference On Trust, Security And Privacy In Computing And Communications (TrustCom)*, New York, NY, USA. IEEE.

Sivanathan, A., Sherratt, D., Gharakheili, H. H., Radford, A., Wijenayake, C., Vishwanath, A., and Sivaraman, V. (2017). Characterizing and classifying iot traffic in smart cities and campuses. In *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 559–564.

Turner, A. (2011). Tcpreplay. *http://tcpreplay. synfin. net/trac/*.

Ullah, I., Shah, M. A., Wahid, A., Mehmood, A., and Song, H. (2018). ESOT: a new privacy model for preserving location privacy in internet of things. *Telecommunication Systems*, 67(4):553–575.

Wagner, I. and Eckhoff, D. (2018). Technical privacy metrics: A systematic survey. *ACM Comput. Surv.*, 51(3):57:1–57:38.

Zeitz, K., Cantrell, M., Marchany, R., and Tront, J. (2018). Changing the game: A micro moving target IPv6 defense for the internet of things. *IEEE Wireless Communications Letters*, pages 1–1.

Zhu, T., Feng, D., Wang, F., Hua, Y., Shi, Q., Liu, J., Cheng, Y., and Wan, Y. (2017). Efficient anonymous communication in SDN-Based data center networks. *IEEE/ACM Transactions on Networking*, 25(6):3767–3780.