



FEDERAL UNIVERSITY OF SANTA CATARINA
TECHNOLOGY CENTRE
POSTGRADUATE PROGRAMME IN COMPUTER SCIENCE

Fabricio Bortoluzzi

**Cloud Telescope: an ephemeral, distributed, cloud-native architecture for collecting
Internet Background Radiation**

Florianopolis
2025

Fabrizio Bortoluzzi

**Cloud Telescope: an ephemeral, distributed, cloud-native architecture for
collecting Internet Background Radiation**

Thesis submitted to the Postgraduate Programme in Computer Science to the awarding of the title of Doctor of Computer Science.

Supervisor: Prof. Carla M. Westphall, Dr.

Supervisor: Prof. Barry Irwin, PhD.

Florianopolis

2025

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Bortoluzzi, Fabricio

Cloud Telescope: an ephemeral, distributed, cloud
native architecture for collecting Internet Background
Radiation / Fabricio Bortoluzzi ; orientador, Carla Merkle
Westphall, orientador, Barry Irwin, 2025.
170 p.

Tese (doutorado) - Universidade Federal de Santa
Catarina, Centro Tecnológico, Programa de Pós-Graduação em
Ciência da Computação, Florianópolis, 2025.

Inclui referências.

1. Ciência da Computação. 2. Medição de tráfego da
Internet. 3. Análise de Tráfego de Rede. 4. Telescópios de
Rede. 5. Telescópio de Nuvem. I. Westphall, Carla Merkle.
II. Irwin, Barry III. Universidade Federal de Santa
Catarina. Programa de Pós-Graduação em Ciência da Computação.
IV. Título.

Fabricio Bortoluzzi

**Cloud Telescope: an ephemeral, distributed, cloud-native architecture for collecting
Internet Background Radiation**

The present doctorate thesis was assessed and approved by the examination board comprised
of the following members:

Prof. Jean E. Martina, PhD.
Federal University of Santa Catarina

Prof. Mauricio de Campos, Dr.
University of Vale do Itajai

Prof. Cesar Albenes Zeferino, Dr.
University of Vale do Itajai

Prof. Barry Irwin, PhD.
Supervisor
Noroff University College - Rhodes University

We certify this is the **original and final version** of the thesis,
deemed suitable to the awarding of the title of Doctor of Computer Science.



Documento assinado digitalmente
Marcio Bastos Castro
Data: 02/10/2025 13:22:17-0300
CPF: ***.707.720-**
Verifique as assinaturas em <https://v.ufsc.br>

Prof. Márcio Bastos Castro, PhD.
Programme Leader



Documento assinado digitalmente
Carla Merkle Westphall
Data: 02/10/2025 15:04:11-0300
CPF: ***.150.339-**
Verifique as assinaturas em <https://v.ufsc.br>

Prof. Carla M. Westphall, Dr.
Supervisor

Florianopolis, 2025.

To my parents, Eunice and Odolir, *in memoriam*.

To Julia, João and Luca.

ACKNOWLEDGEMENTS

I would like to express my gratitude to Professor Carla Merkle Westphall. Our collaboration began many years ago, when she first reviewed my Bachelor's thesis in the Computer Science Programme at the University of Vale do Itajaí. That project, supervised by Professor Gilberto Luy (in memoriam), combined ipfw, tcpdump, and Snort to block in real time nmap scan attempts targeting hosts behind a custom-built FreeBSD bridge. Her confidence in my work later opened the door for my admission to the Master's Programme in Computer Science at UFSC, where, under her supervision, I investigated the use of another FreeBSD-based stack to apply root cause analysis to rank the most relevant attacks reaching Univali's /20 network. This thesis thus marks the third occasion on which she has generously spent her time and expertise to guide my work. Professor Carla, this thesis and indeed my career would not exist without your belief in my potential as a researcher. I dedicate this work to you, in recognition of your support and in admiration of your career.

Professor Barry Irwin equally plays an important role in my life. He interviewed me in 2020, to join Noroff University College and teach the most diverse courses in Cyber Security. Our first year working together was the most productive year I ever had. His moderation over my work greatly contributed to shaping my understanding of the teaching and learning relationship, and still today we adopt a peer-based review of teaching duties that greatly enhances the quality of the work delivered to our students. Professor Barry is an authority in using Network Telescopes to capturing and analysing the events discussed in this thesis, and I hope our newly introduced Cloud Telescope meets his expectations in terms of a fair and novel contribution to the field of Internet Background Radiation analysis. The format, shape, content and rationale of this thesis was consistently influenced, reviewed, and directed by Professor Barry, for whom I carry the utmost admiration as an individual, friend and research colleague. I am particularly happy with the fact we both prefer data-driven, data-centric, objective, and opinion-free research, where the observed events speak by themselves. Thank you for influencing my work in this direction, and thank you so much for the many hours spent fixing and reviewing this work, marking from silly typos to major weaknesses before the submission was made. Thank you for the countless hours reviewing and expanding the unfolding papers. I hope we write many more in the years to come.

I would like to express deep gratitude to Professor Cesar Albenes Zeferino, who served as an external reviewer of this thesis. He is one of the most inspiring individuals I have had the pleasure to work with. Professor Zeferino consistently encouraged me to pursue the doctoral degree and created many research opportunities that allowed me to expand my skills and knowledge. His explicit and often hidden support pushed me to aim for higher and more ambitious research goals. His ability to balance research, teaching, and administrative responsibilities has been a source of inspiration for me and many others striving to succeed in academia.

Completing this thesis made me get closer to the remarkable team behind the new MultiLab/IoTecLab at Univali. I am especially grateful to Professor Mauricio de Campos, who

kindly agreed to serve as another external reviewer. His valuable feedback has been instrumental, and I hope it will serve as a foundation for future collaborations that may connect the research programmes at Univali, UFSC, and Noroff. I am also thankful for his time and for hosting the *viva voce* in the fantastic facilities newly deployed on campus.

I am grateful to Professor Jean Everson Martina for his unique perspective on the Cloud Telescope, shared as an internal reviewer during both the qualification and defence sessions. His insights offered the critical reflection I needed to make this research more coherent, meaningful, and worthy of a doctoral degree. I also look forward to exploring opportunities for collaboration that may bring together LabSec, LRG, and LEDs in future research endeavours.

I would like to acknowledge the fruitful interaction and productive qualification sessions I had with Professors Valderi Leithardt and Guilherme Arthur Geronimo. Thank you for your time evaluating the early stage versions of the thesis and providing straightforward feedback that was used to evolve and enhance the depth of discussion found in this final work.

The implementation of the Cloud Telescope using Terraform was made in 2023 by Lucas Silvero Beiler as a result of the supervision process in Computer Science at Univali. Lucas made important implementation decisions that resulted in robust code. Lucas welcomed and expanded each and every optional/desirable telescope feature, and expanded them beyond my request, including the implementation of his own one-catches-all honeypot. This expanded the telescope with an optional feature that talks back to the probing sources, enabling precise malware and botnet pattern recording that later can be attributed to emerging threats including Mirai and variants. Thank you Lucas. I rest assured you will earn the Master's degree under Professor's Carla supervision at UFSC. The cyber security community needs more hacker-researchers like you!

A big shout to the outstanding students who have shared the completion of the mandatory post-graduate courses with me, especially Israel Efraim de Oliveira, who greatly contributed to expand my understanding of Algorithmic Complexity in addition to the comprehensive lectures delivered by UFSC's academic staff. Professor Rafael de Santiago, you rock! Thank you for the encouraging words.

Another shout to my former Computer Networks Laboratory partners with whom I shared around one decade of teaching, researching, and fun activities: Professors Adhemar Maria do Valle Filho and Paulo Roberto Oliveira Valim. I hope we can have more *Tuesday evening meetings* one day.

I also would acknowledge, thank, and dedicate this work to the following academics, staff or friends who have helped me with this achievement:

- Andre Borges Goedert
- Anita Maria da Rocha Fernandes, Dr
- Douglas Rossi Melo, Dr
- Eduardo Alves da Silva, soon Dr

- Ewerton Eire de Morais Alonso, soon Dr
- Fabiane Barreto Vavassori Benitti, Dr
- Jordan Passinato Sausen
- Luis Carlos Martins, Esp
- Marlei Salete Antunes Klein
- Nelson Murilo Rufino - You Shøt the Sheriff
- Seifedine Kadry, PhD

Three institutions directly contributed to the accomplishment of this thesis:

- Federal University of Santa Catarina: Thank you for hosting such a great Computer Science Post Graduate Programme and for implementing a remote-friendly lecture delivery and high-quality research activities. This work would never be possible otherwise. I hope you stay firm and strong against student movements willing to undermine this productive, modern, and friendly study format.
- University of Vale do Itajai: Thank you for retaining me as an academic staff member during all these years of remote contribution. This was only possible because Luis Carlos, Eduardo, Lucas Debatin, Felipe Viel, Anita, Cesar, Mauricio, and ultimately, Rector Professor Valdir Cechinel, Dr, trusted, tolerated, and implemented the paper work to enable this special collaboration setup. Thank you so much for doing this.
- Noroff University College: NUC has contributed with important support without which this research would not be possible. The institution sponsored the coverage of cloud costs related to the three telescope deployments. Noroff also sponsored or covered the majority of the costs related to presenting intermediate research at Cisti 2023 (Aveiro, Portugal), IEEE CloudNet 2023 (New York Area, US), and Cisti 2024 (Salamanca, Spain). Thank you for the funding and for creating a work setup that allowed me to pursue this degree concurrently with the teaching activities at the Institution.

Finally, I thank my family, including my wife Julia, sons João and Luca. You were the most impacted by the decision of pursuing this title and the time spent writing will never return. Thank you siblings Christiane and Lucas, your extended families, and my parents, Odolir and Eunice, *in memoriam*. My mother passed away days after the *viva* session, and before this final version was submitted. Thank you mother for your unconditional love. I am glad we have had the chance to celebrate this achievement together.

Too much time on my hands,
I got you on my mind
Can't ease this pain so easily
When you can't find the words to say,
hard to make it through another day
And it makes me wanna cry,
throw my hands up to the sky

So, understand
Don't waste your time always searching for those wasted years
Face up, make your stand
Realize you're living in the golden years.

Adrian Smith (1986)

RESUMO

Radiação de Fundo da Internet (*Internet Background Radiation* - IBR) é o termo usado para descrever o tráfego não-solicitado que chega à computadores diretamente conectados à Internet. A IBR é de baixa intensidade, mas incessante. Ela chega a dispositivos para os quais um ou mais endereços IP públicos foi atribuído. O conteúdo típico achado em amostras dessa radiação inclui erros de configuração de sistemas e serviços de rede, atividades de varredura benigna como a realizada por plataformas de censo, enumeração de computadores e portas abertas realizada por criminosos, tráfego de retroespalhamento (*backscatter*) e propagação de *malware* e de *botnets*. Por mais de vinte anos, pesquisadores de cibersegurança dedicados à área de medição de tráfego usam as amostras de radiação de fundo para detectar, quantificar e qualificar o surgimento de ameaças em escala global, do Blaster ao Conficker, à ascensão de botnets que exploram a insegurança inerente aos dispositivos que perfazem a Internet das Coisas. O mecanismo padrão para coletar radiação de fundo é o Telescópio de Nuvem: um computador habilitado para capturar de forma promíscua qualquer evento de rede conforme estes chegam ao bloco de endereço IP atribuído a suas interfaces, gravando em formato PCAP para análise posterior. Os Telescópios de Rede são limitados a regiões específicas dos institutos que hospedam esse mecanismo. Essa limitação é física e também lógica, quando considera-se as regras de concessão de endereços IP. Esses fatores somados à escassez de endereços IPv4 exerce pressão para novos mecanismos de captura sejam pesquisados. Essa tese documenta o *Cloud Telescope* (Telescópio de Nuvem): uma arquitetura efêmera, distribuída e nativa de nuvem para coleta da Radiação de Fundo da Internet. O Telescópio de Nuvem é efêmero: pode ser operacionalizado por qualquer pesquisador que deseje lançar seu próprio experimento de captura. É distribuído por natureza, dado que tira proveito da capilaridade geográfica de um provdor de serviços de computação em nuvem. É também cloud-native: descrito na linguagem Terraform e está pronto para operacionalização na Amazon Web Services, onde captura radiação chegando simultaneamente à mais de 26 regiões e países do mundo. Este trabalho discute os resultados alcançados com três lançamentos do Telescópio de Nuvem. Um para validar as decisões de projeto e implementação e que resultou em uma análise de 21 milhões de eventos. Um segundo para explorar o comportamento de um respondedor ativo, do tipo *honeypot*, que resultou em 10 bilhões de eventos maliciosos, muitos dos quais revelando padrões de propagação da *botnet* Mirai. O terceiro, de longo prazo, perfazendo cinco meses, foi usado para uma análise profunda da distribuição do tráfego, com agregações por camada de rede e de transporte como fundamento para determinar possíveis padrões de influência geopolítica como motivador para certos ataques. Essa tese resultou em quatro publicações intermediárias em conferências, uma extensiva publicação em *journal*, três datasets publicados, duas apresentações em conferências de mercado e dezenas de trabalhos de conclusão de curso em Ciência da Computação e Cibersegurança em três instituições de ensino no Brasil e Noruega, além de duas dissertações de mestrado na Universidade Federal de Santa Catarina, em andamento.

Palavras-chave: Medição de tráfego da Internet. Análise de Tráfego de Rede. Telescópios de Rede. Telescópio de Nuvem.

ABSTRACT

Internet Background Radiation (IBR) is the term used to describe the unsolicited portion of traffic arriving to hosts directly connected to the Internet. IBR is low-traffic, but incessant. It arrives to any computer or host for which one or more public IP address has been assigned. Typical contents found in radiation samples include misconfiguration of systems and services, benign scanning activities as performed by Internet census platforms, malicious host and port enumeration scanning attempts performed by unknown actors, backscatter traffic, malware propagation and botnet spreading activity. The cyber security research community focused on internet measurements has been active for more than twenty years analysing radiation samples that can help detect, quantify and qualify the rise of Internet scale threats ranging from Blaster, to Conficker, to the rise of planetary scale botnets exploiting the insecurity inherent to billions of Internet of Things devices. The standard mechanism to collect Internet Background Radiation is the Network Telescope: a computer enabled to promiscuously record PCAP-formatted network events as openly arriving to the network address block assigned to its interfaces. Network Telescopes are inherently bound to a single geographical region of the world, both physically and in terms of logical IPv4 address allocation. This combined with the undesired homogeneity, limited IP address coverage and general IPv4 address scarcity, configure the motivation to evolve the mechanism used for collecting radiation samples. This thesis documents the Cloud Telescope: an ephemeral, distributed, cloud-native architecture for collecting Internet Background Radiation. The Cloud Telescope is ephemeral, and can be deployed by any researcher willing to run their own experiments. It is distributed by nature, as it takes advantage of the geographical capillarity of a cloud computing provider. The Cloud Telescope is also cloud-native, and exists as an automated Terraform Infrastructure-as-Code artefact ready to work on more than twenty-six regions and countries covered by Amazon Web Services. This work discusses the results achieved with three Cloud Telescope deployments. One to validate the design and implementation decisions, which resulted in 21 million events. A second to explore the custom honeypot-like behaviour as implemented by active responders, resulting in a dataset of 10 billion malicious events, much of which revealing Mirai botnet propagation patterns. The third, long-term capture, lasting five months, was used for a deep dive into traffic distribution, network-layer and transport-layer aggregations and as an attempt to reveal the geopolitical influence motivating malicious activity. This thesis resulted in four conference publications, one extensive journal paper publication, three published datasets, two industry conference speeches, dozens of computer science and cyber security undergraduate investigations and two master theses underway. **Keywords:** Internet Measurement. Traffic Analysis. Network Telescopes. Cloud Telescope.

RESUMO EXPANDIDO

Esta tese está organizada em oito capítulos, individualmente resumidos nesta seção.

Introdução

O primeiro capítulo apresenta o campo de Medição de Tráfego e explica a importância de caracterizar o conteúdo da Radiação de Fundo da Internet (Internet Background Radiation - IBR), as questões de cibersegurança que podem ser respondidas com esses estudos e as abordagens atuais para coleta e análise de seu conteúdo. O objetivo geral desse trabalho é fundamentado sob a perspectiva de aquisição de inteligência sobre ameaças cibernéticas a partir de uma nova arquitetura geograficamente distribuída, nativa de nuvem, denominada Telescópio de Nuvem (Cloud Telescope). Três objetivos específicos são documentados: um levantamento do estado da arte em mecanismos de coleta da IBR; o projeto e implementação do Telescópio de Nuvem, que coleta radiação junto a um provedor de serviços de nuvem; e uma avaliação do mecanismo proposto utilizando métodos quantitativos e qualitativos. Isso resultou em um plano de pesquisa separado em três fases, iniciado em 2023 e finalizado no primeiro semestre de 2025, desdobrado em 17 passos metodológicos, incluindo uma revisão sistemática da literatura, uma versão preliminar do mecanismo de captura, um experimento prova de conceito, um experimento em escala completa, um experimento de longo prazo, um fluxo de trabalho para processamento dos dados coletados, uma discussão preliminar de resultados, cinco subprodutos de pesquisa, a escrita da tese e três contribuições técnicas: um artefato de Infraestrutura como Código em Terraform, os datasets resultantes em formato PCAP e o ferramental para enriquecimento de dados apoiado por fontes de inteligência sobre ameaças cibernéticas.

Fundamentação Teórica (Background)

A seção de fundamentação teórica provê uma descrição de cada conceito que precisa ser entendido antes da apresentação do Telescópio de Nuvem enquanto mecanismo relevante para este campo de pesquisa. A Radiação de Fundo da Internet é caracterizada sob uma perspectiva histórica, seu conteúdo típico é ilustrado e comparativamente discutido com base em trabalhos de referência no assunto. O Telescópio de Rede é então apresentado como o método padrão para aquisição da IBR. Suas propriedades gerais são especificadas em torno do nível de interatividade oferecido (passiva, baixa, média e completa), seguida de uma discussão do tamanho de abertura, contiguidade de endereçamento IP, juntamente com as questões de duração do experimento e requisitos de armazenamento. A tese propõe quatro limitações principais aplicáveis aos Telescópios de Rede que justificam a necessidade de pesquisa para avançar o estado da arte: 1. Escassez de endereços IP, 2. Homogeneidade de endereços, 3. Monitoramento estático e 4. Influência geopolítica tendenciosa. A computação em nuvem é então formalizada como abordagem viável para hospedar uma variante geograficamente distribuída do Telescópio de Rede, nomeada Telescópio de Nuvem.

Revisão da Literatura (Literature Review)

Este capítulo inicia com uma discussão do procedimento de revisão sistemática da literatura empregado para identificar os trabalhos relacionados, seguida de uma discussão de onze experimentos baseados em Telescópio de Rede, suportando e contrastando com a variante aqui proposta. Esses trabalhos foram usados para suportar decisões fundamentais do projeto do Telescópio de Nuvem, influenciando seu projeto, implementação, propósito e resultados esperados. A questão principal de pesquisa é apresentada como “É relevante coletar inteligência

sobre ameaças cibernéticas, na Radiação de Fundo da Internet, por meio de uma arquitetura geograficamente distribuída baseada em nuvem?”. Então, é apresentada uma estratégia contendo os quatro passos utilizados para direcionar a seleção de palavras-chave que permitem a reconstrução de uma linha de tempo contendo as contribuições mais importantes na área de medição e análise de tráfego de rede com o uso de telescópios de rede. A estratégia de seleção é então avaliada com um segundo conjunto de quatro critérios desdobrando em análise de trabalho seminais, palavras-chave, terminologias sinônimas e a definição do uso do operador OR em determinadas situações, resultando no seguinte texto de consulta: “internet background radiaton” OR “internet background noise” OR “network telescope” OR “network telescopes” OR “darknet” OR “darknets”. Os repositórios de pesquisa priorizados são três: ACM Digital Library, IEEE Xplore e Science Direct, resultando em 1.972 textos desde o ano de 2022. A seleção de artigos e o procedimento de *snowballing* são descritos com concentração em sete trabalhos históricos seminais, combinados com quatro trabalhos recentes, que foram documentados em detalhe e então contrastados e sumarizados em uma tabela comparativa. O capítulo termina com um mapeamento e discussão de três esforços de pesquisa concorrentes que surgiram logo antes ou durante o desenvolvimento do Telescópio de Nuvem, com motivação similar, porém objetivos e comportamento diferentes do mecanismo apresentado neste texto.

Telescópio de Nuvem (Cloud Telescope)

Este quarto capítulo contém o projeto e a implementação do Telescópio de Nuvem. O projeto é fundamentado sob seis argumentos: expectativas do método distribuído, vantagens da coleta simultânea de radiação de fundo em múltiplas regiões do mundo, reuso de endereços IPv4, priorização de baixo custo de coleta, relevância estatística de telescópios de menor abertura e a natureza efêmera e reproduzível da abordagem baseada em código. O projeto do sensor e da frota de sensores é explicado do ponto de vista de uma das regiões de nuvem, incluindo seus prerequisites, como o gateway de rede, abertura do grupo de segurança, configuração da Rede Virtual Privada, o uso de scripts de automação das tarefas desempenhadas por cada nó e o uso de um único repositório de armazenamento, *bucket*, localizado na região de nuvem da Suécia. Há também uma discussão sobre a seleção da instância computacional T3.nano sob o regime de precificação *spot* e a necessidade de configurar o sistema operacional para reagir ao evento de término de instância, sempre que o sensor contratado é vencido pelo sistema de leilão. Então, oito requisitos funcionais são apresentados para documentar os principais recursos implementados durante as fases de programação e de lançamento, juntamente com cinco requisitos não funcionais e uma regra de negócio acessória. Há ainda uma discussão do caráter ético das capturas de amostras da radiação de fundo da Internet, onde pondera-se riscos e benefícios associados a esse procedimento e onde argumenta-se que a natureza essencialmente passiva do experimento, combinada à uma coleta essencialmente centrada nos endereços IP de origem e destino dos eventos, não incorre em risco de exposição de informações pessoalmente atribuíveis a indivíduos. A arquitetura do Telescópio de Nuvem é então formalmente explicada, seu código em Terraform é brevemente apresentada dentro do conceito de infraestrutura como código, juntamente com instruções de como ela pode ser lançada diretamente do GitHub por qualquer pesquisador interessado. Há também o foco nas decisões de ordem técnica e de implementação: escolha do sistema operacional Alpine Linux, preparo e uso de uma versão servidora do programa de coleta tcpdump, automação da obtenção da lista de regiões de nuvem disponibilizadas pelo provedor, imposição do esquema de cobrança *spot*, definição do número de sensores por região, definição dos volumes de armazenamento de dados, acesso emergencial por ssh, definição das chaves criptográficas de acesso e instruções de lançamento no provedor de serviços de nuvem Amazon Web Services. O orçamento de referência é apresentado e discutido com base em estimativas pré-lançamento e contabilização pós-lançamento, revelando que

uma frota de 130 sensores espalhados por 26 regiões do mundo incorre no custo mensal de 345 dólares americanos. O capítulo finaliza com uma comparação entre Telescópios de Rede e o Telescópio de Nuvem centrada nos aspectos operacionais de lançamento, cobertura de endereços IP, capilaridade da coleta, requisitos computacionais e de armazenamento, bem como uma avaliação da transição de custeio baseado em despesa de capital para despesa operacional.

Resultados Preliminares (Preliminary Results)

Este capítulo apresenta o resultado intermediário gerado pelos subprodutos de pesquisa antes do experimento principal. Ele descreve os quatro artigos preliminares e contém uma caracterização dos três datasets resultantes dos lançamentos do Telescópio de Nuvem. O primeiro artigo descreve a proposta de um *framework* nativo de nuvem para coleta geograficamente distribuída e subsequente análise da Radiação de Fundo da Internet, onde uma primeira versão da arquitetura foi apresentada na forma de um texto curto, de quatro páginas, visando obtenção de críticas ou feedback metodológico para aperfeiçoamento do mecanismo antes de sua operacionalização efetiva. O segundo artigo documenta o experimento de prova de conceito, que capturou 21 milhões de eventos de rede e que foi analisado e usado para a publicação que introduz formalmente o Telescópio de Nuvem enquanto arquitetura distribuída para captura de Radiação de Fundo da Internet. O formato empregado nesse texto herda a abordagem dos trabalhos relacionados, em que se descreve a composição global do tráfego, seguida de análise TCP, UDP, ICMP e o ranqueamento dos países que mais emitem tráfego não solicitado. Ao método, é adicionada a análise dos destinos mais atacados, que é possível em razão da natureza geograficamente distribuída da arquitetura. O terceiro texto documenta o segundo experimento, em que se emprega comunicação interativa, honeypots, em portas selecionadas, resultando em 10 bilhões de pacotes gerados em apenas 45 dias de captura por uma frota com 260 sensores, dez por região de nuvem, abordando detalhamentos de camada de aplicação que revelam intensa atividade causada pela botnet Mirai e suas variantes Mozi e Sora. O quarto texto contém o terceiro experimento, onde o telescópio foi lançado de forma passiva para capturar uma amostra da IBR de longo prazo, documentando detalhadamente tanto o método quando os resultados que ele produz, servindo de base para o capítulo seguinte.

Experimento de longo prazo (Long-term experiment)

Este capítulo contém uma discussão detalhada dos resultados alcançados com o experimento principal de lançamento do Telescópio de Nuvem. Ele começa com uma explicação de como o experimento principal foi configurado para cobrir vinte e seis regiões de nuvem da Amazon Web Services, com uma frota de 130 sensores, durante o período de cinco meses, ou 151 dias, iniciando em 1 de outubro de 2023. O dataset resultante é documentado contendo 530 milhões de eventos de rede e com um pico de atividades em novembro onde seis milhões de eventos atingiram a frota de sensores em um único dia. A estratégia usada para análise de dados inclui o ranqueamento das origens da radiação, revelando que os Estados Unidos da América figuraram como o emissor mais prevalente, seguido do Reino Unido e China, enquanto a análise individualizada dos protocolos TCP, UDP e ICMP mostrou atividade relativamente intensa originada também na Bulgária, Suécia e Holanda. A análise centrada em portas de destino TCP confirmou que a porta 23/TCP permanece sendo a mais ativamente testada. No UDP, as portas mais intensamente testadas são a 5900 (VNC), 5060 (SIP) e DNS (53), enquanto a análise ICMP mostrou prevalência de pacotes contendo requisições do tipo 8 (echo request). A segunda metade do capítulo apresenta novos tipos de análise resultantes da abordagem distribuída. A radiação é quantificada e avaliada conforme afetando diferentes regiões de nuvem do mundo, onde os Estados Unidos da América figuram como território mais testado, seguido do Bahrein e Indonésia,

enquanto o detalhamento TCP/UDP/ICMP demonstra tráfego substancial atingindo também o Japão, Índia, Suécia, África do Sul e outros.

Discussão (Discussion)

O capítulo de discussão avalia o Telescópio de Nuvem do ponto de vista qualitativo, como implementado e em contraste às perguntas de pesquisa originalmente formuladas. A discussão confirma o crescimento sustentado na geração de radiação ao longo dos anos. O Telescópio de Nuvem capturou 27 mil pacotes por hora, por endereço IP, 68% a mais do que qualquer outro experimento anterior. O tráfego TCP confirma prevalência de ataques voltados à Internet das Coisas (Internet of Things – IoT), seguido de outros padrões de ataque, incluindo rajadas destinadas às portas 10000-10002 que carregam significado especial e deve fomentar desdobramentos de pesquisa futuros. A discussão do UDP é apresentada contrastando tendências emergentes na exploração do VNC, SIP e DNS em oposição ao RPC. O comportamento de tráfego ICMP é discutido como errático e concentrado em atividades de pico vistas a partir da segunda metade de novembro até o início de dezembro. Uma investigação especulativa é então apresentada mostrando que esse tráfego está ligado a eventos geopolíticos motivados fora do ciberespaço. Em seguida, discute-se as implicações da natureza distribuída das capturas, onde os países mais testados são comparados com os menos testados, revelando discrepâncias de até 200%. Ataques de amplificação do NTP são usados para mostrar que é possível aprofundar-se em protocolos de interesse específicos, assim como é feito nos trabalhos relacionados. Então a natureza efêmera do Telescópio de Nuvem é reavaliada em propriedades chave, como os gastos operacionais, o tempo de captura, reuso e reciclagem de endereços IP e em comparação com os trabalhos que usam o Telescópio de Nuvem da CAIDA/UCSD. O design nativo de nuvem é apresentado em sua forma final, seguido de uma discussão das bibliotecas, ferramentas e tecnologias empregadas na sua arquitetura. Há então uma discussão das particularidades da fase de coleta de radiação, do processamento dos arquivos PCAP e da capacidade de se inferir consultas, bem com um detalhamento da geração de saída CSV, saída como tabelas formatadas para o $\text{\LaTeX}$, o uso do Bing Maps para geração de mapas de calor e as possibilidades futuras com o uso do Security Onion, Seaborn e Neo4J. O capítulo termina com um estudo especulativo que compara a radiação de fundo afetando os Estados Unidos da América em comparação ao resto do mundo, conforme observado pelo Telescópio de Nuvem. Ele introduz um método onde define-se um país de interesse e utiliza-se o dataset para geração das tabelas sumarizadoras do tráfego que o afetam, para então compará-las ao tráfego de referência, composto por todos os países restantes. A partir disso são feitos exercícios de suporte à análises que evidenciam motivação geopolítica e cujas propriedades, países de origem, países de destino, temporalidade e tipo de tráfego podem ser atribuídas a ataques conforme descrito no noticiário. No caso exercitado, revela-se que o Bahrein e a Indonésia foram alvos de ataque de recusa de serviço em datas idênticas ao noticiado em relação às campanhas do coletivo *Anonymous* em resposta ao conflito entre Hamas e Israel, iniciado durante a vigência do experimento principal.

Conclusão (Conclusion)

Este último capítulo conclui o esforço de pesquisa. Cada capítulo anterior é revisitado com uma reflexão dos ganhos obtidos em cada etapa prevista no plano de pesquisa. Uma conclusão é feita em relação a originalidade do projeto, a hipótese da pesquisa e em suporte a cada uma das questões de investigação inicialmente estabelecidas, em especial, relativa às limitações dos Telescópios de Rede como motivação para projetar e implementar o novo Telescópio de Nuvem. A revisão da literatura é então revisitada para avaliar a influência dos trabalhos relacionados, tanto em relação aos formatos utilizados para reportar resultados, quanto sobre os

resultados em si, comparativamente. O aspecto ético da coleta de dados também é brevemente revisto para sustentar que o conteúdo dos datasets de fato não inclui informações pessoais identificáveis. Conclusões são então apresentadas sobre como a infraestrutura definida por código do Telescópio de Nuvem pode ser considerada mais democrática, seguida de uma contabilização dos subprodutos de pesquisa alcançados, incluindo um artigo resumido e três artigos completos apresentados em conferências, um artigo completo em journal, três datasets abertos, uma monografia de bacharelado dedicada à implementação do telescópio, um fluxo de trabalho extrair-transformar-carregar que dá suporte às análises, uma dissertação de mestrado na UFSC, 15+ projetos de bacharelado na Noroff University College, duas apresentações em conferências de indústria: You Shøt the Sheriff e BSides Las Vegas e uma série de workshops de análise de tráfego e colóquios realizados no Brasil e na Noruega. Esta tese finaliza com a indicação de doze trabalhos futuros propostos em três áreas de concentração: expansão da captura, enriquecimento dos datasets e indexação de pacotes para suporte a novas análises.

Palavras chave: Medição de Tráfego. Análise de Tráfego. Telescópios de Rede. Telescópio de Nuvem.

LIST OF FIGURES

Figure 1 – Sample of Internet Background Radiation.	37
Figure 2 – Network Telescope Topology.	38
Figure 3 – Results of Pang.	49
Figure 4 – Results of Wustrow.	51
Figure 5 – Results of Irwin.	52
Figure 6 – Results of Benson.	54
Figure 7 – Results of Shaikh.	56
Figure 8 – Results of Zakroum.	58
Figure 9 – Results of Zolotykh.	59
Figure 10 – Results of Kori.	60
Figure 11 – Results of Anh and Nakamura.	61
Figure 12 – O’Hara’s cloud-native capture mechanism.	65
Figure 13 – Pauley’s architecture for DScope.	66
Figure 14 – García-Peñas’ architecture for HoDiNT.	68
Figure 15 – Cloud Telescope Architecture	74
Figure 16 – Terraform compute instance launch definition.	80
Figure 17 – The Cloud Telescope as built	82
Figure 18 – Infection attempt captured by the responder on port 23	100
Figure 19 – Architecture of the Final Experiment.	111
Figure 20 – Temporal traffic distribution.	112
Figure 21 – Top radiation sources	114
Figure 22 – TCP radiation source countries	116
Figure 23 – UDP radiation source countries	117
Figure 24 – ICMP radiation source countries.	118
Figure 25 – Top radiation destinations.	125
Figure 26 – TCP radiation target countries.	127
Figure 27 – UDP radiation target countries.	130
Figure 28 – ICMP radiation target countries.	132
Figure 29 – Hourly network events per allocated IP address.	135
Figure 30 – Daily traffic distribution targeting TCP ports 10000,10001 and 10002.	137
Figure 31 – Daily traffic distribution targeting prevalent UDP ports.	138
Figure 32 – Daily traffic distribution targeting prevalent UDP ports.	138
Figure 33 – Daily traffic distribution of prevalent ICMP types.	139
Figure 34 – NTP amplification attack exploiting CVE-2013-5211.	141
Figure 35 – Delving into the ICMP spike.	142
Figure 36 – The Cloud Telescope work flow.	144
Figure 37 – Towards a fully cloud-native analysis workflow	145
Figure 38 – Data Extract-Transform-Load workflow.	146

Figure 39 – ICMP type 8 traffic patterns. 155

Figure 40 – Compared background radiation protocol breakdown 158

CONTENTS

1	INTRODUCTION	29
1.1	RESEARCH AIM	32
1.2	RESEARCH GOALS	32
1.3	DOCUMENT STRUCTURE	33
2	BACKGROUND	35
2.1	NETWORK TELESCOPE	37
2.2	NETWORK TELESCOPE LIMITATIONS	39
2.3	CLOUD COMPUTING	41
2.4	SUMMARY OF CONCERNS	43
3	LITERATURE REVIEW	45
3.1	RESEARCH QUESTION	45
3.2	KEYWORD SELECTION STRATEGY	45
3.3	RELATED WORK SELECTION STRATEGY	46
3.4	SELECTION OF RELATED WORKS	47
3.5	MOORE’S CHARACTERISATION OF IBR	48
3.6	PANG’S IBR CAPTURE AND ANALYSIS	48
3.7	WUSTROW’S LONG-TERM CAPTURE	50
3.8	IRWIN’S TELESCOPE AT RHODES UNIVERSITY	52
3.9	BENSON’S FRAMEWORK	53
3.10	SHAIKH’S CORRELATION: THE IOT AFFECTS IBR	55
3.11	CHINDIPHA ON SMALL-APERTURE TELESCOPES	56
3.12	ZAKROUM’S PREDICTION MODEL	57
3.13	ZOLOTYKH’S LOOK FROM THE RUSSIAN SEGMENT	58
3.14	KORI’S ANALYSIS OF AN ACTIVE NETWORK TELESCOPE	59
3.15	ANH AND NAKAMURA ON MIRAI AND HAJIME	60
3.16	DISCUSSION OF RELATED WORKS	61
3.17	CONCURRENT RESEARCH EFFORTS	63
3.18	O’HARA AND WEBER’S CLOUD-BASED NETWORK TELESCOPE	64
3.19	PAULEY’S DSCOPE	66
3.20	RODOLFO GARCÍA-PEÑAS’ HODINT	67
3.21	COMPARATIVE ANALYSIS OF CONCURRENT WORKS	69
4	CLOUD TELESCOPE	73
4.1	CLOUD SENSOR DESIGN	75
4.2	FUNCTIONAL REQUIREMENTS	77
4.3	NON-FUNCTIONAL REQUIREMENTS	77

4.4	BUSINESS RULES	78
4.5	ETHICAL CONSIDERATION	78
4.6	ARCHITECTURE IMPLEMENTATION	79
4.7	DEPLOYING AND REPRODUCING THE CLOUD TELESCOPE	81
4.8	BASELINE BUDGET	82
4.9	NETWORK TELESCOPE AND CLOUD TELESCOPE COMPARED	83
5	PRELIMINARY RESULTS	87
5.1	EXPERIMENTAL RESULTS	88
5.2	EXPERIMENT #1: PROOF-OF-CONCEPT	88
5.2.1	Global Traffic Composition	90
5.2.2	TCP Traffic Composition	90
5.2.3	UDP Traffic Composition	91
5.2.4	Most active sources	92
5.2.5	Geographic Distribution	92
5.2.6	Experiment #1 Conclusion	93
5.3	EXPERIMENT #2: FULL-SCALE, INTERACTIVE CAPTURE	95
5.3.1	Global Traffic Composition	96
5.3.2	Top TCP Ports	97
5.3.3	Top UDP Ports	98
5.3.4	Top radiation sources	98
5.3.5	Global Radiation Distribution	99
5.3.6	Port 23 Analysis	100
5.3.7	Port 80 Analysis	101
5.3.8	Experiment #2 Conclusion	103
5.4	EXPERIMENT #3 OUTLINE	104
5.5	DATASET OUTPUT	105
5.5.1	Dataset #1	105
5.5.2	Dataset #2	106
5.5.3	Dataset #3	107
5.6	ACCESSORY RESULTS	107
6	LONG-TERM EXPERIMENT	111
6.1	EXPERIMENT SETUP	111
6.2	RESULTING DATASET	112
6.3	RADIATION SOURCES	113
6.3.1	TCP Radiation sources	115
6.3.2	UDP Radiation sources	115
6.3.3	ICMP Radiation sources	117
6.4	RADIATION TARGETS	117

6.4.1	Most probed TCP ports	119
6.4.2	TCP flag combination analysis	120
6.4.3	Most probed UDP ports	121
6.4.4	Most frequent ICMP messages	123
6.5	DISTRIBUTED ANALYSIS	123
6.5.1	Overall traffic dispersion	124
6.5.2	TCP radiation: target countries	125
6.5.3	TCP radiation: target cloud regions	126
6.5.4	UDP radiation: target countries	128
6.5.5	UDP radiation: target cloud regions	129
6.5.6	ICMP radiation: target countries	130
6.5.7	ICMP radiation: target cloud regions	130
7	DISCUSSION	135
7.1	TCP TRAFFIC	135
7.2	UDP TRAFFIC	137
7.3	ICMP TRAFFIC	139
7.4	IMPLICATIONS OF THE DISTRIBUTED APPROACH	139
7.5	IMPLICATIONS OF THE EPHEMERAL METHOD	143
7.6	IMPLICATIONS OF THE CLOUD-NATIVE DESIGN	144
7.7	EXPERIMENTING GEOPOLITICAL INFLUENCE	145
7.7.1	TCP analysis	147
7.7.2	UDP analysis	150
7.7.3	ICMP analysis	151
7.8	CLOSING REMARK	157
8	CONCLUSION	159
8.1	ASSUMPTIONS	160
8.2	ETHICS IN CLOUD IP ADDRESS REUSE	162
8.3	THE CLOUD TELESCOPE AS <i>CODE</i>	163
8.4	RESULTS	164
8.5	FUTURE WORK	167
8.5.1	Capture expansion	167
8.5.2	Dataset enrichment	168
8.5.3	Packet indexing and analytics	169
	BIBLIOGRAPHY	171

1 INTRODUCTION

For more than twenty years, a reasonably small, yet persistent group of cyber security researchers have been looking into ways to answer one inspirational question: “What threats affect the internet at large?”. Moreover, what kind of cybernetic activity would be present in unsolicited packets arriving to computers publicly facing the Internet? Ultimately, how can this knowledge contribute to a safer Internet?

The area in computer science aiming at finding answers to such questions is *Internet Measurement*. This field focuses on the collection, analysis, and interpretation of data that quantitatively describes the structure, performance, usage, and behaviour of the Internet and its underlying networks (Ziviani, 2006). It discusses a range of methods and tools designed to make the otherwise opaque workings of the Internet observable and understandable, supporting both academic research and practical network management, providing input that can be used to shape next generation of network appliances, packet filters and intrusion detectors.

Research efforts in Internet Measurement intersect two computer science domains: computer networks and cyber security. These are backed with basic statistics to measure five main network properties:

- Network topology: Arrangement and interconnection of routers, switches and long distance links;
- Routing dynamics: How routes change over time , including the impact of failures or policy changes;
- End-to-End performance: The experience of users or applications, supported by sending test traffic;
- Traffic flows: Aggregated or per-flow statistics on data movement, including protocol usage and application-layer behaviour; and
- Security events: In alignment with this research, this refers to the quantitative analysis of events carrying some sort of malicious activity combined with qualitative methods that could explain inner-workings and external motivation.

The available methods to conduct research into quantitative-qualitative analysis of security events vary, but can be summarised into three main approaches, in accordance to how much noise is generated by the introduction of the data collection probe in the topology.

- Passive Measurement: Observing existing traffic at strategic points in the network without introducing additional load;
- Active Measurement: Injecting test traffic or answering to incoming requests and measuring these interactions to infer network and security properties they carry;

- **Hybrid Measurement:** Combining active and passive techniques to extract more comprehensive insights.

Choosing the ideal data collection approach, especially for the measurement of security events, constitutes a challenge in terms of scalability, coverage, accuracy and privacy. The size of the internet, combined with its decentralized nature, turns the measurement of relevant malicious activity particularly difficult as data collection is biased towards the region being measured, and there is no *de jure* standard to ensure minimal parameters of tool calibration, data quality, interoperable output format, and result normalization. Research is therefore needed in Internet Measurement to advance the understanding of the Internet, its behaviour, evolution and dominant security risks. Research activity can also focus on network protocol improvement, and new communication technologies to promote enhanced security, reliability and better user experience world-wide. This is ultimately a rigorous data-driven discipline, aimed at making the Internet's structure, performance and usage transparent, using both scientific inquiry and practical improvements to evolve networked systems.

Among the types of network events that can be measured, one is of particular interest to cyber security researchers: unsolicited packets. These refer to the principle that any IP address, associated or not with productive traffic, will be constantly targeted by probing activity, and there is a chance the incoming traffic carries some sort of malicious purpose, such as exploiting security breaches on the target. The common name used to refer to this kind of traffic is Internet Background Radiation.

Internet Background Radiation therefore can be described as the non-productive, often malicious portion of traffic exchanged on the public Internet (Bortoluzzi; Irwin; Westphall, 2025). This traffic arrives in small or large portions to any computer or device assigned with one or more IP addresses. Most devices, especially the ones implementing network address translation, such as domestic routers, ignore this traffic as it won't match locally-initiated connections, freeing these users from most of the threats it imposes. One question however remains: "What content do such packets carry?".

As it is typical of the scientific method, the interest to characterize Internet Background Radiation traffic led to the development of tools and mechanisms meant to capture its propagation and shed light on its properties. On this perspective, the Network Telescope arose as the mechanism designed to capture the radiation (Pang et al., 2004). It is implemented as a computer system, enabled to capture and record network packets routed to it. The Network Telescope uses commodity hardware combined with a toolset centred on tcpdump to produce pcap-formatted files containing the events of interest that can be later analysed either with tshark and Wireshark or piped through a fancier customised data processing pipeline (Irwin, 2012). Data analysis tends to include port aggregations by protocol type, most actively probed port destination and other forms of protocol breakdowns. Datasets can be enriched with external intelligence providers - libraries and APIs - to determine the geographical source of event generators and source host profiling, among many other forms of open source intelligence tech-

niques.

Much can be learned with the use of Network Telescopes to capture Internet Background Radiation. Since its inception in the early 2000's, radiation analysis contributed to the detection, quantification and qualification of major malware outbreaks, including now ancient SQL Slammer, Code Red and Conficker, aiming to explore vulnerabilities associated with Microsoft's ecosystem, to the rise of modern, distributed botnets exploiting poorly protected IoT devices to enable massive denial-of-service attacks, being Mirai, Mozi and Sora remarkable examples (Torabi et al., 2022).

The Network Telescope operated by the Center for Applied Internet Data Analysis based at the University of California's San Diego Supercomputer Center - CAIDA/UCSD Network Telescope - is the benchmark passive monitoring system especially designed to capture and analyse Internet Background Radiation resulting from misconfiguration, malware, scans, and cyberattacks. Historically operated as a /8 network, it monitors a /9 and /10 IPv4 network, processing 1,25 Mbps of traffic on average, with datasets spanning petabytes of traffic (Irwin; Nkhumeleni, 2015).

CAIDA's Telescope, as much as any other Network Telescope, come with some inherent characteristics, or limitations. Three are of particular concern:

1. Geographical homogeneity. IP addresses are usually allocated in contiguous blocks. This can contribute to reduced detection of globally distributed threats.
2. Predictable IP addresses. IP addresses allocated to Network Telescopes can be discovered. One of the implications is that malicious actors can map large Network Telescopes and avoid their network blocks when designing scanner and subsequent global scale attacks.
3. Address scarcity. IPv4 address blocks are long exhausted and there is no primary allocations available any more. This dramatically reduces the number of researchers with access to the operation of Network Telescopes.

Conversely, cloud computing arose with the promise of pseudo-infinite resource pooling (Mell; Grance, 2011). There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction, such as the country in which virtualised systems are deployed. Any system thought to operate within cloud computing environments can be design to benefit from unique advantages:

- Distributed service availability. Cloud computing is distributed by design;
- Coverage of multiple regions of the world. Cloud service providers operate in multiple countries where there is economic activity to justify an additional point of presence;

- Ephemerality: Deploying systems in the cloud can be done with scriptable actions, which are reproducible by nature. There is no need for physical access to privileged facilities.

This doctoral degree thesis takes into consideration the perceived limitations within standard Network Telescopes, in combination with the potential benefits offered by cloud computing service providers to propose the Cloud Telescope: an ephemeral, geographically distributed and cloud-native architecture for collecting Internet Background Radiation.

1.1 RESEARCH AIM

This research evaluates the hypothesis of using a cloud-native architecture, termed Cloud Telescope, to acquire cyber threat intelligence from Internet Background Radiation.

1.2 RESEARCH GOALS

The research aim is supported by three research goals (RG).

- RG1 To assess the state-of-the art in mechanisms to collect Internet Background Radiation;
- RG2 To design and implement a capture mechanism that can collect Internet Background Radiation within a cloud computing service provider; and
- RG3 To evaluate the cyber threat intelligence that can be learned from the approach.

RG1 refers to the need to systematically review the state-of-the-art in capture mechanisms to collect Internet Background Radiation samples and common practises used by peer researchers to document findings and report results, including data aggregation, data processing and data visualization techniques.

RG2 documents the cloud-native architecture designed to collect Internet Background Radiation. This is presented with the use of standard software engineering artefacts, including functional and non-functional requirements, business rules and cost estimation. The resulting architecture is presented from a reproducibility perspective and implemented as Terraform Infrastructure-as-Code package.

RG3 delves into a quantitative and qualitative discussion of the findings that can be achieved with the mechanism. The discussion is based on three distinct deployments conducted to validate core aspects of the Cloud Telescope.

This research was conducted in three main phases, each aiming at answering or producing input to comply with each predetermined research goal, as follows.

- Phase 1: Assessment of the state-of-the-art in Internet Background Radiation capture mechanisms. This unfolded into the systematic review, a draft version of the capture architecture, and a proof-of-concept experimentation carried out in May 2023;

- Phase 2: The design and implementation of the Cloud Telescope. This focuses on the steps required to describe the cloud-native architecture to collect radiation and the deployment of three capture experiments to validate different properties of its operational capacity; and
- Phase 3: Findings and results. This encompasses a quantitative-qualitative discussion of the results achieved with each of the experiments. Both the method to report results and the results themselves count as the intended results to the research field.

A summary of these research activities is presented on the Activity Schedule in Table 1. It also considers the time taken to producing intermediate research output, in particular, the write-up of four conference articles and one thorough journal paper publications, each focusing on one key contribution to the field. The first covering a methodologically sound, cost-oriented initial version of the architecture prior to actual implementation. The second contains both an architectural description of the Cloud Telescope and a discussion of the preliminary findings, used to validate both the method and the approach to reporting results. The third is in the form of a botnet-centric analysis, revealing distributed patterns employed by Mirai and its variants to self-propagate using vulnerable targets on the Internet. The fourth is in the form of a thorough documentation of the whole project, including a careful inspection of the achievements with the third and long-term capture experiment. All were accepted and published. A fifth was submitted to the flagship event in internet measurement and is under review.

In addition, time was allocated to design and supervise the actual implementation of the Cloud Telescope, in Terraform. Finally, the required data enrichment pipeline was coded by this candidate, to serve as a foundational ground on top of which future research and supervision can be conducted.

1.3 DOCUMENT STRUCTURE

The remainder of this document is organized as follows. Chapter 2: Background, formally presents Internet Background Radiation as a phenomenon, the Network Telescope as a capture mechanism, properties of interest applicable to Cloud Computing and a summary of concerns this research desires to address. Chapter 3: Literature Review, formalizes the research question and the criteria utilized to conduct the systematic review of literature, after which a selection of eleven related works ending in a rich discussion of how they contribute to the decision making process. Chapter 4: Cloud Telescope, explains the design and implementation of the Cloud Telescope. Design properties are described and supported by diagrams and a collection of three software engineering artefacts discussing functional requirements, non-functional requirements and business rules. This is followed by insights into ethical considerations towards user privacy and how the architecture was actually implemented in Terraform. The chapter also compares Network Telescopes and the proposed Cloud Telescope, ending with a presentation of budget and baseline costs associated with its deployment within a cloud computing service

provider. Chapter 5: Preliminary Results, summarises the results achieved with preliminary deployments of the Cloud Telescope, focusing on two intermediate experiments and the characterization of the two generated datasets to validate elementary assumptions of its operation before the final launch. Chapter 6: Long-Term Experiment, delves into the presentation of detailed results achieved with the main experiment. This unfolds into five sections: Experiment Setup, Resulting Dataset, Radiation Sources, Radiation Targets and the newly introduced Distributed Analysis, which depicts how different regions of the world are affected by Internet Background Radiation. Chapter 7: Discussion, provides an analysis of selected properties from the previous chapter to discuss qualitative aspects of TCP/UDP/ICMP traffic, and the implications of the ephemeral method, combined with the cloud-native design and the enrichment techniques used to build the data analysis pipeline. This chapter also contains one unfolding experiment where traffic destined to the United States of America was compared to the rest of the world. The work ends with Chapter 8: Conclusion, which summarises this research and provides indicators of upcoming unfoldings towards expanding the capture mechanism, enhance datasets, and advance studies in the geopolitical influence motivating Internet Background Radiation generation.

Table 1 – Research Activity Schedule

Research Activity Schedule											
Year	2023				2024				2025		
Quartile	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	
Research Goal 1: State-of-the-art in Internet Background Radiation capture mechanisms											
Systematic literature review	•	•									
Draft architecture		•									
Experiment #1: Proof of concept			•								
Research Goal 2: The design and implementation of the Cloud Telescope											
Experiment #2: Full-scale and interactive			•								
Long-term experimentation				•	•						
Implementation of the data processing pipeline					•						
Research Goal 3: Cyber threat intelligence acquisition											
Discussion of preliminary results						•					
Presentation of long-term results							•	•			
Research output: Publications											
Publication #1: Research intent		•									
Publication #2: Architecture and initial findings				•							
Publication #3: Botnet-centric analysis						•					
Publication #4: Long-term analysis							•	•	•		
Publication #5: Geopolitical analysis									•	•	
Thesis write-up			•	•	•	•	•	•	•	•	
Research by-product: Technical contributions											
Terraform Infrastructure-as-Code		•									
PCAP datasets					•			•			
Data enrichment pipeline			•		•		•				

2 BACKGROUND

The term used to describe the unsolicited portion of traffic targeting hosts directly connected to the Internet is Internet Background Radiation (IBR) (Moore et al., 2004). It is an analogy, or a reference to Cosmic Background Radiation, an electromagnetic radiation that fills the space, including the Cosmic Microwave Background. The IBR is incessant, even though low-traffic. It arrives to computer, host or device for which a public, Internet-facing IP address has been assigned (Irwin, 2011; Izhikevich et al., 2023; Bortoluzzi et al., 2023; García-peñas; Rodríguez-gómez; Maciá-fernández, 2024). Unused, unproductive, and previously unassigned public address ranges are also affected (Wustrow et al., 2010).

The analysis of IBR has been helping both the academic community and commercial venturers to detect, quantify and qualify malicious traffic linked to important threats affecting the Internet globally. Each incoming packet can be considered either benign or potentially malicious, depending on the reasons motivating its generation. An observed network packet is considered benign if it falls under one of the following conditions:

- **Misconfiguration.** This refers to misconfigured systems and network services that attempt to initiate connection to non-existent hosts. It also refers to faulty hardware and software that sends packets to arbitrary destinations or to bad Network Address Translation (NAT) implementations incorrectly mapping RFC 3330 (Iana, 2002) addresses into public ones (Pang et al., 2004).
- **Census platforms.** A number of packets seen in radiation samples can be clearly linked to platforms aiming at conducting continuous *Internet census* assessments. They map the amount of computers currently connected to the public Internet. These platforms usually perform host and port scanning to determine the host's architecture, operating system, up-time, open service ports, and relevant metadata, including source country, domain name, and autonomous system number, among others. Shodan and Censys are two popular examples of platforms profiting from this type of activity (Ilg et al., 2023).

The benign nature is determined when packet inspection is unable to detect malicious intent, even if attackers will make use of census as part of their *Open Source Intelligence* acquisition strategy. Otherwise, it is assumed that the majority of IBR traffic is inherently of malicious nature (Irwin, 2012), in which case it is categorised as follows:

- **Host and port scanning.** One of the most prevalent activities refers to a range of automated tools and techniques employed by attackers to enumerate live hosts potentially exploitable due to known or recently disclosed vulnerabilities (Zhao et al., 2023). The analysis of scanning-related background radiation enables researchers to learn about enumeration patterns, ultimately contributing to the development of more effective firewall rules and intrusion detection signatures to quickly suppress emerging threats;

- **Backscatter.** This is the traffic resulting as the monitored address space being used for spoofing elsewhere, most often as decoy scans, as denial-of-service attacks, or as a result of misconfiguration. Backscatter primarily comprises certain ICMP types, RST-enabled TCP datagrams, or other forms of oddly set TCP flag combination (Irwin, 2011; Pouget; Dacier; Pham, 2008);
- **Malware propagation.** A worm is a particular type of malware capable of propagating itself and infecting new hosts without human intervention. The self-propagating nature of this attack class results in considerable amounts of radiation attributed to malware. Throughout the years, IBR analysis has helped to reveal propagation patterns attributed to CodeRed, Blaster, Sasser, SQL Slammer, Conficker, and many others (Bortoluzzi et al., 2023); and
- **Botnet spreading.** Botnets ultimately aim at causing large-scale denial-of-service attacks for various reasons, ranging from simple hacktivism to business profit. It may carry geopolitical and nation-state interests. The building of a botnet fleet requires the infection of as many hosts as possible, turning them into *zombies*, which will, in turn, accept attack orchestration commands. Botnet infection attempts generate a considerable amount of vestiges in the background radiation and configure a major reason for conducting new IBR analyses. The most prevalent botnet is currently the Mirai family, including variants Moobot, Sora, and Mozi (Bortoluzzi et al., 2024).

A sample of the Internet Background Radiation, as captured by the Cloud Telescope in cloud region sa-east-1, Sao Paulo, Brazil, is depicted in Figure 1.

The figure highlights four interesting event sets as decoded and displayed by Wireshark: Packets in 1 belong to probing activity generated by well-known benign Internet scanners: Censys, BufferOverRun, and the Recyber Project. Intelligence about the origin is gathered by a simple reverse DNS query using the captured public IP source address as the argument. Arrows in 2 point to packets originated in systems within Google Cloud Platform, apparently conducting a distributed-source scan against the cloud telescope sensor node, targeting non-linear enumeration of potentially open TCP ports. The enabled SYN flag indicates the use of the common SYN-scan technique. Activity in 3 reveals continuous Network Time Protocol (NTP) exploitation attempts against one of its implementations, allowing the attacker to achieve an amplification factor of 100x or more. The amplification factor is the metric used to represent the ratio of the server-generated response size in contrast to the number of bytes spent by the attacker to generate the request, as documented since CVE-2013-5211 (NVD, 2025). Lastly, 4 reveals the arrival of RST-flagged TCP datagrams, evidencing typical forged-source backscatter activity.

Figure 1 – Sample of Internet Background Radiation.

Source	Destination	Protocol	Info
www.internet-albedo.net	ec2-15-228-3-135	UDP	
scanner-22.ch1.censys-scanner.com	ec2-15-228-3-135	TCP	
237.149.216.162.bc.googleusercontent.com	ec2-15-228-3-135	TCP	
manila.scan.bufferover.run	ec2-15-228-3-135	SSH	
b.ntp.br	ec2-15-228-3-135	NTP	NTP Version 4, server
89.248.163.57	ec2-15-228-3-135	TCP	
89.248.163.57	ec2-15-228-3-135	TCP	[TCP Retransmission] 58524 → 29424 [RST] Seq=1 Win=1200 Len=6
69.150.216.162.bc.googleusercontent.com	ec2-15-228-3-135	TCP	
94.102.61.42	ec2-15-228-3-135	TCP	
219.211.203.35.bc.googleusercontent.com	ec2-15-228-3-135	TCP	
123.211.203.35.bc.googleusercontent.com	ec2-15-228-3-135	TCP	
area-yxfr.poshflawed.com	ec2-15-228-3-135	TCP	
recyber.net	ec2-15-228-3-135	TCP	
recyber.net	ec2-15-228-3-135	TCP	
b.ntp.br	ec2-15-228-3-135	NTP	NTP Version 4, server
129.150.216.162.bc.googleusercontent.com	ec2-15-228-3-135	TCP	
183.136.225.42	ec2-15-228-3-135	TCP	
220.211.203.35.bc.googleusercontent.com	ec2-15-228-3-135	TCP	
205.210.31.64	ec2-15-228-3-135	TCP	
198.235.24.21	ec2-15-228-3-135	TCP	
99.211.203.35.bc.googleusercontent.com	ec2-15-228-3-135	TCP	
161.210.203.35.bc.googleusercontent.com	ec2-15-228-3-135	TCP	
b.ntp.br	ec2-15-228-3-135	NTP	NTP Version 4, server
141.211.203.35.bc.googleusercontent.com	ec2-15-228-3-135	TCP	
198.235.24.128	ec2-15-228-3-135	TCP	
80.211.203.35.bc.googleusercontent.com	ec2-15-228-3-135	TCP	
199.149.216.162.bc.googleusercontent.com	ec2-15-228-3-135	TCP	
176.111.174.82	ec2-15-228-3-135	TCP	
176.111.174.82	ec2-15-228-3-135	TCP	[TCP Retransmission] 49608 → 1237 [RST] Seq=1 Win=1200 Len=6
scanner-12.ch1.censys-scanner.com	ec2-15-228-3-135	TCP	
45.83.64.57	ec2-15-228-3-135	TCP	
117.211.203.35.bc.googleusercontent.com	ec2-15-228-3-135	TCP	
72.149.216.162.bc.googleusercontent.com	ec2-15-228-3-135	TCP	
100.210.203.35.bc.googleusercontent.com	ec2-15-228-3-135	TCP	
43.154.8.109	ec2-15-228-3-135	TCP	
254.149.216.162.bc.googleusercontent.com	ec2-15-228-3-135	TCP	
204.211.203.35.bc.googleusercontent.com	ec2-15-228-3-135	TCP	
89.248.163.72	ec2-15-228-3-135	TCP	
89.248.163.72	ec2-15-228-3-135	TCP	[TCP Retransmission] 43928 → 17556 [RST] Seq=1 Win=1200 Len=6
scanner-14.ch1.censys-scanner.com	ec2-15-228-3-135	TCP	
116.62.188.148	ec2-15-228-3-135	TCP	
116.62.188.148	ec2-15-228-3-135	TCP	[TCP Retransmission] 52848 → 31390 [RST] Seq=1 Win=1200 Len=6
scanner-22.ch1.censys-scanner.com	ec2-15-228-3-135	TCP	

Source: The author.

2.1 NETWORK TELESCOPE

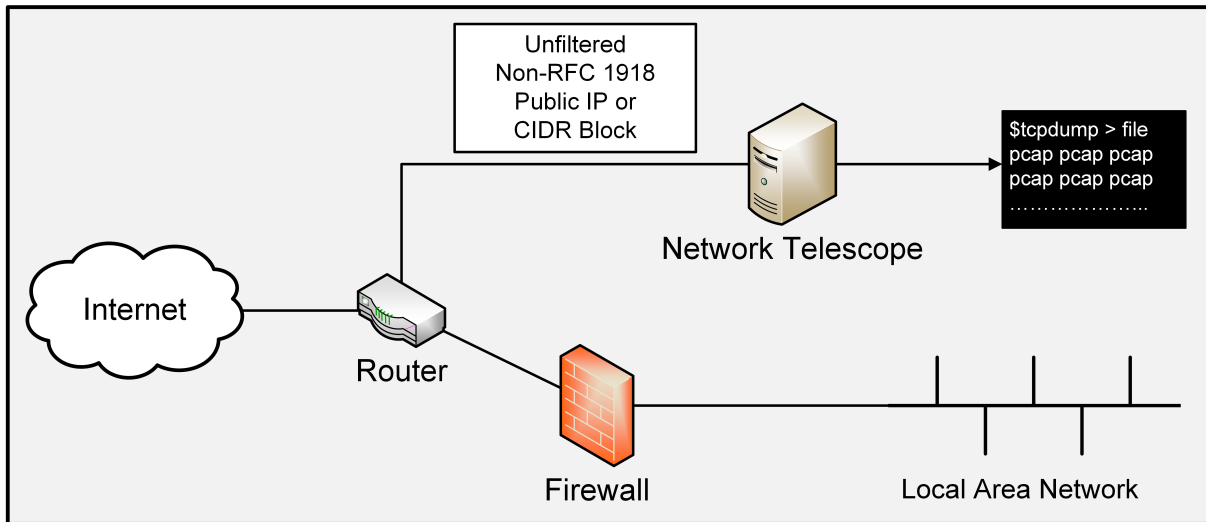
The standard mechanism for Internet Background Radiation collection is the Network Telescope. The Network Telescope is a combination of hardware and software put together to passively receive and record network packets routed to it (Moore et al., 2004; Irwin, 2011; Houmz et al., 2021; Anand et al., 2023). The appliance is strategically inserted onto a network topology to collect unsolicited traffic, as illustrated in Figure 2.

The figure shows the Network Telescope as a standard computer server or workstation to which a range of public IP addresses were assigned. Productive traffic is routed through the firewall towards the Local Area Network. Only the idle portion of the IP address range is actually forwarded to the Network Telescope, which records arriving packets in the PCAP file format, on secondary storage, for later analysis.

Ordinary computer hardware can be used for the deployment of a Network Telescope. The requirement is that the system should be powerful enough to avoid packet drop events: a situation that results from the inability to capture a packet due to a busy system, or due to insufficient memory buffer to handle incoming traffic by the network stack. The literature evidences the use of the following software components (Charan et al., 2023):

- A Unix-like operating system, commonly FreeBSD or Linux;

Figure 2 – Network Telescope Topology.



Source: The author.

- Libpcap-based tcpdump utility, running in promiscuous mode, as a daemon, enabled to record incoming traffic into the PCAP file format;
- Storage facility comprised of a combination of large internal drive units, and external network-attached mount points; and
- Secondary systems used to process the PCAP-formatted recordings, which can include the common Unix toolset, including Bash scripts, TShark, CSV (Comma Separated Values) manipulation tools, and Perl or Python-based transformation pipelines for all sorts of data enrichment procedures.

A Network Telescope is essentially passive. However, it can be tailored to certain interaction levels to enable deeper cyber threat intelligence acquisition, better fitting the behaviour of a Honeypot system. This will affect the Network Telescope's noise - the traffic that will be detected by radiation sources, potentially revealing its existence - and will impact CPU and storage needs. Honeypots can be considered as active-response counterparts to Network Telescopes. Network Telescopes and Honeypots have many characteristics in common, but they vary in the level of interaction implemented, as discussed by (Irwin, 2012).

- **Passive.** The standard operational mode for a Network Telescope. All incoming traffic is recorded and none is answered. This is undetectable and requires the least amount of processing power and storage;
- **Low interaction.** This is deployed with a configuration that will answer the TCP three-way handshake to evaluate how the radiation source behaves right after that event. Once the first payload packet arrives, the established connection is simply dropped by the telescope;

- **Medium interaction.** These implement the standard behaviour of a honeypot. Portions of selected application layer protocols are implemented, such as HTTP's GET method, or FTP's list command, to learn what exploiting actions attackers would take upon successful connection to the target; and
- **Full interaction.** This variant contains a virtualised implementation of intentionally vulnerable services, allowing the observer to record the infection commands taking place beyond transport-layer handshake, i.e. during protocol message exchange and data/payload transfer.

Regarding the infrastructure that is required to operate the Network Telescope, key aspects described in the literature (Chindipha; Irwin; Herbert, 2019) include:

- **Aperture size.** The telescope must be allocated with a public IP address block. The address range can vary from small subnets within a single /24 CIDR block, referred to as a *Small Aperture Telescope*, to one or more /8 (≈ 16 million) address assignments, such as the one operated by CAIDA, in which case, it is called a *Large Aperture Telescope*;
- **Address continuity.** Any combination of multiple IP addresses will work as a source to collect radiation. For example, a full /24 address block containing 256 addresses, as well as the use of multiple /28 subnets within one or more IP ranges. These decisions might, however, affect specific continuity analyses, but are otherwise equivalent in terms of statistical significance;
- **Storage.** The amount of storage required by medium to full interaction honeypots is many times greater than the less interactive approaches. Post-processing tasks also demand more CPU power and will take much longer to complete; and
- **Capture duration.** Related works report a capture length ranging from a few weeks up to a time frame of five years. Capture length is impacted by costs related to running the research-oriented infrastructure and the use of scarce IPv4 addresses for non-productive, non-business-oriented purposes. Duration also impacts the Network Telescope in terms of storage needs, and more importantly, in terms of computing power needed to transform, enrich and load events onto an analytics platform from which meaningful information can be ultimately revealed.

2.2 NETWORK TELESCOPE LIMITATIONS

The Network Telescope is the *de facto* mechanism to collect Internet Background Radiation. The rise of both the phenomenon, (the radiation), and the instrument, (the network telescope), are linked together.

Any conceivable method intended to observe cybernetic activity affecting the Internet at large will necessarily incur some form of limitation, essentially because it is not possible to

observe all existing IP addresses at all times. In principle, it is only possible to observe idle addresses, i.e., which are not currently being put to productive use. In practice, researchers can only observe a very small, ever-decreasing idle fraction of the addressable IP space (Livadariu et al., 2018; Chindipha; Irwin; Herbert, 2019).

Inherent limitations of a Network Telescope must consider the following key points for a discussion:

1. Limited IP address coverage;
2. Homogeneous IP address distribution;
3. Static monitoring locations; and
4. Biased geopolitical influence.

Primary IPv4 allocation is generally no longer available because the address field of the protocol header only accommodates 32 bits (Postel, 1981a). This leads to roughly 4 billion addresses theoretically. This is significantly less than required to uniquely address the 5.4 billion Internet users and 30 to 50 billion devices making up the Internet-of-Things (Petrosyan, 2024). Throughout the years, address exhaustion has been tackled in two ways: IPv6 and NAT. IPv6 extends the routeable address space by orders of magnitude. While it represents the ultimate natural solution for the problem, IPv6 adoption is comparably small and frequently problematic (Richter et al., 2015). Despite community attempts to promote its usage, it demands complex transition mechanisms to ensure compatibility. Alternatively, technical manoeuvres such as Network Address Translators (NAT) and Carrier-Grade NAT have been successfully implemented to cope with exhaustion.

The only way to deploy a new Network Telescope currently is through the acquisition of an after market subnet range, which is either expensive, unachievable, or only feasible for researchers linked to a few institutions willing to share their idle address space for such luxurious, or even questionable use.

Network Telescopes are constrained, bound, and tied to a statically allocated, limited IP address range. Consider the moment a hacker writes their first port scanner to assess how many hosts have an open Remote Desktop Protocol port at 3389/TCP. They are likely to implement the scanner in a sequential fashion, i.e. after probing 203.0.113.20, the next obvious address is 203.0.113.21. Nmap operates the same way as a default, as in `nmap -sS -p 3389 203.0.113.0-255`, which would sequentially scan that whole /24 segment. In case the Network Telescope is configured to listen to the full 203.0.113.0/24 subnet, it would capture the scan attempt completely and sequentially. However, what does that tell about all other port scans taking place beyond the captured block? In this case, the non-contiguous allocation could provide better statistical significance. Non-continuity remains bound to the geographic region, country and continent, and could be biased.

To summarise, the combination of IPv4 address scarcity, ever-shrinking address availability, pressure to put remainder IP addresses to productive use, and the fact idle, observable IPv4 address blocks are only available to few researchers within few institutions worldwide, configure the essential motivation to look at novel, distributed ways of collecting Internet Background Radiation.

2.3 CLOUD COMPUTING

Cloud computing has become a highly viable service model for convenient, on-demand access to compute, storage and network resources globally. The remarkable characteristics of cloud-based systems deployment include (Mell; Grance, 2011; Ponnumani; Partheeban, 2022):

- **On-demand self-service.** The cloud provider should contain enough resources to provide consumers with the perception that they have access to infinite computing power. This includes the need to possess more IPv4 addresses than currently allocated to the clients in each of the Cloud Service Provider (CSP) regions;
- **Broad network access.** The Internet stands as the common channel to both service configuration and service access, regardless of how powerful the client side is, ranging from mobile phones to fully-fledged work stations;
- **Resource pooling.** The CSP reuses computing, storage and network power to serve multiple consumers using a multi-tenant model. Resources are dynamically assigned and reassigned according to consumer demand. Clients are not fully aware of the exact place location where their systems are actually deployed;
- **Rapid elasticity.** Cloud-based systems are elastic in the sense they can scale in and out in terms of computer configuration, computer fleet size, storage space and networking capabilities; and
- **Measured service.** Cloud users are billed for the exact amount of services actually consumed, and system monitoring and reporting are available to both the provider and the consumer for the sake of transparency.

Clients expect the CSP to provide them with the physical infrastructure required to deploy Internet-wide information systems and applications. The amount of involvement expected from the cloud provider dictates the service model that should be adopted:

- **Software-as-a-Service.** Ready-to-use systems, for which the CSP transparently provides all the required infrastructure in exchange for a periodic service bill;
- **Platform-as-a-Service.** The CSP provides the consumer with all the hardware-based stack and some of the software-based stack, commonly including the operating system,

the database of choice and the programming libraries required to build an application; and

- **Infrastructure-as-a-Service.** This service model gives access to the configuration of the most fundamental aspects of computing resources. The consumer can fully manage computing power, storage needs and network functions, including the Virtual Private Cloud (VPC) design: IP addressing, DHCP setup, and NAT-based, or fully routed public exposure, with public IP addresses fetched from a common address pool.

The Cloud Computing model's characteristics suggest it could be a viable foundation for developing a distributed architecture to collect Internet Background Radiation data. The most notable aspects of Cloud Computing linking to such possibility encompasses the following key properties:

- Distributed by design;
- Geographical capillarity;
- Ephemerality; and
- Idle capacity.

Service provisioning within CSPs is distributed by design. Computing power is provided in the form of virtualised compute instances, for which CPU models are divided into generation and class. Object-based storage is available through programmable repositories, such as the *bucket*. Network resources are distributed. For each cloud region, there is a Virtual Private Cloud enabled with cloud-based counterparts to routers, firewall state-tables, internet gateways, and one or more *elastic*, public IPv4 addresses that can be temporarily or permanently attached to the virtualised compute instance.

Cloud Service Providers operate in many regions worldwide. They offer unmatched capillarity when compared to on-premises deployments. For the experiment reported here, Amazon Web Services (AWS) enabled twenty-six *cloud regions* in the following Countries: The United States of America, Canada, Germany, Ireland, England, France, Sweden, Italy, Switzerland, Spain, China, Japan, India, Indonesia, Singapore, Bahrain, UAE, Israel, South Africa, Brazil, Australia and New Zealand.

The deployment of cloud-based systems is programmatically achievable via modern, service-oriented, declarative languages, termed Infrastructure-as-Code (IaC). CloudFormation and Terraform are notable examples and the most frequently used. IaC enables programmers to orchestrate the end-to-end life-cycle of cloud-based distributed systems in a programmatic, reliable fashion. These deployments are ephemeral. They are launched upon request and will endure for as long as the cloud architect wants, or for as long as there is a budget to keep the systems running. From a research perspective, IaC architectures are reproducible in the sense

other researchers can relaunch the same architecture and compare or validate their method and results.

Because CSPs must comply with “on-demand self-service”, they must hold more resources than the current sum of their consumers’ demands. This leads to a business phenomenon of *auctioning the idle capacity*. Cloud providers offer their idle capacity, especially in terms of computing power, at heavily discounted prices, of up to 90%, as long as the consumer accepts to enter a bidding scheme - whoever pays the most - to temporarily consume these otherwise more expensive resources. In order to benefit from the scheme, the cloud architect must accept the sudden termination of the allocated systems in a very short time frame, i.e., two minutes, and prepare to bid slightly higher to restore the desired computing capacity.

2.4 SUMMARY OF CONCERNS

This work addresses three primary arguments concerning the reliance on Network Telescopes as the only method for acquiring Internet Background Radiation.

1. IPv4 address availability is scarce. There are very limited IPv4 address blocks allocable for IBR capture and analysis. Not only are the addresses scarce, but they are constrained to very few researchers within limited research institutions assigning IP addresses to radiation collection and analysis;
2. The Network Telescope is inherently bound to a specific geographical location. Even when assigned with fractions of multiple IP subnets. It is bound to the Country or IANA (Internet Assigned Numbers Authority) region it belongs to. This impacts, to a debatable extent, the diversity of cyber threat intelligence that can be acquired; and
3. There is little to learn about the geopolitical influence of malicious Internet traffic by investigating country-bound radiation. There is research to support the fact it is only a matter of time until the Network Telescope will be the target for an Internet-wide malicious threat, such as in (Chindipha; Irwin; Herbert, 2019). However, until today, there is no clear way to establish a cause-and-effect relationship between nation-state-sponsored campaigns and how this affects the targeted countries, as there is no effective distributed method to capture the radiation they emit to provide the grounds for analysis.

The research documented onward focuses on two main contributions: 1) a detailed description of the Cloud Telescope, which takes advantage of the geographical capillarity of cloud computing service providers to enable distributed capture of Internet Background Radiation, in the form of a functioning cloud architecture described as Infrastructure-as-Code, and 2) the results of a five-month-long capture experiment using a distributed sensor fleet of 130 sensor nodes across 26 world regions to validate the method.

3 LITERATURE REVIEW

This chapter contains a discussion of the systematic procedure employed to identify relevant works and a discussion of eleven related Network Telescope experiments, to support and contrast with our proposed model. These works were used to support most decisions around the Cloud Telescope’s architectural design, implementation, aim, goals and expected results. Firstly, seminal papers formalising IBR capture and analysis were identified, selected, and summarised. They were used as a starting point to a forward snowballing, in accordance with the fundamental aspects of the method as established by (Kitchenham et al., 2007).

3.1 RESEARCH QUESTION

The fundamental question when designing the selection criteria to pick up the most pertinent, relevant, and productive related works, summarises as “Is it relevant to collect cybernetic threat intelligence on Internet Background Radiation, by means of a geographically distributed, cloud-based architecture?”. In support, “how does the method or data presented in the candidate paper directly or indirectly endorse, or refute, the use of a cloud-native approach to this purpose?”. The following selection criteria are discussed:

- The strategy used to choose the most suitable keywords, the query string, to submit to each research databases;
- The strategy used to select each related work, and each viable research database;
- An accounting of the number of publications found; and
- The inclusion or exclusion criteria used as the final filter to select the most relevant papers discussed along this section.

3.2 KEYWORD SELECTION STRATEGY

The initially selected search engine keywords were extracted from a straightforward four-step criterion:

1. The use of existing keywords found on the seminal papers;
2. Use of quotes with all compound terms, such as in “Internet Background Radiation”;
3. Identification of synonym terminologies; and
4. Use of the OR operator to assess the effective amount of research publications associated with the purpose of this research.

Keywords were then collected from primary works:

- Pang et al. (Pang et al., 2004): internet background radiation, network telescope;
- Wustrow et al. (Wustrow et al., 2010): darknet, internet background traffic, network pollution;
- Benson et al. (Benson et al., 2015): internet background radiation; network telescope; opportunistic network analysis; and
- Chindipha et al. (Chindipha; Irwin; Herbert, 2019): internet background radiation, network telescope, threat intelligence, network scanning.

The result is the following main search string:

```
"internet background radiation" OR
"internet background noise" OR
"network telescope" OR
"network telescopes" OR
"darknet" OR "darknets"
```

3.3 RELATED WORK SELECTION STRATEGY

The search strategy aimed at two main goals. First, the rebuilding of the historical timeline of the most relevant contributions to the Internet Background Radiation field. Second, a structured, time-constrained, selection was conducted to find the most relevant recent advancements in IBR analysis, constrained to the last five years.

The main research databases considered to quantitatively assess the number of works in the field, focused on the most frequently used repositories in Computer Science:

Research Database Repositories

- ACM Digital Library
- IEEE Xplore
- Science Direct

Seminal works considered for the snowballing phase

- Technical report by Moore (Moore et al., 2004);
- Primary research work Pang (Pang et al., 2004);
- Long-term IBR analysis by Wustrow (Wustrow et al., 2010); and
- Long-term IBR analysis by Irwin (Irwin, 2011).

Forward snowballing was performed upon the four seminal works, which enabled the identification of a range of relevant, interconnected research contributions. Table 2 shows the publication count found on each selected research database, considering the most relevant keywords linked to IBR analysis.

Table 2 – Publication count for the most relevant search keywords supporting the Literature Review.

Query string	ACM	IEEE	Elsevier
"internet background radiation" OR "internet background noise"	52	26	10
"network telescope" OR "network telescopes"	137	63	77
"darknet" OR "darknets"	856	1,161	1,845
"internet background radiation" OR "internet background noise" OR "network telescope" OR "network telescopes" OR darknet OR darknets	950	1,200	1,895
5-year cut $\geq$ 2019	727	981	1,654
3-year cut $\geq$ 2022	421	472	1,079

A total amount of 950 papers were found on ACM's Digital Library, out of which 727 fall into the 5-year cut, and 421 were published on or after 2022. IEEE's Xplore sourced 1,200 papers, out of which 981 fall into the 5-year cut, and 472 into the 3-year constraint. Finally, Elsevier's Science Direct indicated the existence of up to 1,895 potentially related works, of which 1,654 are recent contributions. 1,079 falling into the 3-year cut.

3.4 SELECTION OF RELATED WORKS

Any existing related work was considered as a potential candidate to figure in this literature review. A set of four questions was used to prioritise an inclusion:

Does the candidate related work:

1. present a characterisation of Internet Background Radiation? Alternatively, does it present a mechanism for its collection?
2. present a quantitative analysis of the captured traffic, the number of packets observed, the capture length and/or an accounting of affected network-layer, transport-layer and/or application layer protocols and service ports?
3. provide some sort of qualitative analysis of cyber threat intelligence acquisition based on IBR-originated data?
4. contain any methodological similarity, or divergence, that could contribute to, or refute, the intent to design a cloud-based, distributed architecture for IBR collection?

All candidate-works figuring among the first 30 (thirty) result items on each research database, for which the abstract indicated reasonable signs alignment with one or more of those questions, were set apart for a critical reading.

Research papers indicated on seminal works, or cited by the seminal works, were also given priority. In those cases, even if the related work did not figure on the three research databases, they were fetched, and included regardless of any constraint.

3.5 MOORE'S CHARACTERISATION OF IBR

Moore et al. (Moore et al., 2004) pioneered the characterisation of the Internet Background Radiation centred on a discussion of a proposed method for its capture. They determined the criteria used to operate a Network Telescope, including its dimension, the procedures for event detection, a discussion of attack duration, and formal method for data evaluation and visualisation. Many practical considerations are given to those interested in reproducing their approach. Most of the considerations revolve around the probabilistic nature of statistical nature shaping the observed data, including the following:

The straightforward implication is that the average time taken to observe at least one packet originating from a source host, at random, is 25 seconds if the telescope observes a whole /8, or 2^{24} IPs.

It should take ≈ 2 hours to observe the same event if capture scope is narrowed down to /16, or 2^{16} IPs (the equivalent to a Class B address block), or up to ≈ 19 days, if capturing over a /24 subnet, containing only 2^8 hosts.

These statistical parameters directly influence the Cloud Telescope's aperture and capture length. Any experiment lasting more than 20 days should suffice in terms of producing minimally viable results.

3.6 PANG'S IBR CAPTURE AND ANALYSIS

Pang et al. (Pang et al., 2004) conducted the first IBR capture ever published in the rigour of a peer-reviewed research paper. Pang's research group obtained access to two distinct network segments, named iSink and LBL.

- *iSink* segment: comprised by a) a whole Class A block, containing 2^{24} observable addresses; and b) two /19 subnets, adding 16K addresses to the sum; and
- *LBL* segment: containing two 10x /24 contiguous network blocks, accounting for 5,120 additional addresses.

The capture was performed during a non-contiguous time-frame, between 11 March and 4 May 2004. The work presents both the method, focused on the Network Telescope, and the achieved results. Regarding the method, they discussed the passive versus active nature of

the capturing device. One single active responder was coded to simulate a vulnerable device prone to SAMR-exe, a malware spreading on-the-wild at the time. Authors provide a detailed quantitative and qualitative analysis of the captured traffic. The more extensive, passive analysis, focused on the following quantitative aspects:

- TCP, UDP, and ICMP, traffic distribution, including port and ICMP packet type aggregations;
- An accounting of the unique radiation sources sending one or more packets to each of the network segments;
- A ranking of the most probed application-layer applications.

Figure 3 shows the most relevant findings.

Figure 3 – Results of Pang.

A	iSink		LBL		iSink		Popular Ports – LBL segment		
	UW-I		LBL-P		Class A		TCP Port	# Source IP (%)	# Packets (%)
	Rate	%	Rate	%	Rate	%			
TCP	928	95.0%	664	56.5%	130	88.5%	445	43.4%	19.7%
ICMP	4.00	4.2%	488	39.6%	0.376	0.3%	80	28.7%	7.3%
UDP	0.156	0.8%	45.2	3.8%	16.5	11.3%	135	19.1%	30.4%
							1025	4.3%	5.8%
							2745	3.2%	3.6%
							139	3.2%	11.1%
							3127	2.7%	3.2%
							6129	2.2%	2.4%

B	iSink		LBL	
	UW		LBL	
	#SrcIP	Percentage	#SrcIP	Percentage
TCP	759,324	87.9%	586,025	90.0%
ICMP	109,135	12.6%	64,120	9.8%
UDP	4,273	0.5%	4,360	0.7%

C		
TCP Port	# Source IP (%)	# Packets (%)
445	43.4%	19.7%
80	28.7%	7.3%
135	19.1%	30.4%
1025	4.3%	5.8%
2745	3.2%	3.6%
139	3.2%	11.1%
3127	2.7%	3.2%
6129	2.2%	2.4%

Source: Adapted from Pang et al. (2004)

A shows TCP as the most prevalent captured traffic, accounting for 56% and 95% of the received packets. B shows the number of unique IP addresses generating radiation against the observed network segments. C shows the most probed TCP ports, highly concentrated towards 445/TCP, the Windows-based file-sharing protocol suite (SMB/CIFS), followed by 80/TCP, which serves HTTP; and by Remote Procedure Call, RPC, which initiate authentication sessions for Server Message Block, SMB. The three together accounted for the majority of traffic captured in the LBL network segment.

From a qualitative perspective, the authors analysed the captured radiation against potential reasons for its emission. The smaller, active portion of the capture enables for payload inspection, in particular to enable an analysis of TCP datagram payloads. They focused on the three most exploited ports at the time to conclude the following:

- **Port 445:** Contained six prevalent exploitation forms, all aiming at accessing potential file-sharing resources without authentication, as in
`\\<dst-IP>\IPC$ \samr`
- **Port 80:** Index acquisition attack patterns, such as `GET /`, followed by attacks against HTTP's `SEARCH` method to exploit badly formed strings, such as in
`SEARCH /\x90\x02\xb1\x02\xb1...`
- **Port 135:** Also presenting six exploitation patterns, out of which three were clearly linked to malware propagation belonging to Blaster, Welchia, and Exploit1624a.

While the results presented are no longer directly relevant due to the time which has passed since the capture, the method and approach used to report the results are of interest:

- The quantitative analysis focuses on traffic distribution and aggregation of busiest network-layer and transport-layer protocols;
- Consideration is given to the recoverable metadata that can be found based on the radiation's source IP address; and
- The most prevalent radiation target ports found in the quantitative analysis may influence the deployment of an active responder on selected protocols and destination ports.

3.7 WUSTROW'S LONG-TERM CAPTURE

Wustrow et al. (Wustrow et al., 2010) extended Pang et al.'s method (Pang et al., 2004) to conduct a two-step analysis that resulted in two distinct datasets. The first was in the form of a 5-year-long capture of the IBR, targeting a full, unused /8 network block. The second being a short-term capture, following a weekly pattern, containing three fractions of another recently allocated /8 network block.

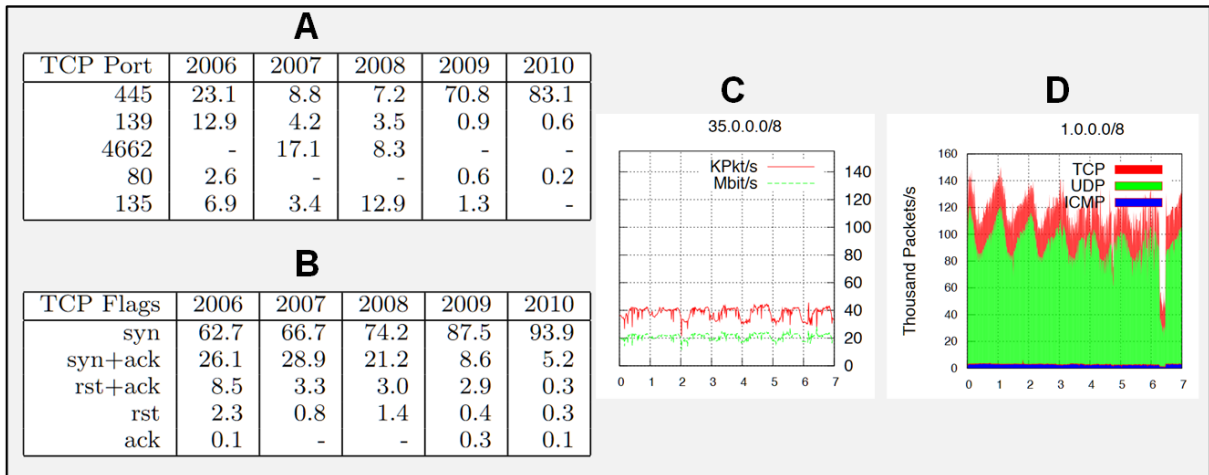
Their work provide several tools, practices, and procedures that system administrators could use to conduct new observations. Key contributions include:

- Unique characterisation of IBR as received by an unallocated /8, for which, *a priori*, no traffic should ever get routed to;
- A differential analysis of the traffic arriving to the unused network block, in opposition to the traffic affecting the productive network range;
- Long-term analysis of the unallocated network block, revealing pattern and behavioural changes throughout the years;
- Discussion and differentiation of the most prevalent activities, including port scanning, from other forms of traffic, combined with a clear definition of *back-scatter*; Focus on scanning *versus* back-scatter affecting TCP. Formalisation of flag state, such as:

- **Port scanning** → TCP SYN
- **Backscatter:** → TCP SYN+ACK,
TCP SYN+ACK, RST, RST+ACK, ACK
among other considerations, including the use of forged IP addresses.

A summary of results is shown in Figure 4.

Figure 4 – Results of Wustrow.



Source: Adapted from Wustrow et al. (2010)

A, shows the quantitative distribution of traffic destined to the most frequently probed TCP ports. There is a decline of interest for port 445 starting in 2007, followed by a new surge in 2009, achieving its peak in 2010. It is also possible to perceive a spike of interest for port 4662, which matches the increasing interest for peer-to-peer networks as implemented on that port by the family of protocols including emule, edonkey, and e2k.

B, aggregates TCP traffic by detected header flag state. SYN scans grew considerably during the period, but anomalous or uncommon flag combinations were also detected, and match, to a large portion, the behaviour of *backscatter* activity.

C shows a fairly stable packet rate arriving to 35.0.0.0/8 during a one-week long time span. The sample denotes an average captured packet count of $\approx 40K/s$ (forty thousand packets per second), while the consumed bandwidth was of $\approx 20Mb/s$ (twenty megabits per second), for the entire set of ≈ 16 million addresses, resulting in an average bandwidth baseline of $\approx 1,25bit$ per second, per IP address, also indicating a little less than 1 packet per second, per destined IP address.

Finally, D plots the protocol distribution between TCP, UDP, and ICMP, arriving to 1.0.0.0/8. The chart shows TCP as the most prevalent, while ICMP is seen to minor extent. Application-layer exploitation was clearly more attempted than denial-of-service based attacks, which tend to generate loads and spikes of ICMP and UDP noise.

This study reinforces the use of aggregation tables in combination with speculative charts to plot the radiation in interesting manners.

3.8 IRWIN'S TELESCOPE AT RHODES UNIVERSITY

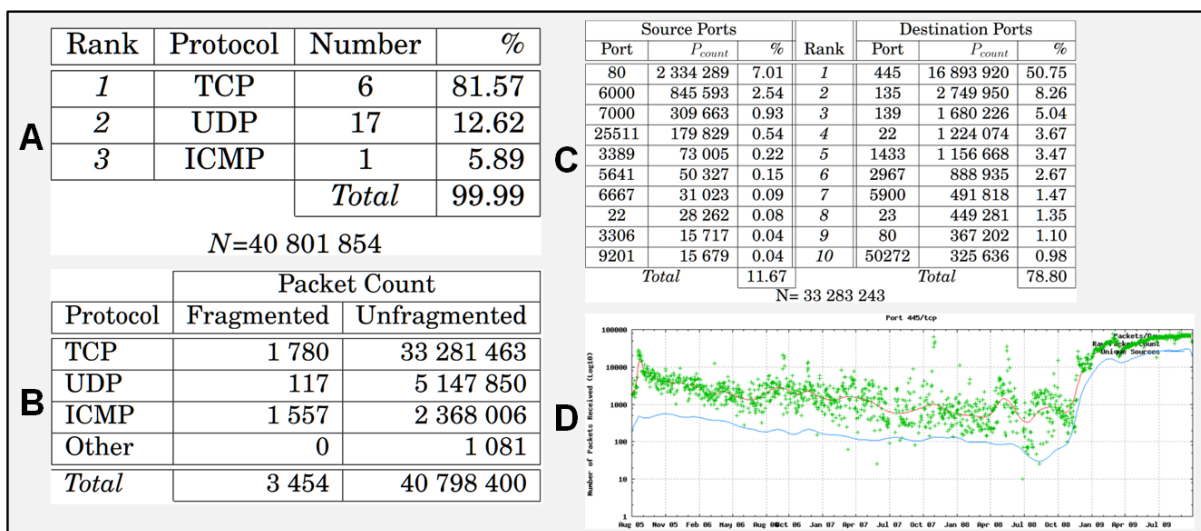
Irwin (Irwin, 2011) thoroughly documented the deployment of a Network Telescope at Rhodes University for the consecutive period of 50 months: from August 2005 until September 2009. The telescope was prepared to capture and observe one full /24 address block belonging to the University, within AFRINIC's address space.

The author reported the collection of a resulting dataset containing 40 million network events. They present different traffic analysis methods for each relevant network-layer, transport-layer and application-layer protocol. At the same time, they provide the introduction to geopolitical analysis influencing radiation emission. There is a discussion around the need for new approaches to data summarising and visualisation, culminating with two visualisation tools, one thought to plot and condense observed data over time, and another more towards near real-time analysis.

The results include a quantitative analysis of the produced dataset, and a qualitative analysis focusing on the rise of Conficker, a major self-propagating worm that used to propagate on port 445/TCP, and was first detected by the telescope fairly before public awareness, on the second half of 2008.

Figure 5 aggregates some of the most interesting findings.

Figure 5 – Results of Irwin.



Source: Adapted from Irwin (2011)

A shows the same TCP/UDP/ICMP summarising as seen in previous works. On this report, 81% of packets belonged to TCP, followed by UDP, 12%. ICMP was seen in 5% of the incoming traffic.

B shows packet fragmentation. Fragmentation is a compute intense task, as the target needs to spend CPU cycles on packet reassembling, which might lead to denial-of-service. Only 3,454 matched such criterion ($\approx 0.01\%$), but the ability to extract this measurement is a remarkable example of cyber threat intelligence that can be acquired from radiation samples.

C ranks the most frequently probed destination ports, preceded by a source port analysis. The source port analysis enables to infer and detect intentionally erratic behaviour, such as source ports <1024 being used to establish connections. Port 80/TCP was used as the source port in 7% of the packets. The most probed destination ports contain half of the radiation towards 445/TCP, followed by other Windows-server file-sharing related ports: 135/TCP, 139/TCP. SSH (22/TCP) also figures on the list.

D shows a plot of traffic distribution targeting 445/TCP. The spike starting at the beginning of 2009 is linked to Conficker, for which a comprehensive analysis was provided in the manuscript.

This work confirms that each Internet Background Radiation capture should reveal trends and patterns that can be linked to certain vulnerabilities and attack types on-the-wild during the observed time frame. New observations, in combination with new architectures and new data visualisation techniques, should reveal new threat intelligence and new perspectives over the observation of the phenomenon.

3.9 BENSON'S FRAMEWORK

Benson et al. (Benson et al., 2015) proposed a framework towards expanding the understanding of the motivation behind radiation emitters and used that knowledge to infer the *macroscopic properties* of the Internet, as perceived through Internet Background Radiation collection.

Using *darknet* as a central terminology, authors revisit, update, and refresh IBR collection methods, clearly defining or redefining relevant terminology, including:

- Darknet: the IP address block configured to run in passive mode, occasionally returning “unreachable” messages (such as a TCP-RST datagram) to radiation sources;
- IP gray space: used to refer to devices or network setups intermittently performing IBR capture and productive Internet access. The goal is to learn how the radiation is affected by the existence of addresses assigned to productive use;
- Honeypot: reinforces the honeypot as an active responder, to a certain interaction level (low, medium, full/high), while also referring to them as “distributed trap-based network monitoring system”, for data collection and analysis; and
- Greynet: A network that is populated with active IP addresses interspersed with darknet addresses, combining darknets, honeypots and productive computers altogether.

Among the most relevant methodological aspects of their work, they shed light on key aspirations of the Cloud Telescope, including:

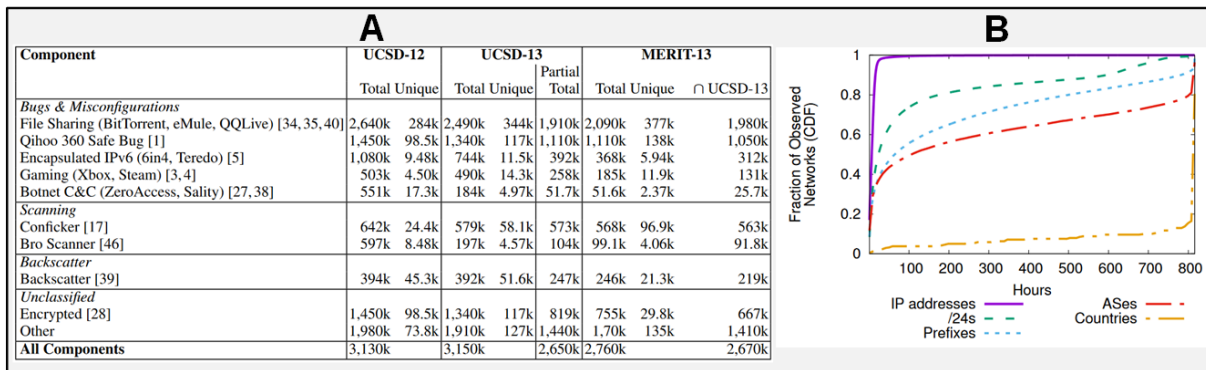
- The use of datasets coming from more than one source;

- Data analysis ranging from July 2012 to August 2013;
- 10 terabytes worth of compressed, PCAP-formatted dataset;
- Hybrid collection combining *spike* detection within traffic, associated with heuristics designed by (Dainotti et al., 2014), and later improved in (Dainotti et al., 2016), to identify source address spoofing, to map and aggregate radiation sources to the ASN they belong to, and the inclusion of geolocation metadata, with the support of Maxmind’s GeoIP database;

Authors conclude that the precision of the results with darknet observations vary in proportion to the telescope’s aperture and capture duration. They also state they have received traffic from virtually all the countries in the world ($N = 245$), and from all existing autonomous systems at the time ($\approx 46K$), reinforcing the perception that telescopes tend to provide a fair, evenly distributed sample of what threats affect the Internet at large.

Figure 6 summarises one tabular and another chart-style result.

Figure 6 – Results of Benson.



Source: Adapted from Benson et al. (2015)

A summarises the received traffic in three different categories. UCSD-2012 refers to the quantitative analysis performed on UCSD’s telescope during the year of 2012 at University of California in San Diego. UCSD-2013 refers to the same place, but during 2013. MERIT-2013 refers to the collection that took place at Merit Networks the same year.

Findings were categorised into four sections: Misconfigurations, Scanning, Backscatter and unclassified. Traffic attributable to P2P was highly prevalent under the misconfiguration aggregation. Conficker, alongside with Bro network scanner, were the main source of radiation affecting the scanning category. Backscatter was individually aggregated as the third most prevalent activity, followed by encrypted transit, among others.

B reveals the number of hours it took to obtain the first packet associated with each observed IP address, autonomous system, /24 network block, Country and Internet prefix. It took around 800 hours, or 33 days, to collect at least one packet from each existing radiation source on the internet. This is around double the amount of time predicted by (Moore et al.,

2004) eleven years earlier, but still a very short time frame, in case the telescope is kept running for a decent amount of time. They also provide some relevant discussion on:

- the use and exploitation of open DNS servers to reversely learn about radiation sources; including their system names and geographic location;
- reverse scanning, in which each radiation source is scanned by nmap to find counter-intelligence regarding source's uptime, fingerprint and open communication ports and services; and
- an evaluation of how aperture size and capture length affect IBR analysis.

Among the conclusions, three are of interest to the design and implementation of a Cloud-based Telescope:

- The larger the telescope's aperture, the least relevant is an individually observed IP address;
- Halving a darknet's size shall retain 89% of its fidelity in the ability to capture a fair sample of what radiation is affecting the Internet at a given time. This shall hold true for smaller darknets; and
- The comparison of radiation properties affecting two distinct Network Telescopes promotes a better understanding of Internet Background Radiation. Comparisons promote more consistent measurement procedures, allowing better inference to extrapolate observed samples to what should happen at a global scale.

3.10 SHAIKH'S CORRELATION: THE IOT AFFECTS IBR

Shaikh et al. (Shaikh et al., 2018) predicted that the low-power, low-cost, and rushed time-to-market nature of the Internet-of-Things would foster a new generation of attacks tailored to exploit such inherent design characteristics. They proposed the use of Internet Background Radiation analysis to passively and actively measure the presence of IoT-enabled devices on the Internet, along with a method to account the radiation targeting this class of computer devices.

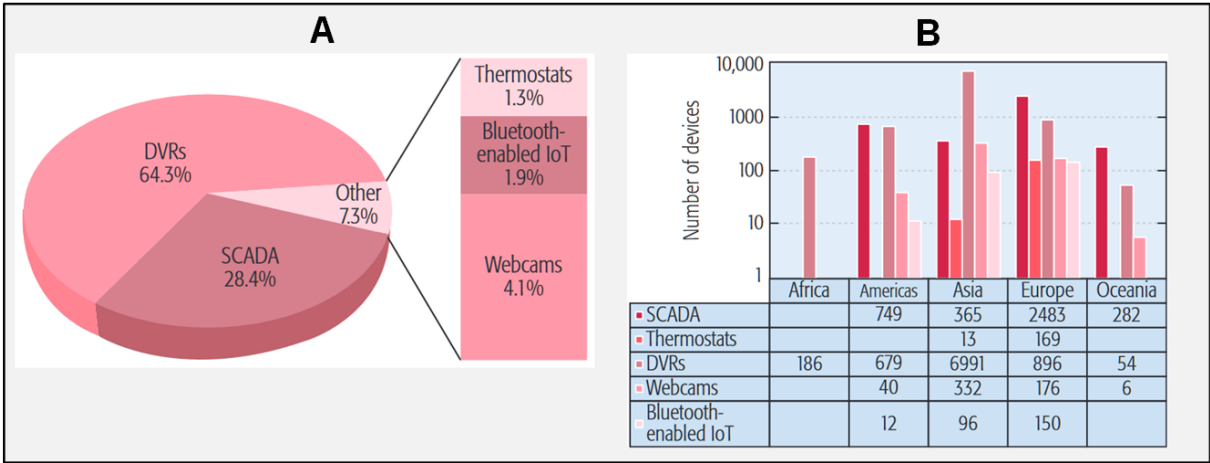
Their work is based on data collected by the Center for Applied Internet Data Analysis - CAIDA's telescope. The capture range was two months, starting in December 2016, and lasting until the end of January 2017. They analysed dataset worth the amount of 1,5 terabyte of compressed, PCAP-formatted capture recording. This hybrid (active, passive) analysis consisted of the individual inspection of each packet in the dataset against patterns, fingerprints, or footprints, that could have been generated by typical, constrained resource, embedded devices, carrying minimal firmware or operating system. Dataset entries had the source IP address enriched with data available on two open source intelligence platforms: Shodan (Shodan, 2024)

and Censys (Censys, 2024). These continuously scan the entire public IPv4 address space and collect any available metadata, including the hardware and software stack embedded to the host.

The effort resulted in the identification of 285,478 IoT devices within the dataset. These were aggregated under five classes: Digital Video Recorders, DVRs, SCADA, Thermostat, Webcam (or IP camera), and Bluetooth-enabled IoT.

Results are shown in Figure 7.

Figure 7 – Results of Shaikh.



Source: Adapted from Shaikh et al. (2018)

A shows that more than half of the attack patterns found on the dataset aimed at exploiting vulnerabilities targeting digital video recorders. SCADA systems figure on the second position, followed by thermostats, Bluetooth-enabled IoT and IP cameras.

B shows attack type distribution per world region. There is a high concentration of attacks aiming at exploiting SCADA vulnerabilities in Europe, which is the same target for most of the thermostat, or smart home, oriented attacks. Asia faced the majority of attacks targeting surveillance devices, DVR and webcams included.

3.11 CHINDIPHA ON SMALL-APERTURE TELESCOPES

(Chindipha; Irwin; Herbert, 2019) quantified the accuracy of smaller aperture Network Telescope configurations. The main argument of the authors is that radiation capture by means of large darknets is no longer feasible due to pressure to put unallocated IPv4 addresses to productive use. They made use of capture samples taken from two very large network blocks, 146/8 and 155/8. Dataset down-sampling was generated by segmenting each /8 into many /24. These were then sub-divided into smaller samples ranging from /27 until /30, named “subnet equivalents”.

Their statistically sound method provide an strategy to establish the error margin associated with each “aperture stop”. This involves four calculations, three of which are relevant to present:

- MAPE: The Mean Absolute Percentage Error;
- SMAPE: The Symmetric Mean Absolute Percentage Error; and
- MASE: Mean Absolute Scaled Error, or the measurement to predict the accuracy of MAPE. The closer to zero, lower the error margin.

Accuracy is reported as a percentage relative to the amount of observed IP addresses in the whole /24 block against the subnet equivalent under consideration, as reported in Table 3.

Table 3 – Smaller aperture telescope accuracy (Chindipha; Irwin; Herbert, 2019)

Subnet	MAPE(%)	SMAPE(%)	MASE
/24	100.00	100.00	0.00
/27	92.03	92.41	0.68
/28	89.43	90.41	0.87
/29	89.22	90.42	0.89
/30	83.48	90.88	1.48

According to Chindipha (Chindipha; Irwin; Herbert, 2019), down-sampling telescope aperture from /24 to /27 shall retain 92% accuracy, which is represented by a low MASE score of 0,68. Down-sampling to /28, or even /29 shall retain 89% accuracy. A down-sample to /30 (from 256 to only 8 observed IP addresses), shall retain 83% accuracy.

This work provides a safe ground for the research hypothesis in the sense that experimenting with smaller-than-usual IPv4 address ranges should still result in meaningful findings. A longer time frame should also compensate smaller apertures, as previously discussed.

3.12 ZAKROUM'S PREDICTION MODEL

(Zakroum et al., 2022) uses Internet Background Radiation data to propose a framework enabling to predict global cyber threat activity based on current and past traffic behaviour.

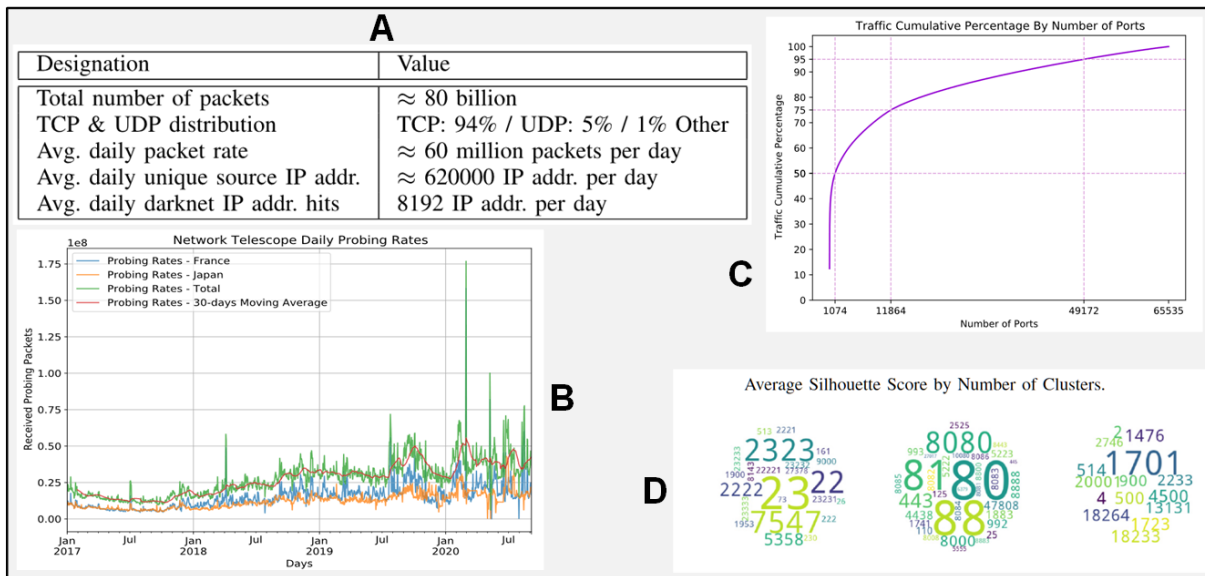
The approach proposes the use of inferring groups containing transport-layer ports grouped by scanning frequency and by temporal and semantic similarity. The scanning/probing rate arriving on each grouped transport-layer port is used to feed a neural Long-Short Term Memory (LSTM) network. This results in a prediction model determining the probability of an attack within a one-hour time-frame and for the next twenty-four hours.

They have used a dataset comprised of IBR arriving to two /20 networks, 8,192 IP addresses, one located in France and another in Japan. Data was collected for three and a half years, between 2017 and 2020.

The most relevant findings are displayed in Figure 8.

A shows a dataset summary, which contained 80 billion packets, out of which 94% TCP and 5% UDP. Average daily packet rate was of $\approx 60\%$ million. All destinations were probed at least once a day during the capture time-span.

Figure 8 – Results of Zakroum.



Source: Adapted from Zakroum et al. (2022)

B shows a reasonably stable traffic growth rate towards the two darknets. The green line stacks the sum of both telescopes. The spike in March 2020 accounted 1,75 billion packets.

C shows the range of destination ports probed during the building of the dataset. This confirms the empirical thought that a small portion of all destination ports are frequently probed, while many unassigned ports are of much lower interest for radiation emitters.

D shows a word cloud, or port number cloud, containing the most frequently probed ports for each semantic-temporal, inferring group aggregation:

- The leftmost group refers to remote administration services, for which traffic on ports 22 (SSH) and 23 (Telnet) were more prevalent. Frequently used alternate ports, such as 2222, and 2323, also stand out;
- The central group shows ports related to HTTP-based services, highlighting the prevalence of ports 80, 81, 88, 8080, 443, among other; and
- The rightmost group aggregates many VPN protocol and service ports as implemented by L2TP, PPTP, IPSec, among other.

Tools, procedures and aggregation techniques found on this work shall contribute to drive decisions regarding how to handle and aggregate the datasets envisaged by this research effort.

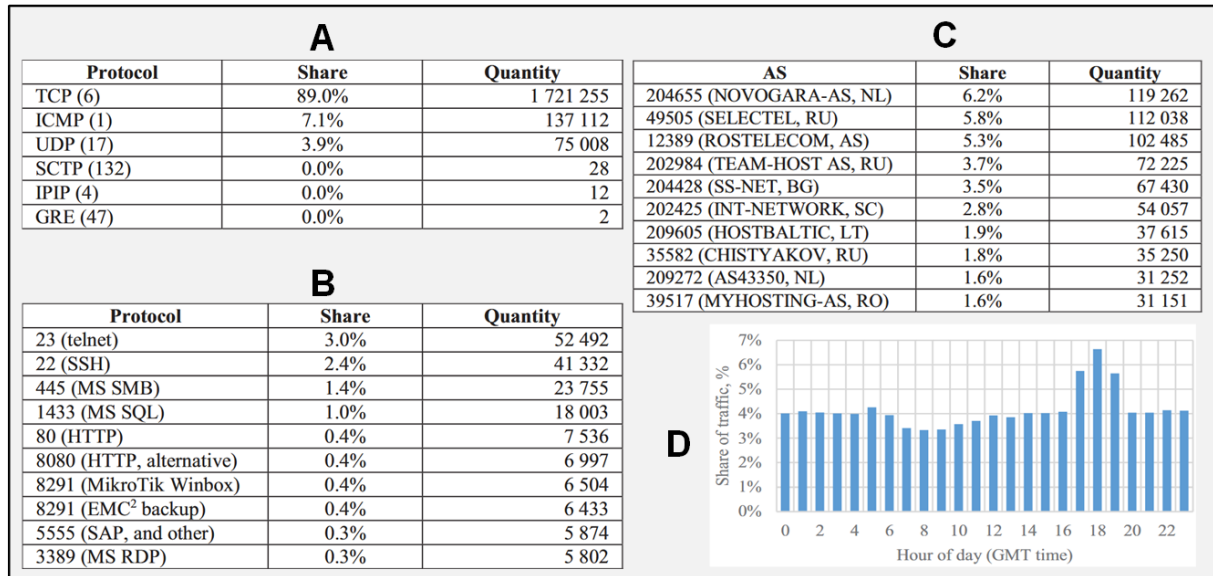
3.13 ZOLOTYKH'S LOOK FROM THE RUSSIAN SEGMENT

Zolotykh (Zolotykh; Nikulin, 2020) reported what they termed the dark collector: “trap for background noise”. Different from a honeypot, the trap is comprised of two elements: a

collector and an analyser. The collector is set up with two interface cards, one of which is publicly exposed, and receives background noise. The analyser is implemented bound to the second, private network card, on top of which a toolset is built, including reverse-DNS solvers, which contributes to identifying the most active noise generators.

Their capture experiment was conducted during four consecutive months, between late 2019 and early 2020. A summary of findings is reproduced in Figure 9.

Figure 9 – Results of Zolotkyh.



Source: Adapted from Zolotkyh & Nikulin (2020)

A reveals the Russian segment under observation is mostly hit by unsolicited TCP traffic, comprising 89% of their sample, followed by ICMP (7%) and UDP (4%). B breaks-down TCP traffic per destination port, revealing that the telnet port is the most frequently probed (3%), followed by SSH and Microsoft's SMB and SQL service ports accumulating at least 1% of the traffic each.

Back to overall traffic, C lists the most active radiation sources, identified by the Autonomous System owner, revealing that a Netherlands-based AS has generated 6% of the noise, followed by a mix of Russian-based and Baltic-based ISPs which together accounted to 24% of all seen packets, in other words, locally attributed radiation corresponds to one fourth of all seen traffic.

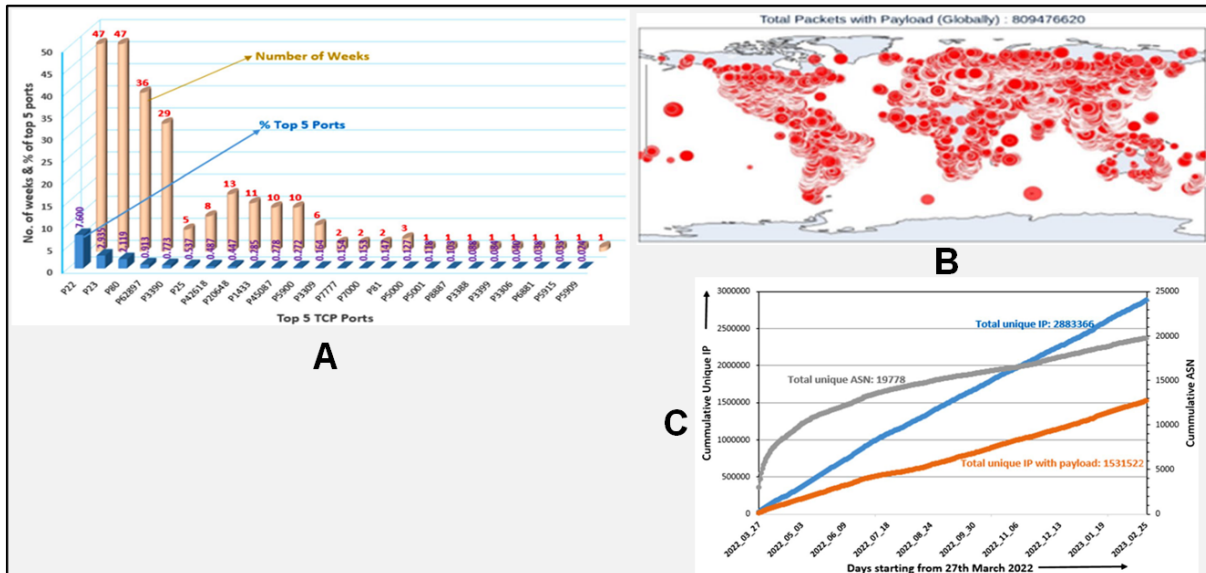
The time of the day was also evaluated. In D they indicate a peak of incoming traffic arriving to the capture mechanism between 5pm and 7pm GMT.

3.14 KORI'S ANALYSIS OF AN ACTIVE NETWORK TELESCOPE

Kori (Kori et al., 2024) recently reported the deployment of an Active Network Telescope to which 210 IP addresses within a /24 network block were assigned. The experiment was deployed in March 2022, and endured for one year. They focused on providing a TCP-centric

quantitative and qualitative analysis of the received radiation. 1.1 billion network events were recorded. Relevant findings are reproduced in Figure 10.

Figure 10 – Results of Kori.



Source: Adapted from Kori et al. (2024)

A shows a combined bar graph reporting the most active TCP destination ports combined with the number of weeks each particular port remained in the Top 5 aggregation. Port 22 SSH was the busiest, followed by Telnet 23, HTTP 80, Xsan Filesystem Access at 62897 and Microsoft RDP 3390, among others.

B shows a plot of radiation sources which actively interacted with the network telescope. The apparently even traffic distribution was accounted in terms of the five most active countries, reported as: United States of America, China, Russia, The Netherlands and Great Britain.

C shows three time lines: The grey curve depicts how long it took to receive at least one packet from each of the 19,778 Autonomous Systems seen. The blue line shows a linear growth in terms of the diversity of IP sources, 2,883,366, exchanging a connection initiation packet with the sensor, while the orange one depicts the subset exchanging one or more payload with the active collector, 1,531,522 (53%).

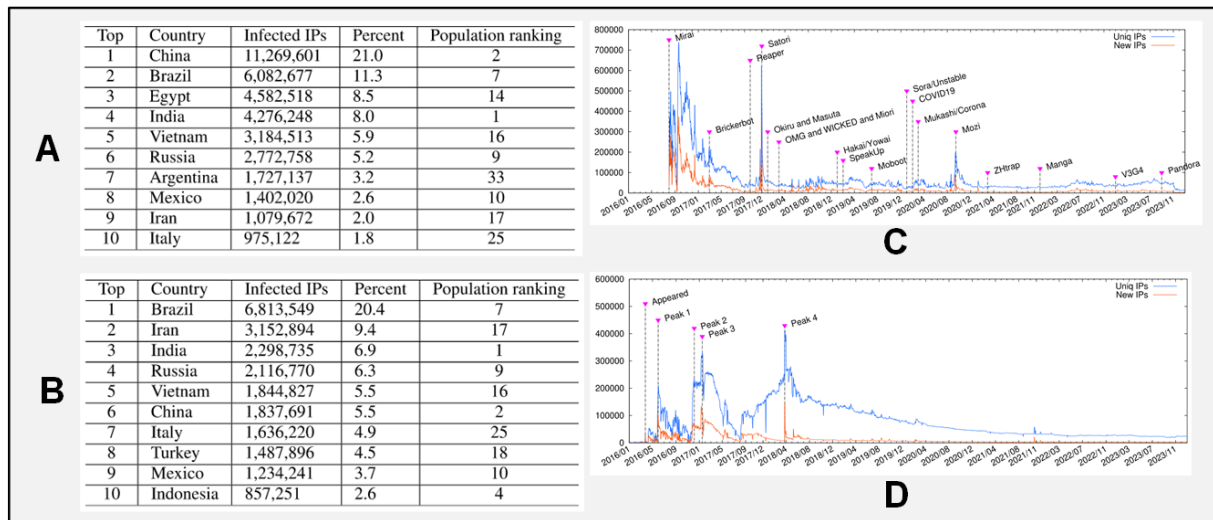
3.15 ANH AND NAKAMURA ON MIRAI AND HAJIME

Anh and Nakamura (Anh; Nakamura, 2024) conducted a baseline investigation into the evolution and prevalence of Mirai and Hajime utilizing datasets as collected by The Network Incident Analysis Center for Tactical Emergency Response (NICTER), operated since 2005 in Japan, and collecting unsolicited traffic arriving to an undisclosed amount of more than 300,000 IP addresses. The dataset was filtered to only contain traffic on ports of botnet prevalence, 23

and 2323. A MySQL-backed method was employed to aggregate data on a daily basis, where malicious sources of interest were added to a list, after which duplicates were removed.

Among the results, researchers were able to rank the most infected countries, by either Mirai or Hajime, and rebuild a timeline of activity belonging to each of these botnets, as aggregated in Figure 11.

Figure 11 – Results of Anh and Nakamura.



Source: Adapted from Anh & Nakamura (2024)

A shows that China, Brazil, Egypt, India, and Vietnam, among others, host the most infected sources attempting to propagate Mirai, as perceived by the Network Telescope. B summarizes the propagation of Hajime, where Brazil, Iran, India, Russia and Vietnam, among others, show up as the most prevalent probing sources.

A Mirai evolution timeline, summarised in C, shows the authors were able to identify the rise of the most important sub-variants, including Moobot, Sora, and Mozi.

D depicts the peak activity for one of the two Hajime strains discussed on the work. This second strain is slowly ceasing activity, whereas the first one was no longer seen by that telescope after April 2018.

3.16 DISCUSSION OF RELATED WORKS

The revised related works confirm the need of a certain agreement between the research entity, an University, for example, with one or more Internet number authorities, IANA or another, to obtain access to an observable IPv4 address range, such as seen in Pang et al. (2004) and Wustrow et al. (2010). Alternatively, one could make use of address ranges already allocated for universities, as in Irwin (2011), or in Zakroum et al. (2022). Finally, it is also possible to make use of openly available datasets, such as the one provided by CAIDA's network telescope, as in Shaikh et al. (2018). It is paramount to consider IPv4 address absolute exhaustion in terms of primary allocation Richter et al. (2015) and the debate around unfair and

unauthorised use of apparently unallocated address blocks, also called *IP squatting* (Salamatian et al., 2023).

Consider the experiment properties as summarised in Table 4. The dataset resulting from each documented research effort directly result from the main factors affecting its quality, size and *resolution*: the telescope’s aperture, the capture length, and the point in time when the capture experiment was conducted. Few works report the final dataset size, using different accounting approaches. Irwin (2011) reported the number of events, in millions, (40M). Benson et al. (2015) reported the collection of a total amount of 10 terabytes worth of compressed PCAPs. Shaikh et al. (2018) made use of a telescope half the aperture of Zakroum et al. (2022)’s, obtaining 1,5 TB worth of data. Zakroum et al. (2022) performed their work eleven years after Irwin (2011), with a 32x larger telescope, and obtained a dataset 2 thousand times larger for very similar time-frames, of 50 and 42 months respectively.

Table 4 – Summary of related works.

Author	Location	CIDR Block size	Dataset size (bytes or packet count)	Timespan	Main quantitative contribution	Main qualitative contribution
Moore (2002)	CAIDA & UC San Diego	1x /8 2x /19	Not reported	Until 2002	How to deploy Network Telescopes	Event probability. Telescope aperture. Capture length.
Pang (2004)	Merit Network University of Michigan	1x /8 2x /19	Not reported	March to May 2004	Traffic characterisation	SMB,RPC,HTTP attacks High-level responders.
Wustrow (2010)	APNIC and ARIN subnets.	1x /8 + 1x /8 (partial)	Not reported	5 years. 2006 - 2010	Traffic characterisation	Backscatter characterisation. P2P traffic.
Irwin (2011)	Rhodes University	1x /24	40 million events	50 months. 2005 - 2009	Traffic characterisation	Packet fragmentation. Conficker. Visualisation techniques.
Benson (2015)	University of California San Diego	1x /8 + 2/3 of 1x /8	10 terabytes	July 2012 August 2013	Traffic characterisation	Definition of Darknet, Gray space Honeypots, and Greynet. Telescope comparison.
Shaikh (2018)	CAIDA dataset	1x /8	1.5 terabytes	December 2016 January 2017	IoT-oriented quantitative analysis	IoT as a motivation for malware propagation.
Chindipha (2019)	Within AFRINIC	5x /24	Not reported	Not reported	Accuracy of Telescope aperture down-sizing.	Viability of small-aperture network telescopes.
Zolotykh (2020)	Russia	2 IP addresses	2 million packets	4 months (2019-2020)	Traffic characterisation	Prevalence of Port 23 Influence of the time of the day.
Zakroum (2022)	Dataset	2x /20	80 billion packets	3.5 years (2017-2020)	Traffic characterisation	Port similarity aggregation. Behaviour prediction with LSTM.
Kori (2024)	France and Japan	210 IP addresses	1,1 billion packets	1 year (2022-2023)	TCP-centric traffic characterisation.	SSH as the busiest port. Source country analysis. Active Network Telescope.
Anh (2024)	NICTER Japan	300,000 IP addresses	Unspecified subset	From 2016 to 2023	Mirai and Hajime infection attempts.	Reconstruction of Mirai’s and Hajime’s timeline Focus on ports 23 and 2323 TCP
This thesis (2024)	26 cloud regions	1-10 sensor per region	10 billion packets 200 GB	1-5 months (2023-2024)	Traffic characterisation	Comparative analysis of the method. Distributed architecture. Geopolitical influence.

Source: The author.

Each related work has reported an objective, quantitative analysis. The accounting approach revealed the most essential characteristics of each Internet Background Radiation capture. This can be considered the minimal standard on any work intending to contribute to the field, typically in the form of Top 10 rankings containing the most active radiation sources, the most scanned transport-layer ports, the most intensely seen ICMP packets, and so on. In most cases some sort of data visualisation technique was also employed to provide visually appealing artefacts, such as time-series distribution, pie and bar charts, and other interesting forms of radiation interpretation.

The most prevalent form of traffic seen across all resulting datasets belongs to TCP.

This has to do with the fact the majority of core Internet services require transmission control reliability features found on this key transport-layer protocol. It is reasonable to assume most of the malware wanting to self-propagate in global scale will attempt to exploit the most popular (and vulnerable) services widely implemented on the Internet. On the other hand, amplification attacks targeting UDP-based services, such as the DNS, is also detected and discussed on the literature.

The inexistence of denial-of-service-oriented attacks is surprising to an extent. The past and current decades saw a range of elaborated attacks against major Internet content providers, such as Spamhaus, in 2013, Github in 2015 and 2018, Dyn in 2016, and Amazon Web Services, in 2020 (Pelloso et al., 2018; Cloudflare, 2024). One plausible explanation could be that these attacks are very well crafted against particular targets, and therefore, out of the range observable by Network Telescopes.

Qualitative-wise, the analyses vary. Less recent works, ranging from 2004 to 2011 tend to discuss one or few key malware-spreading activity on-the-wild during the capture time-frame, from Sasser to Blaster, from SQL Slammer to problematic Conficker, which caused immense loss at the time (Shin; Gu, 2010) and until Windows-based computers worldwide were effectively patched. Recent works focus on botnet propagation activity, while linking the motivation for this shift to the rise of the Internet-of-Things, and how it negatively impacted the Internet from a cyber security perspective. Newer works also make use of newer computer techniques, able to produce more meaningful data visualisation methods in reasonable computing time.

This research therefore proposes a cloud-native architecture to enable distributed capture of Internet Background Radiation, by discussing potential advantages of this approach, and by providing fresh, updated quantitative and qualitative data fostering method comparison. The ultimate goal is to provide a novel, perhaps better approach to IBR-based cyber threat intelligence acquisition, under an ephemeral, reproducible fashion.

3.17 CONCURRENT RESEARCH EFFORTS

Another aspect of the Literature Review at doctoral level refers to the need to continually track the rise of external research efforts that could aim at similar methods or results and compromise originality. Gill & Dolan (2015) proposes that “a good doctoral thesis should clearly outline the different ways in which the work is original. A succinct, focused, critical appraisal of the specific contribution to knowledge is preferable to an exhaustive list.”.

The seek for concurrent research efforts escapes from the time constraints established during the systematic review phase and was continuously kept during the 3,5 years of work, accepting any publication source and format beyond the ACM/IEEE/ScienceDirect repositories.

What follows in this section is a report of identified concurrent research efforts arising right before or during the research time allocated to this thesis. The identified works did not influence the directions took on any methodological aspect of the Cloud Telescope. The candidate took notice of their existence sat them aside for a revisit after all other research steps were

finished, in late Q2 2025. This resulted in the identification of three research efforts with different degrees of similarity. The first is in the form of a Master’s dissertation authored by Joseph O’Hara under supervision of Dr. Stefan Weber at the School of Computer Science and Statistics, Trinity College, Dublin. The second refers to a thesis and subsequent paper article authored by Eric Pauley, a PhD candidate within the Department of Computer Sciences at the University of Wisconsin-Madison, under supervision of professors Paul Bradford and Patrick McDaniel, while the third is in the form of an ingenious home-based approach by Rodolfo García-Peñas, a PhD student in the Department of Signal Theory, Telematics and Communications at the University of Granada, Spain, under supervision of Professors Rafael A. Rodríguez-Gómez and Gabriel Maciá-Fernández that uses a Raspberry PI and Linux stack to collect radiation arriving after enabling the DMZ (DeMilitarized Zone) feature of domestic routers.

3.18 O’HARA AND WEBER’S CLOUD-BASED NETWORK TELESCOPE

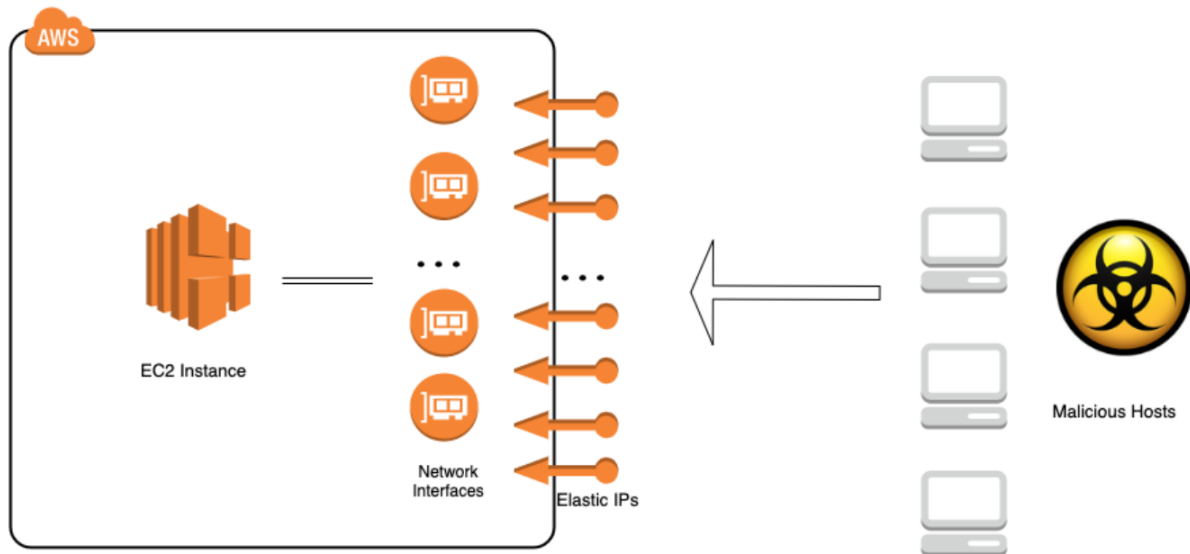
O’Hara (2019) wrote a master’s dissertation proposing a *Cloud-based network telescope for Internet Background Radiation collection*. The author explains that IPv4 scarcity and limited access to standard network telescope datasets served as the main drivers to motivate a shift to cloud-based deployments. The stated novelty of their approach is the employment of a cloud service provider’s diverse IP address ranges to enjoy the benefits of a distributed cloud-based network telescope without the need to actually construct a distributed system.

O’Hara’s aim, to determine if a small cloud-based network telescope with a very diverse IP address range can perform better at the task of botnet analysis than a traditional and larger network telescope, unfolded into four objectives: to compare the number of botnet hosts detected by a large network telescope and the proposed cloud-based solution, the detection of number of scans from the botnet hosts, the overlap between the two approaches and a comparison of traffic patterns as observed by the two systems.

The cloud-based solution was designed to form a /24 network in order to observe and collect Internet Background Radiation, mixing the properties of a distributed telescope and a greynet. The IP address diversity is described as similar to when a group of organisations collaborate to build a distributed network telescope. The architecture deployed on Amazon Web Services is depicted in Figure 12.

They describe it as a single EC2 instance deployed in a single Virtual Private Cloud, with multiple subnets, one per interface, and with one elastic (public) IP address assigned to each. O’Hara attached the maximum number of network interfaces one can attach to that single instance, which ultimately impacted the overall costs of the deployment and the effectively implemented system. One reason stated to explain the one-to-many relationship between the VPC and associated subnets is the intent to avoid issues caused by asymmetric routing which may occur when a machine has multiple network interfaces attached to it. The author also explains that among the 130+ machine types offered by AWS, they have opted to make use of a single, reasonably powerful a1.4xlarge instance type, allowing for the assignment of up to

Figure 12 – O’Hara’s cloud-native capture mechanism.



Source: Reproduced from O’Hara (2019)

240 IP addresses, later explained as effectively 204 simultaneously allocated addresses, rather than using multiple cheaper machine types, such as t3.nano, situation which would require 64 instances to reach a same purpose.

The capture computer was initialised directly via the AWS web console. The computer was configured with a set of python scripts to evoke the addition of 7 subnets and associated network cards required to reach the desired number of IP addresses. The deployment was intended to run for 10 days at an estimated cost of 117 US Dollars. The capture system effectively ran for 7 days and effectively charged 80 US Dollars for the machine instance. The cost of the elastic IP addresses was of additional 177 US dollars. The total cost including data storage, backups and other minor expenses resulted in 257 USD, or approximately 36 USD/day, or 0,17 USD per day, per listened IP address.

The cloud-based experiment resulted in a total amount of 1.9 gigabytes worth of incoming traffic, which was translated to approximately 1828 bytes per second against the system (their calculation), revealing a traffic intensity three times larger than the normalized traffic arriving to a much wider /16, on-premises network telescope counterpart, deployed within HEAnet in Ireland. Results include a comparison of Mirai activity affecting the two telescopes, where the cloud-native deployment accounted 104 thousand unique propagation sources, followed by a series of charts revealing the most active source countries, the most probed TCP ports and a map of the world showing the location of Mirai infected hosts.

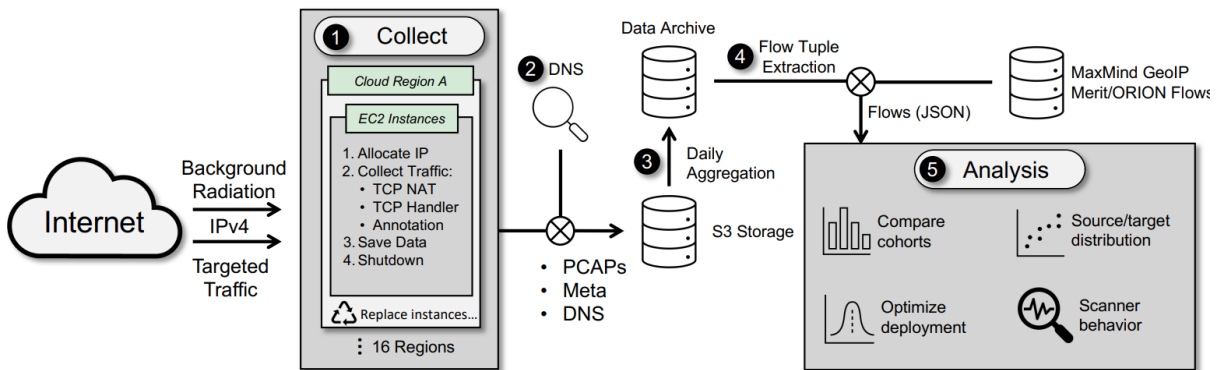
The method used to report results aggregates their findings with both the cloud-native and the on-premises telescopes. Together they saw Egypt, China, Jordan, Greece, Brazil, Russia, Japan, Ireland, India and the United States of America as the ten most active Mirai-specific radiation sources, while the most actively exploited TCP ports were 23 (telnet), 80 (http), 5901 (vnc-alt), 81 (http-alt), 4145 (cctv), 22 (ssh), 445 (netbios), followed by others.

3.19 PAULEY'S DSCOPE

Pauley et al. Pauley, Barford & McDaniel (2023) proposed DScope as a research effort towards the an interactive system for large-scale, distributed cloud-native monitoring aiming at understanding malicious and unwanted activities that target these critical infrastructures. DScope is defined as a cloud-based Internet telescope, able to collect traffic on unused IP addresses from a cloud provider by provisioning low-cost compute instances under those addresses and measuring inbound traffic. Their main paper documents the design considerations, cloud service provider of choice, collection and analysis methods, and resulting dataset characteristics.

Deployed within AWS, DScope takes advantage of the spot price scheme, in which instances can benefit from the service provider's idle capacity, to launch multiple capture instances and start collecting radiation, which is later moved to an aggregating S3 bucket. DScope can accept incoming TCP connections and complete them up to the banner-level information exchange phase. The general workflow built around this capture mechanism is depicted in Figure 13.

Figure 13 – Pauley's architecture for DScope.



Source: Reproduced from Pauley, Barford & McDaniel (2023)

The Collect phase in 1 references the launch of EC2 instances deployed across multiple regions of Amazon Web Service's infrastructure. In 2, DNS related data is enriched to each network event upon arrival, real-time. Data is aggregated daily (3) and then goes through additional enrichment steps, such as the TCP flow tuples enrichment/annotation (4), resulting in exported JSON data products that can be analysed with a custom analysis routine (5).

DScope is reported to have been running since 2021, while the deployment across multiple regions started in October 2022. They have collected data on 4.6 million AWS IPv4 addresses spanning 105 thousand /24 subnets, but opted to constraint the dataset reported on the experiment to one week, from October 11-17, 2022 for a comparative analysis with Merit's telescope.

Their results reveal the use of sixteen AWS regions, including five in Asia-Pacific, five in Europe, five in North America, of which four in the US and one in Canada, and one in South

America. The structure used on the report unfold into four main contributions: The first refers to the quantity of observed scanners, where Orion's large number of unique IPs stands as a benefit, but revealing that DScope receives cloud-targeted traffic not seen by Merit's counterpart. The second refers to telescope avoidance and cloud targeting, where they measure scanner detection split by telescope type: by Merit's Orion alone (92,7%), DScope alone (2,8%) and by both (4,5%). The third aspect of their results investigates the distribution of random scanning, leading to the insight that *actual scanner behaviour is strongly consistent with random untar-geted scanning*, among other valuable insights. The fourth aspect of exercises the advantages of using cloud-backed recycled IP addresses, resulting in 136 thousand distinct source addresses probing DScope, against 190 thousand contacting Orion, among a range of valuable statistical insights, including the following findings:

- The cloud-based mechanism received 73% more traffic than the on-premises counterpart;
- Cloud IP traffic is 95% more variable than darknet traffic;
- Scanners target cloud IP ranges and/or avoid darknet ranges;
- Scans targeting standard telescopes are primarily random;
- Traffic increases over time after instance deployment, to a point;
- Short-lived use of IP addresses maximizes economical yield of new behaviour;
- Extended measurement on a given IP is not necessary to achieve high coverage of events.

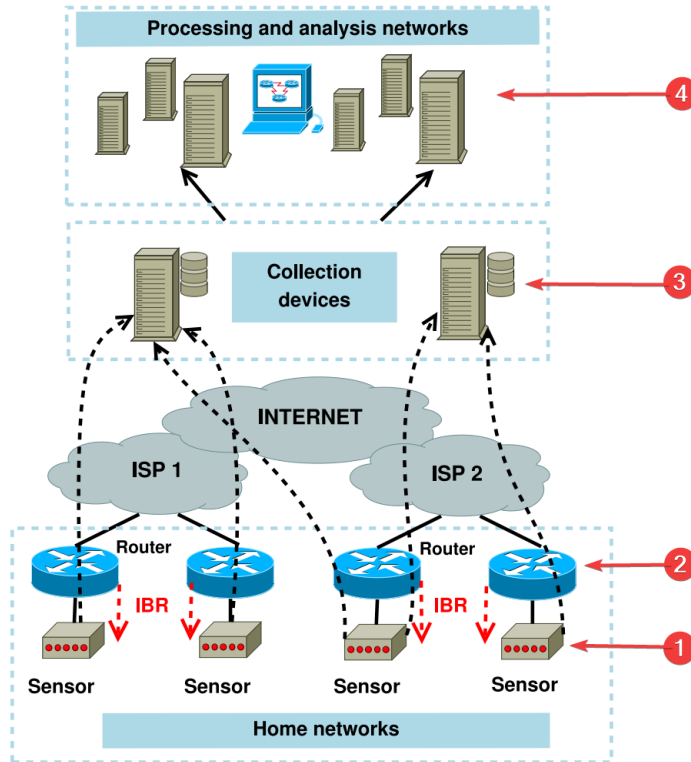
3.20 RODOLFO GARCÍA-PEÑAS' HODINT

In 2024, García-Peñas, supervised by Rodríguez-Gómez and Maciá-Fernández proposed HoDiNT: HOMe DIstributed Network Telescope - A distributed architecture for the acquisition of Internet Background Radiation that implements a low-cost, easy to install, two-phase attack detection mechanism. It is based on the installation of probes within domestic routers and able to report traffic in a distributed manner to a collection back-end (García-peñas; Rodríguez-gómez; Maciá-fernández, 2024). The HoDiNT architecture is shown in Figure 14.

1 depicts each capture sensor responsible for receiving IBR traffic from the router, processing it real-time, selectively responding to traffic of interest, and uploading the captured data to one of the Collection Devices (3). Raspberry Pi devices are used as the proposed standard collection mechanism.

Each capture device reacts to traffic sent by the domestic router on which it is inserted (2). Traffic that is not initiated by uses connected behind the NAT-enabled Local Area Network is considered unexpected, hence Internet Background Radiation. Forwarding is possible due to the existence of a DMZ (DeMilitarised Zone) feature available on most home routers. DMZ

Figure 14 – García-Peñas’ architecture for HoDiNT.



Source: Reproduced from García-Peñas, Rodríguez-Gómez & Maciá-Fernández (2024)

allows all unsolicited traffic to be routed to a LAN IP address of choice, and fully enable radiation capture. Alternatively, authors suggest reverse port opening via Universal Plug and Play (UPnP), or manual port forwarding as a last resource.

3 depicts a distributed server back-end, these computers receive the recordings from the sensor fleet and save for future analysis, while 4 reveals the existence of an analytics back-end where data is processed and analysed. Authors report the use of the ELK (ElasticSearch, Logstash and Kibana) stack to allow for visual data querying and analysis.

Interestingly, they proposed a Python/Scapy active responder to exchange TCP, UDP and ICMP packets with probing sources. This is called “active mode” and simulates the existence of honeypot-like services to collect additional attack information. They evaluate the solution from four perspectives: performance analysis with `hping3` and synthetic traffic; a traffic comparison using passive and active modes; port analysis breakdown; and attack detection patterns as detectable by standard rule-sets implemented by intrusion detection systems. Traffic was collected for 31 days using 10 sensors, 5 in active mode and 5 in passive mode.

The TCP destination port breakdown analysis, of particular interest to this project, reveals a familiar pattern, indicating a prevalence of port 23 (Telnet) as the most exploited on both passive and active sensors mode, followed by 445 (SMB), 22 (SSH), and web ports including 80, 443 and 8080, as the preferred targets, among others.

HoDiNT authors cite the Cloud Telescope, as documented in Bortoluzzi et al. (2023), describing it as “an alternative method for creating network telescopes using the cloud as a

back-end”.

3.21 COMPARATIVE ANALYSIS OF CONCURRENT WORKS

An attempt to compare these three concurrent research efforts against the Cloud Telescope must consider key aspects in terms of method, results and timing. The first of the three works turned public was O’Hara’s, in April 2019. The work stands as a documentation of a research effort in which the infrastructure of a Cloud Computing Service provider is used to host an open capture mechanism willing to collect radiation arriving to any of their interfaces, assigned with a multitude of public IP addresses. The work therefore validates the hypothesis that radiation collection can be implemented with at least one provider, even when considered the default-deny posture typically adopted by them, or the provider’s need to block malicious traffic as part of the cloud provider’s shared responsibility model where they may implement low-level defences to protect the intra-structure before any content is actually routed towards the consumer’s virtual private segment. They also shed light on what to expect in terms of baseline costs associated with the deployment including the capture mechanism and accessory costs related to IP allocation, storage and the related.

O’Hara’s work was not used as a related work essentially because it escaped the search filters implemented by the systematic approach as master theses tend to only get published within each institution’s e-print systems or equivalent mechanisms. That aside, O’Hara’s approach to building the capture fleet comprising the cloud-based mechanism dramatically differs in fundamental ways than intended by this Cloud Telescope. The most notable difference is that O’Hara decided to make use of a single capture instance, bound to a single location within AWS, and has added to that single instance the maximum amount of allocable IP addresses AWS will let one bind to a single virtual computer. The Cloud Telescope documented on this thesis, on the other hand, implements a geographically distributed capture fleet on as many different regions of the world as made possible by the cloud providers regions and availability zones. Only the Cloud Telescope captures traffic arriving to two or more different countries simultaneously, enabling unique ways to infer geopolitical trends or motivation.

An attempt to compare DScope and this Cloud Telescope, as built, indicates they have many similarities. DScope is documented to work within AWS’s infrastructure, so does the Cloud Telescope. DScope is reported to work within sixteen regions of the world, while the three deployments made with the Cloud Telescope this far captured radiation arriving to twenty-six different regions. DScope offers the ability to capture TCP-based application-layer data exchange up to the service banner level, using a customised responder one-catches-all service ports. The Cloud Telescope does a very similar job. It implements a responder on one port and a NAT technique is also used to collect application-level traffic arriving on selected ports, notably the ones typically exploited by active botnets, such as Mirai.

Some similarities between the two mechanisms are remarkable. They both reported an initial estimate of around 70 US Dollars as a baseline cost to get the architecture running,

DScope for one week, the Cloud Telescope, for one month. DScope interacted with 190 thousand different source IP addresses on the main experiment, while the Cloud Telescope received traffic from 270 thousand sources on the first 30-day proof-of-concept run. DScope’s operation is documented in five steps starting with the capture, going through some real-time enrichment procedures, then an aggregation phase that includes the use of GeoIP and a final step that outlines an analysis framework. The Cloud Telescope is documented as a four-step procedure that begins with the same capture phase, followed by an aggregation phase, an enrichment phase also based on GeoIP and a visualization step at the end of the pipeline. Both architectures were implemented to react to the Instance Termination Event sent by the cloud back-end whenever outbid as imposed by the cheaper Spot price model.

On the other hand, DScope and the Cloud Telescope differ in a range of particularities. DScope is thoroughly documented based on statistical methods, where the collected data is compared with a benchmark, darknet telescope, providing ways to contrast traffic hitting the cloud-based sensor fleet in opposition to the traffic arriving to a Merit’s telescope. The Cloud Telescope makes use of a descriptive approach where captured data is aggregated in ways that enables a discussion of the cybernetic properties the most relevant events carry. Data aggregation and dataset split are used to support the idea that it is possible to cross-link radiation events captured by the mechanism with geopolitical events taking place outside the cybernetic realm. DScope is mainly coded in Python and Go, while the Cloud Telescope is orchestrated in Terraform and Bash to implement a very similar behaviour. DScope is described to export events in JSON format. The authors claim that the dataset *allows researchers to effectively measure adversarial behaviour targeted at cloud infrastructures*, but do not mention wherever the dataset or JSON exports are available to public scrutiny. They also mention the existence of a DScope binary that is copied onto each capture sensor, but no access to the binary or supporting scripts is actually provided on the reference paper. The terraform architecture used to deploy the Cloud Telescope exists as a publicly-accessible GitHub repository since its inception, while the three datasets produced by each Cloud Telescope launch were published on IEEE Dataport beginning February 2024.

Finally, HoDiNT is the most recent contribution to the field. It arose in 2024 and proposes a distributed method that relies on home routers rather than the cloud. It is inspiring in many ways. It resembles the spirit once implemented by SETI@Home. a scientific experiment, based at UC Berkeley, that used Internet-connected computers in the Search for Extraterrestrial Intelligence (SETI) in the sense that home computer capacity can be used to solve complex, time-consuming problems. Both HoDiNT and SETI@Home can be considered classic distributed systems at inspirational and practical levels. The investment to build the distributed HoDiNT fleet differs from the cloud-based solutions in the sense that it only requires the acquisition of a Raspberry PI device (capital expense), assuming that the Internet connection pre-exists for productive reasons. A comparison of each concurrent work against two main publications resulting from this thesis is provided on Table 5.

The chronology of concurrent research events, as detectable by the candidate, is itemised

Table 5 – Comparison of concurrent works

Key properties	O’Hara (2019)	Bortoluzzi (2023a)	Pauley (2023)	Bortoluzzi (2023b)	García-Peñas (2024)
Publication date	April 2019	June 2023	August 2023	November 2023	February 2024
Document format	Master Thesis	Short Paper	Full Paper	Full Paper	Full Paper
Environment	Cloud	Cloud	Cloud	Cloud	Residences
Location	Single Region	Multiple Regions	Multiple Regions	Multiple Regions	Distributed
Device	Single virtual	Multiple virtual	Multiple virtual	Multiple virtual	Multiple physical
Expense	Operational	Operational	Operational	Operational	Capital
Architecture	Documented	Architectural plan	Documented	Documented and Reproducible	Documented as firmware image
Dataset	Documented Not available	No dataset	Documented	Documented and Available	Documented
Infrastructure as Code	Not specified	Planned	Not specified	Available and Reproducible	Downloadable firmware

below.

- 2019 April: Joseph O’Hara publishes “Cloud-based network telescope for Internet background radiation collection” on Trinity’s College repository;
- 2023 March: This candidate’s paper “A cloud-native framework for globally distributed capture and analysis of Internet Background Radiation” (Paper #1) is accepted at CISTI 2023;
- 2023 March: The first Cloud Telescope capture is deployed (Dataset #1). Duration: 30 days.
- 2023 June: Paper #1 is presented at the University of Aveiro, Portugal.
- 2023 June: This candidate’s paper “Cloud Telescope: A distributed architecture for capturing Internet Background Radiation” (Paper #2) was submitted at IEEE CloudNet.
- 2023 August: Paper #1 is indexed on IEEE Explore;
- 2023 August: The second Cloud Telescope capture is deployed (Dataset #2). Duration: 45 days.
- 2023 August: Eric Pauley et al. present “DScope: A Cloud-Native Internet Telescope” at Usenix Security Symposium;
- 2023 September: Paper #2 was accepted at IEEE CloudNet.
- 2023 October: The third, long-term Cloud Telescope capture is deployed (Dataset #3). Duration: 151 days.
- 2023 November: Paper #1 is presented at Stevens University in New Jersey, USA.
- 2024 February: Datasets #1 and #2 are pushed onto IEEE Dataport;

- 2024 February: García-Peñas publishes “HoDiNT: Distributed architecture for collection and analysis of Internet Background Radiation” on Elsevier’s Computer Networks;
- 2024 March: The third, long-term capture ends (Dataset #3).
- 2024 April: This candidate’s paper “Quantitative analysis of botnet activity within Internet Background Radiation as observed by a Cloud Telescope” (Paper #3) is accepted at CISTI 2024.
- 2024 June: Paper #3 is presented at the University of Salamanca, Spain.
- 2025 March: This candidate’s paper “Cloud Telescope: An Ephemeral, Distributed, and Cloud-Native Architecture for Collecting Internet Background Radiation’ (Paper #4) is accepted and published in IEEE Access.
- 2025 May: This very section (3.17) is compiled and written.

It is worth mentioning that the candidate was able to document and submit the plan to implement *a cloud-native framework for globally distributed capture and analysis of Internet Background Radiation* five months before Pauley’s work was presented on the Usenix conference, in August. The candidate’s paper outlines the work-in-progress intending to launch an instance fleet of up to 1,500 EC2 sensors distributed across fifteen regions of the world and making use of the spot price scheme to lower down costs at a minimum. Also, the detailed study describing the results of the first Cloud Telescope deployment was submitted by this author to IEEE CloudNet 2023 in June, two months before Pauley presented their work at Usenix.

One fair conclusion is that the originality criteria was never broken in any of the cases. O’Hara proposed the use of a Cloud back-end for the first time in 2019, but using a single compute instance, while Pauley and this candidate concurrently worked in blind until each one emerged, in 2023, with a geographically distributed architecture able to capture radiation arriving to multiple regions of a cloud service provider.

To conclude, the chronology observed in the series of events related to the publication of this candidate’s Paper #1 and acceptance of Paper #2 before others, can be used to sustain the preservation of the originality criterion demanded by a terminal degree research study. More importantly, the fact many research groups are looking into ways to distributedly capture Internet Background Radiation mutually reinforces the understanding of a shared goal based on common needs.

4 CLOUD TELESCOPE

This chapter describes the architecture of the Cloud Telescope. It presents the key functional and non-functional requirements, the most significant business rules for the deployment within a cloud service provider, and insights into operational costs. Some of the terminology follows Amazon Web Service naming scheme, which can be adapted for other cloud service providers with same functionality.

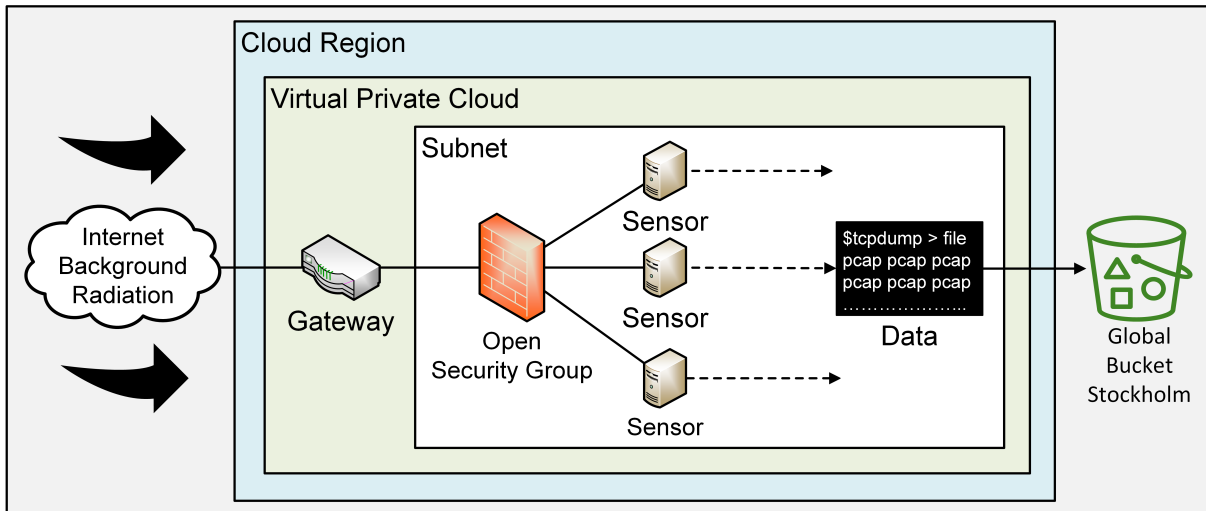
The Cloud Telescope is fundamentally based on the following assumptions:

1. A cloud-based architecture shall enable for Internet Background Radiation collection in a similar, yet substantially different approach, than used for the deployment of a standard Network Telescope;;
2. The geographic topology of a CSP shall provide unmatched, distributed cyber threat observability, of events taking place in different regions of the world; A Cloud Service Provider materialises an ideal approach in terms of globally distributed capillarity for any purpose, including unsolicited packet capture; The distributed architecture, made possible by the operational nature associated with the geographical dispersion of cloud providers, should enhance the detection of events that are also carried out in a distributed fashion. An example is the botnet, aimed at compromising targets on a global scale, potentially motivated by geopolitical events originating outside the cyberspace.
3. The availability of IPv4 network blocks is scarce. Primary allocations are no longer available and there is pressure to put idle blocks to productive use (Livadariu et al., 2018);
4. Idle computing capacity inherent to the cloud computing model shall provide a relevant mix of geographically distributed IPv4 address ranges for radiation collection, at acceptable cost;
5. Small aperture telescopes should retain acceptable statistical relevance for IBR capture, as long as increased observation duration compensates for potential address deficit (Chindipha; Irwin; Herbert, 2019); and
6. The ephemeral and reproducible properties inherent to the design of the Cloud Telescope enables peer validation of its implementation.

Figure 15 summarises the final Cloud Telescope Architecture. IBR capture sensors are deployed within each cloud region to enable the capture in a distributed manner, from both geographical and topological perspectives.

Each viable cloud region is configured to contain the right set of architectural elements to support the distributed capture. The most relevant setup aspects include the Virtual Private Cloud, the Internet Gateway, Routing Table, Security Group, Access Control List, sensor setup, and persistent storage setup, with the use of one single global cloud bucket. AWS

Figure 15 – Cloud Telescope Architecture



Source: The author.

terminology is used when needed, i.e. bucket instead of blob storage, but the architecture itself is designed vendor-neutral. The VPC is enabled to logically group virtualised capturing infrastructure resources, and to which the routing table, access control list and security settings should be applied. A single, internet-facing Internet Gateway is enabled and attached to the VPC. This should be configured with the addition of the corresponding route to sub-net's route table that directs unsolicited internet-bound traffic to the sensor sitting in the sub-net's range.

The capture sensor, which must be virtualised in a cloud environment, receives a non-public, internal address, according to RFC 3330 (Iana, 2002), as an imposed default. It is optional to enable one or more external, public, or *elastic* IP addresses. As it pertains to the capturing platform, every sensor should receive one IPv4 address from the CSP region's address pool. The question of whether each sensor should have only one public IP address is open for experimentation and is mainly implemented as such due to a cost-related perspective. From a performance perspective, one sensor should be able to handle IBR affecting a number of public IP addresses, especially when considering the performance associated with the variety of virtual computer classes offered by cloud providers. Launching a number of small, individually-sized sensors with a single public address each is significantly more cost-effective than assigning multiple IP addresses to a single instance, especially when dealing with tiny-sized virtual machines where the cost of an additional public IP address is greater than that of launching a second instance.

The approach for the security group and related ACLs in the cloud is not as permissive as the on-premises counterpart, in which typical network telescopes operate under a default-allow policy posture. Internet-Gateway-enabled VPCs default to a security group that blocks all internet-originated traffic, with the exception of certain web and remote admin port ranges: 22/TCP (SSH), 80/TCP (HTTP), 443/TCP (HTTPS), and 3389/TCP (RDP). Another exception is the traffic that is originated within the very VPC's subnet, which is usually allowed, for many

reasons, including serving end-user applications, systems update, among others. Radiation capture demands a fully-open, default-allow rule-set. The sensor must get fully exposed to incoming, unsolicited Internet traffic. Therefore, a fully-open security group must be declared and attached to the VPC so that:

- All IP types are allowed in via the public interface. This should include all forms of exotic and unexpected traffic, also SCTP and IPSec packets, among others;
- All ICMP traffic is allowed in via the public interface. This should include all known types, free from any sort of rate limiting technique, either imposed by the Cloud Provider or implicitly defined as a best practice default;
- All transport-layer protocols are allowed in via the public interface. While TCP and UDP are the most prevalent, the capturing scheme should remain open for exotic types, such as the emerging QUIC protocol, among others. State filtering should be rejected, in the sense that malformed TCP flags are allowed. No kind of normalisation or traffic conditioning, such as the normalisation procedure thoroughly documented by (Openbsd, 2025) should be implemented, either by explicit user-accessible commands, or by enforced measure resulting from the shared responsibility model within the Cloud Service Provider;
- The Cloud telescope inherits the behaviour of a Network Telescope, in which no application-layer communication is effectively established. Therefore, no special configuration is required for the application layer. As a default, there is no engagement with any sort of application-layer. However, it shall be easy to implement high-level responders as widely documented (Moubarak; Bassil, 2020; Zeid; Moubarak; Bassil, 2020; Soro et al., 2023); and
- Conversely, in case there is a need to implement active, application-layer responders, this should be possible with the use of commodity software stacks, including bash, python and/or rust-based solutions configurable via Infrastructure-as-Code.

4.1 CLOUD SENSOR DESIGN

The design choices for the Cloud Sensor revolve around decisions affecting instance type, machine class, pricing model, and the ephemeral nature of the method. These are accompanied by software component choices including the selection of an operating system, application stack and automation of the capture life-cycle.

AWS offers a low-cost, general-purpose *T3* instance class that is powerful enough to handle Internet Background Radiation acquisition. The *T3.nano* class, equipped with only 512 MB of RAM, provides an acceptable performance at the lowest cost. The main challenge associated with this decision is to ensure no data loss occurs due to small memory size, leading to potential packet capture drop. To minimise the associated risk, Alpine Linux, rather than

Amazon Linux, was assigned as a suitable light weight solution. Alpine is a minimalist Linux distribution based on musl, a lightweight libc interface implementation, and the compact busy-box user-land, instead of a regular fully-fledged Linux stack, resulting in a practical choice for operating system virtualisation and containerisation (Tazaki et al., 2021).

To keep sensor costs at a minimum, EC2 virtual instances must be acquired under the *spot* pricing model. The scheme provides predictable prices that adjust to current offer and demand, per cloud region. Typical savings are of 70-90 percent when compared to the standard *on-demand* offer (Pary, 2023). This comes with the need to properly handle termination requests on the Cloud Service Provider's (CSP) side, where the price currently paid for an instance is outbid by concurrent clients or a momentary lack of spot instances. Each sensor is therefore prepared to receive and handle the EC2 Spot Instance Interruption Notice, available through each instance's internal metadata service. Upon interruption request, and during the 2-minute grace period, the sensor will adequately terminate the capture process. It will then move the last chunk of recorded data to the external persistent storage: the global S3 bucket. Orchestration will subsequently launch a replacement instance, updating the winning bid price accordingly, returning the collection platform to the desired operational state. Despite the inevitable momentarily continuity loss caused by the method, it is found to be acceptable as a newly launched instance replaces the older sensor. The fact that the new sensor will acquire a new IPv4 address contributes to a wider diversity of observed addresses.

During normal operation, each sensor is configured to comply with the following fundamental requirements:

- R1 A fixed number of sensors should be deployed in each viable region of the cloud service provider.
- R2 There should exist a single, global, and private storage target, in the form of cloud storage, or *bucket*, to which recorded data must be moved to;
- R3 `tcpdump` is demonised, and shall rotate capture files every 24 hours;
- R4 Capture files should be uploaded to the target bucket immediately after rotation; and
- R5 SSH service is reassigned to port 65535/TCP and should be omitted from the capture.

The resulting fleet of N sensors per region, multiplied by the number of viable cloud regions should simultaneously listen and record the arriving radiation. The Stockholm region was chosen as the target for the unified data storage because Sweden implements strong GDPR (General Data Protection Regulation) enforcement, regulated by *Datainspektionen* (Integritetsskyddsmyndigheten, 2024), within the European Union. This can be changed as an environment variable. Implementation of the PCAP file rotation can be achieved via `-G` parameter in `tcpdump`. Capture chunk upload is done by activating a native API interaction with the global bucket, backed by internal role-based access and authorisation, preventing unnecessary

public bucket exposure. Finally, in case administrative access is needed to one sensor, for occasional inspection, that can be done via port 65535/TCP, using SSH configured to require strong key-based authentication.

The design and implementation of the Cloud Telescope is described as a set of functional requirements, non-functional requirements, one business rule, and key aspects of its architectural implementation.

4.2 FUNCTIONAL REQUIREMENTS

These are the key functional requirements applicable to the Cloud Telescope.

- FR1 All filtering capabilities provided by the cloud service provider must be disabled, including security group standard posture and default-deny access control lists implemented as part of the cloud computing shared-responsibility model;
- FR2 One cloud-native object storage facility: a cloud bucket, must receive and aggregate all traffic recorded by each sensor within the sensor fleet;
- FR3 Any packet arriving to a cloud telescope sensor must be recorded, except the traffic eventually generated by the sensor itself, when uploading packet recordings to the aggregation bucket;
- FR4 Packet capture recordings must be compressed and uploaded to the aggregation bucket every 24 hours, or configurable shorter time period;
- FR5 Port 65535/TCP should be assigned to the local, emergency-only SSH service, and kept out of radiation capture; and
- FR6 The format of the capture recording must be the PCAP file format, compressed with gzip, by the sensor, before bucket submission.
- FR7 A full EC2 metadata content snapshot dump should be recorded to a text-file to enable future investigation, including the public IP address, cloud region, and number of packed drops, if any;
- FR8 Each sensor date and time should be set to Universal Time Coordinated, UTC;

4.3 NON-FUNCTIONAL REQUIREMENTS

The non-functional requirements refer to the most important technical decisions made to optimise the sensor fleet's operation, in particular, the measurements put in place to take advantage of the spot pricing model allowing for up to 90% discount in instance/sensor deployment.

- NF1 Low-cost AWS EC2 t3.nano instance type should be employed to run the sensor stack;
- NF2 Low footprint Alpine Linux with lightweight musl library should be used to prevent packet drop;
- NF3 Sensors must implement code to respond to the “EC2 spot instance interruption warning” event.
- NF4 The sensor fleet should be launched under the spot pricing scheme.
- NF5 Architecture must be described in vendor-neutral language Terraform.

4.4 BUSINESS RULES

There is one relevant business rule: Upon receiving the interruption warning notice, the sensor must stop the capture procedure, compress the last recorded chunk and upload the resulting file to the aggregation bucket. A new instance should be immediately launched by winning the highest bid of the spot pricing model again at a slightly higher hourly cost rate.

4.5 ETHICAL CONSIDERATION

Internet traffic collection, as performed by Network Telescopes and as intended by the Cloud Telescope, require responsible data handling, as these capture mechanisms intentionally record network packets arriving to sensors. The metadata that is required to describe a network packet always gets recorded. This necessarily includes: the source IP address, the target IP address, source and destination ports, if TCP or UDP, source and destination ports, if TCP or UDP, and message type, if ICMP. Payload collection, on the other hand, may or may not occur, depending on the strategy that is adopted to implement the capture, and varies according to the desired level of interaction. A fully passive sensor only captures these *envelope metadata* of each incoming packet, whereas a high-interaction, honeypot-style responders record a number of application-layer messages actively exchanged with the attacking source, revealing far greater details about the apparent attacking source.

Researchers at CAIDA (Kenneally; Claffy, 2009) have proposed a comprehensive Internet data sharing framework aiming at balancing privacy and utility of network traffic measurements. We have conducted the experiment reported on this paper in alignment with the ethical considerations discussed by the framework, acknowledging its key properties:

Motivation. Internet researchers lack a legal framework that prescribes, incentivises or forbids the sharing of network measurement data. We share our results, in the form of this paper and the companion dataset as an effort to pragmatically contribute to a better understanding of current cybernetic threats, to foster the development of better defence mechanisms, and towards a generally safer cyberspace.

Challenges. Privacy is hard to quantify, as is the utility of empirical research. Both variables are dynamic and lack normative understanding among domain professionals. Our understanding follows peer researchers in the sense that IBR or darknet traffic essentially contains non-productive, malicious, often artificially generated network packets. This is debatable, but supported by the literature.

Sharing risks and benefits. The framework proposes the addressing of three key questions to balance risks and benefits in this type of research:

- Who is at risk when network data is shared? This refers to persons and their PII (personally identifiable information). ISO 27018 defines PII as any information, such as name, social security number, physical and email addresses, certain biometric markers, which can be used to establish a link between the information and a natural person. Benign network traffic may contain PII, whereas malicious traffic is unlikely to contain such sensitive data;
- Which traffic data components are privacy-relevant? This splits into *first-order* and *second-order* identifiers. The former refers to data that functionally distinguishes an individual: first and last name, social security number, government-issued and other account identifiers, physical and email addresses, certain biometric markers, and possibly the same information about immediate family. The latter refers to an IP address (IPA), machine access code (MAC) address, host name, birthdate, phone number, zip code, gender, and financial, health, or geographic information; and
- Under what conditions do these data types pose risk? Network traffic measurement data can present a privacy risk when information in packets can directly expose non-public information about persons - such as health, sexual orientation, political affiliation, religious affiliation, criminal activity, associations, behavioural activities, physical or virtual location.

The passive-only nature of the experiment constrains the ability to obtain personally identifiable data from the companion dataset. The enrichment procedure used in our intelligence extraction data pipeline is limited to the use of *Open Source Intelligence* libraries and APIs, enabling for some second-order identifiers, notably, the source and destination countries, obtained with libgeoip and the targeted AWS region, with the awsregions Python module. All results discussed ahead are aggregated by a high-level filter, such as by Country, or by AWS Region. There is no linking-back to any form of first-order attribute that can identify an individual or company, or region of the world in closer detail.

4.6 ARCHITECTURE IMPLEMENTATION

The Cloud Telescope is described as Infrastructure-as-Code (IaC) using a combination of Terraform, and Bash. IaC is a by-product of the Devops field. The concept refers to the

purpose of using code to describe infrastructure elements, including virtual machines, and their CPU, RAM and storage needs, along with network cards, switches, and cloud resources. Cloud resources include the setup of a Virtual Private Cloud, internet gateway, DHCP server, access control list, filtering/firewalling rules, and cloud-native resources, such as a storage bucket or a load balance, for instance (Smith, 2022).

The operating system of choice, along with configuration, package installation, setup and tweaking, can all be accomplished via IaC, using either operating system native scripting language, such as Bash and Powershell, or some sort of orchestration-specific language, such as Ansible. The Cloud Telescope is implemented in the Terraform IaC Language, tailored to operate within Amazon Web Services as a CSP. The resulting Terraform Infrastructure-As-Code artefact reflecting this architecture was implemented by Beiler (Beiler; Bortoluzzi, 2024), and is available.

Figure 16 illustrates the compute instance launch configuration for one of the cloud regions.

Figure 16 – Terraform compute instance launch definition.

```

1  module "spot_us_east_1" {
2      source = "../modules/ec2"
3
4      instance_count      = 5
5      iam_instance_profile = local.iam_instance_profile
6      user_data            = local.user_data
7      spot_price          = local.spot_price
8      instance_type       = local.instance_type
9      volume_size_gb      = local.volume_size_gb
10     ssh_keypair_name     = local.key_name
11     ssh_public_key       = tls_private_key.private_key.public_key_openssh
12
13     providers = {
14         aws = aws
15     }
16 }

```

Source: (Beiler; Bortoluzzi, 2024)

Line 1 indicates the configuration applies to the spot pricing scheme applicable to us-east-1 region in the United States of America. Common resources are imported in Line 2. The configuration then describes the need to launch five instances into that region, on Line 4. Lines 5 to 11 feed the configuration with global variables, including an Identity and Access Management profile, the bash script containing operating system automation, under `user_data`. Spot price allocation is negotiated as rules within `spot_price`, CPU and RAM are described as a T3.nano instance class within `instance_type`, storage needs, in the amount of 8 GB, under `volume_size_gb`, and SSH keypair in `key_name`, followed by the corresponding public key onto variable `ssh_public_key`.

Terraform code is used to describe the building, launching and instance orchestration across the range of cloud regions, along with the dedicated cloud bucket allocation in eu-north-1. Bash is employed to script the required automation within the virtualised operating system belonging to each sensor, including:

- SSH port redirection;
- Time-zone set up to UTC;
- Disabling unnecessary services that could generate unwanted traffic or noise;
- A daemon process that implementing metadata collection and reacting to potential instance termination requests;
- A demonised, start-stop script for the tcpdump binary, enabling re-initialisation in case of abnormal termination;
- A customised packet drop monitoring service;
- Automation of data chunk uploading to the bucket every 24 hours, at most.

Terraform automatically fetches the current list of available cloud regions directly from the Cloud Service Provider’s back-end, so the telescope deployment can be set up to dynamically adjust to region growth if desired. Researchers wanting to reproduce the experiment should only need to setup a new IAM key/secret pair, followed by Terraform native commands to have the same architecture launched under their control. Code is as vendor-neutral as possible, however, changes will be required if moving from AWS to another Provider, especially in terms of bucket terminology, instance type and instance acquisition model, among other vendor-related technical and financial decisions.

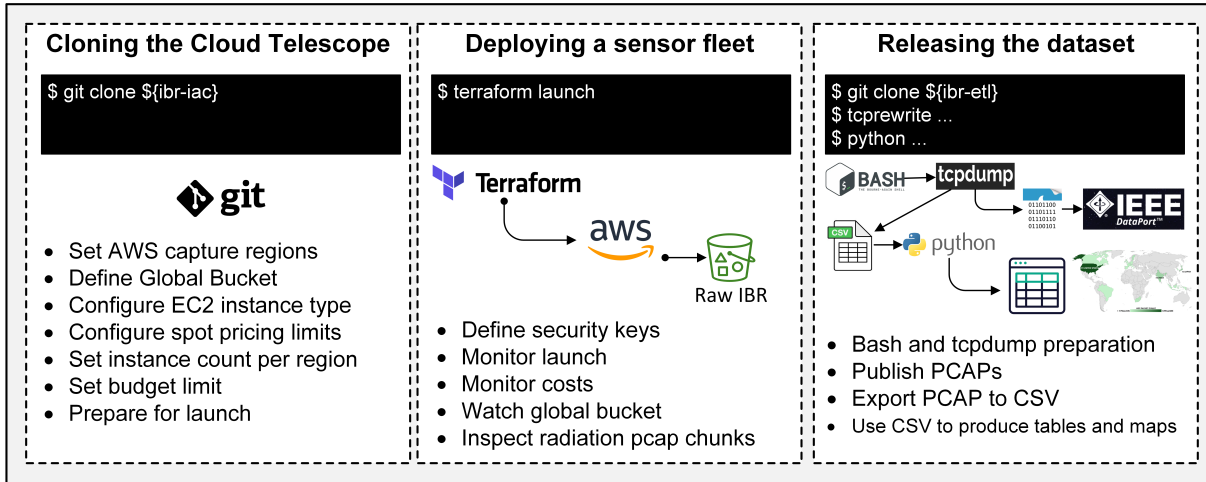
4.7 DEPLOYING AND REPRODUCING THE CLOUD TELESCOPE

A key aspect of the cloud-native design is reproducibility. The Cloud Telescope is ready to be redeployed either by the research group behind its inception or by any other interested party. This idea is illustrated in Figure 17.

A researcher willing to conduct new Cloud Telescope deployments must focus on three main steps:

1. The first step is to clone the Cloud Telescope architecture described in Terraform. This is done with git. Most of the configuration involves setting variables to express the desired deployment regions, setting the global destination bucket to receive capture chunks, confirm the preference for tiny-sized EC2 instances following the spot price regime and prepare to launch;

Figure 17 – The Cloud Telescope as built



Source: The author.

- Effective deployment takes place with Terraform on AWS. The initial step is to create and configure the security keys Terraform uses to operate on the user's account within the CSP. The architecture is then ready for deployment, via ordinary console commands using the terraform system package. Right after launching, AWS will react to terraform commands and radiation chunks will be sent out to the globally defined storage bucket. This phase endures for as long as the defined capture time frame and it is important to monitor cloud usage, especially systems uptime, bucket storage growth and budgeted vs effective cloud service costs.
- Once deployment is terminated, the dataset can be worked towards enrichment and data extraction or simply prepared for publication. Enrichment is achieved by a combined use of bash scripts and python programs that will read data as it exists on a previous step and produce new data with some sort of enrichment, such as geographical position of attacking sources. The result is then used to produce latex-formatted tables, heat maps and more, as desired by the researcher. On the other hand, enriched PCAPs can be compressed and sent directly to publication. The publication venue of choice for Cloud Telescope IBR datasets is IEEE DataPort. They will host compressed pcap-formatted files of sizes up to 2 terabytes per project and generate corresponding digital object identifiers to link back to companion research by-products.

4.8 BASELINE BUDGET

Operational costs related to running the required sensors to support the Cloud Telescope are mainly influenced by the pricing model being used, fleet size, capture duration, and the bucket-based persistent storage, to a smaller extent. AWS Pricing Calculator spreadsheet-based calculation tool was used for an estimation of costs associated with launching a given architecture. The final cost associated with running the month-long observation of IBR using

130 sensors per region, across the 26 commercially available regions of the Amazon Cloud was of 345.80 United States Dollars (USD) per month, as detailed in Table 6.

Table 6 – Architecture budget - monthly costs

Sensor cost	Per sensor	Fleet of 130
Hourly	0.0038	0.494
Monthly	2.65	344.50
Storage	0.01	1.30
Total	2.66	345.80

The final cost associated with launching the Cloud Telescope was, therefore, estimated at 1,729 USD for five months of capture. In February 2024, AWS announced they would start charging for the first public IPv4 associated with any deployed virtual machine, which doubled the cost of the architecture for that month. The effective final cost for the full-scale, 5-month-long (151 days) experiment, was of 2,429.29 USD, averaging to 16.08 USD per day.

4.9 NETWORK TELESCOPE AND CLOUD TELESCOPE COMPARED

Network Telescopes and the Cloud Telescope both aim at the same purpose: to collect internet background radiation. The key difference is that the Cloud Telescope is geographically distributed, and meant for a deployment within cloud service providers. A summary of differences is itemised in Table 7 and discussed ahead.

Table 7 – Network and Cloud telescopes compared

	Network Telescope	Cloud Telescope
Typical deployment	Research institutes	Public CSPs
IP address range	CIDR block	One IP per sensor
Capillarity	One physical location	Multiple cloud regions
Capture Device	Physical computer	Ephemeral VMs
Storage target	Local HDD, NAS	Cloud bucket
Expenditure	Capital expense	Operational expense

Network Telescopes are usually deployed within research institutes and universities. This is not only due to the fact Internet Background Radiation is a research subject, but also because these are the places most likely to hold unused IPv4 address blocks, either partially or in full, and can afford to purpose these addresses for research intents. The Cloud Telescope is meant to be deployed within public Cloud Service Providers. This means any individual or organisation can benefit from the idle IP address ranges the CSP must hold as a business rule, and launch sensor fleets to conduct new observations.

As reviewed, the IPv4 address ranges assigned to Network Telescopes can vary from one or more large /8 blocks to interleaved portions of a single /24. On the other hand, the Cloud

Telescope counterpart has been experimented with an IP address count as small as “one IP address per each of the 26 available cloud regions” (26 IP addresses) or as large as an aperture of “ten IP addresses per available cloud region” (260 IP addresses). The former is constrained by the network block assigned by the local Internet Authority, ultimately, IANA. The latter is constrained by the budget that will be allocated to support the Cloud Telescope deployment, and should appear to come from an infinite IP address pool provided by the cloud company chosen to host the experiment.

In terms of capillarity, Network Telescope deployments are physically bound to a single host institution and ultimately to a single country to which the network block was initially assigned. The Cloud Telescope, on the other hand, benefits from the ability to distributedly listen and capture unsolicited traffic arriving to a mix of IP addresses belonging to many different regions of the world. This variation is relevant when discussing two potential aspects driving the background radiation: the coding practises used in the implementation of host scanners, malware and related threats, and geopolitical motivation. A common approach to the implementation of cybernetic threats is to probe the next host as increment to the current one. For example, after probing 203.0.113.20, the next obvious target is likely to be 203.0.113.21. After 203.0.113.255, the next is probably 203.0.114.1, and so on. While Network Telescopes listen to contiguous network blocks, they are likely to capture such scan attempts in whole, whereas the Cloud Telescope counterpart is more likely to capture only fragments of the same activity, but in increased quantity, as the IP ranges are scattered around the globe. Assuming that actors behind internet-wide attacks might be interested in one or few particular countries, the radiation sample coming from multiple regions of the world, as provided by the Cloud Telescope, is meant to enable the comparison and contrasting of how differently each region is attacked.

A discussion on Network Telescope deployment costs typically revolves around defining a budget to the hardware acquisition, in the form of a single capturing device, and attached on-premises data centre costs, usually shared with other, production oriented, server fleets, and attached costs such as electricity, cooling and physical security. Yearly rates applied to the concession of IP address blocks may also apply. These are typically diluted among common capital and operational expenses within the hosting institution’s needs, and hardly detailed in peer research. On the other hand, the Cloud Telescope fits the “pay-as-you-go” consumption model that defines Cloud Computing. The cost is proportional to the sensor fleet size, or to the telescope’s aperture. In essence, the CSP charges for the virtualised compute instance, for which the cost vary according to CPU specification, RAM memory space, primary storage size, and the total amount of bytes exchanged over the public internet during the period. The target final storage bucket aggregating the captured sample is also of special interest, and charged by the amount of gigabytes effectively spent on a monthly basis. Accessory costs may include the cloud native secrets management service, traffic surcharge related to uploading data chunks to a bucket placed in a region other than the one hosting the sensor, and other management-related services. There is no charge for many pre-requisites, including the Virtual Private Cloud setup, Security Groups, Access Control Lists, Internet Gateway, and Terraform-related deployment

and shutdown actions.

5 PRELIMINARY RESULTS

This chapter summarises the intermediate results achieved with the design and implementation of the Cloud Telescope as a means to collect Internet Background Radiation. The initial design of the Cloud Telescope was documented in the form of a four-page, short paper and sent to the doctoral-degree-oriented track of the 18th Iberian Conference on Information Systems and Technologies, held in June 2023, in Aveiro, Portugal. The purpose of this submission aimed at opening the Cloud Telescope architectural design to public, peer-review scrutiny before allocating time to its implementation. The paper’s abstract is reproduced next.

Among the existing methods for analysing internet traffic, one focuses on unsolicited and often harmful packets, referred to as Internet Background Radiation (IBR). IBR data can be captured by using cloud-based computing instances to listen for incoming packets and recording the received headers and contents in files using the PCAP format. No services is run on these hosts so all traffic can be regarded as non-legitimate. This work presents the research plan for capturing a year-long sample of the IBR arriving to up to 1500 geographically distributed sensors across a major service cloud provider. In combination with collection, a data analysis pipeline will be constructed to enable the querying of most relevant quantitative and qualitative aspects of the resulting dataset. The ultimate goal is to answer “how is the IBR characterised when captured within the context of cloud computing?” Results will include packet distribution according to the properties of the network, transport, and application layers, together with an evaluation of traffic linking to botnet activity such as *Mirai* and *Moobot* (Bortoluzzi; Irwin; Westphall, 2023).

Full reference is as follows:

F. Bortoluzzi, B. Irwin and C. M. Westphall, “A cloud-native framework for globally distributed capture and analysis of Internet Background Radiation” 2023 18th Iberian Conference on Information Systems and Technologies (CISTI), Aveiro, Portugal, 2023, pp. 1-4, doi: 10.23919/CISTI58278.2023.10211290.

The paper introduced Internet Background Radiation by providing a review of the most relevant contributions in the field to the date, while pointing out a research method fitting under what Wazlawick (2020) defines as “non-experimental research¹, in which there is no systematic influence by the researcher”. It is further categorised as an objective investigation, where “any competent observer should agree with the findings and the conclusions resulting from the observation”.

¹ This research is non-experimental in the sense there is no data collection other than the background radiation itself. This work is, therefore, free from any sort of user-based opinion, where facts are coldly analysed by what they are, and represent.

The paper also documents the research plan in which two main forms of Internet Background Radiation analysis would be conducted. One of quantitative nature, as clearly indicated on related works, including traffic volume categorisation by network-layer, transport-layer and application-layer protocols, a drill-down into each form of port scanning activity and the analysis of any botnet-related propagation pattern, particularly ones belonging to Mirai and Moobot; and another, yet to be formalised and implemented, in which the PCAP recordings would be ingested into some form of data analysis platform, potentially including AWS-native Redshift, Time Stream, and/or Security Lake, among other foreseen contributions.

5.1 EXPERIMENTAL RESULTS

A total amount of three experiments were conducted aiming at providing a characterisation of the background radiation when captured by means of this method. Two of the experiments have been completed, documented, and published. The third experiment has been finished, however a discussion of the findings is underway and will figure on the final version of the thesis.

Each experiment is summarised under the following naming convention scheme:

- Experiment #1: The first experiment is referenced as “Proof-of-Concept Experiment”, due to the fact it was the first time the Cloud Telescope was actually launched, and for which preliminary insights and validation were obtained;
- Experiment #2: The second experiment is called “Full-Scale Experiment” and is also referred to as the “Interactive Experiment”. These has to do with the fact this experiment had the larger aperture, and the fact that active responders were implemented on selected ports; and
- Experiment #3: The third experiment is named “Long-Term Analysis” , also “Main” or “Final” experiment, because it ran for the largest amount of time, five full months, enabling for more solid conclusions on IBR characterisation. This is thoroughly detailed as the main result to support this thesis, in Section 6 “Main Experiment”, starting on page 111.

5.2 EXPERIMENT #1: PROOF-OF-CONCEPT

The first, proof-of-concept experiment was thought as a conservative approach to launching the Cloud Telescope, aiming at providing an overview analysis of the primary characteristics of IBR as captured by the mechanism. The experiment was set to launch on the first hour of May 1st 2023 and was set to shutdown at the last hour of the same month, on the 31st, resulting in 31 full days of capture time.

A total amount of 26 sensors were launched, one per cloud region, and a same configuration set was used for all the three experiments conducted, as follows:

- All the twenty-six commercially available cloud regions of Amazon Web Services were used;
- Each sensor had one single public, external IPv4 address assigned to it;
- Launch procedure was made in the form of Terraform launch command slightly before the announced capture date. Termination procedure followed equivalent procedure. Packets outside the announced time-frame were excluded during post-processing.
- AWS S3 bucket was allocated in Stockholm and has worked flawlessly in all cases;
- The possibility of packet drop was monitored at all times. No packet drop event has occurred.
- tcpdump was set to rotate each 24 hour block;
- Traffic related to capture uploads towards the bucket was filtered out via tcpdump argument; and
- Emergency remote access to the sensor, via SSH, was reassigned to port 65535/TCP, excluded from capture recording, and never used throughout the three experiments.

The experiment was described as an 8-page full paper documenting both the implemented Cloud Telescope architecture and the proof-of-concept results, and accepted at the 2023 IEEE 12th International Conference on Cloud Networking (CloudNet), held in November 2023 at Stevens University, in New York area, in the United States of America.

Paper abstract is reproduced next.

Internet Background Radiation (IBR) comprises a range of unsolicited traffic directed towards Internet hosts. In general, this type of traffic is characterised by high levels of port scanning activity, malware propagation, application exploits, system misconfiguration and denial-of-service attacks. IBR capture is typically undertaken by a system termed a network telescope. This records unfiltered incoming internet traffic for a specific CIDR block in the form of a packet capture (PCAP) file for analysis. This work proposes a novel, cloud-native approach to capturing IBR by the deployment of an ephemeral and reproducible architecture, described as code, distributed across all regions of a cloud service provider. In this paper we discuss the technical and financial viability of using a fleet of small-sized compute instances, in a spot price auction model, to maximise platform collection, capillarity and duration. We also present an overview analysis of the primary characteristics of IBR as collected during a month long proof-of-concept experiment across 26

regions of a cloud service provider in May 2023. Our analysis discusses the aspects of the dataset in quantitative terms: traffic aggregation per protocol, top TCP and UDP ports, top radiation sources and radiation distribution per cloud region. We also provide an overview of the most relevant threats detected. Our results include a formalisation and validation of the cloud telescope, with the corresponding supporting architecture described in Terraform and Ansible. The aggregate dataset amounted to 2.2 GB, and 21.8 million packets. Composition by protocol was 78% TCP, 14% ICMP and 8% UDP (Bortoluzzi et al., 2023).

Full reference is as follows:

F. Bortoluzzi, B. Irwin, L. S. Beiler and C. M. Westphall, “Cloud Telescope: A distributed architecture for capturing Internet Background Radiation” 2023 IEEE 12th International Conference on Cloud Networking (CloudNet), Hoboken, NJ, USA, 2023, pp. 77-85, doi: 10.1109/CloudNet59005.2023.10490018.

The approach used in order to report quantitative findings reproduces the method seen in related works: traffic distribution per relevant network and transport protocols: TCP, UDP and ICMP; followed by an accounting of the most frequently observed ports for TCP and UDP. Three new data aggregation formats only possible in a distributed architecture were introduced: a) regional traffic aggregation, b) country-level analysis of top IBR generators and c) top radiation targets.

5.2.1 Global Traffic Composition

Traffic composition by protocol is presented in Table 8. The observed traffic was predominantly TCP, accounting for over three quarters of the capture volume. This was followed by ICMP and UDP. No other protocols were observed.

Table 8 – Global Traffic Composition

Protocol	Packet count	%
TCP	17 008 753	77.9
ICMP	3 076 855	14.0
UDP	1 770 418	8.1
Total	21 856 713	100.0

5.2.2 TCP Traffic Composition

The top 10 TCP ports by volume is shown on Table 9.

Port 23/tcp (telnet) is the most prevalent target constituting nearly twice the volume of the next position. In second place is 9000/tcp used by a range of services, including Docker,

Table 9 – Top 10 TCP Ports

Rank	TCP Port	Packet count	%
1	23	466 859	2.74
2	9000	241 822	1.42
3	22	202 482	1.19
4	6379	176 404	1.04
5	80	156 494	0.92
6	443	130 650	0.77
7	8080	90 350	0.53
8	445	84 081	0.49
9	37125	74 719	0.44
10	3030	67 063	0.39
	Σ_{top10}	1 690 924	9.93

PHP-FPM and Elixir. Port 22/tcp (ssh) was ranked third, but at a lower volume, followed by 6479/tcp (redis). The triplet of web-related ports: 80/tcp, 443/tcp and 8080/tcp account for nearly 2.2% of traffic. Port 445/tcp (ms-smb) ranks 8th, and lower than has been reported in other studies. The final two ports are related to exploitation activity in Huawei routers. The last port is commonly associated with a range of backdoor implementations.

The top 10 observed services covering less than 10% (and only four targets attaining >1%) of the total protocol packet count indicates somewhat more diverse traffic than previous studies, where the top 10 have had a higher representation.

5.2.3 UDP Traffic Composition

The top 10 UDP ports is showed on Table 10. The ranking alone accounts for nearly 79% of all the traffic using this transport-layer protocol.

Table 10 – Top 10 UDP Ports

Rank	UDP Port	Packet count	%
1	9000	876 347	49.50
2	12000	358 295	20.24
3	5060	48 988	2.77
4	123	24 126	1.36
5	53	21 967	1.24
6	30303	18 977	1.07
7	1900	14 241	0.80
8	161	11 667	0.66
9	6500	10 908	0.62
10	30311	10 363	0.59
	Σ_{top10}	1 395 879	78.85

The most prevalent UDP target is port 9000/udp which accounts for nearly half of all UDP traffic. This service is officially registered by IANA as ‘cslistener’. However, it is also used by Cisco Webex and UDPCast, among other streaming services. Port 12000/udp, in

second place is used both by productive protocols, such as ClearCommerce, and in a range of malware/backdoor implementations.

Port 5060/udp is used in SIP/VoIP communication, and commonly abused, although lower than expected. Queries to 123/udp (NTP), 53/udp (DNS) and 161/snmp (SNMP) are commonly abused for amplification attacks. Ports 30303/udp (Ethereum P2P communication) and 1900/udp (primarily used by the UPnP service) were in the middle of the list, the former being a surprise ranking. Port 6500/udp is commonly used as an alternative to VNC's standard port, 5900/udp. Finally, 30311/udp has an unknown purpose, lacks an official IANA record, and should deserve a closer look. An advantage of UDP is that packets contain some payload, which allows for further investigation and validation of the application layer protocol, and possible intent.

5.2.4 Most active sources

The top 10 most active radiation sources are shown in Table 11.

Table 11 – Top 10 radiation sources

IP	Country	Description	Packets	%
45.155.204.x	Russia	IT Resheniya LLC	287 100	1.31
148.66.12.x	Hong Kong	SIMCENTRIC-HK	141 266	0.65
44.213.45.x	United States	Amazon Web Services	116 520	0.53
139.87.112.x	United States	Oracle Cloud	80 059	0.37
89.248.163.x	Seychelles	Recyber Project	78 626	0.36
89.248.162.x	Seychelles	Recyber Project	71 806	0.33
89.248.165.x	Seychelles	Recyber Project	70 369	0.32
89.248.165.y	Seychelles	Recyber Project	70 241	0.32
89.248.165.z	Seychelles	Recyber Project	70 194	0.32
89.248.163.y	Seychelles	Recyber Project	70 114	0.32
Σ_{top10}			1 056 295	4.83

Final octets of sources have been blinded.

To produce this report, the obtained data was aggregated by each IP source and sorted by the number of packets hitting any of the sensors in the fleet.

Russia is the origin of the most active single source of IBR within the data, responsible for 1.31% of all traffic received, 287,100 packets. The second most active source is located in Hong Kong. The third and fourth most frequently observed IP source addresses are located in the United States, within AWS and Oracle Cloud. From that point, a range of IP addresses belonging to the Seychelles base “Recyber Project” show up as the most active radiation source, with a total of 6 IP addresses constituting 431,350 events, or $\approx 2\%$ packets.

5.2.5 Geographic Distribution

The overall distribution of collected IBR per cloud region is shown in Table 12.

Table 12 – Radiation distribution by Cloud Region

Region	Packets	TCP	UDP	ICMP
af-south-1	672 453 (3.08%)	555 381	19 535	97 491
ap-east-1	939 501 (4.30%)	832 207	24 626	82 627
ap-northeast-1	1 138 415 (5.21%)	642 135	376 764	119 473
ap-northeast-2	809 672 (3.70%)	699 774	16 980	92 882
ap-northeast-3	792 242 (3.62%)	684 376	33 458	74 368
ap-south-1	638 624 (2.92%)	542 794	19 472	76 317
ap-south-2	901 945 (4.13%)	532 210	17 775	351 915
ap-southeast-1	765 701 (3.50%)	675 418	23 636	66 610
ap-southeast-2	777 809 (3.56%)	661 188	17 089	99 492
ap-southeast-3	756 014 (3.46%)	593 017	17 865	145 089
ca-central-1	847 884 (3.88%)	708 802	18 571	120 458
eu-central-1	954 519 (4.37%)	835 907	28 279	90 283
eu-central-2	899 713 (4.12%)	563 121	16 206	320 342
eu-north-1	660 199 (3.02%)	574 429	18 652	67 067
eu-south-1	625 629 (2.86%)	569 280	17 083	39 234
eu-south-2	1 066 224 (4.88%)	708 327	18 238	339 616
eu-west-1	719 110 (3.29%)	572 132	30 942	115 989
eu-west-2	1 984 947 (9.08%)	1 024 879	901 792	58 228
eu-west-3	796 437 (3.64%)	717 438	19 915	59 044
me-central-1	743 999 (3.40%)	582 001	16 831	145 115
me-south-1	873 265 (4.00%)	669 375	18 256	185 596
sa-east-1	643 405 (2.94%)	519 336	17 901	106 121
us-east-1	344 916 (1.58%)	303 739	9 741	31 407
us-east-2	866 359 (3.96%)	787 438	22 257	56 619
us-west-1	720 278 (3.30%)	640 443	28 752	51 022
us-west-2	917 453 (4.20%)	813 601	19 332	84 450
Total	21 856 713 (100%)	17 008 748	1 769 948	3 076 855

It is possible to perceive a fairly even amount of background radiation captured in each cloud region. eu-west-2, in London, was hit by the highest amount of packets, with almost 10% of the entire capture, while us-east-1 was only hit by 1,58% of the observed radiation. TCP was the most prevalent protocol in all cases, however, the number of UDP packets arriving to London almost matched TCP, and is at least 3 times greater than seen in any other region. ICMP was captured in greater amount in ap-south-2, Hyderabad, India, with 351K hits.

5.2.6 Experiment #1 Conclusion

The first experiment enabled to conclude that the Cloud Telescope technique stands as a meaningful perspective for capturing Internet Background Radiation, supported by a cloud-native architecture. Due to the distributed topology, the method substantially differs from those discussed in related works. The Cloud Telescope effectively takes advantage of Amazon Web Service’s global capillarity to provide an ephemeral, low-cost, reproducible design.

Researchers willing to replicate the experiment can launch new sensor fleets and con-

tribute to improved cyber threat intelligence acquisition through IBR analysis. Both the architecture implementation and the resulting dataset were made available to individuals wanting to validate these results. They can also investigate better ways to explore, enrich and analyse the dataset.

The findings also shed light on how future works could unfold, into three main fields: capture expansion, dataset enrichment, and data visualisation.

- Capture expansion 1: Capture can be expanded in the number of sensors that are used to collect IBR. In this regard, this research group plans to conduct a new observation that should last 6 months and expand the cloud telescope to 10x magnitude by deploying 260 sensors, 10 in each region. A multi-cloud architecture expanding to all regions of the three major cloud service providers is also on the road-map.
- Capture expansion 2: Another form of expansion refers to the level of interaction implemented by the sensor. We seek to analyse payloads, which already exists with UDP in the dataset, as much as to implement a generic TCP responder able to establish three-way-handshakes of TCP and capture the first application-layer dialogue, such as GET / for HTTP.
- Capture expansion 3: Higher-level application interaction, such as the ones implemented by honeypots, could be considered for the most active TCP and UDP ports seen, although an impact on machine class, storage, and thus, deployment price, should be expected.
- Dataset enrichment: It is necessary to investigate approaches to perform data enrichment using open source threat intelligence platforms. To date, we have relied on local DNS servers for the massive IP-to-name resolution of network addresses in the PCAP. This work has made use of a local geolocation library to determine the country in which each IP address is located. However, there should be more creative methods for enriching the dataset. One method could be to gently scan back radiation sources to identify which device type is generating malicious radiation. Alternatively, an open source intelligence platform, such as Shodan, which actively and continuously scans the internet, could also be considered.
- Data visualisation 1: A look at related works reveals that data visualisation techniques currently used to acquire cyber threat intelligence from IBR captures do not fully consider the advent of newer technologies that could be used for faster and more relevant analysis. Packet indexing, such as implemented by Arkime, backed by Opensearch, should be tested for performance and fit for purpose.
- Data visualisation 2: ElasticSearch could be used in conjunction with Kibana to provide an interesting insight into data observability. The cloud itself provides new, scalable and elastic solutions for data analysis.

- Data visualisation 3: A cloud-native Extract, Transform, Load (ETL) pipeline should be considered, making use of solutions such as AWS Glue and Athena, for fast data querying and visualisation.
- Data visualisation 4: Lastly, AWS TimeStream could be used to evaluate the suitability of a time-series database to provide insights and to forecast emerging internet-scale threats, while real-time data analysis remains as a possibility for the future.

5.3 EXPERIMENT #2: FULL-SCALE, INTERACTIVE CAPTURE

The second, full-scale experiment was build on top of the findings of the first experiment. The most relevant decisions took based on previously observed data were:

1. The realisation that the most prevalent traffic, belonging to port 23/tcp, very likely should contain payload data that could be traced to malware and/or botnet propagation activity;
2. In order to capture the payload, active responders would have to be implemented on selected ports.
3. It should be possible to expand the Infrastructure-as-Code to include a tailor made, transport-layer responder capable of completing the three-way TCP handshake, fitting the concept of a medium-interaction Honeypot; and/or
4. It should also be possible to include tailor-made active responders for few selected ports, including telnet, but also, ssh and http.

The experiment was set to launch on the first hour of August 1st 2023 and was initially set to run for six months. However, it was interrupted at the 45th day due to the massive data volume acquired, ultimately impacting both the allocated budget and the data processing pipeline under construction at the time. The experiment was shut down right after a total amount of 10,000,000 (ten billion) packets were recorded.

A total amount of 260 (two hundred and sixty) sensors were launched, ten per cloud region, and the same configuration set was used than the previous experiment, except:

- Medium interaction responders were implemented on a range of selected ports;
- High interaction responders were implemented on ports 22, 23, 80 and 443;

The experiment was described into an 8-page full paper documenting both the implemented Cloud Telescope architecture and the achieved results. It was accepted at the 19th Iberian Conference on Information Systems and Technologies, held at University of Salamanca, in Salamanca, Spain, on 25-28 June 2024.

Paper abstract is reproduced next.

The constant activity observed in Internet Background Radiation - IBR reveals a range of cybernetic traffic of interest for cyber security researchers. The use of network telescopes, honeypots and, more recently, cloud telescopes enables the capture of billions of network packets that can be analysed to discover attack patterns and detect emerging threats. This work reports a second Cloud Telescope experiment where 260 sensors were deployed in twenty-six different regions of Amazon Web Services to simultaneously capture unsolicited traffic for 45 consecutive days starting in August 2023. We used a combination of default-passive behaviour with active application-layer responders on TCP ports 23 and 80 to acquire detailed cyber threat intelligence on botnet activity, revealing attack patterns belonging to Mirai, Mozi and Sora botnets. We analyse traffic composition by protocol (TCP, UDP, ICMP), top TCP ports, top UDP ports, top radiation sources, overall traffic volume per cloud region, the most frequent payload download attempts, most frequent binary infection commands, most frequently used HTTP request methods and user-agents. A total amount of ≈ 10 billion packets were collected. Our findings indicate a prevalence of botnet activity exploiting highly critical vulnerabilities on IoT devices disclosed in recent years. The cloud architecture is available as a Terraform artefact, while the dataset is available on IEEE Dataport. Our conclusion is that the Cloud Telescope provides an effective approach to capturing global samples of Internet Background Radiation (Bortoluzzi et al., 2024).

Full reference is as follows:

F. Bortoluzzi, B. Irwin, L. Beiler, and C. M. Westphall, “Quantitative analysis of botnet activity within Internet Background Radiation as observed by a Cloud Telescope” 2024 19th Iberian Conference on Information Systems and Technologies (CISTI), Salamanca, Spain, 2024, yet to be indexed on IEEE Xplore.

The approach used in order to report quantitative findings also reproduce the method seen in related works: traffic distribution per relevant network and transport protocols: TCP, UDP and ICMP; followed by an accounting of the most frequently observed ports for TCP and UDP. However these were expanded to also contain the most frequent payload download attempts, the most frequent binary infection commands, the most frequent http request methods and the most frequent http user-agents, as explained on each subsection.

5.3.1 Global Traffic Composition

The experiment resulted in a total amount of 10 billion packets captured. This includes all incoming connection requests arriving at any TCP or UDP port, as well as ICMP requests and responses, as detailed on Table 13.

Table 13 – Traffic composition by protocol

Protocol	Packet count	%
TCP	9,990,374,104	98,96
UDP	89,592,730	0,89
ICMP	15,472,149	0,15
Total	<i>10,095,441,382</i>	<i>100</i>

The vast majority of traffic belongs to TCP. This is due to two reasons. First, TCP is usually more prevalent. The previous experiment had 77% of arriving packets as TCP datagrams. Second, we implemented high-level TCP responders, which increases this protocol volume as the sensor also generates TCP datagrams when answering communication. The fleet received ICMP packets in a greater number than UDP messages, indicating either intense probing activity at the network layer, or potential fragments of availability attacks, i.e. denial-of-service.

5.3.2 Top TCP Ports

The top TCP ports are shown on 14.

Table 14 – Top TCP ports

Rank	Port	Count	%
1	23	7,462,725,339	74,70
2	25565	124,574,312	1,25
3	30120	114,247,932	1,14
4	5900	65,023,371	0,65
5	80	60,072,031	0,60
6	22	59,892,707	0,60
7	3306	49,221,898	0,49
8	21	48,445,918	0,48
9	53	43,211,435	0,43
10	25	42,123,026	0,42
	Total	<i>9,990,374,104</i>	<i>100</i>

Port 23 was again, the busiest, with over 7 billion packets exchanged, also for two reasons. First, there seems to be an intense scanning activity attempting to exploit the telnet service. This is linked to botnet activity looking for vulnerable IoT devices directly connected to the Internet (Griffioen; Doerr, 2020; Affinito et al., 2023). Second, the high-level responder interacts with the connection requests, generating request/response packets that get recorded by the PCAP-capture mechanism, tcpdump. Port 25565 is assigned to Minecraft, while port 30120 is associated with the multiplayer server back-end for GTA V. Port 80, HTTP, only received half a per cent of the background activity despite its popularity and the fact a high-level responder was implemented on it. Remainder ports belong to well-known Internet services, including VNC (5900), SSH (22), MySQL (3306), FTP (21), DNS (53) and 25 (SMTP).

5.3.3 Top UDP Ports

The top UDP ports are shown on Table 15.

Table 15 – Top UDP ports

Rank	Port	Count	%
1	389	12,163,107	13,58
2	11211	7,017,792	7,83
3	500	955,307	1,07
4	5060	639,967	0,71
5	53	559,973	0,63
6	2041	529,166	0,59
7	123	358,633	0,40
8	30311	314,706	0,35
9	80	240,568	0,27
10	36373	199,938	0,22
	Total	89,592,730	100

The most significant portion of the UDP traffic was aimed at the LDAP service on port 389, potentially related to recent vulnerabilities found at the time: CVE-2023-41580 “PHP LDAP IPAM exploit”, CVSS score 7.7, or CVE-2023-28853 “Mastodon LDAP hack”, CVSS score 7.5. Port 11211 is used by Memcached, known for CVE-2018-1000115, which allows obtaining an amplification factor greater than 1:50,000x against reflected targets. Remaining ports include the DNS service (53) and NTP (123), among other less-known service ports in smaller quantities.

5.3.4 Top radiation sources

The most active sources probing the cloud telescope sensor fleet during the experiment are presented on Table 16.

Table 16 – Top radiation sources

Source IP	Country	ASN Owner	$N \times 10^6$	%
185.224.128.191	Netherlands	Alsicon B.V.	3560	35,27
85.209.134.96	Netherlands	Stark Industries	455	4,51
2.59.254.155	Netherlands	Delis LLC	285	2,82
193.42.32.57	Netherlands	Delis LLC	247	2,45
5.181.80.119	Belgium	Tamatiya EOOD	96	0,95
1.32.249.43	Hong Kong	BGPNET Global ASN	42	0,42
194.169.175.201	Netherlands	Matrix Telecom	41	0,41
60.23.156.211	China	CHINA UNICOM	37	0,38
122.96.31.248	China	CHINA UNICOM	28	0,28
202.182.120.152	Japan	The Constant Company	27	0,27
		Total	10095	100

One single source IP address generated a massive amount of 3.5 billion packets towards the telescope. Elementary data enrichment reveals that the Netherlands served as the largest noise generator, occupying the top four rank positions, with three Dutch ASN behind 45% of all seen traffic. Other countries contributed with much lower IBR traffic amounts, including Belgium, Hong Kong, China and Japan.

5.3.5 Global Radiation Distribution

The traffic volume targeting each cloud region is summarised in Table 17.

Table 17 – Radiation distribution by Cloud Region ($N \times 10^6$)

AWS region	%	Count	TCP	UDP	ICMP
af-south-1	6,18	624	620	3	0,5
ap-east-1	4,12	416	412	3	0,2
ap-northeast-1	3,72	375	371	3	0,5
ap-northeast-2	2,27	228	225	2	0,5
ap-northeast-3	4,97	501	498	3	0,4
ap-south-1	5,28	532	528	3	0,7
ap-south-2	5,87	592	589	3	0,3
ap-southeast-1	5,86	591	587	3	0,7
ap-southeast-2	1,64	165	161	2	0,9
ap-southeast-3	6,02	608	605	2	0,8
ca-central-1	2,08	209	205	3	0,2
eu-central-1	1,93	194	191	2	0,3
eu-central-2	4,06	409	405	3	0,5
eu-north-1	2,41	243	240	2	0,3
eu-south-1	2,77	279	275	3	0,3
eu-south-2	5,90	595	591	2	0,5
eu-west-1	2,45	247	244	2	0,3
eu-west-2	4,40	444	440	3	0,2
eu-west-3	3,01	304	301	3	0,2
me-central-1	2,22	223	219	3	0,5
me-south-1	4,60	464	458	3	2
sa-east-1	4,08	411	408	3	0,3
us-east-1	3,38	341	338	2	0,3
us-east-2	3,70	373	357	14	0,9
us-west-1	3,65	368	365	2	0,7
us-west-2	3,44	346	343	2	0,4
Total	100%	10095	9990	89	15

A reasonably even traffic distribution is seen across the twenty-six AWS regions. The region most affected by background radiation was af-south-1, in South Africa, with 6,18 per cent of the captured traffic volume, closely followed by ap-southeast-3, in Indonesia. Remainder regions stayed very close to the expected simple average, ranging from 1,93% (eu-central-1: Germany) to 5,90% (eu-south-2: Spain). South Africa received the largest amount of TCP packet count. Australia, the lowest. Ohio, in the US received the largest number of UDP data-

grams, 3.6x more than Japan, in second place. California, also in the US, the lowest. Bahrain received the largest ICMP message count, three times more than Australia. France, the lowest.

5.3.6 Port 23 Analysis

The port-based analysis strategy for Port 23 was performed by simple aggregation of Telnet-transported operating system commands issued right after connection establishment between the attacker and the application-layer responder two-fold: the most frequent payload download attempts and the most frequent binary infection commands.

To illustrate the analysis process, consider the network packet depicted in Figure 18.

```
> Frame 17125: 320 bytes on wire (2560 bits), 320 bytes captured (2560 bits)
> Ethernet II, Src: 06:bd:5f:0c:ca:68 (06:bd:5f:0c:ca:68), Dst: 06:8d:e2:4d:4e:74 (06:8d:e2:4d:4e:74)
> Internet Protocol Version 4, Src: 172.31.25.14, Dst: 64.92.8.11
> Transmission Control Protocol, Src Port: 23, Dst Port: 39221, Seq: 22072, Ack: 2177, Len: 254
> Telnet
  -Data: 95.214.26.112/wget.sh -O- | sh;/bin/busybox \r\r\n
  -Data: tftp -g 95.214.26.112 -r tftp.sh -l- | sh;/bin/busybox ftpget 95.214.26.112 ftpg\r\r\n
  -Data: et.sh ftpget.sh && sh ftpget.sh;curl http://95.214.26.112/curl.sh -o- | sh\r\n
  -Data: Connecting to 95.214.26.112 (95.214.26.112:80)\r\n
```

eu-west-2a-3.8.158.21-20230727-0208.pcap

Figure 18 – Infection attempt captured by the responder on port 23

The responder in eu-west-2, in London, is asked to execute a bash script containing a wget command to fetch binary *arm7*, a Mirai payload aiming at embedded ARM devices, such as home routers. Table 18 shows the most frequent payload download attempts.

Table 18 – Most frequent payload download attempts

URL	Count	Botnet
185.224.128.191:80/(...)/cutie.x86	6,543,206	Mirai
185.225.74.161:80/jklx86	1,013,433	Mirai
2.59.254.155:80/x86	387,215	Mirai
185.225.74.161:80/shellx86	363,966	Mirai
200.59.72.26:42530/i	240,306	Mozi
200.59.81.5:42801/i	206,979	Mozi
123.14.116.147:59304/i	190,695	Mozi
60.23.156.211:44915/i	182,670	Mozi
5.181.80.119:80/(...)navebusyoo.x86	166,744	Mirai
100.82.210.85:59687/i	158,895	Mozi

Upon connection, botnet infection scripts attempt to download a payload that could turn the device into a new zombie. Mirai and variants focus on IoT devices, frequently using a popular combination of a Linux kernel and the Busybox command-set user-land. As a result, 32-bit binaries aiming at embedded, low-cost, low-power architectures are frequently seen in combination with binaries that aim at Intel's x86 hosts.

Data enrichment provided by Shodan reveals that the Netherlands are the most prevalent country hosting botnet payload distribution servers. The most significant amount of IP address sources belongs to Dutch companies found to provide anonymisation in the form of VPN servers. Argentina, Bulgaria, China and the United States complete the list.

Each infection command can be linked to the malware attempting to propagate it. These commands also reveal the zombie device acting as the source of the attack. All the ten sources attempted to infect the sensors with either Mirai or Mozi (Sahota; Vlajic, 2021).

Table 19 shows the most frequent binary infection commands.

Table 19 – Most frequent binary infection commands

Filename	Count	Botnet
i	10,442,700	Mozi
pedalcheta/cutie.x86	6,558,315	Mirai
jklx86	1,013,784	Mirai
x86	530,205	Mirai
shellx86	363,966	Mirai
bin.sh	226,788	Mozi
sora.sh	217,964	Sora
0xh0roxxnavebusyoo.x86	174,786	Mirai
ohshit.sh	119,260	Mirai
s	83,794	Mirai

The medium interaction, honeypot-like behaviour, implemented by the responder faked the acceptance of infection attempts, partially revealing attackers' *modus operandi*.

In the sample, the most frequently seen payload is a binary named *i*. It is linked to Mozi, described as a novel P2P botnet in the wild, targeting numerous devices. It is remarkably different in many aspects involving self-proliferation, network control, and distribution situations (Wang et al., 2022). *bin.sh* is also linked to Mozi. Sora, another Mirai variant, was detected to a smaller extent by the *sora.sh* execution attempt. All other binaries refer to standard Mirai, in the wild for almost ten years.

5.3.7 Port 80 Analysis

Traffic affecting the sensor fleet on Port 80 was analysed in terms of the most frequently observed HTTP request methods and the most frequent user-agents informed by the source.

Table 20 shows the former.

The most prevalent HTTP request method is GET. Argument “/.env” refers to the attempt to fetch web server's environment variables that could reveal *open source intelligence* about the target's configuration, web server implementation and underlying operating system. Argument “/?q=ultrasurf” should refer to Ultrasurf VPN, which seems to conduct active internet-wide scanning. The HEAD parameter combined with “Login.asp” refers to a potential vulnerability affecting the “.Net Core” web development framework for ASP.NET. Remainder

Table 20 – Most frequent HTTP request methods

Request	Count
GET /.env	480,435
GET /?q=ultrasurf	111,575
HEAD /Core/Skin/Login.aspx	111,050
GET /info.php	24,520
GET /_profiler/phpinfo	24,275
GET /.git/conf	22,570
GET /laravel/.env	21,108
GET /app/.env	20,490
GET /phpinfo.php	19,439
GET /api/.env	19,365

attacks can be linked to similar information-gathering attempts exploiting weak configuration practices related to PHP, GIT, Laravel and general API deployments.

The latter is shown on Table 21.

Table 21 – Most frequent HTTP user-agents

User Agent	Count
Amazon-Route53-Health-Check-Service	1,374,038
Mozilla/5.0 zgrab/0.x	72,965
Expansive, a Palo Alto Networks Company	61,583
Go-http-client/1.1	47,527
Cloud mapping experiment. research@pdrilabs.net	24,149
python-requests/2.31.0	23,954
CensysInspect/1.1;	21,211
Custom-AsyncHttpClient	16,299
Linux Gnu (cow)	8,524
python-requests/2.28.2	7,706

Even though HTTP clients can freely choose what to present as the *user-agent* parameter, it was possible to aggregate relevant quantities of repeated string occurrences linking back to mostly benign measurement services. AWS's DNS infrastructure service *Route 53* conducted unsolicited health checks, as part of their basic instance monitoring tier, corresponding, by far, as the most frequent string seen in the application-layer payload. ZGrab, a fast *banner grabber* used to enumerate attack targets, came in second, followed by Expansive, an *attack surface management* platform recently acquired by Palo Alto Networks. Remainder occurrences relate to a python-coded HTTP client, benign internet census tool Censys, and other custom Linux/bash-based, python-coded clients.

Data enrichment with URLHaus and VirusTotal allowed us to correlate the combination of the most frequently used HTTP methods and the respective user agents to determine what botnet these belong to and which vulnerability they are trying to exploit, as summarised in Table 22.

Exploitation patterns indicate that Mirai, Mozi and Sora actively attempt to take ad-

Table 22 – CVEs linked to malware propagation

Vulnerability	Malware detected
CVE-2016-20016	Mirai, Mozi, Sora
CVE-2020-20601	Kinsing
CVE-2022-40475	Mirai

vantage of the relatively old CVE 2016-20016, CVSS Score 9.8 (Critical), corresponding to a vulnerability found within the embedded Linux stack as used by many large-scale re-branded CCTV and DVR manufacturers, under the name MVPower, models TV-7104HE 1.8.4 115215B9 and TV7108HE. It was possible to detect Kinsing, another emerging botnet, attempting to exploit CVE-2020-20601, another CVSS 9.8 vulnerability found in 2020 on ThinkCMF, a Chinese content management framework forked from ThinkPHP, coded in PHP and MySQL. A Mirai strain was also linked to a more recent vulnerability, CVE-2022-40475, also scoring 9.8, affecting a range of domestic internet appliances branded Totolink.

5.3.8 Experiment #2 Conclusion

The work presented a quantitative analysis of unsolicited internet traffic affecting a reasonably large fleet of capture sensors globally distributed within a cloud service provider. Our distributed, hybrid approach, combining passively recorded traffic with active responders on selected ports, resulted in a unique dataset that can be used for cyber threat intelligence acquisition within acceptable parameters of processing power and cloud service costs. The 10 billion packet count achieved by the fleet in a relatively small time frame resulted in an average radiation of 35,952 unsolicited packets per hour, per IP address. This can be considered the radiation baseline for future hybrid observations. The fact that most observed traffic belongs to TCP sustains the thought that attackers currently aim at exploiting targets with confidentiality attacks, where the objective is to break in, possibly to renter targets into zombies for upcoming denial-of-service campaigns. This is reinforced by the application layer analysis presented. The characterisation of traffic activity affecting port 23, for which three-quarters of the traffic was destined, indicates that the low-cost nature of IoT devices, for which security is often neglected, has become a significant concern. UDP exploitation, frequently linked to availability attacks, was seen to a much smaller extent, essentially aiming either injection or amplification opportunities for well-known, recently discovered critical vulnerabilities.

Even though the cloud telescope is set up to allow all traffic inward, where protection mechanisms are forcefully disabled, AWS might enforce denial-of-service prevention mechanisms beyond the researchers' control as part of the cloud computing shared responsibility model. This should deserve a detailed investigation.

IBR analysis conducted before 2016, such as in Irwin (2012), used to show a prevalence of straightforward malware propagation, Conficker is the most remarkable example. CVE-2016-20016 enabled malicious actors to improve their attack methodology, where IoT devices,

despite low power, are massively turned into zombies, after which availability attacks can be planned and orchestrated. This two-step distributed approach inaugurates an era where the terabit-per-second packet ratio attacks have become a reality.

Future works were thought to consider:

- The launch of tuned versions of the Cloud Telescope aiming at actively interacting with attack sources on different transport-layer ports right after zero-day disclosures;
- Deeper qualitative analysis of the datasets produced so far, where data aggregation can be used to generate visualisations revealing patterns and trends this quantitative approach is unable to establish;
- To index and enrich the PCAPs within observability platforms for fast packet query and establish new data insights and correlation. This is under investigation with software stacks including Elasticsearch, InfluxDB, and the cloud-native combination of AWS Glue, Firehose and Athena. Azure Sentinel is another emerging solution under consideration.

5.4 EXPERIMENT #3 OUTLINE

The third, long-term analysis experiment raised as a need due to the fact the second, full-scale experiment, had to prematurely end after 45 days of radiation collection. This final experiment was thought to return to a fully passive capture, along the same lines of the first observation, however, the fleet size, or telescope aperture size, was set to 5x larger than the first, and half of the size of the second experiment, hence, 5 sensors per cloud region, resulting in 130 compute instances.

The experiment was set to launch on the first hour of October 1st 2023 and was initially set to run for six months. However, it was interrupted at the end of the fifth month, on 29 February 2024, due to a major change in price composition strategy employed by AWS, as documented on their community blog at Nem (2024).

The long-term analysis of this capture is thoroughly documented as the main output of the thesis, throughout Sections 6.2 (Resulting Dataset), 6.3 (Radiation Sources), 6.4 (Radiation Targets), and 6.5 (Distributed Analysis). The results were also submitted and accepted for publication with IEEE Access, in the form of a longer paper entitled “Cloud Telescope: an ephemeral, distributed, cloud-native architecture for collecting Internet Background Radiation”, containing the abstract reproduced below (Bortoluzzi; Irwin; Westphall, 2025).

For two decades, cyber security researchers have been looking to answer one major question: what threats affect the Internet at large? In addition, what malicious traffic patterns would emerge if we could sample the unsolicited traffic - termed Internet Background Radiation (IBR) - arriving at devices directly connected to

the Internet? The standard approach to collecting malicious traffic is the Network Telescope: a computer device assigned with a public IP address range, configured to passively listen to incoming packets. The deployment of Network Telescopes has helped to detect and quantify major cyberspace outbreaks, from the rise of the Conficker malware, to uncovering massive botnet propagation activity, such as performed by Mirai and its variants, against the Internet-of-Things. This paper introduces the Cloud Telescope: an ephemeral, cloud-native architecture, described as Infrastructure-as-Code, enabling for geographically distributed capture of the IBR, along with a discussion of a 5-month-long validation experiment, in which a sensor fleet comprising 130 cloud instances was launched across twenty-six regions of the world. The result is a quantitative and qualitative analysis of 530 million captured packets. This includes traffic breakdown by protocol: TCP (80%), UDP (3%), ICMP (17%), and by source country. We also discuss traffic aggregation by destination country and by affected cloud region, enabling novel forms of geopolitical influence analysis.

Full reference is as follows:

F. Bortoluzzi, B. Irwin and C. M. Westphall, “Cloud Telescope: An Ephemeral, Distributed, and Cloud-Native Architecture for Collecting Internet Background Radiation,” in *IEEE Access*, vol. 13, pp. 45682-45714, 2025, doi: 10.1109/ACCESS.2025.3549623.

5.5 DATASET OUTPUT

Three datasets were collected as by-products of this research effort. The first one was used on the proof-of-concept experiment, while the second was used to produce the second, full-length capture. The third dataset results from a 5-month, long-term capture and configures the main source of the analysis presented ahead.

5.5.1 Dataset #1

The 21-million packet dataset, comprising the first experiment, summarised on Table 23, was published onto IEEE Dataport.

Dataset 1 companion abstract is reproduced next.

This dataset results from a month-long cloud-based Internet Background Radiation observation conducted in May 2023. A sensor fleet comprised of 26 EC2 compute instances was deployed within Amazon Web Services across their 26 commercially available regions, 1 sensor per region. The dataset contains 21,856,713 incoming

Table 23 – Dataset 1 Characteristics

Dataset name	Cloud Telescope - Internet Background Radiation - May 2023
Dataset authors	Fabrizio Bortoluzzi, Lucas Beiler, Barry Irwin, Carla Merke Westphall
Dataset DOI	10.21227/sg9h-ta16
Dataset format	Single compressed PCAP file.
Number of entries	21,856,713 incoming packets
File size	890 megabytes uncompressed
Published date	02 February 2024
Dataset website	https://ieee-dataport.org/documents/cloud-telescope-internet-background-radiation-may-2023

packets, out of which 17,008,753 are TCP datagrams, 3,076,855 are ICMP packets and the remainder, 1,770,418 are UDP messages. Top TCP traffic affected by malicious activity includes ports 23 (Telnet), 22 (SSH), 80 (HTTP), 443 (HTTPS), and 445 (SMB), among others. Top UDP traffic includes port 9000, 123 (NTP) and 53 (DNS), among others. Sensors were deployed in all viable AWS Regions. It is possible to trace IP addresses back to the following regions and countries: af-south-1, ap-east-1, ap-northeast-1, ap-northeast-2, ap-northeast-3, ap-south-1, ap-south-2, ap-southeast-1, ap-southeast-2, ap-southeast-3, ca-central-1, eu-central-1, eu-central-2, eu-north-1, eu-south-1, eu-south-2, eu-west-1, eu-west-2, eu-west-3, me-central-1, me-south-1, sa-east-1, us-east-1, us-east-2, us-west-1 and us-west-2 (Bortoluzzi et al., 2024b).

5.5.2 Dataset #2

The 10-billion, interactive packet dataset, comprising the second experiment, summarised on Table 24, was also published onto IEEE Dataport.

Table 24 – Dataset 2 Characteristics

Dataset name	Cloud Telescope - Internet Background Radiation - August 2023
Dataset authors	Fabrizio Bortoluzzi, Lucas Beiler, Barry Irwin, Carla Merke Westphall
Dataset DOI	10.21227/hs3j-pg74
Dataset format	26 (twenty-six) compressed PCAP files, aggregated one per cloud region
Number of entries	10,095,441,382 interactive network events
File size	290 gigabytes combined (uncompressed)
Published date	02 February 2024
Dataset website	https://ieee-dataport.org/documents/cloud-telescope-internet-background-radiation-august-2023

Dataset 2 companion abstract is reproduced next.

This dataset results from a 45-day Cloud Telescope Internet Background Radiation collection experiment conducted during the months of August and September 2023. A total amount of 260 EC2 instances (sensors) were deployed across all the 26 commercially available AWS regions at the time, 10 sensors per region. A Cloud Telescope sensor does not serve information. All traffic arriving to the sensor is unsolicited, and potentially malicious. Sensors were configured to allow all unsolicited traffic. In this experiment, we implemented high-level responders

on TCP ports 23 and 80, coded in rust, to record the commands issued by bot-nets such as Mirai when they attempt to infect IoT devices. All other TCP ports were configured to only respond to connection requests until three-way handshake completion. This should enable TCP connection state analysis (syn,fin,ack,...). The architecture is reproducible. Terraform Infrastructure-As-Code is available at: <https://github.com/lucasbeiler/ibr-iac> (Bortoluzzi et al., 2024a).

5.5.3 Dataset #3

The third and main dataset was also published on IEEE Dataport, and is summarised on Table 25.

Table 25 – Dataset 3 Characteristics

Dataset name	Cloud Telescope - Internet Background Radiation - October 2023 - February 2024
Dataset authors	Fabricio Bortoluzzi, Lucas Beiler, Barry Irwin, Carla Merke Westphall
Dataset DOI	10.21227/zkyy-gk56
Dataset format	26 (twenty-six) compressed PCAP files, aggregated one per cloud region
Number of entries	530,914,922 passive network events
File size	20 gigabytes combined (compressed)
Published date	30 September 2024
Dataset website	https://iee-dataport.org/documents/cloud-telescope-internet-background-radiation-october-2023-february-2024

Dataset 3 companion abstract is reproduced next.

This dataset results from a 5-month-long Cloud Telescope Internet Background Radiation collection experiment conducted during the months of October 2023 until February 2024. A total amount of 130 EC2 instances (sensors) were deployed across all the 26 commercially available AWS regions at the time, 5 sensors per region. A Cloud Telescope sensor does not serve information. All traffic arriving to the sensor is unsolicited, and potentially malicious. Sensors were configured to allow all unsolicited traffic. This dataset contains 530,914,922 network events recorded in the PCAP format. 425,674,635 (80%) of the datagrams fit the TCP transport-layer protocol, while 86,139,655 (16%) belong to ICMP messages, and 18,982,971 (4%) are UDP datagrams. The architecture is reproducible. Terraform Infrastructure-As-Code is available at: <https://github.com/lucasbeiler/ibr-iac> (Bortoluzzi et al., 2024c).

5.6 ACCESSORY RESULTS

The research into the Cloud Telescope has also produced the following indirect results and research unfoldings. These are detailed ahead.

Computer Science Bachelor's Degree Projects under candidate's immediate supervision at the University of Vale do Itajaí:

1. Internet Background Radiation analysis by means of a distributed, cloud-based telescope. Lucas Silveiro Beiler;
2. An AWS Athena based interface for quantitative analysis of Internet Background Radiation samples. Mateus Ramos Ventura;
3. Evaluation of Microsoft Azure Sentinel as a data analysis platform for Internet Background Radiation. Natan Nilo Martins; and
4. Evaluation of the ELK stack as a data analysis platform for Internet Background Radiation. Enzo Redivo Canella.
5. Evaluation of the Snort Intrusion Detection System as a detection mechanism for Internet Background Radiation samples acquired with the Cloud Telescope. Guilherme Victor Butzke.
6. Web-based platform to enrich and analyse Internet Background Radiation samples as captured by the Cloud Telescope. Matheus Silvestre.

Cyber Security Bachelor's Degree Projects under candidate's immediate supervision at Noroff University College:

1. Analysis of cyber security threats in May 2023 Dataset as detected by Snort. Andreas Skåra;
2. An evaluation of Arkime as a data analysis platform for Internet Background Radiation samples. Mike Deck;
3. Evaluation of the ELK Stack as an observability tool to explore the Internet Background Radiation within the May 2023 dataset. Jan-Terje Vidhammer;
4. Analysis of UDP content within May 2023 dataset containing Internet Background Radiation. Thomas Sannan.
5. Quantifying botnet activity in Internet Background Radiation captured by the Cloud Telescope. Sebastian Walmann.
6. Internet Background Radiation Affecting Sweden: A Contrast with Global Traffic. Nicoleta-Ioana Zapirtan.
7. Cloud-Based Detection of Botnet Activity: A Script-Based Pipeline for Analysing Internet Background Radiation. Mads Førde Folkestad.
8. Isolating cases of IoT botnet activity from a Cloud Telescope IBR dataset. Simen Espeland.

Computer Science Bachelor's Degree Projects co-supervised by the candidate and candidate's supervisor, at the Federal University of Santa Catarina:

- An analysis of botnet activity within Internet Background Radiation found on Cloud Telescope August 2023 Dataset. Bernardo Pasa Ribeiro.

Computer Science Master's Degree Project co-supervised by the candidate and candidate's supervisor, at the Federal University of Santa Catarina:

- A graph-based approach to intrusion detection within Internet Background Radiation. Antonio Silverio Montagner.

Workshop, colloquium, and industry-oriented conference presentations delivered by the degree candidate and research partners:

1. Research Colloquium session held at Noroff University College: "Cloud watching: Research experimentation with a cloud-based network telescope". Fabricio Bortoluzzi and Barry Irwin. 22 March 2024;
2. Workshop session held at University of Vale do Itajaí: "Cyber Threat Intelligence Acquisition with the Cloud Telescope". Fabricio Bortoluzzi and Lucas Silvero Beiler. 18 May 2024;
3. "16th You Shøt the Sheriff" cyber security industry conference, held at Blue Note in São Paulo: "Cloud Telescope: A cloud-native architecture for capturing Internet Background Radiation". 20 May 2024.
4. "BSides Las Vegas 2024" cyber security community conference, held at The Tuscany in Las Vegas: "Insights on using a Cloud Telescope to observe internet-wide botnet propagation activity". 7 August 2024. Recording available at Bortoluzzi (2024a).
5. "BSides Kristiansand 2025" cyber security community conference, held at Noroff University College in Kristiansand, Norway: "Cyber Threat Intelligence Acquisition with the Cloud Telescope". 6 June 2025.

6 LONG-TERM EXPERIMENT

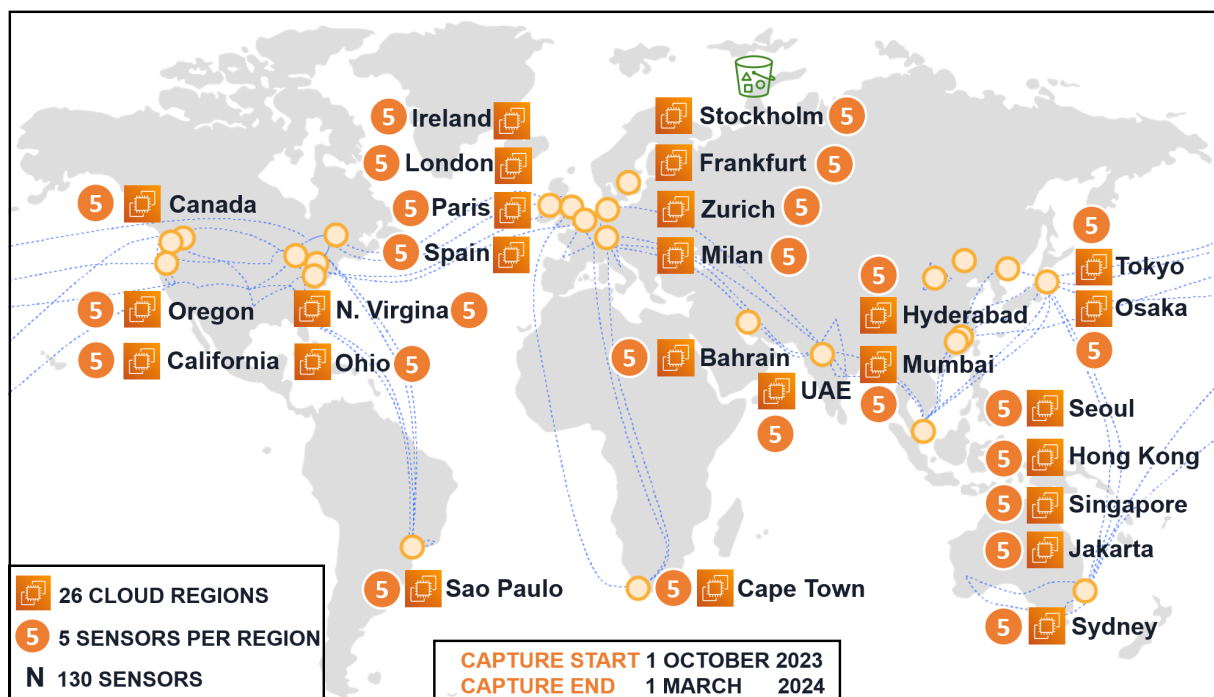
This section documents the third and main experiment conducted towards validating the Cloud Telescope’s design and implementation as a distributed method for collecting Internet Background Radiation.

The following sections explain how the experiment was set up, followed by speculative ethical consideration, a characterisation of the resulting dataset, and the radiation analysis in three perspectives: radiation sources, focusing on the quantitative distribution of probing origins, radiation targets, focusing on network-layer and transport-layer protocol distribution, and distributed analysis, where traffic dispersion is discussed towards the different countries and cloud regions that hosted the distributed capture experiment.

6.1 EXPERIMENT SETUP

The Cloud Telescope was set to launch on October 1st 2023. The capture process lasted for full five months, ending 151 days later, on March 1st, 2024. Figure 19 shows the twenty-six commercially available AWS Regions in which the Cloud Telescope sensor fleet was deployed.

Figure 19 – Architecture of the Final Experiment.



Source: The author.

A total of 130 cloud telescope sensors were deployed to simultaneously collect Internet Background Radiation. Five sensors per each AWS region. Five locations were covered in North America: Canada Central (ca-central-1), US West in Oregon (us-west-2), US West in Northern California (us-west-1), US East in Northern Virginia (us-east-1), US East in Ohio (us-east-2). One region in South America: Brazil/Sao Paulo (sa-east-1). Eight regions in Europe:

Europe/Ireland (eu-west-1), UK/London (eu-west-2), Europe/Paris (eu-west-3), Europe/Spain (eu-south-2), Europe/Stockholm (eu-north-1), Europe/Frankfurt (eu-central-1), Europe/Zurich (eu-central-2), Italy/Milan (eu-south-1). Two regions in Middle East: Bahrain (me-south-1), United Arab Emirates (me-central-1). One region in Africa: South Africa/Cape Town (af-south-1). Nine regions in *Asia and The Pacific*: India/Hyderabad (ap-south-2), India/Mumbai (ap-south-1), Jakarta (ap-southeast-3), Japan/Tokyo (ap-northeast-1), Japan/Osaka (ap-northeast-3), South Korea/Seoul (ap-northeast-2), Hong Kong (ap-east-1), Singapore (ap-southeast-1), and Australia (ap-southeast-2). The aggregation bucket was set up in Sweden.

6.2 RESULTING DATASET

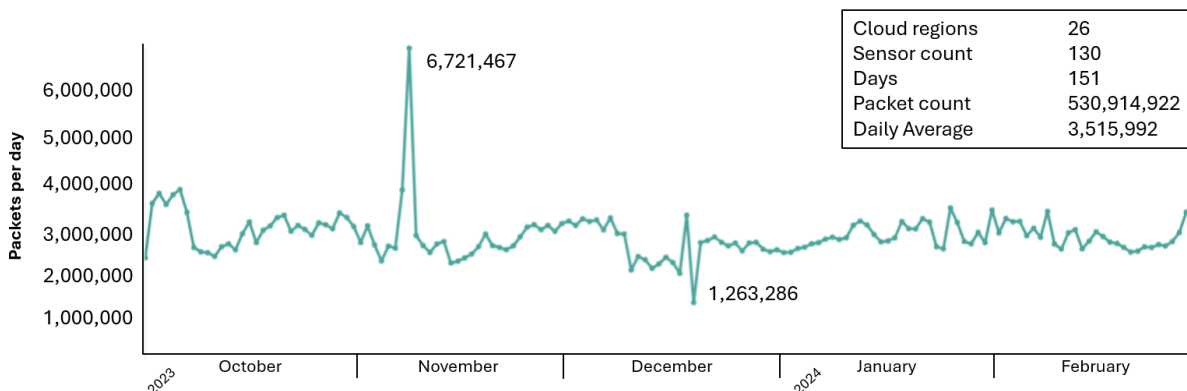
A characterisation of the resulting dataset, available on IEEE Dataport at (Bortoluzzi et al., 2024c), is summarised in Table 26. The experiment resulted in a total amount of 530 million unsolicited incoming network events, of which 425 million were TCP, 86 million consisted of ICMP packets, and 18 million UDP messages.

Table 26 – Protocol Distribution

Protocol	Count	%
TCP	425,674,635	80.18
ICMP	86,139,655	16.22
UDP	18,982,971	3.58
Unknown	117,661	0.02
Total	530,914,922	100.00

The daily distribution of arriving packets behaved as depicted in Figure 20. On 8 November there was a peak of 6,721,467 incoming packets. The average packet count was of 3,515,992 events per day. The lowest daily packet count was seen on 19 December, with only 1,263,286 events.

Figure 20 – Temporal traffic distribution.



Source: The author.

The fact the majority of the captured traffic belongs to TCP indicates a prevalence of infection-oriented attacks, in which the attacking sources aim at finding vulnerable software running on probed targets, either with harmless intent, such as performed by benign scanners, including Censys and Shodan, or for malicious purposes, which is more prevalent and further discussed. This behaviour opposes service exhaustion oriented attacks, such as ping flooding, denial-of-service and amplification attacks, which usually exploit ICMP and UDP-based traffic. One hypothesis to justify this situation is that infection-oriented attacks are effective against any vulnerable target, whereas service-exhaustion attacks usually aim at specific address blocks belonging to each target companies, preventing a telescope placed elsewhere from capturing such events.

Table 27 shows the number of unique IP source and target addresses observed during the deployment phase of the experiment.

Table 27 – Unique Radiation Sources and Destinations

Radiation	Count
Source IP	1,834,887
Destination IP	3,475

A total amount of 1.8 million different IP sources generated at least one network event against one or more cloud telescope sensor fleet instances. The number corresponds to 0.04% of IPv4's theoretical 2^{32} addressable space. EC2-based virtual machines had to recycle their IP addresses so many times that the diversity of observed IP destinations accounted 3.4 thousand distinct target addresses, 130 at any given time.

6.3 RADIATION SOURCES

The approach used to report radiation sources follows the method typically employed in related studies. The result series presented in this section is comprised of the twenty most active radiation source countries, followed by the most active sources propagating TCP-based traffic, UDP radiation sources, and ICMP radiation origins.

Radiation sources were aggregated by country. Tabular results are summarised in Table 28.

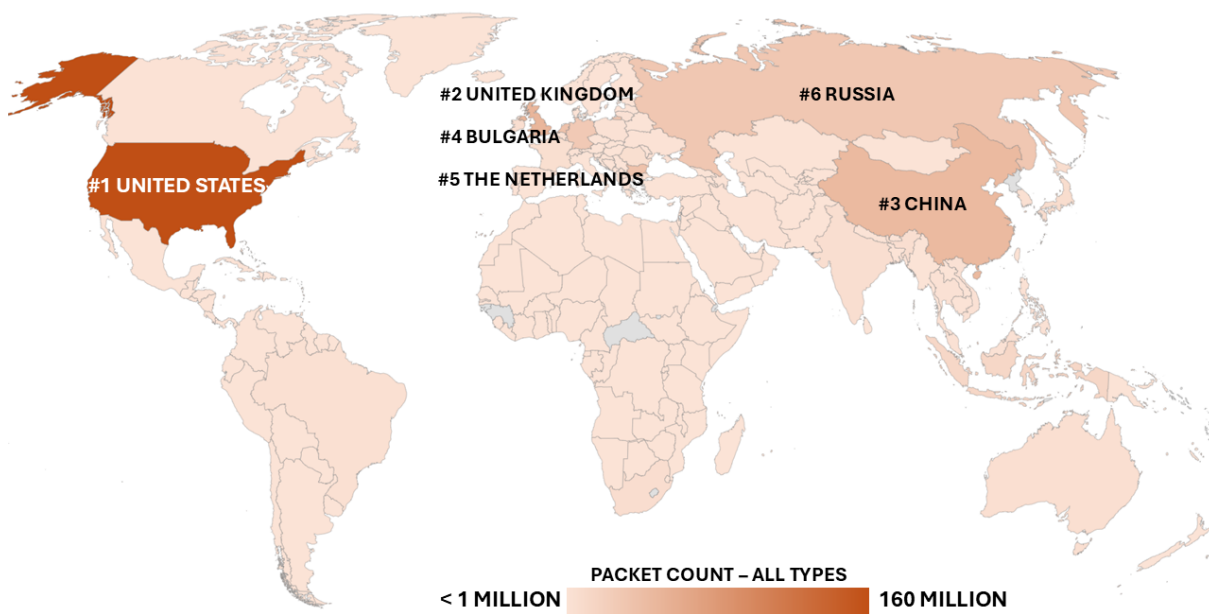
Table 28 shows that a total amount of 169 million network events had the United States of America as the radiation source. This corresponds to almost a third of the entire dataset. The United Kingdom came second, with almost 10%. China, Bulgaria, The Netherlands, Russia, Singapore and Germany were the source of at least 5% of the packets each. Indonesia, Hong Kong, Japan, South Africa and India contributed with 1%-2.5% of the incoming traffic. Other source countries were detected only by smaller fractions $R < 1\%$.

A visual representation of the most active radiation sources, highlighting the ones that generated more than 30 million packets, is illustrated in Figure 21.

Table 28 – Most active radiation sources (Country)

Rank	Country	Count	%
1	United States	169,249,433	31.88
2	United Kingdom	51,570,308	9.71
3	China	47,063,884	8.86
4	Bulgaria	37,510,860	7.07
5	The Netherlands	36,744,264	6.92
6	Russia	31,869,718	6.00
7	Singapore	29,909,254	5.63
8	Germany	29,267,184	5.51
9	Indonesia	13,180,679	2.48
10	Hong Kong	10,005,851	1.88
11	Japan	7,268,133	1.37
12	South Africa	6,420,366	1.21
13	India	5,623,995	1.06
14	Ireland	5,167,944	0.97
15	Romania	5,013,302	0.94
16	France	4,808,674	0.91
17	Australia	4,130,374	0.78
18	South Korea	3,699,848	0.70
19	Brazil	3,325,202	0.63
20	Belgium	2,209,498	0.42
	Sub-total	504,038,771	94.94
	Total	530,914,922	100.00

Figure 21 – Top radiation sources



Source: The author.

6.3.1 TCP Radiation sources

TCP radiation, which constitutes the largest portion of all observed traffic, is ranked in Table 29. The twenty most prevalent sources account for 96% of this traffic.

Table 29 – Most prevalent TCP radiation sources

Rank	Country	Count	%
1	United States	150,732,982	35.41
2	United Kingdom	48,695,027	11.44
3	Bulgaria	37,442,973	8.80
4	The Netherlands	36,157,740	8.49
5	Russia	31,502,161	7.40
6	Germany	27,881,298	6.55
7	China	14,079,601	3.31
8	Singapore	13,419,307	3.15
9	Hong Kong	6,671,238	1.57
10	South Africa	6,327,190	1.49
11	Ireland	4,997,185	1.17
12	Romania	4,965,994	1.17
13	India	4,345,126	1.02
14	Japan	4,307,954	1.01
15	France	4,291,378	1.01
16	Australia	3,212,286	0.75
17	Brazil	3,121,154	0.73
18	South Korea	2,594,847	0.61
19	Belgium	2,121,411	0.50
20	Lithuania	1,998,886	0.47
	Sub-total	408,865,738	96.05
	Total	425,674,635	100.00

The United States alone originated slightly more than one third (35%) of all TCP radiation. The United Kingdom was second, with 11%, followed by Bulgaria, The Netherlands, Russia, and Germany also contributed with significant amounts of collected radiation.

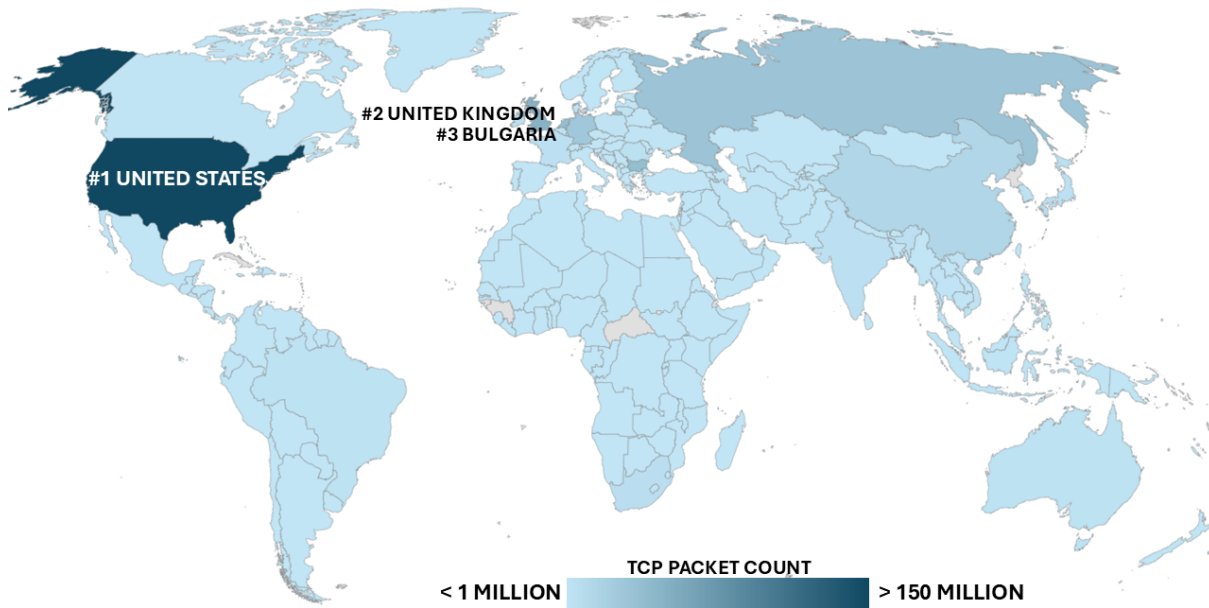
Figure 22 contains a heat map summarising the most active TCP radiation sources in proportion to the amount of traffic originated by each country.

6.3.2 UDP Radiation sources

UDP traffic accounted for approximately 19 million events. 94% of which concentrated on the twenty most active radiation generators, as summarised in Table 30.

The United States of America is the source country for more than half of all UDP radiation, 10 million packets, or 55%. The United Kingdom is the second most prevalent radiation generator, 1.7 million events, or 9%. China, Sweden and The Netherlands have also generated at least half a million packets, contributing each with 2 to 4% of all seen UDP datagrams.

Figure 22 – TCP radiation source countries



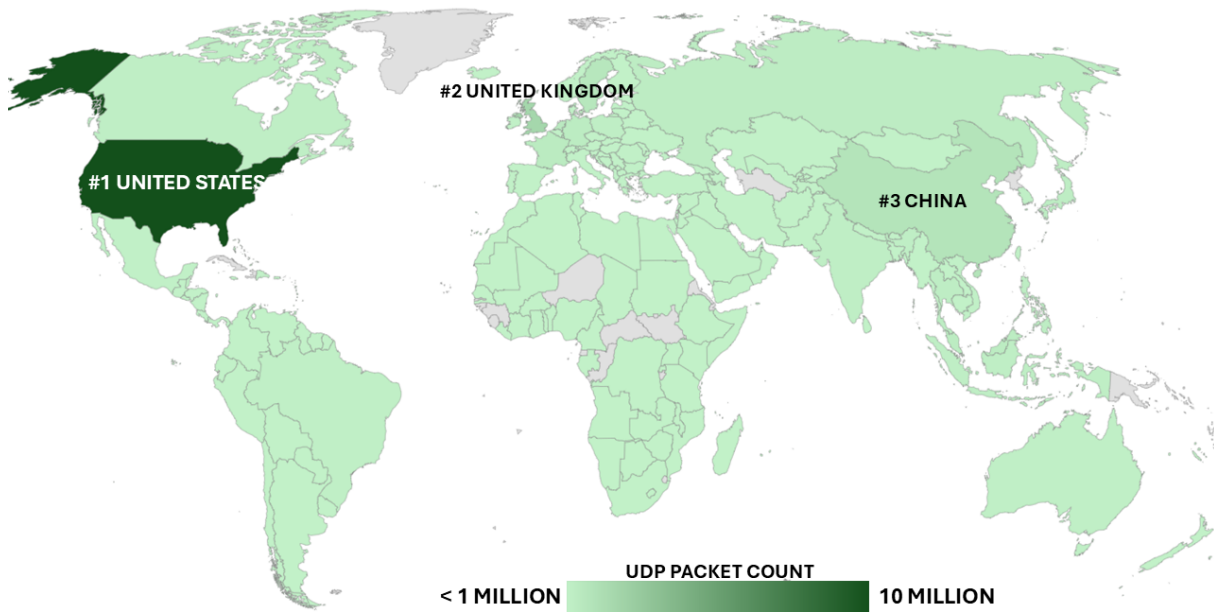
Source: The author.

Table 30 – Most prevalent UDP radiation sources

Rank	Country	Count	%
1	United States	10,509,012	55.36
2	United Kingdom	1,715,522	9.04
3	China	818,985	4.31
4	Sweden	652,682	3.44
5	The Netherlands	501,429	2.64
6	France	395,117	2.08
7	Germany	346,891	1.83
8	Malaysia	335,084	1.77
9	Singapore	313,142	1.65
10	Russia	294,493	1.55
11	Hong Kong	268,899	1.42
12	Japan	248,886	1.31
13	Thailand	248,287	1.31
14	Philippines	229,352	1.21
15	Indonesia	203,811	1.07
16	Canada	201,809	1.06
17	India	178,807	0.94
18	Poland	178,330	0.94
19	Brazil	141,838	0.75
20	Taiwan	126,749	0.67
	Sub-total	17,909,125	94.34
	Total	18,982,971	100.00

Figure 23 depicts traffic distribution on a world map. Darker regions indicate packet counts of up to 10 million events being generated on a highlighted location.

Figure 23 – UDP radiation source countries



Source: The author.

6.3.3 ICMP Radiation sources

ICMP traffic accounted 86 million network events. The ranking containing the twenty most active ICMP radiation sources is summarised in Table 31.

China generated more than one third of this radiation type, 37% of the packets, followed by Singapore, with 18% and Indonesia, 13%. The United States of America were fourth, with 9%, followed by Hong Kong and Japan, with 3% each. Figure 24 illustrates the resulting heat map revealing some concentration in eastern countries.

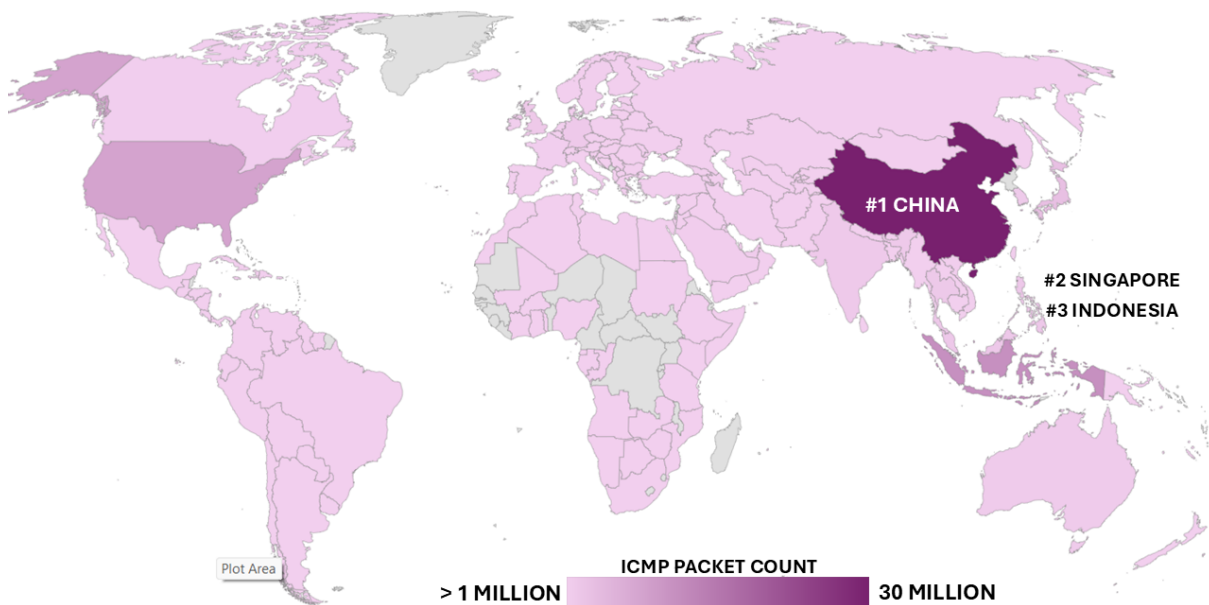
6.4 RADIATION TARGETS

Radiation target distribution uses similar aggregations than seen in related works. The TCP-centric analysis allows to focus not only on the most frequently probed TCP destination ports, but also towards a characterisation of the most commonly exploited vulnerabilities linked to each port, and in terms of investigating the most commonly used protocol flag combinations as recorded by the sensor fleet. UDP radiation is also ranked by the most probed destination port, whereas the ICMP protocol is grouped the most frequently seen message types.

Table 31 – Most prevalent ICMP radiation sources

Rank	Country	Count	%
1	China	32,165,142	37.34
2	Singapore	16,176,805	18.78
3	Indonesia	11,609,410	13.48
4	United States	8,006,679	9.29
5	Hong Kong	3,065,714	3.56
6	Japan	2,711,230	3.15
7	Malaysia	1,294,291	1.50
8	Philippines	1,225,954	1.42
9	Thailand	1,186,495	1.38
10	United Kingdom	1,159,757	1.35
11	India	1,100,040	1.28
12	Germany	1,038,961	1.21
13	South Korea	1,005,728	1.17
14	Australia	819,998	0.95
15	United Arab Emirates	542,462	0.63
16	Saudi Arabia	288,877	0.34
17	Sweden	191,464	0.22
18	Vietnam	173,846	0.20
19	Taiwan	159,836	0.19
20	Ireland	130,207	0.15
	Sub-total	84,052,896	97.58
	Total	86,139,655	100.00

Figure 24 – ICMP radiation source countries.



Source: The author.

6.4.1 Most probed TCP ports

Probing activity using the Transmission Control Protocol aimed the list of destination ports as ranked in Table 32.

Table 32 – Most probed TCP destination ports

Rank	TCP Port	Count	%
1	23	9,536,412	2.24
2	80	8,036,707	1.89
3	10000	5,105,766	1.20
4	10001	5,064,150	1.19
5	10002	4,931,950	1.16
6	443	4,524,434	1.06
7	22	4,521,161	1.06
8	6379	2,889,367	0.68
9	9077	2,794,237	0.66
10	8080	2,388,326	0.56
11	8443	2,198,942	0.52
12	3389	1,731,099	0.41
13	445	1,543,228	0.36
14	8088	1,394,605	0.33
15	2375	957,139	0.22
16	8728	894,006	0.21
17	81	859,420	0.20
18	2222	815,181	0.19
19	8888	789,024	0.19
20	5555	762,072	0.18
	Sub-total	61,737,226	14.50
	Total	425,674,635	100.00

The most targeted TCP port was 23, the Telnet application port. Sensors recorded more than nine million events of this type, corresponding to 2.2% of the capture sample. HTTP, on port 80 was second, with 8 million events, or 1.8%. This prevalence follows the same behaviour as our previous, active, honeypot-enabled experiment, documented in (Bortoluzzi et al., 2024).

Mirai, Mozi, and Sora variants exploit a vulnerability where unauthenticated web-shell access can be obtained within a range of IoT IP Cameras. The Kinsing variant exploits the vulnerability documented in CVE-2020-20601, related to remote code execution against a PHP-based Content Management System, ThinkCMS, widely adopted in China. Mirai also exploits the more recent vulnerability documented on CVE-2022-40475, which enables unauthenticated command injection in a range of Totolink branded IoT devices.

The trends observed on this experiment match and reinforces the findings of (Shaikh et al., 2018), in which they have perceived a quantitative radiation pattern shift, from malware-propagation aiming at ordinary server-based operating systems, such as Windows Server and Linux, towards the exploitation of more abundant, billion-device scale, low-cost, low-power,

fast-time-to-market Internet-of-Things devices, which, by their very nature, cannot afford to implement robust security mechanisms. Researchers in (Griffioen; Doerr, 2020) characterise Mirai as the central actor in an ongoing battle over the IoT, while (Gelgi et al., 2024) provides an a updated discussion of Mirai’s propagation patterns.

TCP ports 10000, 10001, and 10002 accounted ≈ 5 million events each. IANA assigned port 10000 to “Network Data Management Protocol”, however at least 22 distinct application-layer services are implemented on this port (Speedguide, 2024a), including open source tools, such as the PHP-based interface Webmin, backup-related agent Veritas Remote Execution Client, and a range of backdoor access software implemented by some malware.

Port 10001 is linked to eight different uses (Speedguide, 2024b), including the live-streaming service of a range of IP cameras, Veeam Agent, some game servers, and a couple of backdoor-based malware. Port 10002 follows the same pattern.

Port 443 implements the well-known HTTPS service. Unsolicited traffic arriving at it is often benign, includes IP sources traceable to Censys and Shodan, plus AWS’s health checks as part of their cloud monitoring services. Remainder traffic, $N < 1\%$, include widely adopted services, such as Microsoft’s Remote Desktop Protocol (3389), and the Common Internet File System - Server Message Block (CIFS/SMB) at 445.

6.4.2 TCP flag combination analysis

TCP flags, as documented on RFC 793 (Postel, 1981b), allow the differentiation of many states of a controlled transmission, including its beginning, with the SYN flag, acknowledgement, with SYN and ACK flags enabled, transmission, with ACK, re-transmission requests, with RST, and connection termination, combining FIN and ACK flags.

Internet Background Radiation often contains malformed and malicious TCP flag combinations, usually aiming at firewall evasion, service exhaustion, or unexpected flag combination, hoping the target stack will halt due to poor, strict or unhardened flag handling.

Table 33 shows flag combinations of received TCP datagrams. The flag array is depicted using the dotted format, inherited from PCAP and decoded by TShark, masked as RRRNCEUAPRSF. From right to left, RFC 793 defines the original 6 TCP flags. RFC 3168 (Ramakrishnan; Floyd; Black, 2001) introduced the ECN-Echo and Congestion Window Reduced flags. RFC 3540 (Wetherall; Spring, 2003) introduced the Nonce Sum flag. The remaining three bits of the TCP flags field are still marked as Reserved.

The vast majority of traffic, 82%, refer to SYN requests, which is the standard behaviour for any new incoming datagrams. Transmission reset requests, accounting to 17%, can be related to either reset-centric port scanning, or to *backscatter*, where radiation sensors collect reflected traffic generated by victim’s server IP, in response to attack packets with forged source IP addresses. Few packets, $N < 0.5\%$, exploited a range of unexpected and odd flag combinations, including ACK+SYN, used for ongoing data exchange (0.36%) and CWE+ECE+SYN used to initiate a congestion window negotiation (0.28%). Transit of the ACK flag alone, Christmas tree

Table 33 – Most used TCP flag combinations

Rank	TCP Flags	Count	%
1	S.	349,778,700	82.17
2	R..	72,511,053	17.03
3	A..S.	1,511,298	0.36
4	CE....S.	1,176,120	0.28
5	A....	219,834	0.05
6	A.R..	181,857	0.04
7	UAPRSF	72,480	0.02
8	A...F	69,063	0.02
9		62,472	0.01
10	AP...	45,260	0.01
11	.R..C....S.	23,186	0.01
12	U...S.	13,027	0.00
13	UAPR.F	1,801	0.00
14	CE.A..S.	1,368	0.00
15	AP...F	1,306	0.00
16	R...CE....S.	1,256	0.00
17	C...A..S.	875	0.00
18	E.A..S.	676	0.00
19	R.....S.	659	0.00
20	...NCE....S.	397	0.00
	Sub-total	425,672,688	100.00
	Total	425,674,635	100.00

scan, UAPRSF, and null scan (empty flags), were also captured in very small quantities.

6.4.3 Most probed UDP ports

UDP is described as a connection-less transport-layer protocol on RFC (Postel, 1980). The UDP portion of Internet Background Radiation helps to understand attack types unrelated to typical system intrusion, either linked to service exhaustion, that is, denial-of-service attacks, or amplification, the situation where a vulnerable, intermediate service is used as a proxy to generate more traffic towards the final target than initially sent by the attacker to the proxy itself (Kumar; Gowda, 2022).

The twenty most probed UDP ports are ranked in Table 34. This ranking accounts for just less than a third of the entire UDP packet count, 30%, suggesting a higher diversity of probing and attacking patterns.

UDP-based traffic was diverse. 5.91% of the packets aimed at port 5100. This is used by a range of services, including Avaya's VoIP/PBX "Communication Server 1000", registered within IANA as *socalia*. SANS Internet Storm Centre reveals the port has had a surge in exploitation activity starting late 2019 (SANS Internet Storm Center, 2024). Port 5060 (2.91%) is officially registered as the standard Session Initiation Protocol - SIP, the VoIP port. VoIP im-

Table 34 – Top 20 UDP Destination ports

Rank	UDP Port	Count	%
1	5100	1,121,200	5.91
2	5060	551,508	2.91
3	53	541,760	2.85
4	123	410,870	2.16
5	9993	403,120	2.12
6	161	305,246	1.61
7	1900	286,660	1.51
8	5353	237,173	1.25
9	1434	220,238	1.16
10	500	196,574	1.04
11	54321	168,690	0.89
12	6881	156,758	0.83
13	27015	152,184	0.80
14	11211	152,101	0.80
15	1194	148,808	0.78
16	443	133,434	0.70
17	137	133,144	0.70
18	5683	126,760	0.67
19	8080	126,714	0.67
20	8081	124,638	0.66
	Sub-total	5,697,580	30.01
	Total	18,982,971	100.00

plementations, such as in Asterisk, Microsoft Lync, Cisco IOS and Siemens Openstage/Gigaset are vulnerable to a range of exploits widely available on the Internet. The DNS port, 53, was the target in 2.81% of cases. DNS implementations, such as widely-adopted BIND, released by the Internet Software Consortium, ISC, contains multiple critical vulnerabilities. Since 2020, 44 CVE entries have been published reporting situations where BIND can be remotely exploited (Cvedetails.com, 2024a) with publicly available code. Network Time Protocol - NTP exploitation came fourth, with 2.16% of the collected radiation. Implementations of this protocol were subject to a spike of 29 disclosed vulnerabilities around 2017, 19 out of which relate to input validation, and 10 to buffer and heap overflow situations (Cvedetails.com, 2024a). NTP is widely used by attackers as a vector to achieve amplification (Ismail et al., 2021). Other interesting cases are port 9993 (2.12%) linked to the Zero Tier service and network encapsulation stack, 161, the Simple Network Management Protocol port, 1900, used by SSDP/UPnP (Simple Service Discovery Protocol - Universal Plug-and-Play protocol) to enable automatic local service discovery, 5353, related to Multicast DNS, and 1434, the Microsoft SQL Server, where the UDP port provides service management to that database software.

6.4.4 Most frequent ICMP messages

ICMP message exchange from probing sources towards the sensor fleet can be explained as a) malicious motivation when a host scanner must first determine whenever a target IP address is live, b) part of target operating system fingerprinting attempt, c) part of a multi-protocol denial-of-service attack, or d) ping-based host monitoring misconfiguration; among other reasons.

A basic account of arriving Internet Control Message Protocol - ICMP is presented in Table 35.

Table 35 – Top 5 ICMP message types

Rank	ICMP Type	Count	%
1	8	81,592,126	94.72
2	0	4,546,190	5.28
3	13	785	0.00
4	17	245	0.00
5	1	129	0.00
	Sub-total	86,139,475	100.00
	Total	86,139,655	100.00

Captured radiation contains 5 distinct ICMP message types. The vast majority refers to ICMP echo request, type 8, as initiated by the ping command. This was the case in 94% of the sample. ICMP echo response, type 0, usually implemented as a response to a ping request, was seen in 5% of the cases. Network and Cloud telescopes are configured to never issue a ping request, therefore the arrival of a response is unsolicited, and could be a result of a packet forgery attempt conducted by an attacker using the sensor's IP address as a response destination. ICMP types 13 "timestamp request", 17 "address mask request" and 1 "destination host unreachable" were occasionally captured.

6.5 DISTRIBUTED ANALYSIS

The distributed architecture of the Cloud Telescope provides an opportunity to investigate how traffic dispersion affects different countries in the world. Therefore, this subsection quantifies how each country hosting a share of capture sensors was hit by overall traffic, followed by a breakdown of TCP, UDP and ICMP network events. In each case, not only a country-level aggregation, but a Cloud Region aggregation is provided to compensate the existence of more than one Cloud Region in a single country, most notably, the United States of America.

Country-level aggregations are capped at the twenty most affected targets to preserve consistency across all result reported, resulting in a Top 20 ranking for each aggregation. Cloud

Region accounting, on the other hand, follows the number of effective experimented AWS points of presence: 26, therefore, tables provide an absolute dispersion accounting, $N = 100\%$.

6.5.1 Overall traffic dispersion

Overall traffic dispersion is accounted for the twenty most affected countries in Table 36. This is regardless of any network layer or transport layer property. The listed countries answer together for 518 million network events, 97% of the total captured amount.

Table 36 – Top 20 Radiation Target Countries

Rank	Country	Count	%
1	United States	84,327,417	15.88
2	Bahrain	44,167,957	8.32
3	Indonesia	36,191,361	6.82
4	Japan	35,420,141	6.67
5	India	32,512,935	6.12
6	Singapore	29,481,087	5.55
7	South Korea	22,446,256	4.23
8	Hong Kong	21,896,218	4.12
9	Sweden	20,274,200	3.82
10	France	20,016,313	3.77
11	Switzerland	19,528,561	3.68
12	United Kingdom	18,961,152	3.57
13	Germany	18,694,488	3.52
14	Ireland	17,631,056	3.32
15	Canada	17,149,403	3.23
16	Australia	17,063,745	3.21
17	Spain	16,909,712	3.19
18	South Africa	16,014,791	3.02
19	Italy	15,039,747	2.83
20	United Arab Emirates	14,376,969	2.71
	Sub-total	518,103,509	97.59
	Total	530,914,922	100.00

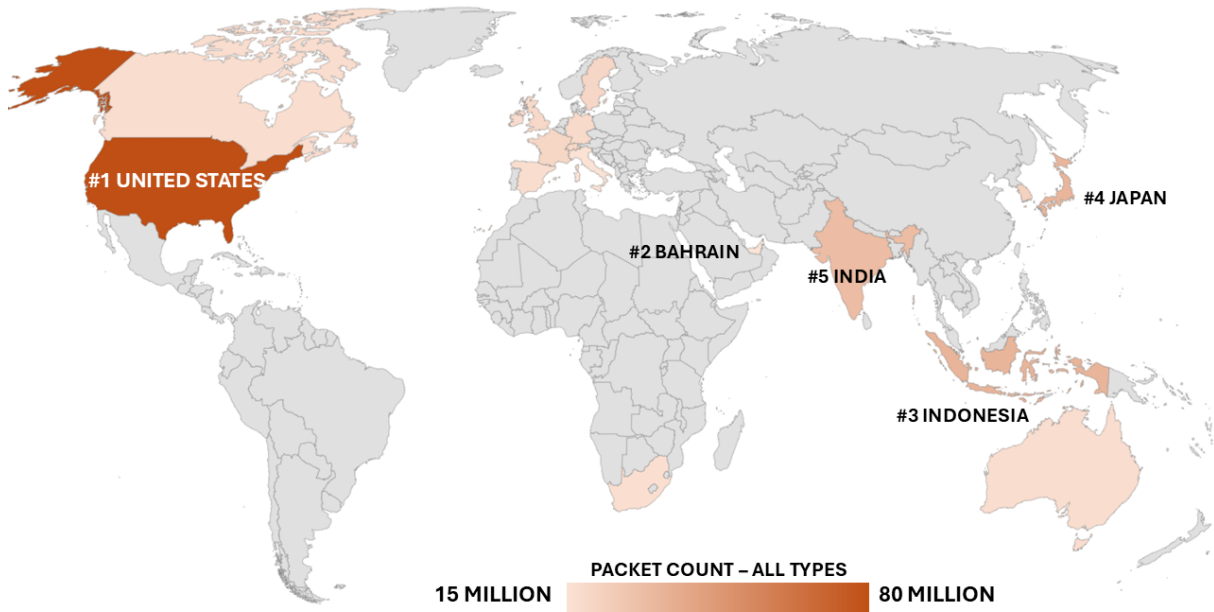
The United States of America is covered with four distinct cloud regions and, as such, was assigned four times the number of baseline sensors deployable in each region, 4 cloud regions $\times$ 5 sensors per region = 20 sensors. The USA was targeted by 84 million network events, corresponding to 15% of all traffic.

The country-level analysis points out Bahrain as the second most affected country, standing as target for 44 million events, or 8% of the entire dataset. Indonesia figures in third position, being the target of 36 million packets, or 6% of the volume. Japan also received a share of 6% of all traffic, followed by India, with 32 million, Singapore, 29 million, and South Korea, with 22 million. Hong Kong was hit by 21 million packets, while Sweden and France

were the last countries receiving at least 20 million packets each. All remainder countries in the list were hit with at least 14 million packets each.

The heat map in Figure 25 depicts the same twenty countries, highlighting the five most affected nations.

Figure 25 – Top radiation destinations.



Source: The author.

To neutralise the influence of countries containing more than one cloud region, consider the radiation distribution individually arriving to each of the twenty-six AWS Cloud Regions, as accounted in Table 37.

This perspective reveals that me-south-1, Bahrain, figures as the most targeted country, with 8% of the traffic, followed by ap-southeast-3, Indonesia, with 6%. us-west-1, in the USA, figures only as the third most affected individual Cloud Region, with 5.91% of the received packets.

Radiation normalization indicates that the theoretical average radiation distribution per Cloud Region should be $R = 1/26 = 3.8\%$. AWS regions hit above the baseline average also include ap-southeast-1 (5%), ap-northeast-2 (4%), ap-east-1 (4%), and eu-north-1 (3.82%). On the other hand, the least targeted region was sa-east-1, with only 2.41% of the captured events.

Despite the discrepancies, radiation distribution can be considered fairly even, with a median of $med(R) = 3.36$ and standard deviation of $dev(R) = 1.35$.

6.5.2 TCP radiation: target countries

The TCP-based radiation accounted for 425 million network events, as summarised in Table 38. The ranking of the twenty most affect nations corresponded to 97% of cases, revealing

Table 37 – Traffic dispersion per Cloud Region

Rank	AWS Region	Count	%
1	me-south-1	44,167,957	8.32
2	ap-southeast-3	36,191,361	6.82
3	us-west-1	31,397,821	5.91
4	ap-southeast-1	29,481,087	5.55
5	ap-northeast-2	22,446,256	4.23
6	ap-east-1	21,896,218	4.12
7	eu-north-1	20,274,200	3.82
8	eu-west-3	20,016,313	3.77
9	eu-central-2	19,528,561	3.68
10	eu-west-2	18,961,152	3.57
11	ap-northeast-1	18,800,745	3.54
12	eu-central-1	18,694,488	3.52
13	us-west-2	18,023,519	3.39
14	us-east-1	17,690,508	3.33
15	eu-west-1	17,631,056	3.32
16	ap-south-1	17,564,914	3.31
17	us-east-2	17,215,569	3.24
18	ca-central-1	17,149,403	3.23
19	ap-southeast-2	17,063,745	3.21
20	eu-south-2	16,909,712	3.19
21	ap-northeast-3	16,619,396	3.13
22	af-south-1	16,014,791	3.02
23	eu-south-1	15,039,747	2.83
24	ap-south-2	14,948,021	2.82
25	me-central-1	14,376,969	2.71
26	sa-east-1	12,810,879	2.41
	Total	530,914,922	100.00

interesting concentration patterns.

The country-level aggregation confirms that the USA as a preferred destination to 72 million TCP events, or 17% of this traffic type, followed by Japan, hit by 31 million events, or 7%, India, with 28 million events, or 6%, and Singapore, with 23 million events, or 5%. Hong Kong was the last nation to be hit by at least 20 million packets.

The heat map in Figure 26 provides a gradient-based colouring scheme denoting the extent to which each country is affected by TCP-based radiation.

6.5.3 TCP radiation: target cloud regions

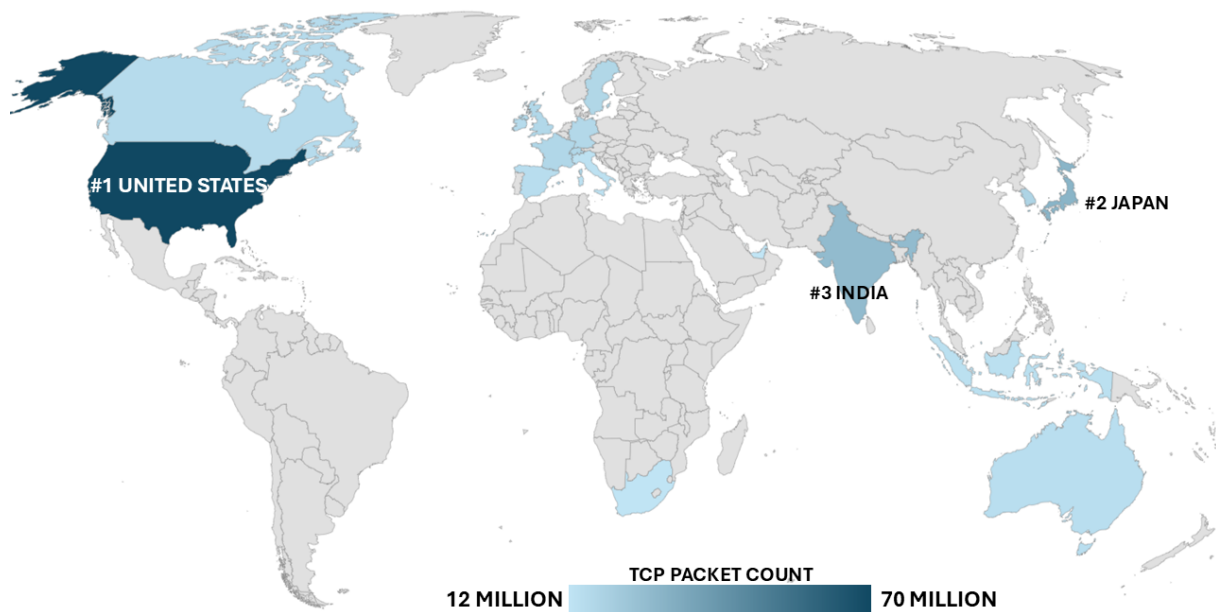
TCP event aggregation as arriving to each individual Cloud Region is accounted for in Table 39. This perspective reveals a fair radiation distribution across each of the twenty-six AWS locations.

This aggregation reveals that not only the United States of America figures first be-

Table 38 – TCP radiation: target countries

Rank	Country	Count	%
1	United States	72,865,034	17.12
2	Japan	31,041,427	7.29
3	India	28,473,594	6.69
4	Singapore	23,720,064	5.57
5	Hong Kong	20,483,122	4.81
6	South Korea	18,928,866	4.45
7	Sweden	17,988,197	4.23
8	France	17,813,716	4.18
9	Switzerland	17,726,802	4.16
10	United Kingdom	17,506,436	4.11
11	Germany	17,350,192	4.08
12	Ireland	16,122,775	3.79
13	Canada	16,103,504	3.78
14	Spain	15,450,831	3.63
15	Australia	14,991,641	3.52
16	Indonesia	14,343,655	3.37
17	Bahrain	14,130,719	3.32
18	Italy	13,873,557	3.26
19	South Africa	12,717,694	2.99
20	United Arab Emirates	12,426,824	2.92
	Sub-total	414,058,650	97.27
	Total	425,674,635	100.00

Figure 26 – TCP radiation target countries.



Source: The author.

Table 39 – TCP radiation: target cloud regions

Rank	AWS Region	Count	%
1	us-west-1	25,408,678	5.97
2	ap-southeast-1	23,720,064	5.57
3	ap-east-1	20,483,122	4.81
4	ap-northeast-2	18,928,866	4.45
5	eu-north-1	17,988,197	4.23
6	eu-west-3	17,813,716	4.18
7	eu-central-2	17,726,802	4.16
8	eu-west-2	17,506,436	4.11
9	eu-central-1	17,350,192	4.08
10	ap-northeast-1	16,934,777	3.98
11	us-west-2	16,447,649	3.86
12	eu-west-1	16,122,775	3.79
13	ca-central-1	16,103,504	3.78
14	us-east-1	15,816,274	3.72
15	eu-south-2	15,450,831	3.63
16	us-east-2	15,192,433	3.57
17	ap-southeast-2	14,991,641	3.52
18	ap-south-1	14,809,750	3.48
19	ap-southeast-3	14,343,655	3.37
20	me-south-1	14,130,719	3.32
21	ap-northeast-3	14,106,650	3.31
22	eu-south-1	13,873,557	3.26
23	ap-south-2	13,663,844	3.21
24	af-south-1	12,717,694	2.99
25	me-central-1	12,426,824	2.92
26	sa-east-1	11,615,472	2.73
	Total	425,674,635	100.00

cause it contains four cloud regions, but also because it hosts us-west-1, the most individually targeted one, with 25 million TCP events, or almost 6% of the captured sample. ap-southeast-1 closely follows, with 23 million packets, or 5.5% of the cases. Regions affected above the baseline average also include ap-east-1, 20 million events, or 4.8%, ap-northeast-2, with 18 million events, or 4.4%, eu-north-1, eu-west-3, eu-central-2, eu-west-2, and eu-central-1, with 17 million packet events each, or $\approx 4\%$, followed by ap-northeast-1, us-west-2, eu-west-1, ca-central-1, and us-east-1, each seeing 3.72%-3.98% of all TCP-based traffic.

6.5.4 UDP radiation: target countries

The relatively small UDP packet count captured in this experiment is listed in Table 40. The list containing the twenty most affected countries represent 98% of all UDP events. Concentration is significantly biased towards the first five nations.

US-based IP addresses received more than one third of all UDP-based radiation. 6.5

Table 40 – UDP radiation: target countries

Rank	Country	Count	%
1	United States	6,490,213	34.19
2	Japan	1,777,701	9.36
3	India	1,476,866	7.78
4	Sweden	992,101	5.23
5	South Africa	786,513	4.14
6	Singapore	739,854	3.90
7	France	628,691	3.31
8	Indonesia	579,119	3.05
9	United Kingdom	520,162	2.74
10	Hong Kong	499,664	2.63
11	Spain	466,502	2.46
12	Bahrain	436,098	2.30
13	Germany	431,049	2.27
14	Switzerland	418,903	2.21
15	Ireland	411,150	2.17
16	South Korea	402,309	2.12
17	Canada	397,087	2.09
18	Brazil	393,717	2.07
19	United Arab Emirates	393,429	2.07
20	Italy	377,770	1.99
	Sub-total	18,618,898	98.08
	Total	18,982,971	100.00

million network events, or 34% of the subset. Japan was hit by 1.7 million UDP datagrams, or 9% of this traffic type. India was the last country to account at least 1 million network events, comprising 7% of the cases. Other relevant targets include Sweden, with 5% of the load, South Africa, 4%, and Singapore, 3.9%, or 730 thousand datagrams.

The heat map in Figure 27 summarises UDP radiation destinations, highlighting the most targeted countries.

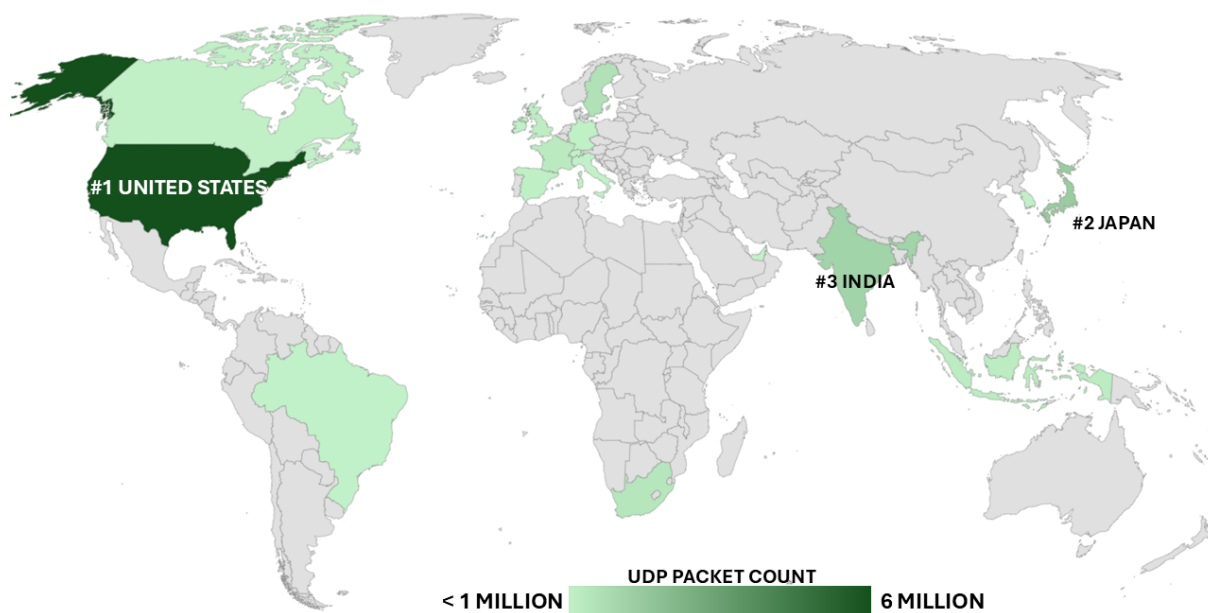
6.5.5 UDP radiation: target cloud regions

The cloud-region-centric analysis of UDP events is supported by the accounting provided in Table 41.

This perspective reveals that one single US based cloud region, us-west-1, was the preferred target. Alone, it was targeted by 27% of the radiation. ap-northeast-3 comes second, with 1.2 million network events, or 6% of the cases. eu-north-1, ap-south-1, and ap-southeast-1 were also hit with a percentage higher than the baseline average, ranging from 3.90 to 5.23% of the UDP-based traffic load.

A query on the dataset, aiming to raise the most prevalent sources to justify this particular aggregation, revealed that UDP-based traffic towards us-west-1 had the USA itself as the

Figure 27 – UDP radiation target countries.



Source: The author.

most prevalent attacking source.

6.5.6 ICMP radiation: target countries

Table 42 ranks the twenty targets most affected by ICMP events. The list answers for 99% of all seen ICMP radiation, revealing a high concentration of two countries as the target of more than half of the packet load.

Bahrain received one third of all captured ICMP traffic, almost 30 million events, or 34% of this event type, followed by Indonesia, with 21 million packets, 24% of the ICMP load. Singapore-based sensors captured 5% of this traffic type, 5 million packets, closely followed by the United States, with a similar amount.

The corresponding heat map, depicted in Figure 28, reveals the ICMP traffic dispersion across the most targeted countries, highlighting the three most prevalent destinations.

6.5.7 ICMP radiation: target cloud regions

The cloud-region-based analysis of ICMP events reinforces the country-level aggregation, as shown in Table 43.

me-south-1 was hit by almost 30 million ICMP packets during the capture, contributing alone with a third of the total load seen for network-layer probing activity. Sensors placed in this region were hit by a total amount of ≈ 44 million packets, indicating that ICMP Type 8, echo request, alone, corresponds to two thirds, or 67% of all seen traffic towards this middle-

Table 41 – UDP radiation: target cloud regions

Rank	AWS Region	Count	%
1	us-west-1	5,155,639	27.16
2	ap-northeast-3	1,247,518	6.57
3	eu-north-1	992,101	5.23
4	ap-south-1	878,820	4.63
5	af-south-1	786,513	4.14
6	ap-southeast-1	739,854	3.90
7	eu-west-3	628,691	3.31
8	ap-south-2	598,046	3.15
9	ap-southeast-3	579,119	3.05
10	ap-northeast-1	530,183	2.79
11	eu-west-2	520,162	2.74
12	ap-east-1	499,664	2.63
13	us-east-2	497,831	2.62
14	eu-south-2	466,502	2.46
15	me-south-1	436,098	2.30
16	eu-central-1	431,049	2.27
17	us-west-2	428,848	2.26
18	eu-central-2	418,903	2.21
19	eu-west-1	411,150	2.17
20	us-east-1	407,895	2.15
21	ap-northeast-2	402,309	2.12
22	ca-central-1	397,087	2.09
23	sa-east-1	393,717	2.07
24	me-central-1	393,429	2.07
25	eu-south-1	377,770	1.99
26	ap-southeast-2	364,052	1.92
	Total	18,982,971	100.00

east country. This drifts considerably from typical TCP-centric dispersion as the average case (as shown in Table 26). The origin of the ping-like activity traces back to China in 40% of the cases (≈ 11 million events), followed by Singapore, with 6.2 million packets or 21%, and the United States of America, with 2.5 million, or 8%.

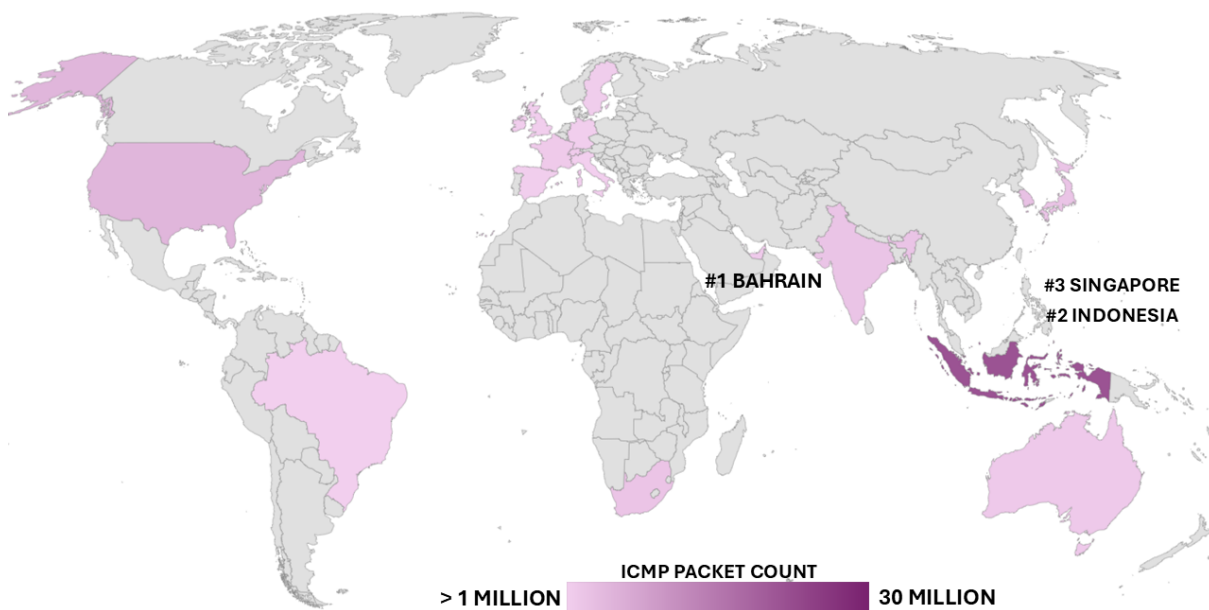
Region ap-southeast-3, was also hit by a large number of ICMP packets, 21 million, representing one fourth of the cases. Considering the total number of 36 million events arriving into that region, ICMP alone corresponded to almost 60% of all incoming packets. Traffic traces back to the same three countries, in order: ≈ 9 million echo requests came from China (42%), while 4 million came from Singapore (20%), followed by US-based IP sources: 1.8 million, or 8%.

Cloud region ap-southeast-3, Singapore, ranking third, showed a behaviour aligned with the average case. ICMP answered for only 17% of the packets, against 80% TCP. Ping-like activity mostly came from Indonesia (26%), followed by China, with 19%. ap-northeast-2, in Korea, in fourth, obtained similar results, where only 13% of the traffic belonged to ICMP,

Table 42 – ICMP radiation: target countries

Rank	Country	Count	%
1	Bahrain	29,601,063	34.36
2	Indonesia	21,268,519	24.69
3	Singapore	5,021,099	5.83
4	United States	4,971,772	5.77
5	South Korea	3,114,983	3.62
6	Japan	2,600,918	3.02
7	India	2,562,315	2.97
8	South Africa	2,510,510	2.91
9	Australia	1,707,994	1.98
10	France	1,573,836	1.83
11	United Arab Emirates	1,556,649	1.81
12	Switzerland	1,382,387	1.60
13	Sweden	1,178,605	1.37
14	Ireland	1,097,061	1.27
15	Spain	992,318	1.15
16	United Kingdom	934,470	1.08
17	Hong Kong	913,365	1.06
18	Germany	913,078	1.06
19	Brazil	801,612	0.93
20	Italy	788,352	0.92
	Top 20 Sub-total	85,490,906	99.25
	Total	86,139,655	100.00

Figure 28 – ICMP radiation target countries.



Source: The author.

Table 43 – ICMP radiation: target cloud regions

Rank	AWS Region	Count	%
1	me-south-1	29,601,063	34.36
2	ap-southeast-3	21,268,519	24.69
3	ap-southeast-1	5,021,099	5.83
4	ap-northeast-2	3,114,983	3.62
5	af-south-1	2,510,510	2.91
6	ap-south-1	1,876,219	2.18
7	ap-southeast-2	1,707,994	1.98
8	eu-west-3	1,573,836	1.83
9	me-central-1	1,556,649	1.81
10	us-east-2	1,525,143	1.77
11	us-east-1	1,466,275	1.70
12	eu-central-2	1,382,387	1.60
13	ap-northeast-1	1,335,721	1.55
14	ap-northeast-3	1,265,197	1.47
15	eu-north-1	1,178,605	1.37
16	us-west-2	1,146,920	1.33
17	eu-west-1	1,097,061	1.27
18	eu-south-2	992,318	1.15
19	eu-west-2	934,470	1.08
20	ap-east-1	913,365	1.06
21	eu-central-1	913,078	1.06
22	us-west-1	833,434	0.97
23	sa-east-1	801,612	0.93
24	eu-south-1	788,352	0.92
25	ap-south-2	686,096	0.80
26	ca-central-1	648,749	0.75
	Total	86,139,655	100.00

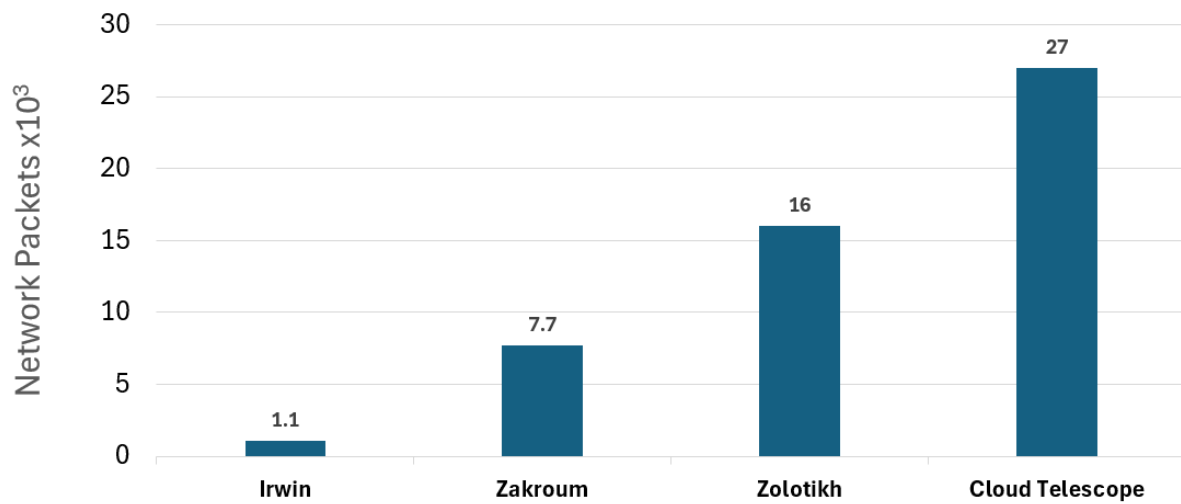
70% of which coming from Singapore.

7 DISCUSSION

This chapter discusses a selection of qualitative properties resulting from the main experiment. A direct comparison with standard Network Telescopes is provided when applicable, followed by some insights only applicable to the properties of the Cloud Telescope alone.

As a baseline, consider the number of captured network events by the Cloud Telescope against standard Network Telescopes. This is possible in many key aspects, including the total number of network packets arriving to each IP address per day, target port aggregation, and the most frequently seen probing/attacking sources. A normalised count of events arriving to three Network Telescopes, per day, per allocated IP address is summarised in Figure 29:

Figure 29 – Hourly network events per allocated IP address.



Source: The author.

In 2011, Irwin (2011) captured 40 million events in 50 months on a /24 subnet, hence 1,1 thousand network packets per day, per IP address. Ten years later, Zakroum et al. (2022) saw 80 billion packets arriving to 8,192 addresses during 42 months, or 7,7 thousand network packets per day, per IP address. Zolotykh & Nikulin (2020) reports the collection of 2 million packets for a single IP address during their four-month-long experiment, or 16 thousand network packets per day. The Cloud Telescope captured 27 thousand packets per day, per allocated IP address, reinforcing the empirical assumption that radiation emission only grows as Internet adoption expands across the world.

7.1 TCP TRAFFIC

The Cloud Telescope had TCP as the most exploited transport-layer protocol, with 80% prevalence. This result aligns with Irwin (2011), who reported 81% of TCP dominance, while Zakroum et al. (2022) reported 94%, and Zolotykh & Nikulin (2020), 89%. Such alignment can indirectly support some empirical assumptions. One is that the traffic affecting cloud-based

sensors, making use of IP address allocations belonging to known cloud service providers does not differ in statistical terms from the traffic usually directed to ordinary, non cloud-based assets. This particular aspect is explored in works such as Wani et al. (2023), where authors aim at determining the significance of IP addresses in cloud services and how their dynamic allocation can impact various forms of data analysis.

A second concern that could have affected the TCP-UDP-ICMP protocol split comes from the fact that it is up to the cloud service provider to protect their infrastructure from most forms of denial-of-service attacks, particularly the ones exploiting flaws in the IP stack unlikely to contain productive services. AWS states that, as part of the shared responsibility model, it is up to the cloud service provider to protect their infrastructure, while cloud users are expected to protect customer data and enable advanced protections as they see fit to their cloud architecture and resources (AWS, 2025b). If these mechanisms were put in place despite the explicit default-allow posture enforced by the Cloud Telescope, very likely the protocol split would drift away from the counterpart experiments, and ICMP and UDP events would be detected to even lower amounts.

TCP-based port aggregation reveals that the Cloud Telescope had the Telnet port as the most intensely probed, followed by other well-known Internet services ranking high, including HTTP, SSH, HTTPS, and others. An attempt to find an intersection of the three most exploited TCP ports figuring high in related works and the Cloud Telescope is summarised on Table 44:

Related work	First	Second	Third
Zolotych et al. (2020)	23	22	445
Zakroum et al. (2022)	22	23	80
Kori et al. (2024)	22	23	80
Cloud Telescope	23	80	443

Table 44 – Intersection of most prevalent TCP ports in related works

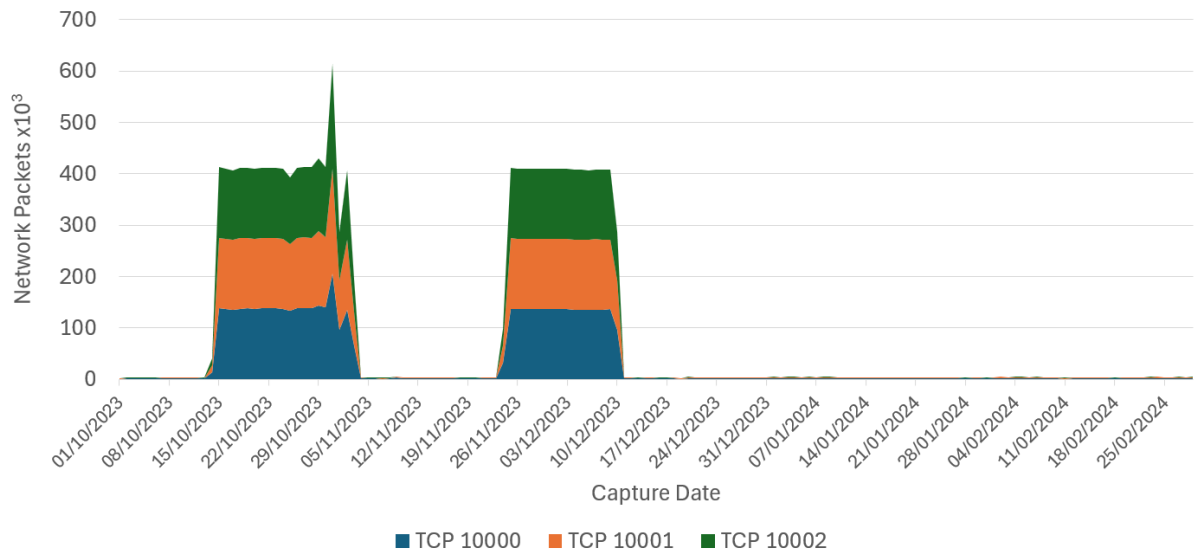
The intersection reveals a general interest for remote administration services (Telnet, SSH), followed by a range of ports that includes web services (80, 443) and then other services.

Zolotych & Nikulin (2020) reported 23, 22 and 445 as the most frequently probed ports, while Zakroum et al. (2022) reported SSH (22), Telnet (23) and HTTP (80) ranking the highest. Kori et al. (2024) reported 22, 23, and 80 as the top three. The Cloud Telescope had ports 23 and 80 as the two most prevalent. The third place drifts from the intersection, having ports 10000, 10001 and 10002 as the most intensely probed. Port 443 (HTTPS) ranking #6 on the Cloud Telescope is the third port intersecting with reference experiments.

The most frequently explanation to justify the prevalence of remote administration ports, in particular, the Telnet port, is the propagation of the Mirai botnet and its variants. This seems to be the trend since the ascension of the Internet of Things, around ten years ago (2015). On the other hand, the Cloud Telescope was the only mechanism reporting high prevalence of traffic destined to the triplet formed by ports 10000-10001-10002 as the third, fourth and fifth most exploited TCP services. An attempt to look further into this drift resulted on the traffic

aggregation depicted in Figure 30.

Figure 30 – Daily traffic distribution targeting TCP ports 10000,10001 and 10002.



Source: The author.

The chart reveals that there is a high concentration of probing events at two different moments of the capture time frame. The first peak started fifteen days after the capture began (15 October) and lasted for eighteen days, until 03 November. There was an interval of twenty days until a second peak started, in 24 November, lasting for seventeen days until 10 December 2023. The amount of packets arriving to each of the three ports on each day was almost the same, indicating that the actor(s) driving these scans were equally interested on learning about potential vulnerabilities residing on any of these ports. The spike activity is uncommon when compared to the usual, continuous behaviour seen for other dominating ports. Consider for example the more ordinary traffic breakdown for the port group intersecting with related works depicted in Figure 31.

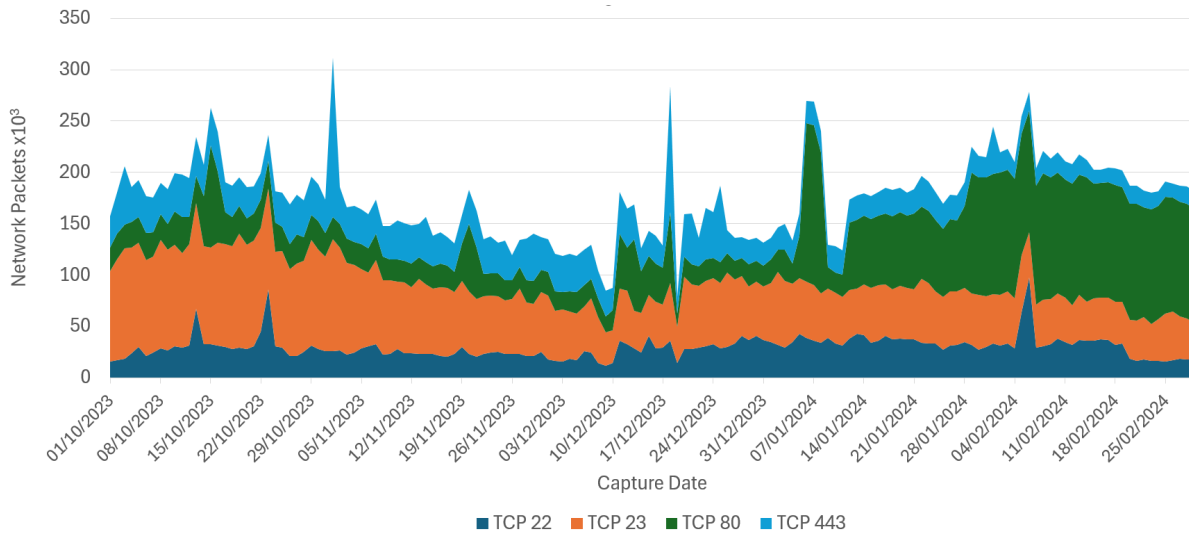
Event distribution of ordinary ports (22, 23, 80, 443) is fairly even across each capture day. The spike of 311,682 events received on 01 November, of which 154,906 (49%) aimed at 443, does not significantly drift away from the average of ≈ 170 thousand events daily.

7.2 UDP TRAFFIC

UDP was seen in much lower quantity both in related works and during the main Cloud Telescope experiment. Table 45 summarises the elements for a comparison.

UDP prevalence was of 12% in Irwin (2011). At the time, top destination ports contained SQL Server related probing activity aiming at 1434, along with Windows RPC-based services on ports 137 and 135. Zakroum et al. (2022) reported 5% UDP traffic, highlighting SIP (5060) as the most targeted. Zolotykh & Nikulin (2020) do not provide an UDP port breakdown, but mention intense targeting of the DNS protocol (53). The Cloud Telescope had

Figure 31 – Daily traffic distribution targeting prevalent UDP ports.



Source: The author.

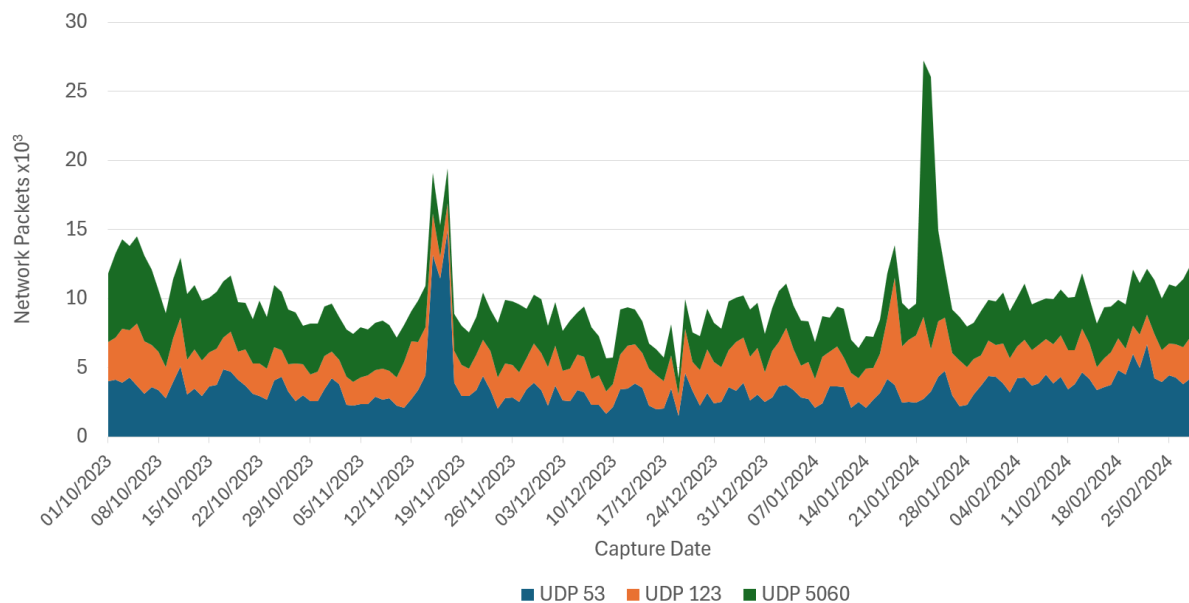
Related work	UDP prevalence	Top port	Other ports
Irwin (2011)	12%	1434	137, 135
Zakroum (2022)	5%	5060	
Zolotykh (2020)	No breakdown	53 (mentioned)	
Cloud Telescope	3.58%	5100	5060, 53, 123

Table 45 – Prevalence of UDP target ports in related works

3.58% of the captured packets belonging to UDP, including the SIP protocol, followed by DNS. Windows RPC-based services also ranked high when using the distributed method.

A look for spikes diverging from average cases is plotted in Figure 32.

Figure 32 – Daily traffic distribution targeting prevalent UDP ports.



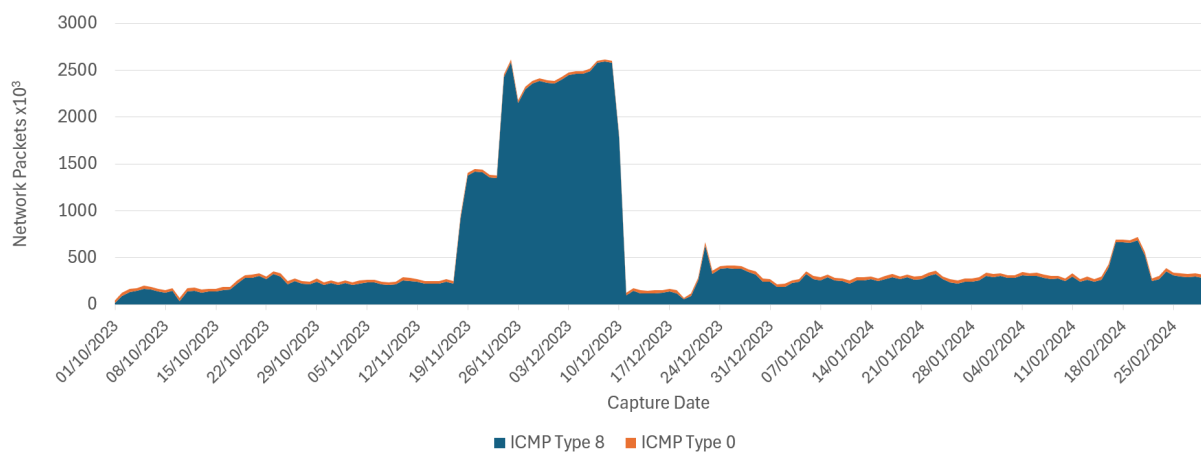
Source: The author.

It was possible to detect two diverging moments. The first burst, aiming at the DNS protocol, started on 15 November, and lasted two days. The average packet count during those days were of ≈ 17 thousand packets/day, out of which ≈ 13 thousand packets/day were destined to port 53 alone. A second spike, this time related to the SIP port, began in 22 January, and reached 18,539 unique events the day after. These were considerably higher than the ordinary average of ≈ 3.2 thousand events received daily on each of the three analysed ports.

7.3 ICMP TRAFFIC

ICMP was captured in low amount in all comparable cases. The Cloud Telescope essentially captured echo requests (95%), while Irwin (2011) had a fair dispersion of types 8 (62%), 3 (24%) and 11 (12%). Zakroum et al. (2022) and Zolotykh & Nikulin (2020) reported $> 1\%$ of ICMP traffic. Figure 33 depicts ICMP distribution along time.

Figure 33 – Daily traffic distribution of prevalent ICMP types.



Source: The author.

ICMP traffic had one intense moment of particular interest. While in the average case only 533 thousand ICMP type 8 packets were captured daily, from 19 November to 10 December, 47,656,002 packets arrived to the sensor fleet, denoting a burst of 2,166,181 network events per day. ICMP type 0, the only other type captured during the experiment, had an average of only ≈ 29 thousand packets/day with no relevant spike or time concentration.

7.4 IMPLICATIONS OF THE DISTRIBUTED APPROACH

The distributed nature of the Cloud Telescope has introduced new forms of dataset aggregation. Consider for example the discussion around overall traffic dispersion, detailed starting in subsection 6.5.1 (page 124 onward). That aggregation enabled to determine the amount of traffic that is destined to each cloud region, hence, each region of the world in which the cloud service provider hosts one or more availability zones. Leaving aside the fact the United States of America was deployed with four times the fleet size of other countries, it is

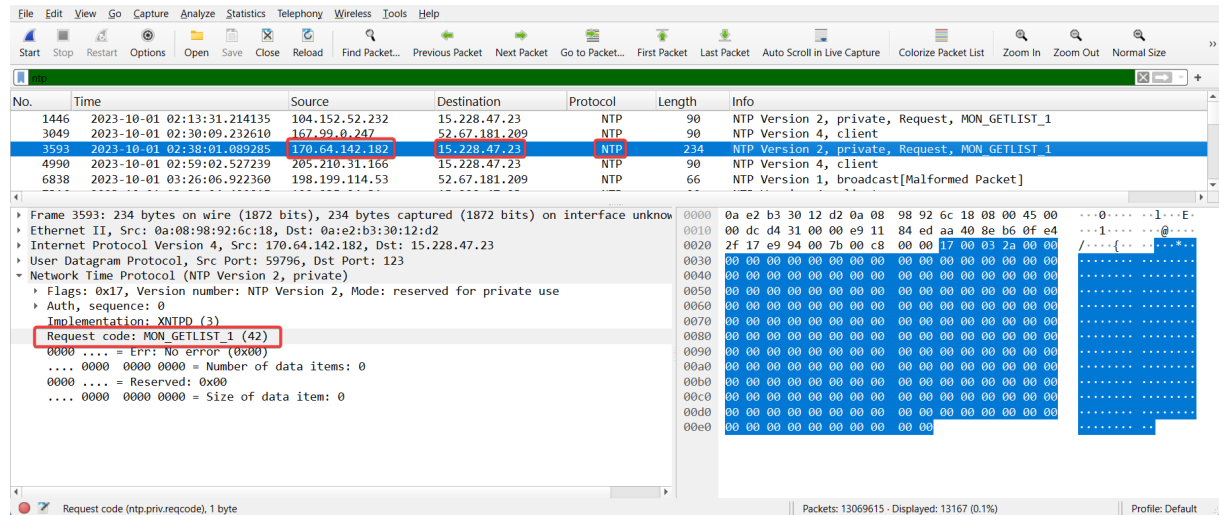
still possible to see Bahrain as the second most preferred target, with 8.3% of all traffic arriving to its single cloud region. That is $\approx 200\%$ more radiation than received in same fleet sizes located in South Africa, Italy, or the United Arab Emirates. It is particularly interesting the fact Bahrain and the UAE are both located in the Middle East and share borders. What could explain the former stands as the most preferred individual target and the latter, the least? The Cloud Telescope provides a start point for an analysis into potential drivers for a motivation that can be potentially linked with events taking place on the *surface*, in the geopolitical realm.

The discussion of protocol types affecting each region of the world, as unfolded in subsections 6.5.2 and onward is relevant because they can be linked with the type of attack intended, potentially narrowing down the search space for attribution. Generally speaking, most of the probing activity using the Transmission Control Protocol can be directly related to application-layer service exploitation. Normally, the intent behind these attacks is to compromise the target and take advantage of its computational resources: CPU, RAM, storage, network capabilities. Data exfiltration may or may not be the final intent, depending on how the target was chosen. Targeted attacks tend to aim at data exfiltration, while non-targeted attacks, the most prevalent ones in background radiation, seem to aim at building computer fleets that will, in the future, trigger denial-of-service attacks against selected targets. That is the case of Mirai and similar botnets. There is a botnet building phase, where attacks are conducted, predominantly against IoT devices, to build a fleet that will respond to a set of attack commands, and there is the botnet attack phase, where network traffic of varied forms will be generated by infected devices and sent out to the target. Internet Background Radiation, as perceived by the Cloud Telescope contains large amounts of traffic related to the building phase. These can be detected with the standard, passive method, but can only be elucidated, studied and understood in reasonable depth with application-layer responders. Experiment #2 (Subsection 5.3, page 95) was very fruitful in this regard, not only confirming the prevalence of botnet attacks as evidenced by related works but also into determining detection patterns in the form of regular expressions that can be applied straight into the pcap recordings and reveal the source addresses of infected devices, and, more importantly, the IP addresses of compromised payload distribution hosts. These were found to be fully based on TCP communication and mostly exploiting vulnerabilities that make more sense on embedded devices than ordinary, fully-fledged computer systems. The presence of traffic belonging to the attack phase as either not captured or has not emerged with the use of the data analysis pipeline that was created to find relevant patterns in the dataset, and should deserve further investigation.

UDP radiation does not easily fit one single pattern. The classic reason to explain the use of a connectionless protocol to generate malicious traffic has been directly linked to denial-of-service campaigns (Campbell, 2005; Chaudhari; Talmale, 2019). Because these attacks are targeted, the Cloud Telescope is not effective at detecting the attack phase of many attack forms, including the ones promoted by botnets such as Mirai, but vestiges of infection attempts do exist in the datasets. The most evident use-case for UDP, as detected by the method, is attack amplification. Consider a researcher who could be interested in quantifying vestiges

of attack amplification exploiting vulnerabilities in NTP, as documented on CVE-2013-5211 and discussed by Czyz et al. (2014). The Cloud Telescope, as much as any Network Telescope, can detect, record and decode the MON_GETLIST_1 command exploited during the attack, as illustrated in Figure 34.

Figure 34 – NTP amplification attack exploiting CVE-2013-5211.



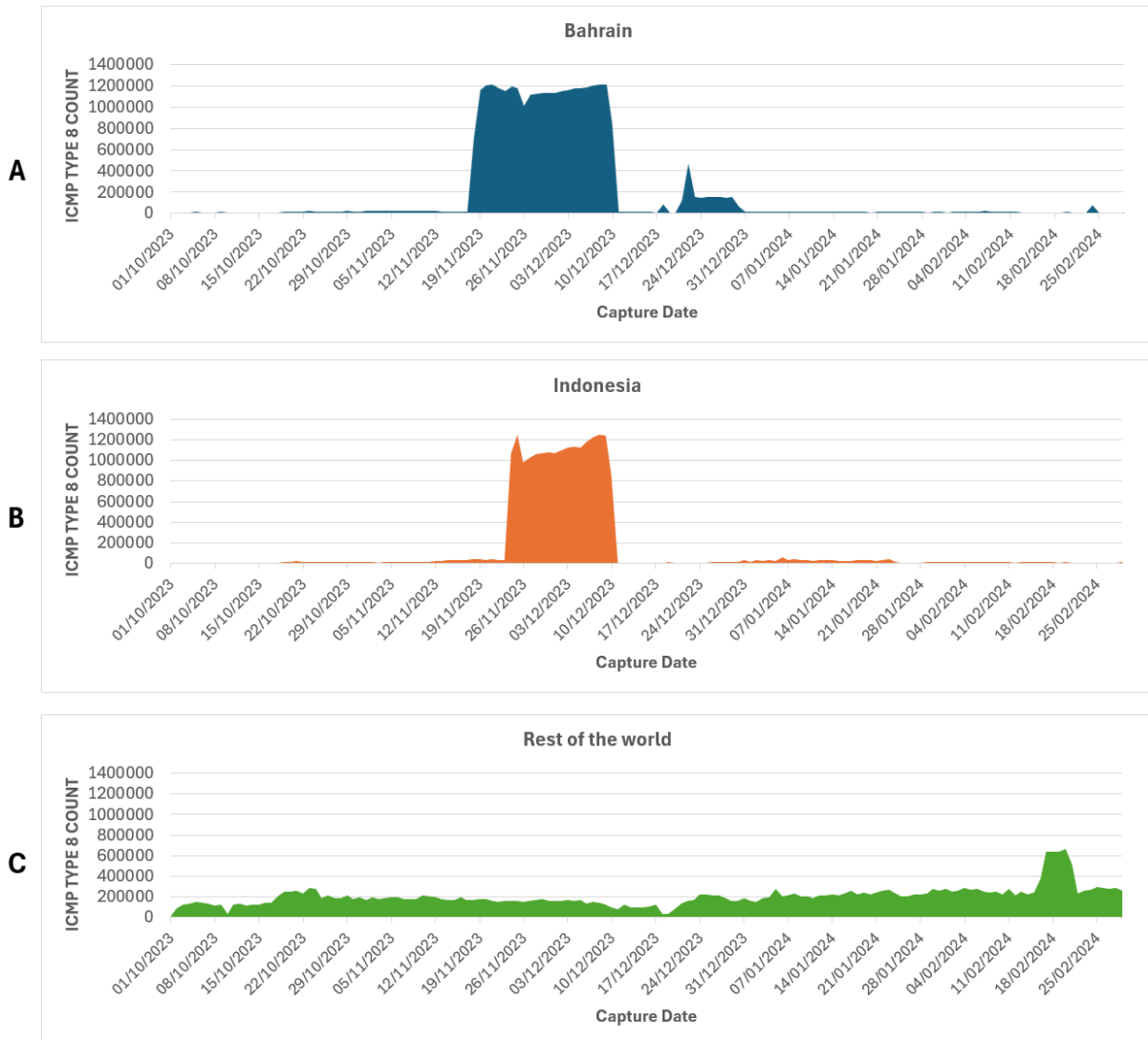
Source: The author.

What turns the Cloud Telescope more meaningful is the ability to quantify the prevalence of this type of attack in different regions of the world. That researcher can compare how differently this particular NTP vulnerability was exploited in, say, Africa (af-south-1) in contrast to South America (sa-east-1). Using `ntp.priv.reqcode==42` as a PCAP filter to detect the request for *getlist* command, one can find that there were exactly 1500 (20%) more attempts to exploit the vulnerability against the African sensor fleet in comparison to the South American one. The point being that once a researcher formulates a new question on potential drifts in radiation patterns affecting different regions of the world, the dataset can be inferred and looked up to support, confirm or refute any hypothesis of geopolitical trend or influence. But still, it is up to the researcher to formulate a meaningful question that can be materialised in the form of a PCAP filter against each file representing each region of the world in the dataset.

The presence of ICMP traffic arriving to the cloud-based sensor fleet can be explained in both benign and malicious terms. On the benign side, the fact the IP addresses used by the sensor fleet are recycled indicates that low-intensity but persistent activity can be attributed to ordinary ping-based system monitoring. In malicious terms, ICMP can be traced to denial-of-service attacks, especially when traffic volume spikes and drifts from a baseline behaviour. Once a suspicious pattern is found, such as the one depicted back on Figure 33, one can query and learn towards which country that particular activity was directed to. This was exercised and plotted on Figure 35.

The approach helps to determine that the ICMP spike was directed against two particular countries: Bahrain and Indonesia. The ICMP volume spike against Barbarian started on

Figure 35 – Delving into the ICMP spike.



Source: The author.

19 November. The sensor fleet was flooded for twenty days with a total amount of 24,422,764 type 8 packets, an average of 1,162,988 events per day and a very low standard deviation of 44,271. The Indonesian fleet recorded a same behaviour starting 24 November, five days later, and lasting until 9 December, date on which flooding ceased on both regions. Indonesia received 17,936,872 ICMP type 8 packets in the period, 1,121,054 events per day. Standard deviation of 80,004. The rest of the world saw a baseline type 8 radiation of only 204,519 per day and a small 5-day spike of 619,891 packets per day starting 17 February, closer to the end of the experiment. The intersection of time, duration and volume of ICMP type 8 activity targeting Bahrain and Indonesia is remarkable.

Section 7.7 “Experimenting geopolitical influence” (page 145) is dedicated to exercising an attempt to link these activities with geopolitical events taking place outside the Internet by cross-linking event dates, radiation destination and online newspapers to find plausible reasons behind anomalous traffic and packet spikes.

7.5 IMPLICATIONS OF THE EPHEMERAL METHOD

The ephemeral nature of the Cloud Telescope implies in the fact its deployment is bound to a constrained time-frame. Experiment #3 is reported as long-term yet it only lasted for five months. Network Telescopes last much longer. Irwin (2011) captured traffic towards a full /24 network address block for more than four years before reporting results, while Zakroum et al. (2022)'s analysis come from a 3,5 years of observation. Kori et al. (2024) listened to 1-year worth of traffic on their study, not to mention CAIDA's / UCSD's telescope, actively listening to one /9 and one /10 (originally one /8) for more than twenty years.

Roughly speaking, investing on the deployment of a standard Network Telescope is seen as a capital expenditure. There is an investment phase, in which one or more computer machines - the telescope - are bought, along with the required network assets: routers, switches, UPS systems, data-centre facilities, *et cetera*, followed by an operative phase, where these resources are put together to collect and record radiation data. More importantly, the Network Telescope must be allocated with a range of IP addresses to listen to. In the past, IPv4 allocations were merely a matter of an additional capital expense with RIPE, ARIN or another local Internet Authority. Primary IPv4 allocations are long gone and the address space is scarce. Access to never-used or lightly-used IP address blocks for the sake of research is less and less likely to happen, limiting research opportunities based on Network Telescopes to a very small group of privileged investigators.

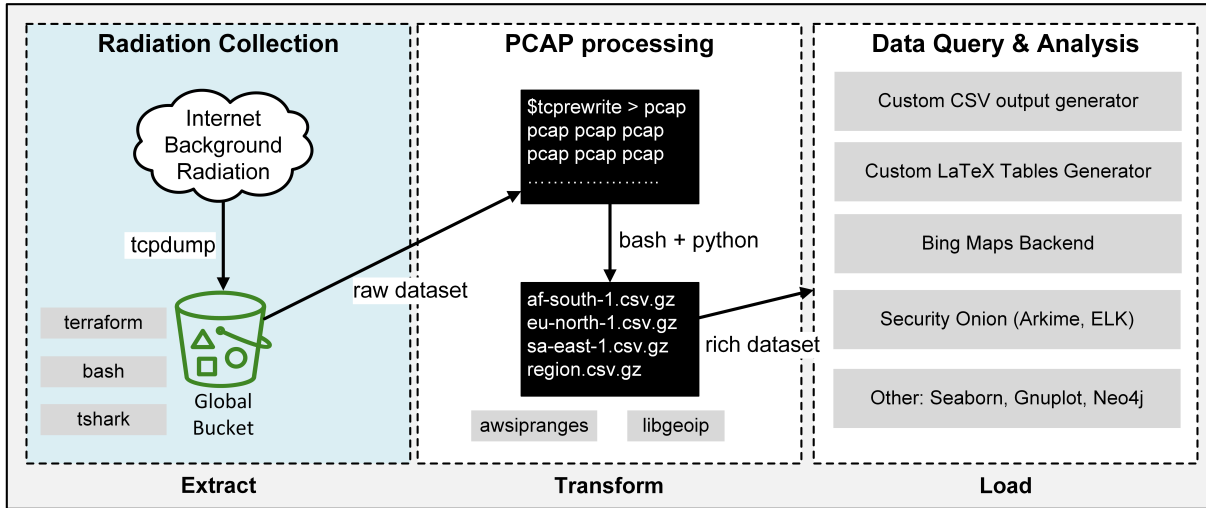
The ephemeral nature of the Cloud Telescope, on the other hand, constitutes operational expense, in which there is no upfront cost related to procurement or hardware purchase. The monthly cost can be forecasted, being directly related to the required or fleet size (telescope aperture). A budget is still needed, but the cost associated with the capture phase is diluted along the capture time frame. Once the capture ends, all computer resources associated with the experiment can be terminated.

Operational expenses can be considered both the weakest and strongest point of the Cloud Telescope. It is the weakest point in the sense that long capture periods will turn the architecture more expensive than Network Telescope counterparts. This is particularly true when considering the amount of dollars spent per allocated IP address. On the other hand, the ephemeral nature has a democratic appealing. Anyone can clone the Cloud Telescope from the GitHub repository, launch a sensor fleet, collect, analyse and eventually publish their results. Repeated launches, conducted by different individuals and research groups, covering a growing range of cloud regions, making use of different IP addresses from the address pool, should contribute to more diverse Internet Background Radiation methods, techniques, practises and findings. Also, when considering the hypothesis that mass scanners operated by malicious actors could be avoiding known address blocks allocated to Network Telescopes, as approached by Camargo, Bertholdo & Granville (2024), the ephemeral variant stands as an advantage as there is no way to predict which exact IP addresses will be allocated to the Cloud Telescope within the known IP blocks allocated to the Cloud Service Provider in question.

7.6 IMPLICATIONS OF THE CLOUD-NATIVE DESIGN

The portion of the Cloud Telescope effectively implemented as a cloud-native artefact, which in itself is part of a larger Extract-Transform-Load workflow, is highlighted in a blue shade on the diagram presented on Figure 36.

Figure 36 – The Cloud Telescope work flow.

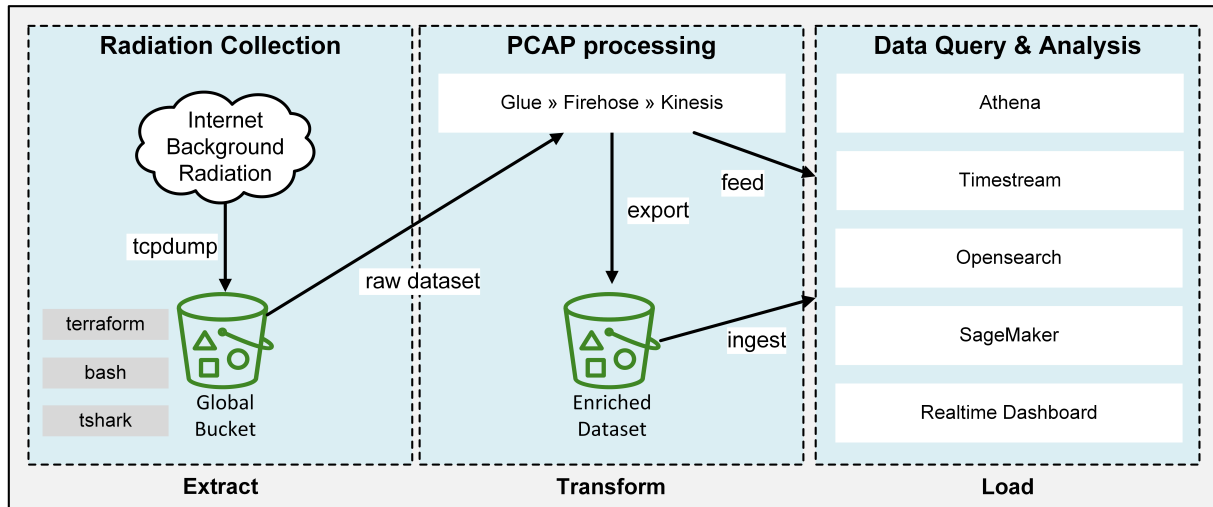


Source: The author.

Radiation, as captured by the Terraform Infrastructure-as-Code artefact from Beiler & Bortoluzzi (2024), is stored on a temporary bucket. The collection of all pcap files comprising one Cloud Telescope deployment configures the raw dataset. The raw dataset is then downloaded to an on-premises workstation and processed with the transformation pipeline available at Bortoluzzi (2024b). Basic enrichment is accomplished with two sources: *awsipranges*, a python library that restores which AWS cloud region a certain destination IP address belongs to, and *geoip*, a combination of python library and local database that will reveal the latitude, longitude, country, state and city name of any given IP address number, as maintained by Maxmind (2023). From there, the enriched dataset can be loaded onto two custom pipelines developed as needed for this research: a comma-separated-values output generator and a $\text{\LaTeX}$ tables exporter; and sent to a range of visualization back-ends, of which two were tested and used to support the results/publication phase: a Bing heat maps generator and the packet indexing, open-source solution, Security Onion, which embeds the Zeek Intrusion Detection System, Arkime indexing, InfluxDB time-series and the Elastic Stack engine altogether. Other tools are under speculative analysis, including python's Seaborn and Neo4J.

The pipeline is effective at producing the analyses on this thesis, and was used to create all number-oriented and chart-oriented results discussed on previous chapters, with exception of botnet-related studies as achieved during the supervision process in Beiler (2023). However, consider the draft how a fully cloud-native architecture should look like as depicted on Figure 37 while using AWS terminology to specify the required key components for operationalisation.

Figure 37 – Towards a fully cloud-native analysis workflow



Source: The author.

The extraction phase would behave the same. Changes start during transformation. Instead of downloading the dataset to an on-premises facility, the raw dataset would be ingested into AWS Glue, which is used as a first step data preparation for any analytics or AI project on that cloud provider. The output is a data lake that can be streamed into Firehose to perform real-time analytics. From there, results can be either exported to an enriched data bucket or the output can become the input for AWS Athena, which enables for an SQL-like interface to query through log data, or for TimeStream, a time-series cloud service, or indexed on the cloud version of Opensearch, implementing the equivalent to the on-premises ELK stack. Data could be fed into SageMaker to enable for a range of machine learning algorithms exploration. This stack could also be configured to culminate in a soft real-time dashboard piping radiation data immediately after capture.

7.7 EXPERIMENTING GEOPOLITICAL INFLUENCE

This section exercises one of many possibilities resulting from the geographically distributed radiation capture. It was taken into consideration the fact the United States of America contained four of the twenty-six regions alone to propose the following research question: “How does radiation targeting the USA differ from radiation destined to the rest of the world?”. This resulted in an interesting geopolitical analysis contrasting the two regional sub-groups: USA vs *rest of the world*, submitted to *ACM Internet Measurement Conference 2025* as *A comparison of Internet Background Radiation targeting the USA vs the rest of the world as captured by the Cloud Telescope*. The method, result and conclusion sections are reproduced below.

The method employed on this unfolding experiment operated as depicted on the workflow in Figure 38.

Dataset #3 was downloaded and divided into traffic targeting to the USA - four files representing us-east-1, us-east-2, us-west-1 and us-west-2 were given one processing environ-

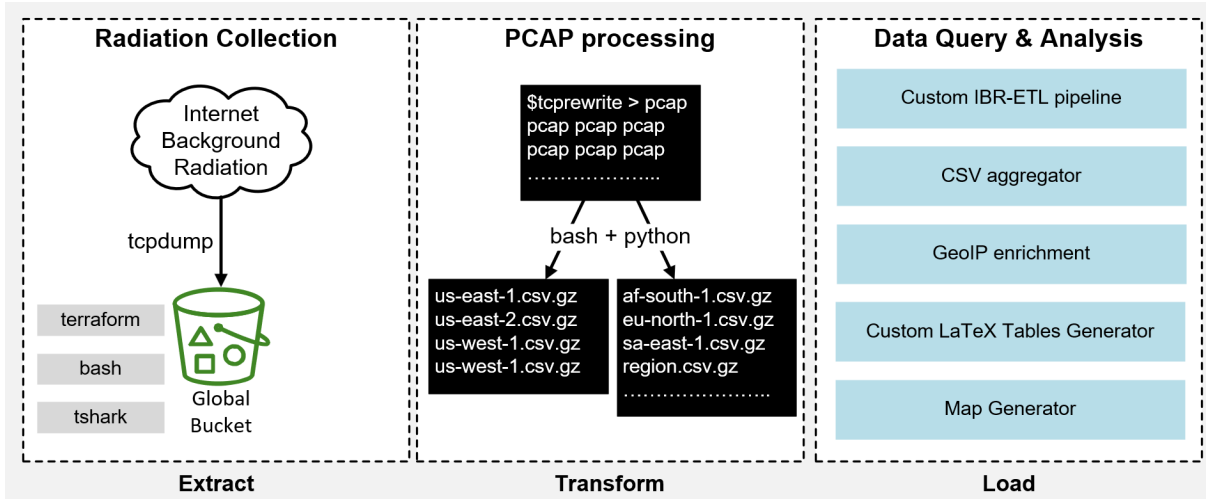


Figure 38 – Data Extract-Transform-Load workflow.

ment, while the remainder 22 recordings went through a second processing environment. This resulted into two sets of CSV files, one reflecting radiation as collected in the USA and another with the remainder cloud regions. Bash, Python and accessory libraries, such as GeoIP were used to enrich these recordings, after which tables and maps were generated, compared and compiled in the structure presented ahead.

Internet Background Radiation arriving to the twenty sensors located in the United States of America (USA) had the protocol split summarised in Table 46.

Table 46 – USA Network/Transport Layer Distribution

Protocol	Count	%
TCP	72,865,163	86.41
UDP	6,490,217	7.70
ICMP	4,971,772	5.90
Unknown	398	0.00
Sub-total	84,327,550	100.00
Total	84,327,550	100.00

A total amount of eighty-four million network events were recorded by the sensor fleet. TCP was the most prevalent, with 86% of the packets. UDP was second, with 7%. ICMP represented almost 6% of the packets.

The 110 remaining capture sensors located in countries other than the United States of America - termed “rest of the world” (rotw) onward, captured the event distribution accounted on Table 47.

TCP represented 79% of the captured packets. ICMP was second, with 18%. UDP was seen in less than 3% of the cases.

On one hand, TCP radiation seems to follow a same quantitative pattern. On the other, there is a remarkable discrepancy of ICMP vs UDP events. The experiment saw a fifth of all traffic directed to the USA as ICMP, while the rest of the world had UDP as the second most

Table 47 – ROTW Network/Transport Layer Distribution

Protocol	Count	%
TCP	352,809,472	79.00
ICMP	81,167,883	18.18
UDP	12,492,754	2.80
Unknown	117,263	0.03
Sub-total	446,587,372	100.00
Total	446,587,372	100.00

exploited protocol.

7.7.1 TCP analysis

TCP traffic targeting the sensor fleet located in the USA had the port distribution summarised on 48.

Table 48 – USA: Top 20 TCP ports

TCP Port	Count	%
80	2,887,249	3.96
10000	2,301,810	3.16
10001	2,294,395	3.15
10002	2,273,258	3.12
23	1,704,316	2.34
443	883,851	1.21
22	772,800	1.06
6379	478,711	0.66
8080	356,030	0.49
8443	316,258	0.43
8088	285,065	0.39
445	283,527	0.39
3389	261,971	0.36
2375	156,809	0.22
1433	155,426	0.21
81	150,253	0.21
8728	141,943	0.19
2222	131,884	0.18
5555	123,220	0.17
8888	104,817	0.14
Sub-total	16,063,593	22.05
Total	72,865,163	100.00

Radiation arriving to TCP ports in the rest of the world had the properties accounted on Table 49.

Table 49 – ROTW: Top 20 TCP ports

TCP Port	Count	%
23	7,832,096	2.22
80	5,149,458	1.46
22	3,748,361	1.06
443	3,640,583	1.03
10000	2,803,956	0.79
9077	2,789,762	0.79
10001	2,769,755	0.79
10002	2,658,692	0.75
6379	2,410,656	0.68
8080	2,032,296	0.58
8443	1,882,684	0.53
3389	1,469,128	0.42
445	1,259,701	0.36
8088	1,109,540	0.31
2375	800,330	0.23
8728	752,063	0.21
81	709,167	0.20
37215	688,301	0.20
8888	684,207	0.19
2222	683,297	0.19
Sub-total	45,874,033	13.00
Total	352,809,472	100.00

While the sensor fleet located in the rest of the world saw prevalent traffic directed to the Telnet port (23/TCP), the USA was more predominantly exploited at HTTP (80/TCP). Dominance of Telnet port exploitation is historical, and currently linked to the rise of the Internet-of-Things (Antonakakis et al., 2017; Kambourakis; Kolias; Stavrou, 2017; Griffioen; Doerr, 2020) as a result of poor security practises implemented on these low-cost, fast time-to-market devices. Exploitation of the HTTP port in USA targets can also be linked to remote management interfaces implemented by the IoT, in particular, domestic routers and other forms of Internet-connected home appliances.

Another particularity of the traffic targeted to American soil refers to the triplet of ports numbered 10000-10002. These together account for almost 10% of all TCP traffic aimed at the USA. These ports together only represent 2% of the traffic directed to sensors located in other countries. IANA assigned port 10000 to “Network Data Management Protocol”, however at least 22 distinct application-layer services are implemented on this port (Speedguide, 2024a), including open source tools, such as the PHP-based interface Webmin, backup-related agent Veritas Remote Execution Client, and a range of backdoor access software implemented by some malware.

Port 10001 is linked to eight different uses (Speedguide, 2024b), including the live-streaming service of a range of IP cameras, Veeam Agent, some game servers, and a couple of

backdoor-based malware. Port 10002 follows a similar usage pattern. The rest of the world was hit by a more commonly found port set, having HTTP, SSH, HTTPS right after Telnet. Port distribution is smooth in all cases, as each remaining target port was hit by a small percentage of traffic, usually of less than one percent.

Probing activity encapsulated in TCP was generated by the countries summarised on Table 50.

Table 50 – USA: Top 20 TCP Radiation Sources

Country	Count	%
United States	27,498,821	37.74
United Kingdom	7,556,426	10.37
Russia	6,148,108	8.44
The Netherlands	5,484,884	7.53
Germany	4,627,620	6.35
Bulgaria	4,559,017	6.26
South Africa	2,309,745	3.17
China	2,235,047	3.07
Singapore	1,379,137	1.89
Japan	1,265,995	1.74
Brazil	1,258,883	1.73
Australia	1,173,810	1.61
Ireland	1,137,344	1.56
Hong Kong	1,036,546	1.42
Belgium	736,951	1.01
India	680,750	0.93
South Korea	450,364	0.62
Lithuania	433,681	0.60
Romania	338,251	0.46
France	324,439	0.45
Sub-total	70,635,819	96.94
Total	72,865,163	100.00

The sensor fleet located in the United States of America recorded 37% of the TCP activity having the very USA as the perceived attacking source. This is followed by the United Kingdom (10%), Russia (8%), The Netherlands (7%), and Germany (6%), among other less prevalent sources.

Same type of attacks directed to other regions of the world had radiation emitted by countries as listed on Table 51.

The USA stood as the most frequent attacking source using TCP as transport, with 34% of all recorded events. The United Kingdom appears as second (11%), followed by Bulgaria, The Netherlands, Russia and Germany with $N > 5\%$.

Table 51 – Top 20 TCP Radiation Sources

Country	Count	%
United States	123,234,161	34.93
United Kingdom	41,138,601	11.66
Bulgaria	32,883,956	9.32
The Netherlands	30,672,856	8.69
Russia	25,354,053	7.19
Germany	23,253,678	6.59
Singapore	12,040,170	3.41
China	11,844,554	3.36
Hong Kong	5,634,692	1.60
Romania	4,627,743	1.31
South Africa	4,017,445	1.14
France	3,966,939	1.12
Ireland	3,859,841	1.09
India	3,664,376	1.04
Japan	3,041,959	0.86
South Korea	2,144,483	0.61
Australia	2,038,476	0.58
Brazil	1,862,271	0.53
Switzerland	1,678,858	0.48
Lithuania	1,565,205	0.44
Sub-total	338,524,317	95.95
Total	352,809,472	100.00

7.7.2 UDP analysis

In terms of UDP traffic, the Cloud Telescope sensors located in the USA captured the destination port distribution summarised on Table 52.

The sensor fleet located in the rest of the world captured a fairly different pattern of port prevalence, as accounted in Table 53.

The most probed UDP port in USA hosts the DNS service, 1.5% of the cases, followed by 7016/UDP, which is officially assigned by IANA to “SPG Controls Carrier”, however, actively used by Backdoor.Win32.Xingdoor to promote denial of service via a family of Chinese malware documented as "System_XingCheng" (Trendmicro, 2013). Both the USA and the rest of the world receive radiation that can be linked to the exploitation of common services, including SIP (5060), aiming at VoIP hijacking, NTP (123), commonly exploiting the amplification attacks documented on CVE-2013-5211 (Cvedetails.com, 2024b), SNMP (161), and others. The rest of the world had a high concentration of exploitation activity aiming at port 5100, almost 9%, of all UDP traffic alone. That port is popularly used by a range of services and applications, including Mac OS X Camera sharing, “ChiliSoft ASP manager for Cobalt RaQ” and, more importantly, Avaya Communication Server 1000, which is vulnerable to a denial of service when parsing requests. By sending a specially-crafted packet, a remote attacker can

Table 52 – USA: Top 20 UDP ports

UDP Port	Count	%
53	101,933	1.57
7016	81,253	1.25
5060	80,762	1.24
123	62,923	0.97
161	44,759	0.69
1900	43,491	0.67
5353	42,102	0.65
1434	39,373	0.61
500	32,540	0.50
6881	29,542	0.46
11211	28,207	0.43
8080	25,239	0.39
8081	24,672	0.38
4000	24,143	0.37
8000	23,838	0.37
1194	23,611	0.36
30301	23,313	0.36
8082	23,288	0.36
1027	23,227	0.36
8083	23,167	0.36
Sub-total	801,383	12.35
Total	6,490,217	100.00

exploit this vulnerability to cause the server to crash (Secunia, 2013). Conversely, in the USA, such port does not figure on the Top 20 at all. It is also noticeable that the UDP traffic towards the USA is more sparse as the TOP 20 ports only reveal 12% of the activity, whereas the rest of the world concentrates 40% of the UDP traffic on these high-prevalence destinations.

UDP radiation aiming at the USA was essentially generated also in the USA (85%). The UK is the source behind 2% of this traffic type, followed by Canada, China and France, contributing $\approx 1\%$ each, as accounted on Table 54.

The sensor fleet located in other countries also registered the USA as the most prevalent UDP radiation generator, corresponding for almost 40% of this traffic type. The UK was second (12%), followed by China and Sweden with 5% each.

The Top 20 UDP Radiation Sources are shown on Table 55.

7.7.3 ICMP analysis

The sensors located in the USA recorded the arrival of ICMP traffic as summarised on Table 56.

ICMP traffic targeting sensors located in the USA was 85% dominated by echo requests (type 8), followed by echo responses (type 0) at $\approx 15\%$. Type 8 prevalence was much

Table 53 – ROTW: Top 20 UDP ports

UDP Port	Count	%
5100	1,121,133	8.97
5060	470,746	3.77
53	439,827	3.52
9993	403,048	3.23
123	347,947	2.79
161	260,487	2.09
1900	243,169	1.95
5353	195,071	1.56
1434	180,865	1.45
54321	168,180	1.35
500	164,034	1.31
27015	134,365	1.08
6881	127,216	1.02
1194	125,197	1.00
11211	123,894	0.99
137	113,382	0.91
443	113,133	0.91
5351	107,076	0.86
5683	106,921	0.86
5632	102,942	0.82
Sub-total	5,048,633	40.41
Total	12,492,754	100.00

higher in ICMP radiation arriving to sensors located in other countries, as accounted on Table 57.

Echo requests accounted 95% of the ICMP traffic arriving to the sensors located in the rest of the world. Usually, TCP and UDP radiation are low-frequency, but steady. ICMP frequently spikes as a reflex of denial-of-service activity. A traffic split containing daily accumulations was accounted and resulted on the two plots presented on Figure 39.

While the USA received a low-intensity and steady ICMP type 8 radiation, the sensor fleet located in other countries captured a spike of ICMP flood activity directed against two particular countries: Bahrain and Indonesia. The ICMP volume spike against Bahrain started on 19 November. The sensor fleet was flooded for twenty days with a total amount of 24,422,764 type 8 packets, an average of 1,162,988 events per day and a very low standard deviation of 44,271. The Indonesian fleet recorded a same behaviour starting 24 November, five days later, and lasting until 9 December, date on which flooding ceased on both regions. Indonesia received 17,936,872 ICMP type 8 packets in the period. 1,121,054 events per day. Standard deviation of 80,004. The rest of the world saw a baseline type 8 radiation of only 204,519 per day and a small 5-day spike of 619,891 packets per day starting 17 February, closer to the end of the experiment. The intersection of time, duration and volume of ICMP type 8 activity targeting Bahrain and Indonesia is remarkable.

Table 54 – USA: Top 20 UDP Radiation Sources

Country	Count	%
United States	5,551,068	85.53
United Kingdom	157,680	2.43
Canada	92,698	1.43
China	88,793	1.37
France	76,033	1.17
The Netherlands	75,410	1.16
Germany	58,300	0.90
Russia	44,489	0.69
Mexico	41,512	0.64
Poland	27,480	0.42
Singapore	22,141	0.34
Brazil	19,940	0.31
Hong Kong	17,894	0.28
India	17,588	0.27
Taiwan	15,680	0.24
Thailand	14,814	0.23
Philippines	10,916	0.17
Vietnam	10,851	0.17
South Korea	10,606	0.16
Portugal	9,669	0.15
Sub-total	6,363,562	98.05
Total	6,490,217	100.00

An attempt to link these activities with geopolitical events taking place *on the surface* resulted in a range of findings, particularly from newspaper publications. (Jain, 2024) reported that hacktivist group Anonymous conducted a series of denial-of-service attacks against north-African and middle-east targets in November and December 2023, including Cairo Airport, the UAE Government Portal and, Bahrain’s E-Visa service. They revealed that the motivation for the attacks was the lack of support from these nations to Gaza, against Israel’s response to Hamas attack on 7 October 2023. According to the white paper from (GCore, 2024), Indonesia figured as the second country from which most of the denial-of-service originated in Q3-Q4 2023. These attacks aimed at the gaming industry, financial sector, gambling services, telecommunications, and infrastructure/cloud service providers, none of which is hosted in the USA.

ICMP radiation targeting USA-based cloud sensors recorded probing sources as accounted on Table 58.

The rest of the world recorded a similar pattern, as presented on Table 59.

China was the most prevalent source probing both the USA and the rest of the world ($\approx 38\%$). Indonesia was the second most prevalent source probing the USA (31%), while the rest of the world had Singapore (19%). USA-to-USA ICMP radiation accounted 14%, ranking third, whereas Indonesia figured on the same place targeting the rest of the world, with

Table 55 – ROTW: Top 20 UDP Radiation Sources

Country	Count	%
United States	4,957,944	39.69
United Kingdom	1,557,842	12.47
China	730,192	5.84
Sweden	645,512	5.17
The Netherlands	426,019	3.41
Malaysia	334,070	2.67
France	319,084	2.55
Singapore	291,001	2.33
Germany	288,591	2.31
Hong Kong	251,005	2.01
Russia	250,004	2.00
Japan	241,393	1.93
Thailand	233,473	1.87
Philippines	218,436	1.75
Indonesia	195,569	1.57
India	161,219	1.29
Poland	150,850	1.21
Brazil	121,898	0.98
Taiwan	111,069	0.89
Canada	109,111	0.87
Sub-total	11,594,282	92.81
Total	12,492,754	100.00

Table 56 – USA: Top 5 ICMP message types

ICMP Type	Count	%
8	4,251,584	85.51
0	719,666	14.48
17	202	0.00
13	160	0.00
165	66	0.00
Sub-total	4,971,678	100.00
Total	4,971,772	100.00

12%. Singapore was the fourth most active source against the USA (8%). Remainder countries contributed with $N < 1\%$ of the ICMP traffic directed to the USA. The rest of the world had the USA ranking as the fourth most prevalent source (9%), followed by smaller concentrations $N < 4\%$ coming from Asia: Hong Kong, Japan, Malaysia, Philippines and Thailand. A series of summarising geographical heat-maps containing compared background radiation dispersion is presented in Figure 40.

Sub-figure 40a shows the most prevalent TCP radiation sources directing traffic to the sensor base located in the USA, followed by the same aggregation, but for the rest of the world in Sub-figure 40d. In both cases, the USA was the most active radiation generator, followed by

Table 57 – ROTW: Top 5 ICMP message types

ICMP Type	Count	%
8	77,340,542	95.28
0	3,826,524	4.71
13	625	0.00
1	112	0.00
17	43	0.00
Sub-total	81,167,846	100.00
Total	81,167,883	100.00

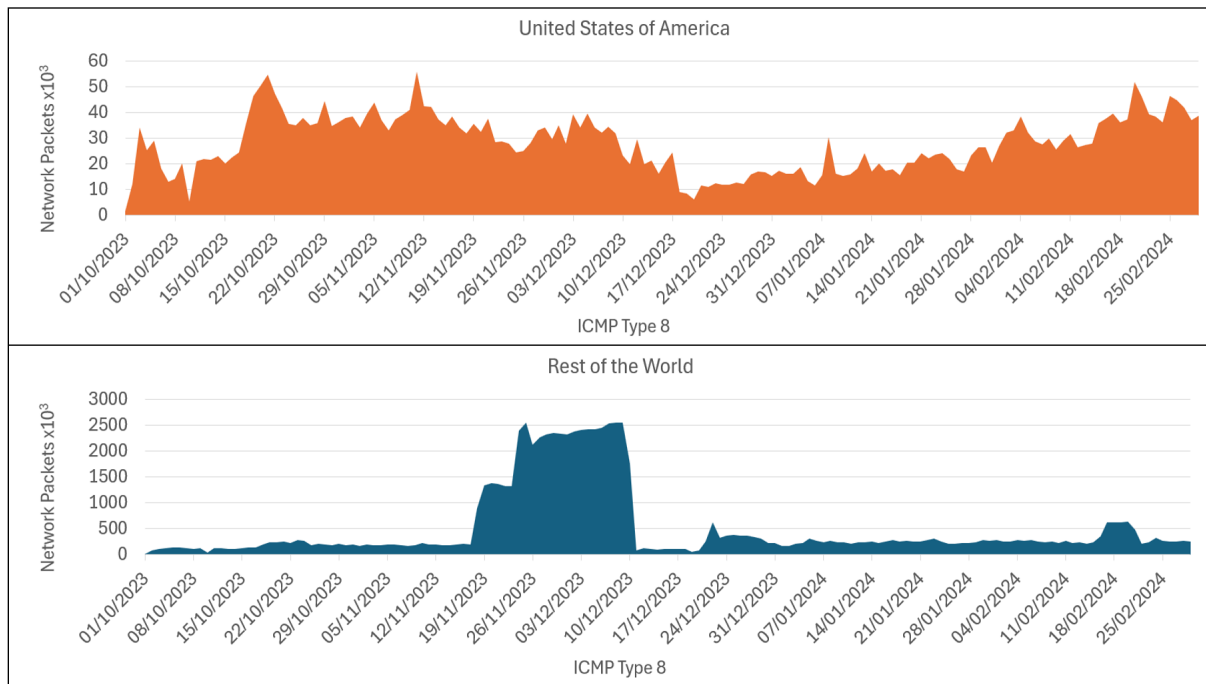


Figure 39 – ICMP type 8 traffic patterns.

the UK, and then Russia or Bulgaria. The average number of TCP events hitting USA-based IP addresses was of 3,643,258 packets per sensor, or 24,127 packets per day. The rest of the world had a slightly lower average of 3,207,358 TCP packets per sensor, or 21,240 packets per day. Sub-figure 40c shows the most prevalent UDP radiation sources directing traffic to the sensor base located in the USA, followed by the same aggregation, but for the rest of the world in 40d. Again, in both cases, the USA was the most active radiation generator, followed by the UK, and then Canada or China. The average number of UDP events hitting USA-based IP addresses was of 324,510 packets per sensor, or 2,149 packets per day. The rest of the world, having 5.5 times the *aperture* of the American fleet, only recorded the average of 113,570 packets per sensor, or 752 UDP datagrams per day. Finally, Sub-figures 40e and 40f compare ICMP traffic directed to the USA versus the rest of the world. The cloud-based IP addresses allocated to the USA recorded 4,971,772 ICMP packets. That is 248,588 packets per sensor, 1,646 per day. The rest of the world had an average ICMP packet count of 737,889 per sensor, resulting in 4,886 packets per day, three times more.

Table 58 – USA: Top 20 ICMP Radiation Sources

Country	Count	%
China	1,919,286	38.60
Indonesia	1,545,250	31.08
United States	740,511	14.89
Singapore	429,324	8.64
Hong Kong	22,930	0.46
Japan	18,237	0.37
The Netherlands	15,464	0.31
Czechia	14,345	0.29
United Kingdom	14,103	0.28
Greece	12,077	0.24
Germany	10,666	0.21
Iraq	10,468	0.21
Russia	9,480	0.19
India	8,279	0.17
France	8,049	0.16
Canada	7,194	0.14
Australia	6,359	0.13
Brazil	6,358	0.13
Italy	6,071	0.12
Spain	5,385	0.11
Sub-total	4,809,836	96.74
Total	4,971,772	100.00

The use of a distributed approach to perform Internet Background Radiation capture provides researchers with a chance to compare, contrast and characterize the different properties of malicious activity directed to one country in relation to others. This enables an opportunistic comparison of potential geopolitical factors that could justify anomalous activity spikes deviating from what can be considered standard radiation activity.

The exercise of splitting apart traffic belonging to the USA is essentially empirical. The most relevant technical motivation to choose the USA as the target for this comparison comes from the fact that the country is uniquely covered with four distinct cloud deployment regions and contained four times the sensor fleet of any other country as implemented by the reference capture experiment. Equally important, the United States of America figures as the most powerful economy in the world (Kose et al., 2017) and one of the most important geopolitical targets, cyberspace included.

The discrepancy of radiation events affecting the USA in contrast to the rest of the world is intriguing. It is relatively unsurprising that China stands as the busiest ICMP radiation emitter against the USA, as it only validates empirical assumptions on their political and economical divergence. However, it is surprising that the USA figures as the busiest TCP and UDP radiation source globally, followed by the politically aligned United Kingdom to only then having Russia as an emitter.

Table 59 – ROTW: Top 20 ICMP Radiation Sources

Country	Count	%
China	30,245,856	37.26
Singapore	15,747,481	19.40
Indonesia	10,064,160	12.40
United States	7,266,168	8.95
Hong Kong	3,042,784	3.75
Japan	2,692,993	3.32
Malaysia	1,291,315	1.59
Philippines	1,224,026	1.51
Thailand	1,182,626	1.46
United Kingdom	1,145,654	1.41
India	1,091,761	1.35
Germany	1,028,295	1.27
South Korea	1,003,924	1.24
Australia	813,639	1.00
United Arab Emirates	538,655	0.66
Saudi Arabia	284,368	0.35
Sweden	187,975	0.23
Vietnam	170,188	0.21
Taiwan	155,446	0.19
Ireland	127,643	0.16
Sub-total	79,304,957	97.70
Total	81,167,883	100.00

The ability to relate perceivable traffic anomalies with events taking place outside the cyberspace was intended and pursued as a clear goal during this study. The ICMP traffic spike discussed in Subsection 7.7.3 can be considered the most meaningful result, if not by the content it reveals, for the method it proposes, and its application to explain future geopolitical incidents.

7.8 CLOSING REMARK

As a closing remark, future deployments could focus at finding more straightforward drivers to explain many patterns found in background radiation. They should consider the geopolitical motivation influencing known recent attacks, as encouraged by ENISA's Threat Landscape for DoS Attacks report (ENISA, 2023). They report a general trend to use the Internet as a stage to primarily disturb european government administration. 46% of the attacks aim at disturbing public services made available to large audiences, such as an entire country's population. Two third of all relevant attacks make use of cyber warfare to drive, compel or affect activist agendas. Service interruption is fully achieved from the attacker's perspective in at least eight percent of the known cases. While most countries avoid the use of public clouds to host

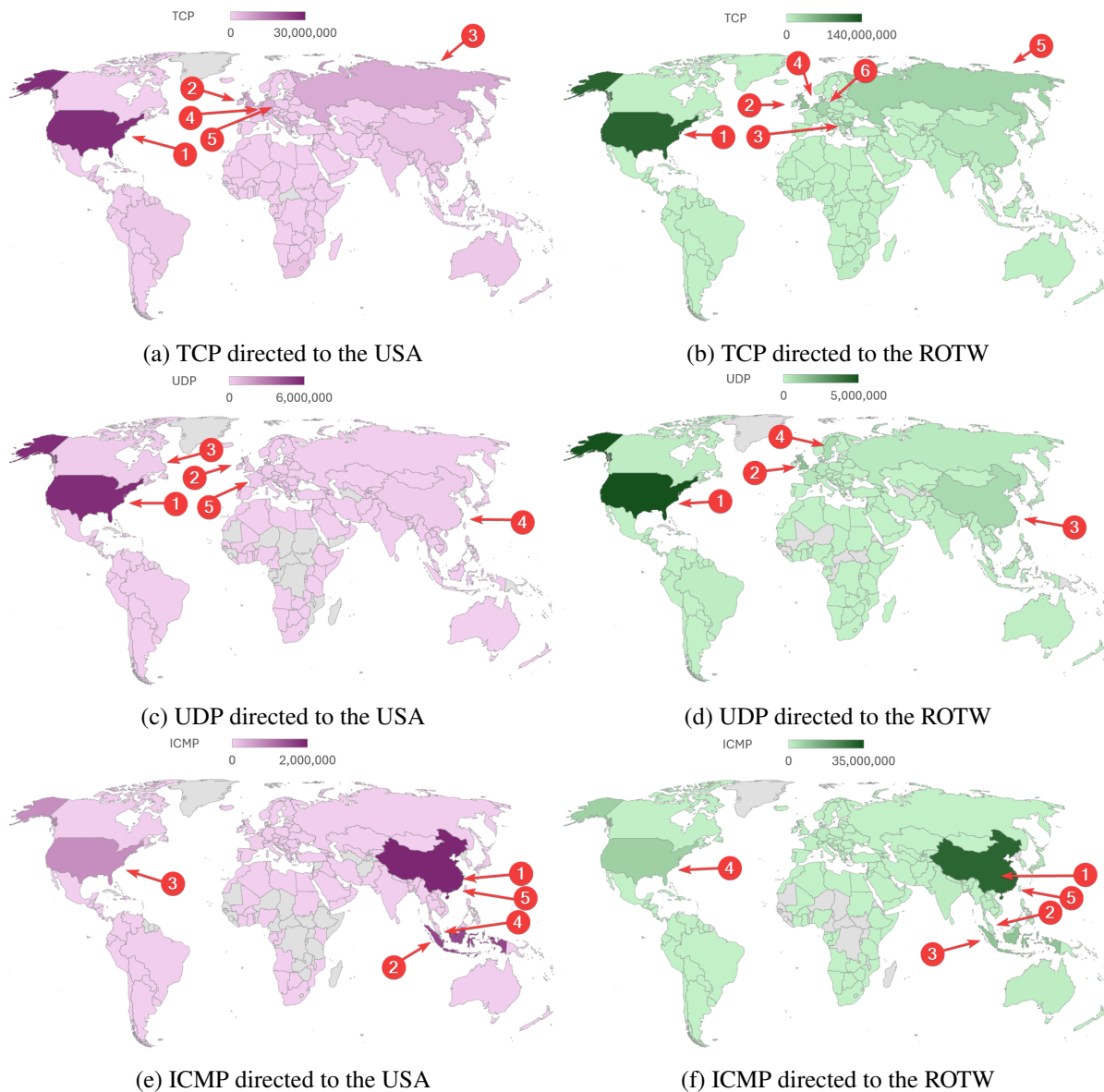


Figure 40 – Compared background radiation protocol breakdown

their systems - a debatable claim on its own right, termed *digital sovereignty*¹, IP address blocks of cloud computing service providers stand as valuable vectors for denial-of-service campaigns, botnet fleet building, and attack amplification, thus, vestiges are likely to be found on upcoming cloud-backed radiation datasets.

¹ Digital Sovereignty as intended by the European Innovation Council, refers to a nation's ability to develop, control, and maintain its own technology infrastructure, reducing reliance on foreign providers. In this context, the ultimate goal is to build a cloud infrastructure that "complies with EU data protection laws and prevents unauthorized access by non-EU entities". Examples include Gaia-X and other smaller EU-based cloud providers (European Innovation Council, 2025).

8 CONCLUSION

This research introduced the Cloud Telescope, an ephemeral, distributed and cloud-native mechanism to collect Internet Background Radiation. The hypothesis was that the collection of radiation simultaneously arriving to multiple cloud regions, located in multiple countries, would enable the establishment of a degree of geopolitical influence that could justify radiation generation, motivation and attribution. The Cloud Telescope initially arose as an empirical assumption naturally arising, in 2022, when reading seminal papers dating twenty years back. The initial assumption was that the forward snowballing process would find a range of distributed approaches as literature review approached recent works in the telescope mechanisms, and that the thesis would build on top of currently existing distributed methods, cloud computing or not, to find its point of original contribution. The result of this perception is documented on Chapters two and three. Chapter 2 ensured that each relevant concept was clearly introduced and explained, this includes the very definition of a Network Telescope, followed by an argumentation of why they are unable to perceive geographically sparse radiation traffic associated with operational concerns, in particular, IPv4 address scarcity. A straightforward introduction to Cloud Computing was also provided. This was done by describing generic expectations around the mechanisms combined with the reasons it was believed that cloud-based infrastructures could stand as means for IBR acquisition.

The literature review documented on Chapter 3 used the systematic approach adapted by Kitchenham et al. (2007) to the field of Computer Science under a Software Engineering perspective. The intended format follows a chronological reconstruction of how relevant works were searched, selected and reviewed to build a rationale that could support the need for a cloud-based distributed mechanism to collect Internet Background Radiation. The first seven related works, from Moore et al. (2004) to Chindipha, Irwin & Herbert (2019) were brought to the discussion to stress how Network Telescopes are typically deployed, how authors usually report results in a quantitative and qualitative fashion, and to evidence that radiation captured in different moments in time tend to reveal different forms of malicious activity, from malware based, to botnet centric, and to sponsored campaigns ultimately. More recent works, from Zakroum et al. (2022) to Anh & Nakamura (2024) were raised to reveal more recent trends and to help build an understanding of the current state of the art. It was during this phase that it became clear that evaluating the efficiency of the Cloud Telescope would have to include a quantitative discussion on the most targeted ports before any qualitative analysis could be drawn. This affected core aspects of preliminary attempts at building a data extraction pipeline that both reproduced previous methods while also accounting data breakdowns that only make sense when dealing with multiple geographical locations, notably, transport layer destination country and destination cloud region aggregations.

At all times, the research group was monitoring the potential rise of any equivalent method that could arise during the main three years of work. BS and MS level degree candidates under the candidate's supervision were consistently refreshing the literature review for

their own works and constantly detecting the emergence of similar works. Many made way into this manuscript, and the ones documented in subsection 3.17 were formally explained and compared. Timing was critical, particularly when contrasting the Cloud Telescope with DScope (Pauley; Barford; Mcdaniel, 2023). The decision, in early 2023, of publishing a short paper containing the method, the intended architecture and the estimated budget before implementation took place, resulted in a robust argument to ensure the required originality of the thesis. Looking at Pauley’s results it is also clear that they were busy building and deploying DScope during the exact same period as this candidate was making plans to accomplish the common goal of capturing Internet Background Radiation as it arrives to a geographically distributed fleet of cloud-based sensors. The statistical rigour of DScope is remarkable and configures its strongest aspect, while the Cloud Telescope can be acknowledged by the open Terraform implementation, enabling reuse and reproducibility of both the mechanism and the resulting datasets as its most remarkable feature set. The fact two research groups were concurrently busy designing, deploying and analysing a similar mechanism, for similar goals, reinforces that the importance of tackling this research gap, a win-win situation for both research groups, degree candidates, and ultimately the whole Internet Measurement community.

The low-cost, non-cloud, home-centric, and distributed IBR capture mechanism presented by García-Peñas, Rodríguez-Gómez & Maciá-Fernández (2024) is relevant in many ways. It initially shows that there is more to distributed IBR capture than cloud computing can do. HoDiNT elegantly tackles the IPv4 scarcity problem by using the same IP address that is assigned to each domestic router to collect unsolicited arriving traffic. While it can be somehow challenging to convince participants to enable the DMZ feature of their routers, the benefits are remarkable. The cost essentially circles around a Raspberry PI on the participant’s side and a modest server fleet on the research host’s side to enable for continuous distributed capture. There is no extra investments related to IP address acquisition, while the geographical capillarity is potentially as sparse as the Internet infrastructure can be. The mechanism, the method and the resulting datasets of HoDiNT are of great scientific interest. Their results should be compared against the cloud-based solutions, this Cloud Telescope and DScope included, in key aspects, including the most obvious ones delving into destination port breakdowns and, more importantly, the distributed geographical properties they carry, to more technical discussions, such as target avoidance and target prioritization, also speculated by related works. The methods applicable to *Darknet Zooming*, as recently documented by Kallitsis et al. (2021), including change detection, clustering techniques, and artificial intelligence algorithms supporting unsupervised learning can be of help to such effort.

8.1 ASSUMPTIONS

The six assumptions established when designing the Cloud Telescope proven mostly true (Chapter 4 page 73). The cloud-based architecture effectively enabled Internet Background Radiation collection in a similar fashion than performed by standard Network Telescopes. There

was an initial concern that the shared responsibility model enforced by Cloud Computing Service Providers could constrain or restrict the amplitude of attacks effectively arriving to the sensor, even when explicitly requesting the cloud back-end to enforce a fully open packet filter policy. This assumption came from the fact AWS does state that they take as a non-negotiable responsibility the protection of each zone's infrastructure regardless of user awareness (AWS, 2025a). The cloud architect can only optionally enable certain AWS Guard Duty features to learn what was effectively filtered out beyond any on/off control on the infrastructure level, while being able to adjust some controls related to the transport and application layers. What is fair to say is that the diversity of observed TCP, UDP and ICMP traffic, including the full range of destination ports, denotes the experiment transparently captured all probing activity destined to the sensor fleet, including some relevant ICMP repetition attacks, denial-of-service style, arriving to the middle-east sensors during the late 2023 Hamas/Israel geopolitical events. The only exception refers to attempts targeting 65535/TCP as this port was left out to stand as an emergency SSH shell into the sensors. There was no effective use of this mechanism and future captures will likely disable this feature to also record packets arriving to that port.

The geographical capillarity did enable cyber threat intelligence of different events taking place in different regions of the world, even though it was only possible to find an entry point to discuss this feature as a by-product when splitting apart radiation targeted to the USA *versus* rest of the world (Section 7.7). As it often happens in long research projects, the drilling down of data to test one hypothesis, resulted in the finding of a phenomenon that explains another open question. On one hand, there is a lucky intersection between the radiation capture time frame of the third experiment and the rise of the Hamas-Israel conflict, which ultimately allowed to modestly establish some degree of geopolitical influence motivating radiation generation. A deeper discussion of geopolitics escapes from the reasoning that makes up this computer science and cyber security experiment, however it is actually saddening to see that the third decade of this century is being marked by concerning armed and non-armed conflicts, particularly from Russia towards Ukraine, from China towards Taiwan, from the USA towards China, Greenland and more subtly towards so many other countries, and of course from Israel towards their middle-east neighbours and vice-versa. An extrapolation from the datasets indicate a high likelihood of a greater deal of cybernetic activity being exchanged between these leading nations, and a lot more to be learned with greater aperture experiments and more refined intelligence extraction and enrichment procedures.

The rationale that the Cloud Telescope stands as the most suitable method to collect radiation while coping with IPv4 scarcity is debatable. The home-centric approach seen with HoDiNT could represent a better solution if address scarcity was the most concerning problem under investigation. To make it worse, AWS decided to implement an extra charge for each public IP address effective February 2024, affecting the second half of the main capture experiment. This measure particularly affects budget allocation when many small EC2 compute fleet is needed over a smaller number of more powerful virtualised machines. T3.nano machines enabled with a public address, initially costing around five dollars per month, were added 3.6

USD - 0.005 USD per hour, 720 hours per month - immediately increasing the experiment cost by a rate of 72%. This is understood as a sign that IPv4 scarcity is very likely constraining the CSP's ability to grow, given that as a standard approach, they never charge for baseline service components, such as Virtual Private Cloud setup, Security Groups (packet filtering), virtual routers, DNS resolvers, DHCP servers, NAT gateways, and, until then, the first public IP address attached to a virtual machine.

On the other hand, taking advantage of the CSP's idle computing capacity was very effective. All three deployments were ran at minimal EC2 machinery costs due to the spot price scheme. The terraform architecture nicely captured all sudden termination events imposed by the bidding scheme, and correctly restored the fleet capacity in a time window of few seconds, also providing some sense of unintended randomness of IPv4 address shuffling on each cloud region. This is taken as an irrelevant factor from a statistical perspective. One way to estimate how intensely a capture experiment was affected by termination events is to compare the sensor fleet size against the total number of public IP addresses assigned by the CSP to the sensors upon restart. For example, Experiment #3 was launched to contain 130 persistent sensors. Therefore, it was expected that 130 public destination IP addresses would show up on the pcap recordings. That experiment actually had 3,508 distinct destination IPs, implying in 3,378 (3,508 - 130) termination events throughout the 151 capture days, an average of 22 terminations per day, or slightly less than one instance relaunch per day, per region, which is actually not much.

The fifth assumption relied on Chindipha, Irwin & Herbert (2019) statistics where increased capture lengths compensate lower aperture radiation capture experiments. The five month, 130 sensor fleet capture captured half a billion events. No other known capture experiment saw these many packets per public IP, per time unit (see discussion around Figure 29 on page 135) confirming radiation growth as Internet expands, and potentially confirming Chindipha's reasoning, yet more statistical rigour would be required to prove the argument and can configure one possible future work.

A final aspect that remains open for future validation is wherever the reproducible aspect - the Terraform description - will be effectively used by peer researchers to cross-validate the mechanism, the results, or any other desired feature it offers.

8.2 ETHICS IN CLOUD IP ADDRESS REUSE

Determining if the Cloud Telescope's recordings contain Personally Identifiable Information configures a small challenge on its own right. On one hand, one can say that once a cloud-based public IP address is assigned to a new cloud user, it is implicitly understood that any past dependency on its use has been removed, fixed or updated by the previous user, particularly in DNS configuration. On the other, often system administration or cloud automation can fail in the cleaning task, potentially resulting in Personally Identifiable Information (PII) to be unintendedly sent out to the capture fleet and end up being recorded on the generated dataset. DScope creator Eric Pauley provides detailed insights into this matter (Pauley et al.,

2022). They discovered that exploitable cloud configurations were both common and in many cases extremely dangerous; they captured over 5 million cloud messages, many containing sensitive data such as financial transactions, GPS location, and PII. They also identified dozens of exploitable software systems spanning hundreds of servers (e.g., databases, caches, mobile applications, and web services) alongside with 5446 exploitable domains. Potential causes for these vulnerabilities include lack of organizational controls, poor service hygiene, and failure to follow security best practices.

Until now, the three datasets produced by the Cloud Telescope were handled by one or more senior researchers experienced with traffic analysis, internet measurement and DNS security, as much as they were used as input to foster around twenty BS-level works, delving into basic quantitative aggregations, qualitative discussion, comparison against standard Network Telescope datasets, botnet classification, support vector machine unsupervised classification, and a range of regular-expression event classification procedures mimicking the software stack seen in classic intrusion detectors, such as Snort and Zeek, to find patterns of interest, particularly, of geopolitical nature. In all cases, no PII was found on the datasets. Due to the passive nature adopted in experiments #1 and #3, none is expected to be found on the first and third datasets, while it is yet to be determined if Dataset #2, containing 10 billion events captured in active mode, contains application-layer data other than botnet or system exploitation activity. It is assumed that “utility overcomes risk”, as proposed by Internet Measurement researcher authorities Kenneally & Claffy (2009) on this particular matter.

8.3 THE CLOUD TELESCOPE AS *CODE*

The strategy used to document the Cloud Telescope involved important decisions regarding the artefacts that could better communicate the many design decisions and a corresponding infrastructure description language that could increase the likelihood of adoption by peer researchers. There are three main aspects to foster a conclusion around these choices. First, the use of architectural diagrams, especially as documented on Figure 15, on page 74. The figure follows the rigour encouraged by AWS where each cloud region are depicted as the aggregating set, containing the Virtual Private Cloud as an intermediate set, and inside it, one or more subnets in which each regional sensor fleet is deployed. The reference to an open security group and the need to include an Internet gateway sends a concise message to individuals willing to reproduce the architecture even if they prefer to code their own solution instead of using our implementation. At a first glance, the fully open security group may look like a great security concern. This requires from the reader a clearer understanding of the whole stack, to perceive that there is no other effective way to implement a promiscuous listener free from such risk. The cloud setup exposes an equivalent attack surface than seen in standard Network Telescopes. The global bucket is depicted as a global resource, where a single file aggregation destination is preferred over a distributed architecture. Concerns related to availability and fault-tolerance can be addressed by the fact these S3 buckets are backed by eleven nines of

durability (99.99999999%), only possible by keeping three hidden copies of each file on three distinct availability zones within the selected cloud region destination.

The second aspect is the choice for using basic software engineering artefacts to explain, high-level, the functional and non-functional requirements that should be observed to materialise the Cloud Telescope as a working mechanism. Those in combination with the single business rule were shared with the 5-year computer science BS candidate Lucas Beiler, who effectively used the artefacts to describe and code the telescope as it exists today. It costed around three months of continued research and programming time to come up with a solution that works as designed. His work contains a handful of observations on the active responders implemented on the second experiment, and delves deep into botnet propagation activity (Beiler, 2023).

The third element is the resulting Terraform code describing the Cloud Telescope architecture combined with the Bash scripts used to automate the capture procedure, from tcpdump initiation, to responding to termination events and sending the pcap recording chunks to the aggregation bucket (Beiler; Bortoluzzi, 2024). They are materialised in fourteen Terraform files containing detailed instructions on region selection, virtual machine type selection, ssh keys, storage volume selection, bucket writing access role, VPC setup and IAM configuration. The `user_data` section is fed with 100+ operating system commands demonizing tcpdump and implementing a one-serves-all active responder to any TCP port selection of interest. This can be tuned to match ports exploited by emerging threats, on the spot, whenever a CVE of interest goes public and will be explored as part of a larger strategy by Lucas in postgraduate works.

8.4 RESULTS

This volume spent three chapters explaining results. Chapter 5 (page 87) was built to gradually accommodate intermediate results as each experiment ended. The intention to initially conduct a small proof-of-concept before any large deployment was quickly identified in early stages of the research, concurrently taking place as the first rounds of literature review were coming out. This resulted in three distinct intermediate works, one to report the plan, a second to assess a preliminary deployment and a third to delve into the realm of active responders and their ability to interact with botnet infection agents. This last feature was not initially planned, as there is no fundamental need to do honeypot-like interaction to extract meaningful intelligence, and also because there is no strong link between these and the geopolitical influence as the main driver. However, a combination of factors did allow for a robust dive into interactive analysis. These factors include the empirical assumption that it would be revealing as a research experiment, followed by a common agreement in the research group, including supervisors and key students under supervision. Another factor that enabled this achievement is the funding granted by Noroff University College, which arrived in time to support this exploration. The initial assumption was that the active responder would generate a reasonable overhead in capture size, file size and network traffic. However, not after two weeks from the

capture start, it was possible to notice that interactivity was generating far more traffic per sensor than the previous proof-of-concept passive capture. The second experiment would become the long-term, half-year-long experiment if it had not so quickly reached the USD 2.000,00 budget limit only 47 days after the capture started. The second experiment was therefore actually aborted on the spot. This decision was taken both from the cost perspective but also from a dataset size view, as it had already reached the count of 10 billion captured packets, including of course many packets belonging to the subsequent data exchange that only occurs if the TCP three-way handshake is completed. The more this research group moves into using unsupervised machine learning models, such as support vector machine, more potential is seen to exploring the results of interactive captures.

It was possible to document the following number of scientific and non-scientific by-products:

- 1 short conference paper publication;
- 3 long conference papers publications;
- 1 extensive journal paper publication;
- 3 open datasets;
- 1 dedicated BS project to fully document the main technical contribution: the Terraform Infrastructure-as-Code artefact;
- 1 Extract-Load-Transform data enrichment and analysis pipeline;
- 1 Computer Science MS dissertation at UFSC;
- 2 Computer Science BS works at UFSC;
- 5+ Computer Science MS dissertation at Univali;
- 15+ Cyber Security BS projects at Noroff University College;
- 2 Industry Conference presentations, one at the renowned You Shøt the Sheriff and another at BSides Las Vegas 2024.
- A handful of traffic analysis workshops and colloquia in Brazil and Norway.

These achievements contribute to thinking about the Cloud Telescope as a mechanism that is useful beyond the sole expectation of standing as a justification to the awarding of one terminal degree in Computer Science, broadening the perspective of having it as a democratic resource that can be used by others willing to capture new samples of the radiation phenomena.

The detailed discussion of the main experiment reported in Chapter 6 (page 111) aimed at showcasing the potential of the Cloud Telescope. The results presented in Radiation Sources,

Section 6.3, reproduced the same method as seen in most related works, while Radiation Targets, Section 6.4 only makes sense to the geographically distributed property of the Cloud Telescope, and took considerably more time and effort to design, develop and explain. The protocol-centric and port-centric analyses focusing on the different countries and cloud regions was complemented with world-map heat-maps visually communicating the aggregated findings and enabling the establishment of speculative patterns that could deserve extra rounds of dataset exploration. Finally, the distributed analysis in Section 6.5 focused on the destination cloud regions and countries to build the foundations to geopolitical influence exploration, yet the more these elements properties were explored, clearer it became that more work is needed towards revealing other hidden geopolitical patterns.

Finally, the discussion chapter (Chapter 7) resulted in an *ad-hoc* exploration of selected insights obtained after considerable reflection on the most relevant properties of the proposed Cloud Telescope. Initially, two concerning questions were addressed: Can the Cloud Telescope confirm the ever growing nature of radiation generation? Equally important: Is it possible to extract at least one or more geopolitically motivated attacks from the combination of method + dataset? Both were confirmed. The former via basic quantitative accounting and the latter via dataset split and analysis. Once again a protocol drill-down was used to produce insights on TCP and UDP patterns of interest, and ICMP supported the argument that geopolitical factors can justify some of the captured events affecting selected regions of the world, raising the possibility of a future work that could provide automated means for pattern detection combining media source news, their metadata including date, textual content, intended audience and potentially linked network packets serving as possible motivation.

The Cloud Telescope is documented as implementing three key properties: distributed, ephemeral, and cloud-native. These properties were discussed as *implications*. The distributed nature was briefly summarised from three perspectives: geopolitical influence, botnet mapping capability and a showcase demonstrating how NTP amplification attack can be analysed from unique angles only possible by a distributed capture. The ephemerality argument attempted to compare the Cloud variant against standard Network Telescopes in terms of how rigid and exclusive they are. The Cloud Telescope is therefore democratic, in the sense that anyone can deploy it. Whenever budget stands as a constraint, then the open datasets can be used to test theories. Comparing existing Cloud Telescope datasets against new HoDiNT deployments can also configure interesting future works for individuals holding modest investment capacity. The third property, being cloud-native, means that the ability to reproduce the telescope must be reasonably well documented. This resulted in additional content to exercise the idea of splitting apart traffic going to one region of interest, the USA, in comparison to the rest of the world. The reader is guided through the process of how to perform such operation hoping that upcoming researchers can extrapolate from there towards the splits of their own interest. BS and MS level students do exercise dataset splitting where they choose regions of their interest and work on baseline quantitative analysis revealing unexpected port rankings yet to be compiled and published. This is another fertile ground to test the versatile properties of the Cloud Telescope.

8.5 FUTURE WORK

This last section maps the upcoming research that can unfold from the Cloud Telescope as a baseline work for future distributed Internet Background Radiation collection and analysis. The envisaged future works can be aggregated into three research branches:

- Capture expansion;
- Dataset enrichment;
- Packet indexing and analytics;

8.5.1 Capture expansion

As a reproducible mechanism, the Cloud Telescope can be deployed in different ways, only constrained by the monthly expenditure to cover Cloud Computing service costs.

Work 1 refers to the situation where peer research groups could afford to spend the estimated amount of 250 US Dollars to continuously maintain a fleet of 130 passive sensors collecting radiation to the aggregation bucket, and from there, confirm or refute the findings achieved this far, as well as detecting pattern change that could stand as an entry point for ad-hoc research on the motivation. One incipient work on change detection applicable to standard Network Telescopes is under way by Smedberg & Irwin (2024) where they make use of descriptive statistical analysis to report “an increase in TCP traffic, a decrease in UDP traffic, and a substantial increase in ICMP traffic, primarily from Russia, while observing a notable decrease in the Russian overall attributed traffic. A sharp decrease in specific destination port targeting for both TCP and UDP traffic suggests broader scanning activity than before”. The same rationale can be applied to the Cloud Telescope datasets, especially once more captures are launched and more temporal data is gathered to support meaningful analyses.

Capture expansion can also consider the rapid growth of the preferred Cloud Service Provider, AWS (Work 2). The three Cloud Telescope deployments were configured to cover all commercially available AWS regions, only excluding the *Government cloud regions* available in the United States of America. The region count in 2023-2024 was of 26 (twenty-six) regions. As of May 2025, a second public region was formed in Canada, Canada West (Calgary), one in Mexico (Mexico Central), one in Israel (Tel-Aviv), two in Asia Pacific: Malaysia and Thailand, and a second in Australia (Melbourne). All these new regions contribute to refined granularity and better geographical radiation capture capillarity, however, the existence of a new cloud region in Israel is of special interest, as there is historical and recent geopolitical motivation to look closer into the cybernetic activity destined towards this country and vice-versa. An updated Cloud Telescope deployment in 2025 would be able to listen to 32 regions, a capillarity increase of 23%, which would raise the baseline monthly cost to ≈ 310 US Dollars per month.

An alternative approach to capture expansion could be the zooming into regions of particular interest (Work 3). Assuming a researcher could be only interested in detecting nation-state cybernetic campaigns in the middle-east, a larger fleet of, say, 20 (twenty) sensors, could be deployed on Bahrain, the UAE and Tel-Aviv. Post-processing analytics could then make use of the GeoIP library to only compute traffic sources tracing back to that same region, and then detailed analyses could lead to the discovery of geopolitical patterns currently hidden below the more frenetic, TCP-based, IoT-focused, port 23 exploitation. The Terraform description can both automatically fetch the list of publicly available regions for upcoming deployments or can be manually configured to only collect radiation from selected regions of interest.

Perhaps the most complex contribution to the field would be the natural evolution from an AWS-centric Cloud Telescope towards a multi-cloud architecture that could collect radiation arriving to the sum of countries and regions covered by the three leading CSPs: AWS, Azure and Google Cloud Platform. Work 4 is a DR-level proposal that would require the candidate to rethink the software engineering artefacts, the terraform implementation, the radiation collection scheme, the data analysis pipeline and the enrichment procedures to extract meaningful intelligence from the IBR as it affects this ultra-granular, super sparse capture sensor fleet and potentially reveal unseen patterns until today.

Table 60 summarises this branch of future research.

Table 60 – Future works on capture expansion

#	Future work	Motivation	Comment	Degree Level
1	New CT deployments	Validate, compare and detect IBR changes	Baseline budget of 250 USD per month	BS, MS
2	Coverage of new AWS regions	Enhance geopolitical granularity.	New regions allow for better capillarity and more refined findings.	BS, MS
3	Zoom in on regions of interest	Go deeper into particular regions of interest.	I.e. Israel can now be observed and should reveal meaningful geopolitical cybernetic activity	MS, DR
4	Go multicloud	Pursue the ideal radiation collection granularity	Deploy an universal Terraform IaC that works cross-cloud in AWS, Azure and the Google Cloud Platform.	DR

The degree level column indicate the perceived complexity of each unfolding research effort: BS for Bachelor's monographs, MS for Master's dissertations and DR for doctoral-level, PhD equivalent theses.

8.5.2 Dataset enrichment

There is significantly more that can be extracted from the three Cloud Telescope datasets than the methods used to perform the basic accounting and aggregations used to document this thesis.

The approach to enrichment techniques currently used with the Cloud Telescope datasets were explained on and around Figure 36, on page 144. A tcpdump extraction pattern is used to produce CSV entries that are enriched with the *awsipranges* and *libgeoip* python libraries. These libraries restore the name of each AWS region while feeding the output data with latitude

and longitude coordinates and country name. While this is enough to answer some research questions, more creative ways exist to enrich these events, particularly in terms of profiling the radiation emitter.

BS-level future works can look into ways to provide one single entry point contribution to the dataset enrichment procedure as it exists on GitHub (Bortoluzzi, 2024b). Each contribution configures one BS-level contribution to the ETL as summarised on Table 61.

Table 61 – Future works on dataset enrichment

#	Future work	Motivation	Comment	Degree Level
5	Enrichment with Shodan	Intelligence acquisition on radiation sources using the Shodan API.	Rich mechanism for source profiling. Limited to 5000 sources per day.	BS, MS
6	Enrichment with Censys	Intelligence acquisition on radiation sources using the Censys API.	Rich mechanism for source profiling. API limitation must be investigated.	BS, MS
7	Enrichment with other sources	Intelligence acquisition on radiation sources using other emerging CTI acquisition platforms.	This refers to mapping the rise of new enrichment backends and test them to enrich the CT datasets.	BS, MS, DR
8	Building a mesh-up interface	Build a customised web mesh-up interface that can serve as a single point of unified cyber threat intelligence acquisition.	Creative web design work that is limited by the many constraints imposed by intelligence providers.	BS, MS

Works 5 and 6 refer to reprogramming the ETL to implement remote calls to Shodan and Censys and enrich the intelligence acquisition mechanism to reveal properties of radiation emitters beyond geolocation. For each given source IP address, Shodan and Censys can provide the following metadata: DNS hostname by reverse lookup, IP block organization name, the Autonomous System Number (ASN), the hardware architecture (i.e. ARM, PC), the operating system stack, the application-layer stack and the list of currently open ports as they continuously scan the whole Internet. This is a high-priority work in the research group and may require one or more dedicated MS-level candidates to perform the work in full extent. Work 7 proposes that any other cyber threat intelligence acquisition source that arises in the future can be considered as a 1-year research effort to expand the Cloud Telescope ETL.

Work 8 refers to the possibility of building a *mesh-up* web platform that can be used to unify all cyber threat intelligence acquisition sources in a single place and allow for real-time enrichment and exploration of small radiation chunks. This can be used for *ad-hoc* radiation exploration as much as for teaching and learning purposes. This is expected to be severely constrained by the web server's processing capabilities and, more importantly by the limits imposed by free or student accounts within intelligence providers via API rate-limited calls. There is currently one student kickstarting this branch within University of Vale do Itajai, and more students are expected to join soon.

8.5.3 Packet indexing and analytics

The third branch of research activities refers to the possibilities around performing packet indexing and benefitting from the performance gains achieved with the method. These are summarised on Table 62.

Table 62 – Future works on packet indexing and analytics

#	Future work	Motivation	Comment	Degree Level
9	Indexing with Arkime	Arkime indexes packets enabling query response in acceptable time.	Arkime is the de-facto standard backend for packet indexing.	BS, MS
10	Using the ELK stack	Packet indexing and exploration using the Elastic Stack.	The ELK stack is backed by the opensearch engine and also contributes to acceptable query response times to freely explore network events.	BS, MS
11	Exploring radiation with Security Onion	Explore Internet Background Radiation making use of open source, SIEM, market-oriented cyber security solutions.	Security Onion includes a packet indexing back-end, an open search back-end, an intrusion detector, and a time-series database to enable visual data exploration.	BS, MS
12	Full cloud-native analytics pipeline	Build a cloud-native radiation analysis platform that continuously capture the IBR and produces analytical results real-time.	This would have to be sponsored with enough cloud credits from one or more CSPs but would provide the ultimate resource for real-time radiation analysis.	DR

Work 9 proposes the use of Arkime, a specialised tool meant to index PCAP recordings and enable almost real-time network event exploration. Currently, without an indexing backed such as Arkime, a researcher spends at least 40 minutes to go through the entire half-billion packets in Dataset 3 if using an ordinary, modern i7 computer with enough RAM and fast SSD storage. It can take more than a full week to apply a regular expression against the entire 10-billion entry second dataset. Indexing therefore, even though time and storage consuming at-first, can ease the process of gradually querying the dataset with more complex questions until a certain desired pattern is found, preferably around supporting or refuting the existence of geopolitical motivations.

Work 10 proposes the use of the ELK stack, comprised by Elastic Search, an open format data indexing back-end, Logstash, a log collector, and Kibana, a visually attractive data exploration interface, to serve as a counterpart to Arkime, in providing opensearch-backed indexing allowing researchers to quickly explore radiation events and come up with new insights. A similar suggestion is Security Onion, proposed as Work 11, providing an unified interface where PCAPs can be ingested once and automatically processed against ELK's Open Search, the Zeek intrusion detection and the Influx DB time-series database back-ends. From there, it is possible to explore network events using Security Onion's own interface as well as Kibana's and Influx's. Dry run experiments performed with this stack revealed promising capabilities, and can be adopted as undergraduate and post-graduate exercise activities to engage more students in the field of distributed Internet Measurement.

The final goal envisaged for the Cloud Telescope is proposed in Work 12: the building of a full cloud-native analytics pipeline that continuously capture the IBR and produces analytical results real-time. The concept, documented around Figure 37 on page 145, requires investment, and would configure the ultimate resource for real-time radiation tracking and analysis, to be done as DR degree level research contribution once sponsorship is agreed with one or more Cloud Computing Service Providers.

BIBLIOGRAPHY

AFFINITO, A. et al. The evolution of mirai botnet scans over a six-year period. **Journal of Information Security and Applications**, v. 79, p. 103629, 2023. ISSN 2214-2126. Available from: <https://www.sciencedirect.com/science/article/pii/S2214212623002132>.

ANAND, A. et al. Aggressive internet-wide scanners: Network impact and longitudinal characterization. In: **Companion of the 19th International Conference on Emerging Networking EXperiments and Technologies**. New York, NY, USA: Association for Computing Machinery, 2023. (CoNEXT 2023), p. 1–8. ISBN 9798400704079. Available from: <https://doi.org/10.1145/3624354.3630583>.

ANH, S. P.; NAKAMURA, Y. A baseline investigation into the evolution and prevalence of mirai and hajime utilizing a network telescope. **IEEE Access**, v. 12, p. 103789–103809, 2024.

ANTONAKAKIS, M. et al. Understanding the mirai botnet. In: **Proceedings of the 26th USENIX Conference on Security Symposium**. USA: USENIX Association, 2017. (SEC'17), p. 1093–1110. ISBN 9781931971409.

AWS. **AWS Shared Responsibility Model**. 2025. Available from: <https://aws.amazon.com/compliance/shared-responsibility-model/>.

AWS. **DDoS protection**. 2025. Available from: <https://aws.amazon.com/developer/application-security-performance/articles/ddos-protection/>.

BEILER, L. S. **Análise da radiação de fundo da internet por captura distribuída em telescópio de nuvem**. 71 f. Monografia (Trabalho de Conclusão de Curso) — Ciência da Computação, Universidade do Vale do Itajaí, Itajai, Santa Catarina, Brazil, 2023.

BEILER, L. S.; BORTOLUZZI, F. **Cloud Telescope Terraform Architecture**. Github, 2024. Available from: <https://github.com/lucasbeiler/ibr-iac>.

BENSON, K. et al. Leveraging internet background radiation for opportunistic network analysis. In: **Proceedings of the 2015 Internet Measurement Conference**. New York, NY, USA: Association for Computing Machinery, 2015. (IMC '15), p. 423–436. ISBN 9781450338486. Available from: <https://doi.org/10.1145/2815675.2815702>.

BORTOLUZZI, F. **BSides Las Vegas 2024 Breaking Ground Fabricio Bortoluzzi Insights on Using a Cloud Telescope**. 2024. Available from: <https://www.youtube.com/watch?v=le2-yszqHL3E>.

BORTOLUZZI, F. **Cloud Telescope Extract-Transform-Load data pipeline**. Github, 2024. Available from: <https://github.com/bortoluzzi/ibr-etl>.

BORTOLUZZI, F. et al. **Cloud Telescope Internet Background Radiation August 2023**. IEEE Dataport, 2024. Available from: <https://dx.doi.org/10.21227/hs3j-pg74>.

BORTOLUZZI, F. et al. **Cloud Telescope Internet Background Radiation May 2023**. IEEE Dataport, 2024. Available from: <https://dx.doi.org/10.21227/sg9h-ta16>.

BORTOLUZZI, F. et al. **Cloud Telescope Internet Background Radiation October 2023 February 2024**. IEEE Dataport, 2024. Available from: <https://dx.doi.org/10.21227/zkyy-gk56>.

BORTOLUZZI, F. et al. Cloud telescope: A distributed architecture for capturing internet background radiation. In: **2023 IEEE 12th International Conference on Cloud Networking (CloudNet)**. [S.l.: s.n.], 2023. p. 77–85.

BORTOLUZZI, F. et al. Quantitative analysis of botnet activity within internet background radiation as observed by a cloud telescope. In: **2024 19th Iberian Conference on Information Systems and Technologies (CISTI)**. [S.l.: s.n.], 2024.

BORTOLUZZI, F.; IRWIN, B.; WESTPHALL, C. M. A cloud-native framework for globally distributed capture and analysis of internet background radiation. In: **2023 18th Iberian Conference on Information Systems and Technologies (CISTI)**. [S.l.: s.n.], 2023. p. 1–4.

BORTOLUZZI, F.; IRWIN, B.; WESTPHALL, C. M. Cloud telescope: An ephemeral, distributed, and cloud-native architecture for collecting internet background radiation. **IEEE Access**, v. 13, p. 45682–45714, 2025.

CAMARGO, A. V. C.; BERTHOLDO, L. M.; GRANVILLE, L. Z. Less is more? exploring the impact of scaled-down network telescopes on security and research. In: **Anais do XLII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC 2024)**. Sociedade Brasileira de Computação, 2024. (SBRC 2024), p. 1050–1063. Available from: <http://dx.doi.org/10.5753/sbrc.2024.1538>.

CAMPBELL, P. The denial-of-service dance. **IEEE Security and Privacy**, v. 3, n. 6, p. 34–40, 2005.

CENSYS. **Censys: Exposure management and threat hunting solutions**. 2024. Available from: <https://www.censys.io/>.

CHARAN, P. S. et al. Dkaas: Dark-kernel as a service for active cyber threat intelligence. **Computers & Security**, v. 132, p. 103329, 2023. ISSN 0167-4048. Available from: <https://www.sciencedirect.com/science/article/pii/S0167404823002390>.

CHAUDHARI, R. S.; TALMALE, G. R. A review on detection approaches for distributed denial of service attacks. In: **2019 International Conference on Intelligent Sustainable Systems (ICISS)**. [S.l.: s.n.], 2019. p. 323–327.

CHINDIPHA, S. D.; IRWIN, B.; HERBERT, A. Quantifying the accuracy of small subnet-equivalent sampling of ipv4 internet background radiation datasets. In: **Proceedings of the South African Institute of Computer Scientists and Information Technologists 2019**. New York, NY, USA: Association for Computing Machinery, 2019. (SAICSIT '19). ISBN 9781450372657. Available from: <https://doi.org/10.1145/3351108.3351129>.

CLOUDFLARE. **Cloudflare: What was the largest DDoS attack of all time?** 2024. Available from: <https://www.cloudflare.com/learning/ddos/famous-ddos-attacks/>.

CVEDETAILS.COM. **Bind Security Vulnerabilities and CVEs**. 2024. Available from: <https://bit.ly/49QVjkz>.

CVEDETAILS.COM. **DNS Security Vulnerabilities and CVEs** <https://www.cvedetails.com/vendor/2153/>. 2024. Available from: <https://www.cvedetails.com/vendor/2153/>.

CZYZ, J. et al. Taming the 800 pound gorilla: The rise and decline of ntp ddos attacks. In: **Proceedings of the 2014 Conference on Internet Measurement Conference**. New York, NY, USA: Association for Computing Machinery, 2014. (IMC '14), p. 435–448. ISBN 9781450332132. Available from: <https://doi.org/10.1145/2663716.2663717>.

DAINOTTI, A. et al. Estimating internet address space usage through passive measurements. **SIGCOMM Comput. Commun. Rev.**, Association for Computing Machinery, New York, NY, USA, v. 44, n. 1, p. 42–49, 2014. ISSN 0146-4833. Available from: <https://doi.org/10.1145/2567561.2567568>.

DAINOTTI, A. et al. Lost in space: Improving inference of ipv4 address space utilization. **IEEE Journal on Selected Areas in Communications**, v. 34, n. 6, p. 1862–1876, 2016.

ENISA. **ENISA threat landscape for DoS attack: January 2022 to August 2023**. LU: ENISA Publications Office, 2023. Available from: <https://data.europa.eu/doi/10.2824/859909>.

European Innovation Council. **Statement on Technological Sovereignty**. 2025. Available from: https://eic.ec.europa.eu/document/download/61d52ef5-5b28-4c00-bfb8-a67e9c22666f_en.

GARCÍA-PEÑAS, R.; RODRÍGUEZ-GÓMEZ, R. A.; MACÍÁ-FERNÁNDEZ, G. HoDiNT: Distributed architecture for collection and analysis of internet background radiation. **Computer Networks**, p. 110570, 2024. ISSN 1389-1286. Available from: <https://doi.org/10.1016/j.comnet.2024.110570>.

GCore. **Gcore Radar Report Reveals DDoS Peak Attack Volumes Doubled in H2 2023**. 2024. Available from: https://assets.gcore.pro/site-media/uploads/gcore_radar_ddos_attack_trends_q3_q4_2023_aec6a31b19.pdf.

GELGI, M. et al. Systematic literature review of IoT Botnet DDOS attacks and evaluation of detection techniques. **Sensors**, v. 24, n. 11, 2024. ISSN 1424-8220. Available from: <https://www.mdpi.com/1424-8220/24/11/3571>.

GILL, P.; DOLAN, G. Originality and the phd: what is it and how can it be demonstrated? **Nurse Researcher**, RCN Publishing Ltd., v. 22, n. 6, p. 11–15, 07 2015. ISSN 2047-8992. Available from: <http://dx.doi.org/10.7748/nr.22.6.11.e1335>.

GRIFFIOEN, H.; DOERR, C. Examining mirai's battle over the internet of things. In: **Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security**. New York, NY, USA: Association for Computing Machinery, 2020. (CCS '20), p. 743–756. ISBN 9781450370899. Available from: <https://doi.org/10.1145/3372297.3417277>.

HOUMZ, A. et al. Detecting the impact of software vulnerability on attacks: A case study of network telescope scans. **Journal of Network and Computer Applications**, v. 195, p. 103230, 2021. ISSN 1084-8045. Available from: <https://doi.org/10.1016/j.jnca.2021.103230>.

IANA. **Special-Use IPv4 Addresses**. RFC Editor, 2002. RFC 3330. (Request for Comments, 3330). Available from: <https://www.rfc-editor.org/info/rfc3330>.

ILG, N. et al. A survey of contemporary open-source honeypots, frameworks, and tools. **Journal of Network and Computer Applications**, v. 220, p. 103737, 2023. ISSN 1084-8045. Available from: <https://doi.org/10.1016/j.jnca.2023.103737>.

Integritetsskyddsmyndigheten. **GDPR Rättslig grund**. 2024.

<https://www.imy.se/verksamhet/dataskydd/det-har-galler-enligt-gdpr/rattslig-grund/>.

IRWIN, B. **A framework for the application of network telescope sensors in a global IP network**. 243 p. Tese (Doutorado) — Rhodes University, 2011.

IRWIN, B. Network telescope metrics. In: **2012 Southern African Telecommunications and Applications Conference (SATNAC)**. George, South Africa: [s.n.], 2012.

IRWIN, B.; NKHUMELENI, T. Observed correlations of unsolicited ip traffic across five distinct network telescopes. **Journal of Information Warfare**, Peregrine Technical Solutions, v. 14, n. 3, p. 1–14, 2015. ISSN 14453312, 14453347. Available from: <https://www.jstor.org/stable/26502727>.

ISMAIL, S. et al. A review of amplification-based distributed denial of service attacks and their mitigation. **Computers & Security**, v. 109, p. 102380, 2021. ISSN 0167-4048. Available from: <https://doi.org/10.1016/j.cose.2021.102380>.

IZHIKEVICH, L. et al. Cloud watching: Understanding attacks against cloud-hosted services. In: **Proceedings of the 2023 ACM on Internet Measurement Conference**. New York, NY, USA: Association for Computing Machinery, 2023. (IMC '23), p. 313–327. ISBN 9798400703829. Available from: <https://doi.org/10.1145/3618257.3624818>.

JAIN, S. **Anonymous Collective Targets Bahrain's E Visa Service**. 2024. Available from: <https://thecyberexpress.com/cyberattack-on-e-visa-service-of-the-bahrain/>.

KALLITSIS, M. et al. **Zooming Into the Darknet: Characterizing Internet Background Radiation and its Structural Changes**. 2021. Available from: <https://arxiv.org/abs/2108.00079>.

KAMBOURAKIS, G.; KOLIAS, C.; STAVROU, A. The mirai botnet and the iot zombie armies. In: **MILCOM 2017 - 2017 IEEE Military Communications Conference (MILCOM)**. IEEE Press, 2017. p. 267–272. Available from: <https://doi.org/10.1109/MILCOM.2017.8170867>.

KENNEALLY, E.; CLAFFY, K. An internet data sharing framework for balancing privacy and utility. In: **Engaging Data: First International Forum on the Application and Management of Personal Electronic Information**. [S.l.: s.n.], 2009.

KITCHENHAM, B. et al. Guidelines for performing systematic literature reviews in software engineering. **EBSE Technical Report EBSE-2007-01**, p. ..., 2007.

KORI, M. et al. Quantitative and qualitative evaluation of tcp target ports through active network telescope. **International Journal of Information Technology**, v. 16, n. 5, p. 3177–3191, 2024. ISSN 2511-2112.

KOSE, A. et al. **The Global Role of the U.S. Economy: Linkages, Policies and Spillovers. Policy Research Working Paper 7962**. [S.l.]: World Bank Group, 2017. Available at: <https://documents1.worldbank.org/curated/en/649771486479478785/pdf/WPS7962.pdf>.

KUMAR, B. S.; GOWDA, V. S. Detection and prevention of UDP reflection amplification attack in WSN using cumulative sum algorithm. In: **2022 IEEE International Conference on Data Science and Information System (ICDSIS)**. [S.l.: s.n.], 2022. p. 1–5.

LIVADARIU, I. et al. Inferring carrier-grade nat deployment in the wild. In: **IEEE INFOCOM 2018 - IEEE Conference on Computer Communications**. [S.l.: s.n.], 2018. p. 2249–2257.

MAXMIND. **Maxmind GeoIP database**. 2023. Available from: <https://www.maxmind.com/en/geoip-demo>.

MELL, P.; GRANCE, T. **Special Publication 800-145. The NIST Definition of Cloud Computing**. . [S.l.], 2011. Available from: <https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-145.pdf>.

MOORE, D. et al. **Network Telescopes: Technical Report**. [S.l.], 2004.

MOUBARAK, J.; BASSIL, C. On darknet honeybots. In: **2020 4th Cyber Security in Networking Conference (CSNet)**. [S.l.: s.n.], 2020. p. 1–3.

NEM, R. **AWS to Charge Public IPv4 Addresses**. 2024. Available from: <https://community.aws/content/2c6cKl0W4KQZIO6dWnnDHs1tHSp/aws-to-charge-public-ipv4-addresses>.

NVD. **CVE-2013-5211 detail**. 2025. Available from: <https://cve.mitre.org/cgi-bin/cvename.cgi?name=cve-2013-5211>.

OPENBSD. **The OpenBSD Documentation Project. OpenBSD PF - Packet Filtering. Traffic normalisation and conditioning**. 2025. <https://www.openbsd.org/faq/pf/filter.html>.

O'HARA, J. **Cloud-based network telescope for Internet Background Radiation collection**. Dissertação (Master's dissertation) — Trinity College Dublin, The University of Dublin, Ireland, April 2019. Available at <https://bit.ly/4iYqsFM>.

PANG, R. et al. Characteristics of internet background radiation. In: **Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement**. New York, NY, USA: Association for Computing Machinery, 2004. (IMC '04), p. 27–40. ISBN 1581138210. Available from: <https://doi.org/10.1145/1028788.1028794>.

PARY, R. **New Amazon EC2 Spot pricing model: Simplified purchasing without bidding and fewer interruptions**. 2023. <https://aws.amazon.com/blogs/compute/new-amazon-ec2-spot-pricing/>.

PAULEY, E.; BARFORD, P.; MCDANIEL, P. DScope: A Cloud-Native Internet Telescope. In: **Proceedings of the 32nd USENIX Security Symposium (USENIX Security 2023)**. Anaheim, CA: USENIX Association, 2023.

PAULEY, E. et al. Measuring and mitigating the risk of ip reuse on public clouds. In: **2022 IEEE Symposium on Security and Privacy (SP)**. [S.l.: s.n.], 2022. p. 558–575.

PELLOSO, M. et al. A self-adaptable system for ddos attack prediction based on the metastability theory. In: **2018 IEEE Global Communications Conference (GLOBECOM)**. [S.l.: s.n.], 2018. p. 1–6.

PETROSYAN, A. **Number of internet and social media users worldwide as of April 2024**. 2024. Available from: <https://www.statista.com/statistics/607716/worldwide-artificial-intelligence-market-revenues/>.

PONNUMANI, R.; PARTHEEBAN, N. Various threats and attack on cloud computing: A review. In: **2022 5th International Conference on Contemporary Computing and Informatics (IC3I)**. [S.l.: s.n.], 2022. p. 1296–1300.

POSTEL, J. **User Datagram Protocol**. [S.l.], 1980. (Request for Comments, RFC768). Available from: <https://doi.org/10.17487/RFC0768>.

POSTEL, J. **Internet Protocol. Technical Report 791**. 1981. Available from: <http://www.ietf.org/rfc/rfc791.txt>.

POSTEL, J. **Transmission Control Protocol**. [S.l.], 1981. (Request for Comments, 793). Available from: <https://www.rfc-editor.org/info/rfc793>.

POUGET, F.; DACIER, M.; PHAM, V. Understanding threats: a prerequisite to enhance survivability of computing systems. **International Journal of Critical Infrastructures**, v. 4, n. 1-2, p. 153–171, 2008.

RAMAKRISHNAN, K.; FLOYD, S.; BLACK, D. **The Addition of Explicit Congestion Notification (ECN) to IP**. [S.l.], 2001. (Request for Comments, 3168). Available from: <https://www.rfc-editor.org/info/rfc3168>.

RICHTER, P. et al. A primer on ipv4 scarcity. **SIGCOMM Comput. Commun. Rev.**, Association for Computing Machinery, New York, NY, USA, v. 45, n. 2, p. 21–31, 2015. ISSN 0146-4833. Available from: <https://doi.org/10.1145/2766330.2766335>.

SAHOTA, J.; VLAJIC, N. Mozi iot malware and its botnets: From theory to real-world observations. In: **2021 International Conference on Computational Science and Computational Intelligence (CSCI)**. [S.l.: s.n.], 2021. p. 698–703.

SALAMATIAN, L. et al. Who squats ipv4 addresses? **SIGCOMM Comput. Commun. Rev.**, Association for Computing Machinery, New York, NY, USA, v. 53, n. 1, p. 48–72, 2023. ISSN 0146-4833. Available from: <https://doi.org/10.1145/3594255.3594260>.

SANS Internet Storm Center. **SANS Internet Storm Center: Data for Port 5100 UDP**. 2024. Available from: <https://www.speedguide.net/port.php?port=5100>.

SECUNIA. **Avaya Communication Server 1000 Denial of Service Vulnerability**. 2013. Available from: <http://www2.packetstormsecurity.org/1104-advisories/sa44213.txt>.

SHAIKH, F. et al. Internet of malicious things: Correlating active and passive measurements for inferring and characterizing internet-scale unsolicited IoT devices. **IEEE Communications Magazine**, v. 56, n. 9, p. 170–177, 2018. Available from: <https://doi.org/10.1109/MCOM.2018.1700685>.

SHIN, S.; GU, G. Conficker and beyond: A large-scale empirical study. In: **Proceedings of the 26th Annual Computer Security Applications Conference**. New York, NY, USA: Association for Computing Machinery, 2010. (ACSAC '10), p. 151–160. ISBN 9781450301336. Available from: <https://doi.org/10.1145/1920261.1920285>.

SHODAN. **Shodan: Search engine for the internet of everything**. 2024. Available from: <https://www.shodan.io/>.

SMEDBERG, M.; IRWIN, B. Changes in observed internet background radiation traffic in the 155/8 netblock. In: . [S.l.: s.n.], 2024.

SMITH, B. **DevOps for the desperate: a hands-on survival guide**. San Francisco, CA: No Starch, 2022. ISBN 978-1-71850-249-9.

SORO, F. et al. Enlightening the darknets: Augmenting darknet visibility with active probes. **IEEE Transactions on Network and Service Management**, p. 1–1, 2023.

SPEEDGUIDE. **Speedguide Port 10000 Details - Known port assignments and vulnerabilities**. <https://www.speedguide.net/port.php?port=10000>. 2024. Available from: <https://www.speedguide.net/port.php?port=10000>.

SPEEDGUIDE. **Speedguide Port 10001 Details - Known port assignments and vulnerabilities**. <https://www.speedguide.net/port.php?port=10000>. 2024. Available from: <https://www.speedguide.net/port.php?port=10000>.

TAZAKI, H. et al. How to design a library os for practical containers? In: **Proceedings of the 17th ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments**. New York, NY, USA: Association for Computing Machinery, 2021. (VEE 2021), p. 15–28. ISBN 9781450383943. Available from: <https://doi.org/10.1145/3453933.3454011>.

TORABI, S. et al. Inferring and investigating iot-generated scanning campaigns targeting a large network telescope. **IEEE Transactions on Dependable and Secure Computing**, v. 19, n. 1, p. 402–418, 2022.

TRENDMICRO. **BKDR_XINGDOOR.A**. 2013. Available from: https://www.trendmicro.com/vinfo/nz/threat-encyclopedia/malware/bkdr_xingdoor.a.

WANG, B. et al. A longitudinal measurement and analysis study of Mozi, an evolving p2p IoT botnet. In: **2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)**. [S.l.: s.n.], 2022. p. 117–122. ISSN 2324-9013.

WANI, A. et al. Data analytics technique of different ip address utilized for cloud services. In: **2023 9th International Conference on Signal Processing and Communication (ICSC)**. [S.l.: s.n.], 2023. p. 274–278.

WAZLAWICK, R. S. **Metodologia de Pesquisa para Ciência da Computação**. 3. ed. [S.l.]: GEN LTC, 2020.

WETHERALL, D.; SPRING, N. **Robust Explicit Congestion Notification (ECN) Signaling with Nonces**. [S.l.], 2003. (Request for Comments, 3540). Available from: <https://www.rfc-editor.org/info/rfc3540>.

WUSTROW, E. et al. Internet background radiation revisited. In: **Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement**. New York, NY, USA: Association for Computing Machinery, 2010. (IMC '10), p. 62–74. ISBN 9781450304832. Available from: <https://doi.org/10.1145/1879141.1879149>.

ZAKROUM, M. et al. Monitoring network telescopes and inferring anomalous traffic through the prediction of probing rates. **IEEE Transactions on Network and Service Management**, v. 19, n. 4, p. 5170–5182, 2022.

ZEID, R. B.; MOUBARAK, J.; BASSIL, C. Investigating the darknet. In: **2020 International Wireless Communications and Mobile Computing (IWCMC)**. [S.l.: s.n.], 2020. p. 727–732.

ZHAO, J. et al. Research on the speed and accuracy of full port scanning. In: **2023 IEEE 6th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)**. [S.l.: s.n.], 2023. v. 6, p. 1159–1162.

ZIVIANI, A. **An Overview of Internet Measurements: Fundamentals, Techniques, and Trends**. University of Technology, Sydney (UTS), 2006. 39 p. Available from: <http://dx.doi.org/10.5130/ajict.v2i1.19>.

ZOLOTYKH, M.; NIKULIN, I. Background noise of the internet: Look from russian segment. In: **2020 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT)**. [s.n.], 2020. p. 539–542. Available from: <https://doi.org/10.1109/USBREIT48449.2020.9117630>.