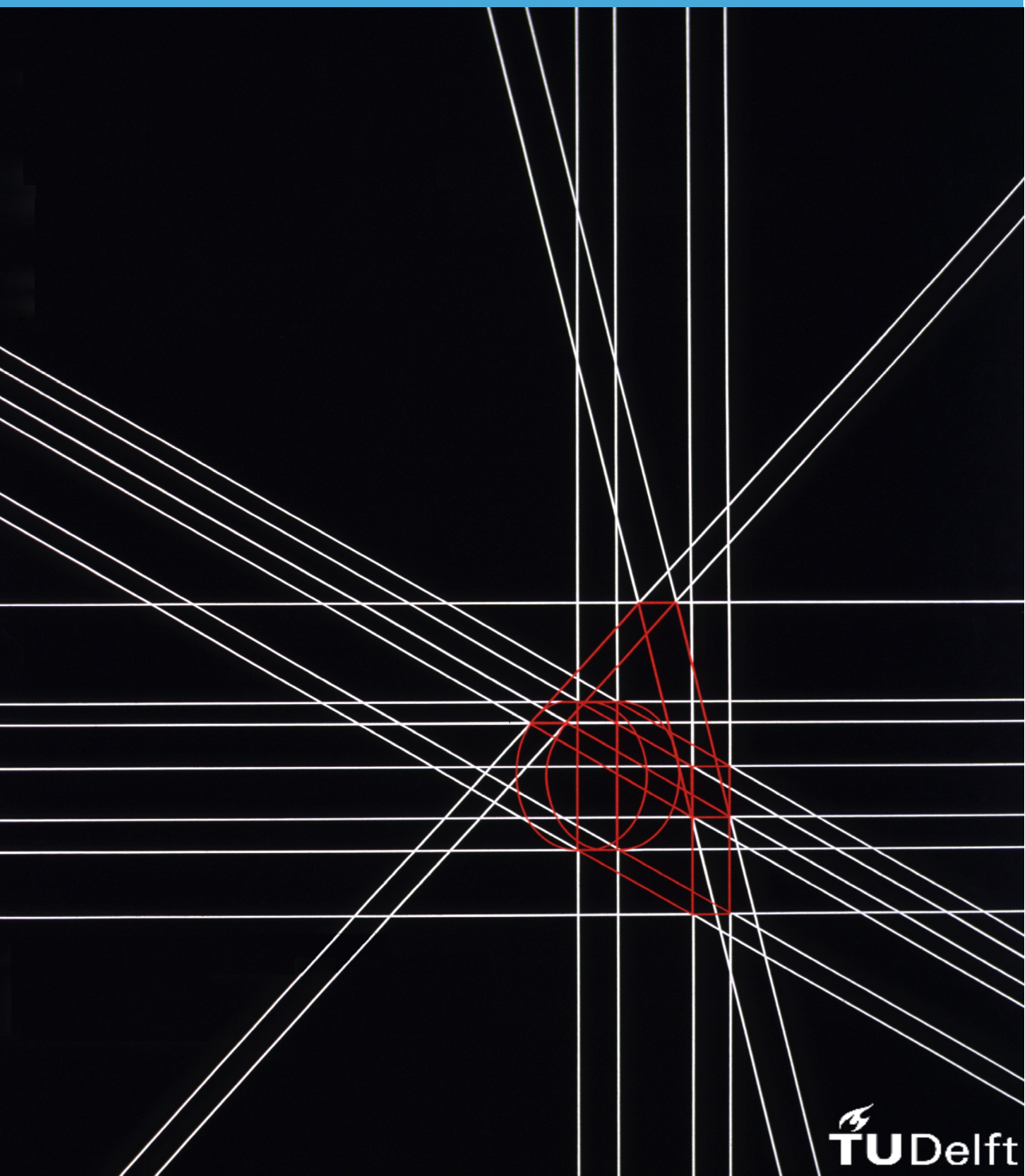
MSc Thesis in Engineering and Policy Analysis

Understanding the Attackers and Victims in IoT-based DDoS attacks

Kevin W. Su

2021



TU Delft

Understanding the Attackers and Victims in IoT-based DDoS attacks

A mixed methodology approach to understanding cybercrime

Master thesis submitted to Delft University of Technology
in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Engineering and Policy Analysis**

Faculty of Technology, Policy and Management

by

Kevin Wang Su

Student number: 4438108

To be defended in public on March 1st 2021

Graduation Committee:

Chairperson:	Prof.dr. M.J.G. van Eeten, section Organisation & Governance
First Supervisor:	Prof.dr. M.J.G. van Eeten, section Organisation & Governance
Second Supervisor:	Dr.ir. B. Enserink, section Policy Analysis
External Supervisor:	Dr.ir. A. Noroozian, University of Amsterdam
External Supervisor:	Ir. M. Aliapoulios, New York University Tandon

Learn from yesterday, live for today, hope for tomorrow.

Albert Einstein

Preface

Here it is, my last deliverable for the Master of Science Engineering and Policy Analysis. It was also the most challenging of them all. My goal for this thesis was to help people in the world deal with a cybersecurity issue so that we can have the trust and safety to rely on digital infrastructure. I truly believe that the digitization has brought humanity much more benefits than costs but we need to stay on top of these risks. Behind these words in this document there was a long struggle to get them on paper. These final documents look clear and on point to outsiders but the process behind it often gets forgotten. I want to make explicit that it was not easy and if you are at that stage now, that it will be ok eventually.

This is the end of my educational career as a student at the faculty of Technology, Policy and Management. I was always confused why we never used the Oxford comma in our name, but as this is the end I might as well. My time spent at Technology, Policy, and Management have taught me so much. We deal with international grand challenges, wicked problems, or complex socio-technical problems, you name it we've had it. These lessons from broad and specialized courses helped me to look at the world and to start understanding it. The colourful cast in our faculty all contributed to this. I could not have been with better people to guide me on the correct path. This research needed their sharpness, perspectives, and support to be made. Arman, Michel, Bert, Max, Qasim, and Elsa: thank you.

Not only the educators but the variety of unique peers I have met amaze me. I started this path with two high school friends but I left with so much more. I want to thank you all. In particular I want to highlight the people who supported and encouraged me during thesis in alphabetical order. Brennen who remains a bright spot in my life with a life vision I can learn much from. The incredibly thoughtful Eva with whom I shared a totally normal work routine. I can truly say I would not have finished this without you but more importantly you were there during some horrible times which I can not thank you enough for. Jason for enabling me to become a bigger thinker than I was and enlarging my world. Jochem for showing me a different world and a source of endless fun discussions. Special shoutout to the EPA family. I am grateful to know people from my cohort, the previous cohort (Mikhail!), and the following cohort (Jin!). Furthermore, let me thank the people in my Dentatus boards, "Dirt Eaters", "Overdue Meetings", "No Thesis Talk", and the wonderful climbing community for allowing me to escape my mind.

I reserve the last paragraph for my real family. 谢谢你我爱你. Lockdown with Just Dance, dumplings, scratch that, I mean all the beautiful food you have prepared, and everything else. I could not be luckier.

Groetjes, *Kevin W. Su*
Rotterdam, February 2021

Executive Summary

To protect critical services in today's society it is necessary to mitigate and prevent risks threatening the reliability of the internet. Internet-of-Things (IoT) devices are the number one attack target on the internet. The situation will become worse as there will be an expected 40 billion IoT devices in 2025. IoT bot activity represented 78% malware network activity or detection events in carrier networks in 2018. The vulnerability and large volume of IoT devices make them a likely target for cybercriminals in distributed denial-of-service (DDoS) attacks. The rise of IoT is increasing the volume of DDoS attacks. A lot of (critical) infrastructure are therefore susceptible being shut down by DDoS attacks.

DDoS attacks are commoditized with booter services, which perform attacks on targets in return for money. This allows a wider audience to utilize DDoS attacks as the only necessary prerequisite is money. These services have increased attack frequencies and attack power of the attacks. The DDoS-as-a-Service landscape has mainly used amplification attacks to take down their victims, however, it is yet unclear if they are also utilizing the growth of IoT for their purposes.

This research will look at the impact of IoT-based DDoS attacks on the victims, with the main research question being: **What patterns of commoditization and victimization can we observe with IoT-based DDoS attacks compared to amplification attacks?**. This research will look into the impact of IoT-based DDoS attacks on underground markets and victims of attacks. Natural language processing techniques such as topic modelling will be used to determine if IoT-based attacks are currently commoditized in Discord communities focussed on DDoS attacks. Furthermore, attack data from AmpPot and Netlab are used to analyze attack characteristics and victimization patterns of amplification and IoT-based attacks respectively. Understanding the impact on commoditization and victimization gives a clearer picture of the threat and risks with IoT-based DDoS attacks.

Firstly, underground marketplaces provide a lot of potential in how attackers behave and what their modus operandi is. This research showed that IoT chatter is frequently used in DDoS-as-a-Service chat communities through natural language processing and clustering of text data. The commoditization of IoT attacks could lead to further growth and development in this area making them even more potent as there is a financial incentive to do so. Furthermore, it shows that we have the ability to filter for new technology in these communities. This will be beneficial for academics and law enforcement to understand how attackers operate and when they change their operations.

Secondly, it becomes clear that there are differences between amplification and

IoT attack characteristics. IoT attacks seem to attack European and Northern American victims more as attacks in Asia are relatively negligible. The victimization pattern also shows a a different trend. While individuals in ISP networks remain the most popular victim for both attack types, IoT attacks hit hosting providers relatively more. An explanation could be that as IoT has not fully matured yet the tools remain at hands of professionals who prefer high profile attacks. As amplification attacks are the standard for booter services their low costs makes them accessible to everyone which lowers the bar and motivation needed for an attack. Limitations are at play as these differences cannot be causally linked to IoT attacks or if these datasets manage to capture the attacks accurately. That is why it is important to research and create tools which allow researchers and practitioners alike to monitor the real situation. These shifts in attacking technology can lead to different behaviour, we need to be prepared for this.

Thirdly, looking at the impact of amplification and IoT on victimization patterns in detail show more interesting results. There are country-level effects at play which explain why certain ISPs are targeted more or less. ICT Development Index and the GDP PPP per Capita are able to explain some of these variances but not all. Some countries even show a completely different pattern than the global pattern, as seen with the Netherlands where attacks on hosting providers dominate attacks on individuals in ISP networks. For the Netherlands it is clear that a few top hosting providers face the majority of the attacks, however, IoT attacks seem to spread their victims across IPs relatively more. This could suggest that they have broadened their attack scope. Furthermore, attacks are often on IPs which share domains making it difficult to find out which domain was truly attacked. This seems to imply that the risk of DDoS attacks have increased attacking new victim groups have increased. Protecting not only individuals more but also preparing organizations should be an important takeaway from these victimization patterns. These policies should be adjusted to the country-level pattern.

Conclusively, vulnerable IoT devices are already a serious threat. They are commoditized and they bring significant differences to DDoS attack characteristics and victimization patterns. As DDoS remains an arms race where adaptation is important, this research showcases a concrete example of how emerging technology can change the existing marketplaces and attack patterns. It also showed its value by looking at IoT from a holistic view to gain understanding of the technical as well as the social impacts. However, more research is needed in this field as the quickly changing field needs to be monitored. Questions still remain which factors can explain the country-level effects in more detail. Expansion of the tools and capabilities to investigate underground chat data would be fruitful as well.

Contents

List of Figures

List of Tables

1

Introduction

1.1. Background

Information and communications technology is a critical part in today's society, as such the United Nations declared access to information and communications technology an important facet of the Sustainable Development Goals (United Nations, 2017). Since the introduction of the information age the world is increasingly putting their trust on digital services and technology. We rely on it for work, communication, entertainment, utilities, transport, and much more. As a global pandemic hit the world in 2020 this reliance on technology has only increased. Lockdowns created physical separation and forced everyone to stay at home. Technology was then used to resume and maintain the *normal* everyday life (Dwivedi *et al.*, 2020). While the long term impact of this sudden transition still needs to be measured, it shows how important reliable digital services and technology are.

In particular, Internet-of-Things (IoT) is being widely adapted by businesses and individuals alike. IoT is described as an interrelated system in which *things* are able to communicate and create data for other *things* but also humans (Atzori *et al.*, 2010; Tan and N. Wang, 2010; Ashton, 2009). This concept can be applied from data collection for manufacturing processes to smart cities to home appliances controlled by your phone. Silverio-Fernández *et al.* (2018) identifies three features for devices to be included as part of IoT: autonomy, context-awareness, and connectivity. Respectively, they suggest that these devices should operate by itself, perceive information from the environment, and establish a connection for communication. Despite their proven usefulness, there are also concerns about the dangers of IoT.

It is estimated that 40 billion IoT devices will be connected in 2025 (Framingham, 2019), with the main drivers being advancements in technological costs

reduction; evolving partnerships and business models; and increase in connectivity and computing power (Rishi and Saluja, 2019). This growth brings unintended consequences with it, as these devices can be misused for malicious purposes. IoT devices are the number one attack target on the internet (Sara Boddy *et al.*, 2019). IoT bot activity represented 78% of malware network activity or detection events in carrier networks in 2018 (Broadband Commission, 2019). It remains difficult to address security challenges in IoT devices. They generally lack the computing power to properly address vulnerabilities. Furthermore, the large pool of heterogeneous devices and the large scale of objects increases the complexity of security as no one-size-fits-all solution exist (Zhang *et al.*, 2014).

Vulnerable IoT devices are often used for Distributed Denial-of-Services (DDoS) attacks (Kolias *et al.*, 2017; Peraković *et al.*, 2015; Vlajic and Zhou, 2018). The basic essence of DDoS attacks is that they render services unavailable by flooding the victim with a lot of traffic to overload the system. IoT devices have the power to generate very powerful attacks with little effort from the attackers making them very dangerous (Lohachab and Karambir, 2018). IoT devices are used as part of a botnet which sends illegitimate traffic for disruption. General DDoS attacks are a credible threat to the reliability of critical information and communication services (Holl, 2015; Cheung, 2017). The biggest DDoS attack recorded was in September of 2017 as Google services[1] were hit with 2.54Tbps, other attacks include AWS[2] in 2020 with 2.3Tbps, Github[3] in 2018 with 1.3Tbps, and Dyn[4] in 2016 shutting down major internet services for Europe and North America (Cloudflare, 2021). These attacks have reached traffic rates in the terabytes and the potency of these attacks keep growing in size. This growth combined with the vulnerabilities of IoT is a threat to the reliability of connectivity in today's society.

The rising threat of DDoS attacks with the use of IoT devices is also facilitated with another process: DDoS-as-a-Service. These type of services are also labelled as *booters, stressors, or flooders*, they perform DDoS attacks in return for money. These have significantly contributed to the danger of DDoS attacks as attack frequencies and attack power have increased (de Santanna, 2017; Noroozian, Korczyński, *et al.*, 2016). Criminals have made DDoS attacks profitable for themselves by renting out their tools and services. These services often make use of amplification DDoS attacks. These amplification attacks make use of vulnerable servers which are able to amplify the attacks sent to them which they redirect with higher bandwidth to a victim. Based on booter websites which were caught and seized, their infrastructure is based on dedicated servers rather than a botnet (de Santanna, 2017; Karami and McCoy, 2013; Zand *et al.*, 2017).

However, there is growing evidence that IoT-enabled DDoS attacks are also being commoditized (Hilt, Kropotov, *et al.*, 2019; Hilt, Mercês, *et al.*, 2020). Tools and techniques related to vulnerable IoT devices are being sold in underground

[1] Google Services include their search engine, email, advertisements, maps, and much more.
[2] Amazon Web Services (AWS) provide cloud computing web services
[3] Github is a hosting platform for software development and version control with git
[4] Dyn is a domain name system (DNS) provider

markets for cybercriminals. In order to prepare and mitigate this threat now and in the future, one aspect would be to expand insights into IoT in DDoS-as-a-Service platforms. It is important to keep track of developments happening in criminal circuits as they can show what might happen in the future. Furthermore, changes of patterns in attack characteristics and victimization are of relevance. Comparing IoT attacks with amplification attacks shows how it might differ from the current modus operandi. As amplification attacks still make up the majority of attacks are popular with booter services (Akamai, 2017). It is important to not only look at the technical aspects of IoT devices. Understanding attackers and victims of these attacks can provide valuable insights how to mitigate this threat with a socio-technical approach. Economics is a way to include a perspective based on incentives rather than technical limitations (Moore, 2010). Security problems are about incentives one way or another and can be explained more clearly and convincingly with economic theories (Anderson, 2001). Using the economic perspective it becomes clear why people who are responsible for protecting systems may not be willing, as they do not bear the consequences (Asghari, M. van Eeten, *et al.*, 2016). Therefore, gaining better insight into the incentives of the actors gives a more comprehensive view of the problem.

1.2. Prior Research

Prior research has been conducted on Internet of Things, DDoS Attacks, and DDoS-as-a-Service.

Prior research into DDoS with the IoT has focused on the technical operations of specific malware such as Mirai and its variants (Kolias *et al.*, 2017; Sinanović and Mrdovic, 2017), showing the potential of IoT as an attack vector. They identified five main reasons why IoT is such a potent threat, namely these devices offer constant and unobtrusive operation, feeble protection, poor maintenance, considerable attack traffic, and non-interactive or minimal interactive user interfaces. There are also case studies of IoT devices which evidently show that these devices have a lot of potential for misuse (Lyu *et al.*, 2017; Vlajic and Zhou, 2018).

As the threat of IoT devices is clear, there is numerous research how to solve this problem. However, there seems to be a bias towards technical solutions in literature surveys on IoT defence (Lohachab and Karambir, 2018; Salim *et al.*, 2020; Vishwakarma and Jain, 2020). Technical solutions such as applying machine learning, blockchain, using edge computing defences, or applying similar malware strategies to notify users instead of infecting their devices (Doshi *et al.*, 2018; Özçelik *et al.*, 2017; Javaid *et al.*, 2018; De Donno, Dragoni, Giaretta, and Mazzara, 2016). These solutions often do not include the socio-technical context of identifying barriers in its implementation. Furthermore, there is also research into the remediation and awareness of infected IoT devices with end-users (Bouwmeester, 2020; Cetin *et al.*, 2019; Verstegen, 2019). Proving that there are more ways to tackle this problem.

However, research into the impact of IoT-based DDoS attacks on commoditiza-

tion and patterns of victimization is limited.

Research into general underground markets show that it has potential to understand attackers, their operations, and incentives. For example, Motoyama *et al.* (2011) looks at cybercrime actors on these forums and map their social network. Others have started by looking at methodologies how to approach analysing underground markets (An and Kim, 2018). Or understanding the overall phenomenon and their value chains and revenue streams (Wegberg *et al.*, 2018). However, analysis into the economics of underground markets or Crimeware-as-a-Service remains difficult. Their closed nature and large volumes of unstructured data makes it difficult to understand what is occurring at scale (Sood and Enbody, 2013; Wegberg *et al.*, 2018). While Hilt, Kropotov, *et al.* (2019) provides evidence of commoditization of IoT in underground forums it remains on a case by case basis.

Through takedowns and seizings from operators of DDoS-as-a-Service the operations beyond the veil were investigated. For example, Karami, Park, *et al.* (2016) and Zand *et al.* (2017) characterizes DDoS-as-a-Service and their internal structures. This helps to look at different interventions such as limiting payment methods, bringing awareness to unknowing co-operators, and avenues for law enforcement to target the market. Building on that, Collier, D. R. Thomas, *et al.* (2019) also evaluated the effects of police interventions on booter services where awareness and takedowns can cause a reduction in DDoS attacks. The literature shows opportunities using broader interventions to mitigate this form of cybcercrime rather than pure technical solutions.

While they looked at DDoS from the attackers perspective, it is also important to consider the victims. Czyz *et al.* (2014) and Noroozian, Korczyński, *et al.* (2016) looked at in-depth investigation and explanation of victimization patterns of booter services using data of amplification attacks. This provides a better understanding of the victimization pattern, as they revealed a democratization of victims and facilitation of crime that is not profit driven but have other incentives.

While prior research has looked into IoT, underground markets, and victimization separately, the impact of IoT on underground markets and DDoS attack patterns have yet to be researched in recent literature. Most of the IoT specific literature have focused on the technical details but not necessarily how it commoditization and victimization are affected. These are core concepts which define the supply and the reason for demand of DDoS attacks. These are valuable insights which can help mitigate the threat.

1.3. Knowledge Gaps

To understand this security issue from both the attacker and victim perspective this research will compare characteristics of IoT-based attacks with more traditional attacks, namely amplification attacks. Based on prior research some key knowledge gaps in our current understanding of the problem will be explained.

The first knowledge gap based on prior research is to provide more clarity how vulnerable IoT devices are being commoditized. Research into the commoditization of IoT-based DDoS attacks remains limited. There are examples given of advertisements related to IoT-based DDoS attacks in underground forums. However, many questions remain on its impact on DDoS-as-a-Service. Its scale and popularity for booter services has not been properly investigated.

The second knowledge gap relates to the impact of IoT-based DDoS attacks on changes in attack characteristics. These patterns have been analyzed before with amplification attacks but a comparative study between both attack types will reveal more about the risks of IoT. As of now it is unclear if these attacks are similar or if they require drastically different techniques for mitigation.

The third knowledge gap lies in understanding the victimization patterns. The victims of DDoS attacks is an area of research that needs to be continually updated as attackers continue to adapt. While there are seasonal updates from the industry, these might be biased towards specific victims (Noroozian, Korczyński, *et al.*, 2016). As these attacks are often sourced from the companies they defend, it is also important to showcase results with different capture methodologies. It remains unclear if IoT-based attacks follow amplification attacks in their targets or if they differentiate.

1.4. Research Questions, Methods, and Flow

The knowledge gaps can be transformed into a cohesive objective for this research. The main focus of this research is to provide an answer to the question: **"What patterns of commoditization and victimization can we observe with IoT-based DDoS attacks compared to amplification attacks?"**. It captures the growth and misuse of vulnerable IoT devices and their role in underground markets and attack patterns. This research targets both attackers and victims, as it is important to understand the incentives from both sides.

The knowledge gap identified reveals there is a need for a holistic approach to understand DDoS attacks due to the commoditization of cybercrime and the Internet of Things trend. To protect the critical ICT infrastructure there needs to be a better understanding of the technical and socio-economic aspects of the security issue.

This question will be dissected into four sub-questions which support the main research objective. They also provide guidance to the scope of the project as it tries to tackle many multifaceted issues of this cybercrime. Hereby limited are the types of attacks that are investigated for these purposes. While DDoS attacks can happen in many different forms the ones investigated are amplification attacks as they remain a significant portion of DDoS attacks in the wild, especially common with booter services (Karami, 2016). Furthermore, IoT attacks of the Mirai variant will be analyzed and not other variants. While other variants have also emerged and are a threat as well, Mirai remains a very popular and effective IoT threat (Safaei

Pour *et al.*, 2020).

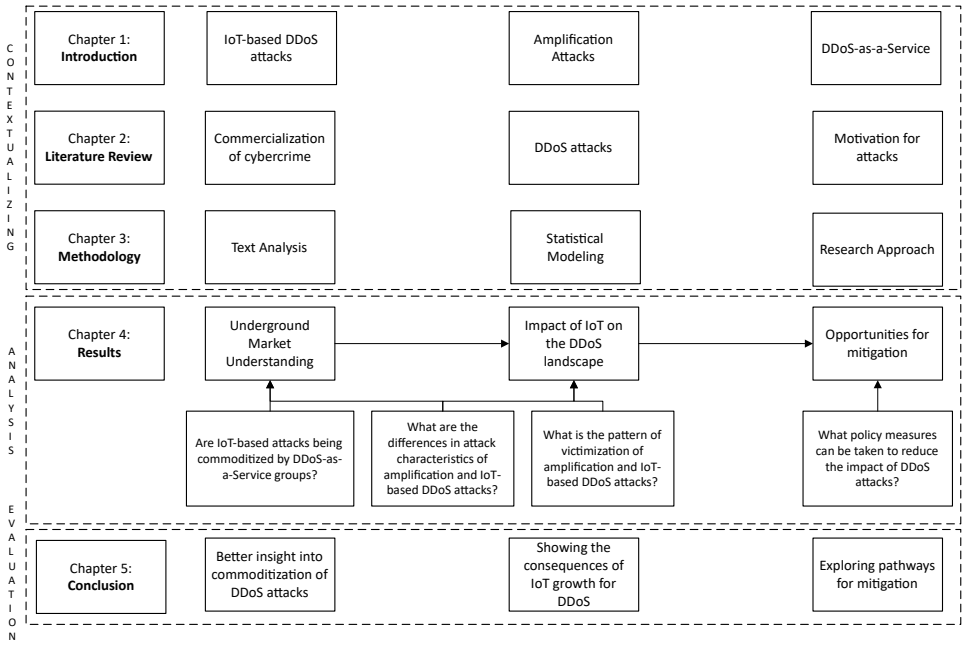


Figure 1.1: Research Flow Diagram

The research flow diagram that is seen in Figure 1.1 details out the structure for the thesis. There are three key areas for this research to look into, namely the insights into the underground markets, impact of IoT on DDoS attack and victimization patterns, and opportunities to mitigate the impact of DDoS attacks.

1. Are IoT-based attacks being commoditized by DDoS-as-a-Service groups?

The first sub-question will look at online underground marketplaces focussed on DDoS-as-a-Service to determine if IoT-based attacks are being commoditized by these groups. These marketplaces are on the popular chat platform *Discord*. It will help increase our understanding of the impact of the growing number of vulnerable IoT-devices. It is important to know about the attackers and their operations given that amplification attacks have been the more popular form. Therefore, the next sub-questions are related to understanding the differences between amplification and IoT-based DDoS attacks.

2. What are the differences in attack characteristics of amplification and IoT-based DDoS attacks?

The second sub-question compares attack characteristics of amplification and IoT-based DDoS attacks. These will be focussed on the quantitative differences measured in their attacks, such as timeline, protocols, attack duration, geographical frequencies. The next sub-question will compare them in more detail.

3. What is the pattern of victimization of amplification and IoT-based DDoS attacks?

The third sub-question is focussed on the victimization patterns of both attacks. While it is important to understand the differences in attacks, it is also important to compare who are being attacked. The victims will be analyzed in more detail as specific categories of autonomous systems (AS). Understanding these in more detail adds better understanding for policy measures as well.

4. What policy measures can be taken to reduce the impact of DDoS attacks?

The fourth and last sub-question will synthesize the previous insights into concrete policies to reduce the impact of DDoS attacks. Through a better understanding of the attackers, attacks, and the victims it will benefit the fight against DDoS attacks and in specific the growth of IoT-based DDoS attacks. The policy measures can focus on all three aspects of the kill chain.

These four research questions help address the main research objective in a logical structure. They help target the identified knowledge gaps about understanding of the commoditization and victimization aspects of IoT-based attacks compared to the current situation with amplification attacks.

1.5. Academic and Societal Relevance

This research adds to the current academic and societal knowledge to maintain a reliable and safe internet. Specifically it targets DDoS attacks which is a rampant security problem for many citizens, organizations, and governments alike.

The academic relevance of this research is that it helps proliferate different avenues of cybersecurity, with a mixed methodology approach of qualitative and quantitative analysis. Not only will it provide knowledge in an already under highlighted part regarding the victims, it will do so applied to an emerging threat with the rise of IoT. Furthermore, it adds an interesting perspective as it also takes into account the attackers side with the use of chat data of these underground markets. Knowledge of these phenomena are hard to come by and any new exploratory insight can help create a better understanding of these opaque worlds.

The societal relevance stems from a bigger picture in which reliable internet services are a crucial part of the modern world. DDoS attacks are a threat to our critical services, therefore also our personal lives. Several key stakeholders in this problem will be able to benefit from the knowledge of this research, such as the national police and the government organization tasked with cybersecurity. It helps identify new avenues of possibilities but also different processes.

1

1.6. Outline

The research proposal is structured as follows, in Chapter 2 an overview of the existing literature will be given to define important concepts and identify knowledge gaps. DDoS attacks and its variants, motivation for attacks, and the commercialization of cybercrime will be discussed. Chapter 3 discusses the most suitable research approach. The design approach, methodology per sub-question, and information on the data used will all be discussed in this chapter. Chapter 4 discusses the perspective of the attackers. This chapter discusses how IoT-based attacks influence underground markets. Chapter 5 discusses the attacks itself. By comparing amplification and IoT attacks DDoS attacks characteristics. The last chapter for the results will discuss the victims of the attacks. Chapter 6 compares victimization patterns of the attacks. Furthermore, Chapter 7 includes limitations of the research and answers the last sub-question on policy recommendations. We conclude this thesis with Chapter 8.

2

Literature Review

This chapter discusses the core concepts related to amplification and IoT-based DDoS attacks. They will address the technical knowledge combined with theoretical concepts that will be addressed in this study.

DDoS attacks
There are a myriad of techniques attackers can use to deny services. This section will evaluate the different characteristics of a DDoS attack. It will set up the difference between amplification and IoT-based attacks. This section also delves into the emerging technology and its dangers. As vulnerable IoT devices are taken advantage of for malicious ends it is important to identify the reasons how and why these devices have become an easy target.

Criminology in cybercrime
To better understand the problem at hand it is important to look at the incentives which enable these operations. Theories based in the domain of criminology might help understand the victimization patterns of the attacks. This section looks at criminology theories, its adaptability to cybercrime, and the motivation of attackers.

Commoditization of cybercrime
An important drive of DDoS attacks is the commoditization of these services lowering the barrier for the general population to perform these attacks. This section will look at their infrastructure and operation, previous studies, and the concept of underground markets and communities.

2.1. DDoS Attacks

This section gives an overview of the current literature on DDoS attacks. Firstly, it explains what DDoS attacks are and the many different characteristics which can be

used to group them under. Secondly, it discusses the amplification and IoT attacks in more detail. Using taxonomies from the first section a comparison will be made between these two attack vectors.

2.1.1. What are DDoS Attacks?

The aim of Distributed Denial-of-Service (DDoS) attacks is to disrupt a network such that the services are not available. While there are many different variants and types of attacks, at its core the attacker uses vulnerabilities to bring multiple machines under their control. The difference between a Denial-of-Service attack (DoS) and DDoS is the fact that multiple machines to perform this task. These compromised hosts then continue to send traffic to overload services so that the victims will not be reachable by legitimate users. The process is visualized in Figure 2.1.

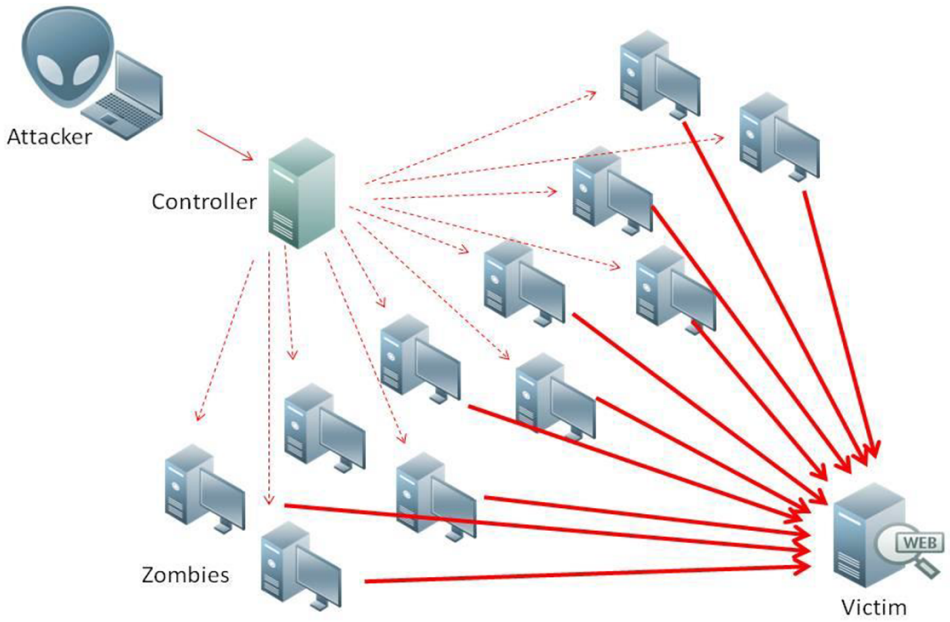


Figure 2.1: Simplified illustration of how DDoS attacks work from Nasanbuyn (2016)

To discuss the different types of DDoS attacks the current literature gives multiple different categorisations or so-called taxonomies. Some common models will be discussed here as a setup to be able to compare between amplification and IoT-based attacks. The OSI model is used commonly to discuss computer networking. It is relevant as DDoS attacks are attacks to disrupt the networking capabilities of services. Furthermore, there are taxonomies specifically related to DDoS attacks and their operations as well.

OSI Model

The OSI reference model is a conceptual model to represent how computer systems communicate with each other. Each layer represents a different function for this communication purpose. It is important to note that this is a conceptualization of networking rather than the precise definition. Each layer could potentially become the cause of a disruption (*DDoS Quick Guide* 2020). Meaning that a service can be disrupted when a single layer is fully exhausted. However, a white paper by Imperva (2014) showed that Layer 3, 4, and 7 attacks were most popular but also the multi-vector approach combining different layers where they combine the following attack types.

Cloudflare, as one of the world's biggest network and anti-DDoS companies, refers to the OSI model to explain three different categories of attacks (*What Is a Distributed Denial-of-Service (DDoS) Attack?* 2020). Namely volumetric attacks, application layer attacks, and protocol attacks. Volumetric attacks are effective like other attacks by sending high volumes of malicious traffic in order to consume the bandwidth of the victim. Application layer (layer 7) attacks refer to the application, specifically the end-user layer in the OSI model. Not only do they utilize bandwidth but also system resources. Protocol (layer 3 and 4) attacks or state-exhaustion attacks attack the underlying network infrastructure, referring to the network and transport layer in the OSI model.

Layer			PDU	Function
Host Layers	7	Application Layer	Data	Human-computer interaction: applications
	6	Presentation Layer		Makes sure that data is in usable format
	5	Session Layer		Controlling ports and sessions for connections
	4	Transport Layer	Segment, Datagram	Transmits data using transmission protocols.
	3	Network Layer	Packet	Facilates data transfer (different network)
Media Layers	2	Data Link Layer	Frame	Facilates data transfer (same network)
	1	Physical Layer	Bit, Symbol	Raw bit stream over networks

Figure 2.2: OSI Model based on the Basic Reference Model ISO/IEC 7498-1

Taxonomies

There are a few different taxonomies created to capture the differences between DDoS attacks. Mirkovic and Reiher (2004) created a taxonomy of DDoS attacks

based on the means to perform the attack, the characteristics itself, and the effect on the victim as seen in Figure 2.3. Douligeris and Mitrokotsa (2004) mentions classification by degree of automation, exploited vulnerability, attack rate dynamics, and impact on the victims. Hoque *et al.* (2015) categorizes it in network/transport or application layer attacks, where they look at degree of automation, exploited vulnerability, attack network, attack rate, victim type, and impact. This is to give a brief overview of the complexity and the many different ways a DDoS attack can differ.

The taxonomy of Mirkovic and Reiher (2004) is the most promising as it features most of the attributes found in the other taxonomies. It does not explicitly differentiate attacks with network layers but that is why it is a good addition to the OSI model.

All these taxonomies are not necessarily related to one DDoS attack type but aimed to show the different attributes of a specific attack. A DDoS attack performed with one specific technique can evolve and change their attack strategy. They could also mix certain strategies together in order to adapt to the defences of their victims. While these taxonomies make them look static, these attack characteristics can change. Its use therefore lies in discussing characteristics that is most commonly associated with an attack type. This makes the taxonomy flexible but also gives us a common language to distinguish the differences between IoT and amplification attacks.

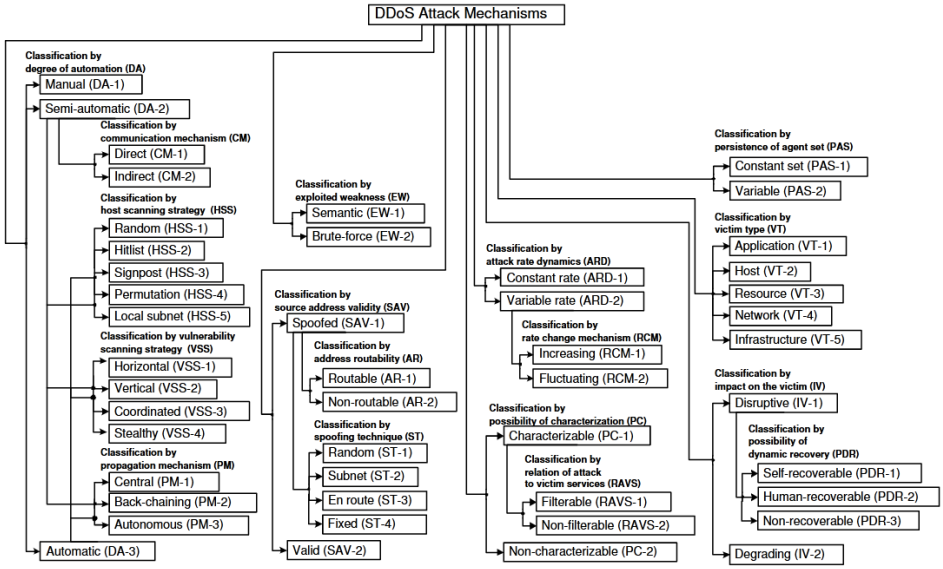


Figure 2.3: DDoS attack mechanisms taxonomy borrowed from Mirkovic and Reiher (2004)

2.1.2. Types of attacks

There are many different types of DDoS attacks. This research will scope itself and discuss the characteristics of two attack types, respectively amplification and IoT attacks.

These attacks were chosen due to their popularity and potential. Amplification attacks make up a significant portion of all attacks in the DDoS ecosystem, namely half in 2017 (Akamai, 2017). At the moment of writing they hold the record of largest DDoS attack ever recorded with 2.3 Tbps on Amazon servers (AWS Shield, 2020). They are also widely used by booter services as an attack mechanism (de Santanna, 2017). Furthermore, IoT attacks are becoming a bigger threat due to the large vector of unsecured devices, as the Mirai malware managed to infect 600,000 IoT devices (Antonakakis *et al.*, 2017). But also more and more devices are infected in a turfwar to sell in the underground of cybercrime (Trend Micro, 2020).

It should be clear that this comparison discusses different techniques but not necessarily mutually exclusive. To be classified as an IoT attack only the source of the attack needs to made up of IoT devices. To be classified as an amplification attack it needs to make use of an amplification technique. It is possible for an IoT attack to make use of an amplification technique in order to harm its victims. The research makes this distinction as it is concluded that IoT attacks very rarely make use of amplification techniques (Antonakakis *et al.*, 2017).

Amplification Attacks

Amplification attacks make use of amplifiers and IP spoofing to strengthen their attacks with minimal resources. The attacks are reflective, meaning that the attacker does not directly send traffic to the victim but goes through amplifiers first. The difference between a reflection as described by (Paxson, 2001) and an amplification attack is the ability to increase the attack volume. These amplifiers are usually public servers which can increase the initial traffic that is sent through them, meaning traffic can quickly ramp up in scale. The fact that these attacks are relayed through other servers means it becomes even more difficult to trace the origin of the attack. To put it simply, the adversary sends a request to the amplifier, however, the adversary makes sure that the amplifier sends the response to the victim rather than the adversary's own machines. They manage to do this through internet protocol (IP) spoofing, so they can impersonate another computing system, tricking the amplifiers that it was the victim that made the initial request and need a response. A schematic of this can be seen in Figure 2.4.

Rossow (2014) identified an initial 14 protocols susceptible to be used for amplification up to a multiplication factor of 4670. DNS amplification seems to be the most popular but there are a variety of other protocols which can be abused. They addressed network protocols such as SNMP, NTP, NetBios, and SSDP but also legacy protocols such as Chargen and QOTD. Furthermore, applications such as P2P BitTorrent and Kad with Quake 3 and Steam as gaming related protocols are susceptible. The availability of numerous protocols which are available also makes

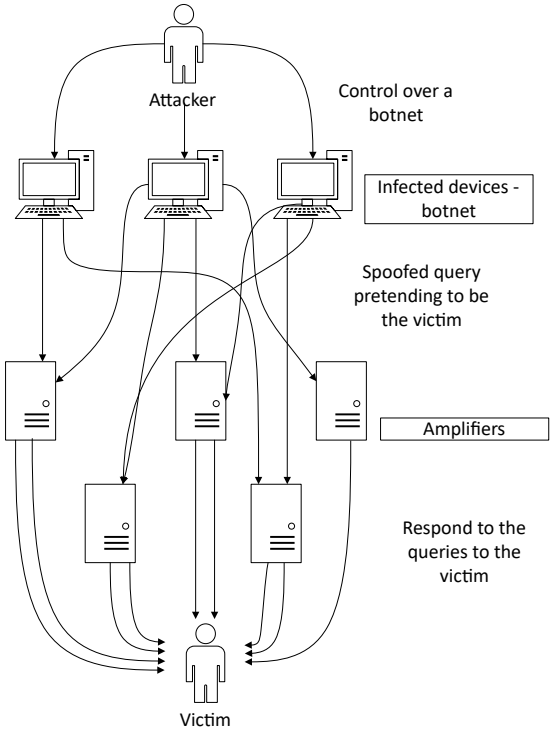


Figure 2.4: DDoS Amplification Attacks Diagram

the threat more dangerous, as new protocols continue being found and abused for DDoS attacks. Ryba *et al.* (2016) lists the ability for amplification attacks to continue to grow and maximize their damages, noting increases in bandwidth and attacks in recent years.

Czyz *et al.* (2014) shows how fast a protocol can rise in popularity to create more potent attacks. The dynamics of this suggests that attackers are quite easily to adapt or find new attack vectors. However, they also report how fast and steadily a protocol can be protected from DDoS attacks by a collaborative effort from the operations community. This shows the arms race between attackers and defenders of amplification DDoS attacks.

Amplification attacks can increase the attack volume and make it more difficult to trace the attacks back to the attacker. They have therefore become quite popular in the DDoS-as-a-Service community, which will be discussed later. Nonetheless, because they use vulnerable public servers they can also be defeated when these servers get the proper maintenance.

Internet-of-Things Attacks

Internet-of-Things attacks are performed through IoT devices, several mechanisms make them an interesting attack vector. While IoT devices often do not have high computational power or bandwidth capabilities compared to traditional infected computers, they are available in high numbers for relatively low effort. In 2020 the number of IoT devices have surpassed the numbers of non-IoT devices for the first time, consisting of 11.7 billion IoT devices out of 21.7 billion (Lueth, 2020). Kolias *et al.* (2017) identified five particular reasons why IoT botnets are getting more popular: constant and unobtrusive operation, feeble protection, poor maintenance, considerable attack traffic, and minimal user interfaces. Their potential as an attack vector, without any interventions, will become ever increasing.

The main drivers of IoT growth can be attributed to technological costs reduction; evolving partnerships and business models; and increase in connectivity and computing power (Rishi and Saluja, 2019). IoT devices are seen as the weakest link in the security chain. They are enabled by a 'winner take all' market structure (Anderson, 2001), meaning that rushing to the market is better than delivering safe products. The enablers of vulnerable IoT devices are: small computing capacity force devices to only run lightweight security applications (Mahmoud *et al.*, 2015); vulnerabilities are not easily patched by manufacturers and end-users alike (Kolias *et al.*, 2017); and the heterogeneity of devices complicate the standardization of security (Ryan and Watson, 2017).

Early instances of IoT malware with DDoS capabilities have been found in 2008, such as *Linux. Hyper* claimed by some to be the progenitor of this specific sort of malware (De Donno, Dragoni, Giaretta, and Spognardi, 2017). The rise of vulnerable IoT devices also led to an increase in malware families, with increased DDoS capabilities and a wider range of devices they target. Some current active botnets are Hajime and IoT_Reaper (Vlajic and Zhou, 2018): Hajime does not have any DDoS attack capabilities and IoT_Reaper does not have any attack activity so far to the knowledge of the author. While they have seen high infection counts they have not been utilized at the moment. Therefore, they are also less popular than Mirai variants. *Mirai* is seen as the most dominant form in the last few years due to its high effectiveness and high-profile cases they have attacked. This makes Mirai a better case study for the development of IoT DDoS attacks.

Mirai is a popular malware targeting IoT devices and using it for DDoS attacks. Some famous DDoS attacks have been perpetrated by this software, creating at the time unseen attack volumes, for example the Dyn (DNS provider), Krebs on Security (famous cybersecurity blog), and OVH (hosting provider) attacks in 2016 were all coordinated with Mirai (Antonakakis *et al.*, 2017).

The source code for Mirai is made publicly available by the creator, which allows for in-depth analysis of the workings behind this botnet (Kolias *et al.*, 2017; Sinanović and Mrdovic, 2017; Kambourakis *et al.*, 2017; Marzano *et al.*, 2018). There are four components which are used for an attack: (1) bot which is the IoT device infected with the malware performing the attack, (2) command and control

(C&C) server, used by the attacker to communicate with its botnet, (3) loader server which connects initially with the IoT device to implement the malware, and (4) a reporter server that collects all the information about the active infected devices.

Using these four components Mirai is able to maintain, grow, and create attacks. They are able to maintain and keep control of the software using the *ki11er module* which shuts down ports 22, 23, and 80 so other malware cannot take control of these devices. Furthermore, it scans and kills similar malware created by other attacks. The botnet is able to keep growing as the *scanner module* uses telnet and randomly generated IP addresses to find other vulnerable IoT devices. They brute-force their way for access with a defined list of default account and password combinations and reports back to the reporting server.

Mirai makes use of the minimal use of interfaces of IoT devices as people by trying to guess the access information, seeing users often do not change the default account and password combination.

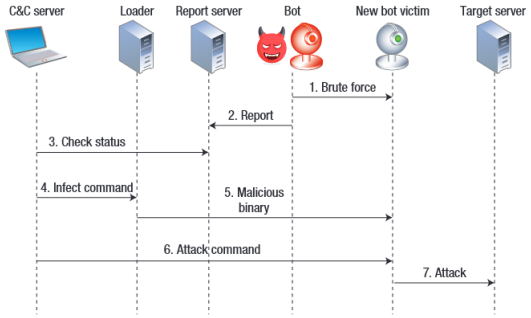


Figure 2.5: Mirai botnet and operation and communication from Kolias *et al.* (2017)

This section showed that IoT devices are continuing to grow, however, there are structural challenges to securing them. The technical issues such as low computing power, no incentives for patching devices, and the heterogeneity complicating the standardization of security. Furthermore, the market also tries to gain a competitive advantage by releasing it first rather than delay it to secure it. They are not personally affected nor are they held responsible for vulnerable devices. While IoT malware has been around for a while, it was not until Mirai that they have become a serious threat. Mirai is able to maintain, grow, and create attacks and preys on vulnerable devices with default user logins. These attacks have also become more popular as the source code of the malware is released allowing everyone to create their own botnet. IoT-based DDoS attacks remain a challenge and its potential seems to be only become bigger.

Comparison of amplification and IoT attacks

These two attack types can hypothetically be utilized in combination with each other, however, the practice show that this does not happen. DDoS-as-a-Service relies heavily on amplification attacks (74%), compared to the fact that only 2.8% of

Mirai's attack commands included amplification (Antonakakis *et al.*, 2017). Meaning that if DDoS-as-a-Service groups adapt IoT attacks, the attack types will be different from their current modus operandi. Therefore, it is important to discuss the similarities and differences between amplification and IoT attacks.

The OSI model shows where each attack type targets their disruption in the communication chain.

Amplification attacks are defined as reflection-based volumetric attacks. These attacks can utilize different layers of the OSI model to cause havoc, however, application and the network layer remain popular with NTP, SSDP, IP fragmentation, and DNS amplification (Majkowski, 2020).

Mirai is able to launch multi-vector attacks, meaning they can utilize volumetric (32.8%), application layer (39.8%), and protocol attacks (34.5%) based on a five month analysis of the botnet by Antonakakis *et al.* (2017). Differing from the standard landscape where volumetric attacks are the majority with 65%. Their most frequent attack types were HTTP Flood, UDP-PLAIN flood, UDP flood, and ACK flood. With the addition of ACK flood (protocol / state-exhaust attack) this means that resource depletion, as the server has to process each ACK packet, becomes a worry for reliable infrastructure (*What Is an ACK Flood DDoS Attack?* 2020).

Using Mirkovic's Taxonomy model to explain differences we are able to look at DDoS-specific attributes. The choice is made to only select these three attributes as there was too much variety in the other attributes to be able to link them to the attack types.

The first point is the degree of automation these attacks are setup with. While the majority of DDoS attacks operate under the understanding of a botnet to increase its efficacy, there is growing research that booter services use dedicated (rented) servers for these attacks (Zand *et al.*, 2017; de Santanna, 2017; Karami and McCoy, 2013). Not to say that amplification attacks are not created by botnets, but the reliability of servers are important considering they would need infrastructure for consistent attacks. Krämer *et al.* (2015) has found that 96.3% of amplification attacks come from a single source, suggesting a booter service rather than a botnet. However, there is a slight degree of automation when they scan for vulnerable ports on the open internet. Mirkovic and Reiher (2004) would characterize these type of attacks as manual degree of automation where the 'recruitment' of devices and planned attacks are manual. Compared to IoT attacks who make use of scanning open protocols to find vulnerable IoT devices. They would operate under some form of Command and Control server to perform their attacks.

Concerning the exploited weakness, this discusses how they are able to find vulnerabilities. They both are able to utilize semantic as well as brute-force attacks. Semantic meaning a specific protocol or application abuse which amounts to excess amount of resources, whereas brute force follows the same volumetric definition.

Source address validity is an important distinction between the two type of attacks. Amplification attacks make use of the ability to spoof the IP address of their

Table 2.1: Overview comparison amplification and IoT attacks

	OSI Model	Degree of Automation	Exploited Weakness	Source Address Validity
Amplification Attacks	Volumetric Attacks	Manual	Semantic/ Brute-Force	Spoofed
IoT Attacks	Multi-Vector Attacks	Automatic	Semantic/ Brute-Force	Valid

victim, making amplifiers reflect their traffic towards them. However, IoT attacks do not necessarily use this technique. While it may seem beneficial for the attackers to hide their bots it can also limit their actions. For instance, it would limit their ability to use HTTP flood as an attack vector as you need an established TCP connection (Mirkovic and Reiher, 2004).

An important note here is that there are many different ways to utilize these attacks, as they do not always fit into clear cut boundaries of these taxonomies. The above classifications are also based on the general understanding of these attacks, however, research is still very novel and this can change very quickly. These other classifications of Mirkovic and Reiher (2004) would have to be analyzed on a case by case basis rather than a general characterization because they way they attack and who would vary too much. These will be discussed at in a later stage of this thesis.

The comparison between amplification and IoT attacks show that their expected pattern are very different from each other. IoT attacks are able to attack a network in multiple ways, whereas amplification remains limited to volumetric attacks. This means that not only bandwidth is a worry but also system resources and underlying network infrastructure. Furthermore, they differ in how they accumulate their raw attack power. Whereas amplification attacks mostly come from a few powerful servers, IoT attacks use a large volume of devices with small computing power. They do share a similar strategy to exploit weaknesses at their victim. Abusing certain protocols for amplification or resource depletion and the ability to use raw bandwidth as well to take down victims are part of their toolset. The last difference is the fact that amplification attacks are spoofed and IoT attacks often are not. However, tracing the attacks back to the large volume of devices they were launched from would not help in mitigation necessarily. These devices are often owned by individuals or organizations whose only fault is negligence in security but not the ones behind an attack.

2.2. Cyber Criminology

These DDoS attacks form a serious problem for the reliability and safety of the internet. However, it remains difficult to fully understand the motivation of the attackers. To be clear, the attacker is hereby defined as the entity that chooses the target of the attack. This is to clear up the confusion between those who create

their own technical infrastructure for an attack and those who rent those services from other people. As it remains difficult to separate these two entities for an attack. This research will draw infer motivation based on the attack targets, therefore, it is important to understand the crime theories and previous research around DDoS attack motivation.

2.2.1. Routine Activity Theory and Cybercrime

Criminology theories at its core discuss why crime happens. This is relevant to understand the motivation of DDoS attacks and its victimization pattern. This section discusses the routine activity theory and its relation to cybercrime.

It is important to notice that the routine activity theory is a traditional motivational theory, meaning it focuses on the motivation of an offence. There are other branches of theories related to self-control and situational opportunities. The importance of motivation in the field of criminology is contested (Greenberg, 2017). As common motivations can be universally applied to everyone, they are not relevant in explaining why or why not people commit crimes (Hirschi and Gottfredson, 2008). That is why some criminal theories can be applied without motivation, such as the self-control theory which is solely based on people's characteristics and (social) control. While it is important to keep in mind that that control theories and motivational theories are at odds with each other, this research will look at motivational theories.

The use of motivation is relevant as we move from a technology centric perspective to the inclusion of social behaviour in order to understand cybercrime (Mandelcorn, 2013). Modelling these attacks through the perspective of the attacker is crucial in coming up practical recommendations to deal with this problem (M. K. Rogers, 2006). The premise of the routine activity theory is that crime events are seemingly unaffected by socio-economic characteristics such as poverty and unemployment but focusses on the event itself.

In the routine activity theory crime occurs when a motivated offender, a suitable target, and absence of a suitable guardian come together (in time and space) (Cohen and Felson, 1979). Parallels to the cybersecurity challenge can be drawn in which increasingly motivated threats, advanced means, and abundant vulnerabilities enable these criminals. Holt, Leukfeldt, *et al.* (2020) applies the routine activity theory to examine the motivation in cybercrime. They have found evidence and further value in examining the relationship between motivation and the victimziation. We have seen with the rise of the Internet of Things that the vulnerabilities and the means are changing, however, it remains unclear what for effect this has on the motives. It is an area of research that can still be further developed in cybercrime (Holt and Bossler, 2015).

However, it is important that the limitations of applying existing criminology theories in the digital world are discussed. Mandelcorn (2013) shows that certain characteristics such as the perception of victimless crimes, accessibility of tools, and

vagueness of cybercrime laws changes certain dynamics. Yar (2005) also argues that the spatio-temporal of virtual environment limits important assumptions of the routine activity theory. Yet, despite its novelty this field has already moved into understanding attackers through motivations.

Furthermore, forensics and analysis of DDoS attacks complicates the matter even more (Lipson, 2002). Reliability of cybercrime statistics due to underreporting but also loosely defined boundaries of cybercrime make it difficult to make data-driven conclusions (Wall, 2007). Attacks are very difficult to be traced back to their true source. The ability for an attacker to hide their identity are important tools any professional will adapt. Therefore, attribution of attacks remain difficult, hence, why motivation of attacks is often linked to the attack target rather than the attacker.

Concluding, this section gives an overview of the routine activity theory and its relation to cybercrime. It is a classic theory which looks at crime from the combination of a motivated offender, a suitable target, absence of a suitable guardian in time and space. However, limitations should be made when applied to cybercrime. As these crimes are not of physical but digital nature. Complications in forensics and analysis makes it difficult to discover the true motivation of an attack. Nonetheless, there is value in examining the motivation and victimization pattern of DDoS attacks. The offender-victim contact is the most relevant for this research and will become the focus.

2.2.2. Overview Motivation Attackers

This section discusses the various motivations of cybercriminals based in the current literature. These motivations help identify reasons for why attackers will participate in illegal activities and therefore help scope measures effective to prevent them from doing so.

An exploratory research using Q-methodology by Cayubit *et al.* (2017) looked at the psychological aspect of the attackers, where they found that personal accomplishment, exploitation, and need satisfaction following their expectancy-value theory.

Previous research on DDoS motivations from Nazario (2008) are explained using the victim's reason for being for the motivation of the attack. They identified four categories: spite or anger, retaliations, financial motivation, and/or political motivation.

Zargar *et al.* (2013) mentions that the study of attackers' incentives offers promising policies leading to the loss of interest by attackers. They mention five main categories: (1) financial/economic gain, (2) revenge, (3) ideological belief, (4) intellectual challenge, (5) cyberwarfare.

Gandhi *et al.* (2011) divides the motivations of attack into political, socio-cultural conflicts, and economically motivated attacks. The social, political, economic, and

Table 2.2: Overview Motivation DDoS Attacks Literature

Paper	Economic Gain	Revenge	Blackmail	Ideology	Personal Challenge	Cyber Warfare
Cayubit *et al.* (2017)					x	
Nazario (2008)	x	x		x		
Zargar *et al.* (2013)	x	x		x	x	x
Gandhi *et al.* (2011)	x			x	x	x
Kumar and Carley (2016)				x		x
Abhishta *et al.* (2020)	x	?	?	x	?	x
Mandelcorn (2013)	x	x	x		x	
Hutchings and Clayton (2016)	x					

cultural (SPEC) background of attackers become more important as our understanding of the attack motivation becomes better.

Some novel research has also tried to link the motivation of attacks using social media analysis. Kumar and Carley (2016), using sentiment analysis they looked at the perceptions each country citizens expressed about another country linking them with DDoS attacks. Implying nations (or their citizens) attack for nationalistic reasons.

Abhishta *et al.* (2020) approaches it from a different angle, by using news databases of high-profile attacks and applying the SPEC values on these attacks it becomes more clear why they are attacked. Giving six categories: *"1) Attacks on large manufacturing companies 2) Attacks targeting public figures and ideological groups 3) Attacks targeting governments 4) Attacks on gaming and gambling platforms 5) Attacks on internet service providers and hosting service providers and 6) Attacks on financial institutions."* from Abhishta *et al.* (2020). Special events or happenings were often related to an attack on an organization. However, it needs to be mentioned that these high-profile attacks are only reported because they are interesting. The DDoS attacks outside of this interest are not recognized.

The general influence model from Mandelcorn (2013) also focuses on the objective goals of the attacker rather than any emotional 'needs'. They mention the following: monetary gain, non-monetary gain, blackmail, revenge, hate, and challenge. While non-monetary gain includes the last four motivations, they were split due to their uniqueness.

Theories from criminology are also used to better understand the incentives of the attackers. The survey done by Hutchings and Clayton (2016) gives insight into the thought processes of people running booter services. Provision of these services is maintained through the fact money can be earned while time and punishment from law enforcement can be kept to a minimum.

This section shows how different each attacker can be in their motivation. The heterogeneity of the attackers motivation shows that we can not talk about a single group of adversaries but rather a large collection of small units each pursuing their own agenda. DDoS attacks can therefore be utilized for many different purposes, such as: economic gain, revenge, blackmail, ideology, personal challenge, and cyber warfare. An attacker could be motivated with multiple goals in mind as well.

2.3. Commoditization of cybercrime

As demand for these attacks are high, it also becomes more interesting to sell these services. Selling illegal online services (such as DDoS attacks) which help buyers conduct cyber crimes is not new. This commoditization of cybercrime helps make the market more organized, automated, and accessible to more people, therefore, increasing the threat and difficulty of mitigation (Sood and Enbody, 2013). Yet, law enforcement and policy makers have not been able to deal with them making them highly prevalent in the DDoS ecosystem.

2.3.1. Overview of DDoS-as-a-Service

According to Zand *et al.* (2017), de Santanna (2017), and Karami and McCoy (2013) the infrastructure of these booter services can be represented as follows. Commonly, there is a front-end website where they are able to market their services to customers in combination with advertisements on underground forums. They usually offer subscription packages, basic ones range from $2 and $15, however, there are also very exclusive ones up to $350. These payments can be provided with multiple payment platforms, such as *PayPal, bitcoin, and paysafecard*. They often offer multiple different attack types, in terms of different protocols but also specific volumetric and application attacks. The most popular form remain amplification attacks due to the ability to have a simple setup with high potency. This means they also maintain a list of vulnerable servers available for amplification. In this case, most of the evidence remains anecdotal as only booters that are caught are analyzed. The case of using botnets to perform large-scale attacks remain prevalent. However, certain booters use dedicated servers (Virtual Private Servers) they rent from hosting providers for their attacks (Karami, Park, *et al.*, 2016).

Zand *et al.* (2017) looked into the infrastructure and attacks of these booter services. They revealed that there is reliability issue in this market as these services have a short lifespan, offer intermittent or no services at all. As devices used for attacks were mostly Linux-based they inferred that dedicated servers and IoT devices are the likely machines being used for attacks. Furthermore, booters have little to no overlap in infected devices meaning these devices are exclusive to specific booters.

The use of these services has also been analyzed by de Santanna (2017). Their sample of 15 booter websites show that the majority of booter customers take little precaution to hide their identity, meaning that most of them are not very technically competent performing these attacks. The causal use of these services is furthermore underlined by the fact that their customers often opt for the cheapest service with attacks less than 5 minutes targeting a few IP addresses. However, a small user group were also very aware of the illegal activities hid their IP address as they performed multiple long attacks.

Karami and McCoy (2013) analyzed the booter service called *TwBooter* which uses dedicated servers for their high computational power and bandwidth instead

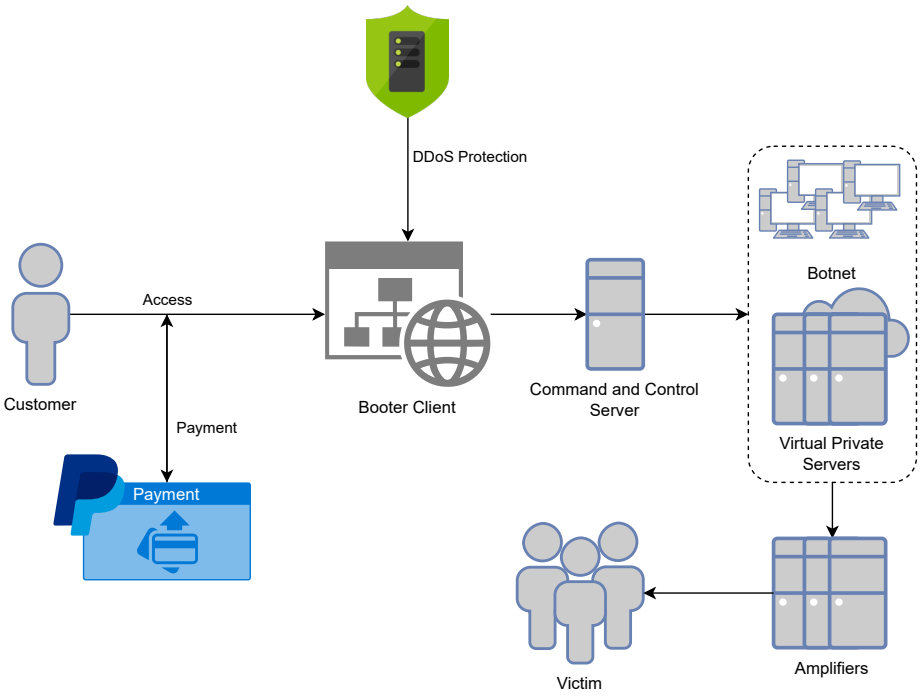


Figure 2.6: Booter Infrastructure Typology

of infected devices. They also identified casual customers using the service for gaming-related advantages and a wider variety of customers targeting diverse websites. In this case, the majority of users do hide their IP address. The discrepancy can be explained as the target audience for each booter service can differ.

Karami, Park, *et al.* (2016) analyses three booters: *Asylum Stresser, Lizzard Stressor, and VDO*, with around 600,000 attacks. Their infrastructure are similar to that of TwBooter, where instead of botnets they would rather use Virtual Private Servers. This paper compared to their previous paper gives a more holistic view of the ecosystem by taking into account the incentives and the stakeholders involved in this enterprise. It discusses the anti-DDoS protection these booter services use themselves, the payment infrastructure, and the amplifiers used for the attacks. The value of this kind of overview has already given multiple avenues for intervention through multiple parties.

All these factors described allow booter services to provide non-technical users to purchase DDoS attacks, it becomes more important to analyze this ecosystem. Where there is demand, there is someone that will offer the service (de Santanna, 2017). Now that the service has been opened to a bigger audience; the number of attacks have also gone up but also the incentives of the players have changed (Chromik *et al.*, 2015; Noroozian, Korczyński, *et al.*, 2016).

2.3.2. Underground Markets and Communities

While booter services seem to be place for little to no interaction, as it is more of an online shop; entire communities are created focussed on DDoS-as-a-Service. These websites often feature social media platforms for customers to be able to reach them. There are examples of them on Twitter [1], TikTok [2], Instagram [3], Youtube [4], and Discord [5]. Footnotes are given to relevant search terms which may change the results depending on the time, however, at the time of writing these platforms do not take posts down regarding DDoS.

While underground ecosystems are an important aspect of growing cybercrime, surprisingly, there is little to no research on communities specific to DDoS-as-a-Service. As seen in previous literature, all papers have dismissed this element as part of the DDoS-as-a-Service infrastructure. However, such a dismissal might become problematic as these communities facilitate the growth and sustainability of cybercrime (K. Thomas *et al.*, 2015). They achieve this by sharing or selling common dependencies for cybercrime and by taking in new actors who will continue to grow the underground market.

The underground ecosystem plays an important role in the distribution and professionalization of these booter services by selling malware kits, and other infrastructure components to create a booter service (Büscher and Holz, 2012). This increases the probability of these findings to be applied to other booter services as they share their way or work.

Collier, Clayton, *et al.* (2020) shares the same sentiments how a small community of highly-skilled actors sell the exploits and vulnerabilities to lower-skilled actors. Therefore, these cybercrime economies play an important role in *"shared, purpose-built illicit infrastructure"*. Meaning that these underground market play an important role in supplying booter services with the knowledge and tools they need.

In order to fully understand the emergence of IoT-based DDoS attacks these platforms give unique insight into the process of the attackers. There is evidence that these communities are keen to learn about IoT devices and ways to misuse them. Hilt, Kropotov, *et al.* (2019) revealed that in five underground communities, from different parts of the world; Russian, Portuguese, English, Arabic, and Spanish, where first attempts are seen to monetize IoT abuse.

While previous research discussed underground markets as a whole, there is another emerging pattern occurring in this space where the black market is moving to purpose-built community sites (Collier, Clayton, *et al.*, 2020; Kamps and Kleinberg, 2018). Discord provides many functionalities over traditional marketplaces,

[1] https://twitter.com/hashtag/DDoS
[2] https://www.tiktok.com/amp/tag/ddos
[3] https://www.instagram.com/explore/tags/ddos/
[4] https://www.youtube.com/results?search_query=how+to+launch+a+ddos+attack
[5] https://disboard.org/servers/tag/boot

as a chat service it is simple to use and extend. Some benefits are its difficulty to take down Discord communities by other rival competitors, ability to screen share and voice chat, functional permission system, and ease of setup (Brewster, 2019). These fast growing communities on chat services are not necessarily new, for example IRC channels are also commonly used by criminals as a marketplace (Herley and Florêncio, 2010). Yet, limited research is performed on these Discord communities so it remains unclear if knowledge on IRC markets can be transferred.

This means that underground markets are not yet fully understood in the context of DDoS-as-a-Service. These markets and communities play an important role in the distribution of tools and services to enable DDoS attacks. To be able to fully understand emergence of IoT-based DDoS attacks the impact on these markets need to be discussed. While Discord hosts many legitimate communities, such as *No Thesis Talk* where a group of people play games and talk about the mental health impact of thesis, they have also attracted many illicit communities. Concluding, little research has been done on underground communities of DDoS-as-a-Service, the impact of IoT on these markets, and communities hosted by Discord in specific. This research will try to tackle all of these factors as they identify the commoditization of IoT-based DDoS attacks.

2.4. Conclusion

This literature review contextualized and defined key concepts used for this research. Revealing the technical details of amplification and IoT attacks and the difficulty of capturing how one attack type operates as it is not rigid. Furthermore, research in criminology and cybercrime gives us frameworks and explanations for why DDoS attacks are performed in the first place. Lastly, a dive into the literature on commoditization of cybercrime shows the professionalization and influence of DDoS-as-a-Service. These three areas help us answer the sub-question 1, 2, and 3 respectively as they all contribute the current state of knowledge on these topics.

3

Methodology

This chapter discusses how the main research objective will be achieved. The first step is the design approach, namely what kind of type this research is and how its philosophy influences the rest of the methodology. Second, each sub-question will be answered separately as they all cover different aspects they need different methodologies. Furthermore, collection of data and its preparation will be discussed in the subsequent sections.

Section 3.1 looks at the design approach of this research, showing the basis of this study. Section 3.2 discusses each research question and how they will be answered. Furthermore, these answers will mostly be based on data. The collection of data will be discussed in 3.3 and the preprocessing in 3.4.

3.1. Research Approach

The knowledge gap previously defined shows that while technical methods are helpful, there is limited information on the incentives despite it being a valuable approach to decrease cybercrime.

A mix of the quantitative and qualitative approaches would be helpful in answering the main research question. According to Denzin (1978) triangulation is ”the combination of methodologies in the study of the same phenomenon” (p. 291). One of the key strengths of triangulation is to be able to capture the *holistic* unit under study (Jick, 1979).

The quantitative methodologies would be used to capture the observed data of DDoS attacks, as valuable information of attacks and victims can be extracted from honeypot data. Data analysis, machine learning, and statistical tests can be of value. Honeypots are ”closely monitored computing resource that we intend to

be probed, attacked, or compromised" (Provos, 2004). They can provide valuable real-time attack data, where they significantly reduce the number of false positives and negatives in the dataset (Nawrocki *et al.*, 2016).

However, they are only useful if they are being attacked. This forms a problem in this context specifically because IoT malware have found ways to identify honeypots and avoid them. The heterogeneity of IoT devices mean that capturing all the data is quite difficult to achieve, because malware can be specific to devices which are not being tracked (Luo *et al.*, 2017). As honeypots are useful for detection and reaction mechanisms, they need supplemental data as well for a better understanding of the problem.

The qualitative approach can be of use in exploring the *ground truth* of the security issue. The data alone cannot explain the motivations of the attackers and victims because a wide variety of factors are not captured (de Bruijne *et al.*, 2017). For example, the data can show which victims have been targeted more than others, however, the exact motivations of the attacks are not explicit. Opposite of that, the data shows that attacks have taken place, however, it is unclear if the victims were aware of it and what they are doing about it. This is where qualitative methodologies could play an important role to understand situational factors. Information on underground markets and attackers can be found through their communication protocols, however, the study would be limited based on the practicality of the available data. The quantitative and qualitative methodologies can therefore be used to verify and validate each others methodologies (Mahoney and Goertz, 2006).

The research will apply an inductive approach to create a better understanding from the real world observations. A mixed methodology approach helps to test several hypotheses caused by the emergence of Internet of Technology in combination with DDoS attacks, such as the differences in attack characteristics and victimization, to understand its effects. To be able to do this, a flexible exploratory research approach is also necessary to find out where the new problems are exactly in the understanding of underground markets and Internet of Things.

Limitations of triangulation could be: replication issues, does not help if the research question is wrong, and the question if the methods are equal (Jick, 1979). The first limitation is valid, however, should not impede on research to better understand the issue. The second limitation is always applicable, but a strong theoretical background in the knowledge gap could reduce the risk of this happening. Lastly, these methodologies should not be seen as contenders but complementary. The research question would strongly benefit from a triangular approach.

In conclusion, the triangulation approach is helpful because the methodologies complement each other to research the problem in a holistic approach; get different viewpoints. The methodologies help to fill in each other's gaps or there could be significant differences, which would indicate unexpected contextual factors.

3.2. Research Questions

The methods will be discussed per each research question.

3.2.1. Underground Markets

The first research question discusses the emergence of IoT-enabled DDoS attacks occurring in DDoS-as-a-Service groups. Based on the literature review new exploits and infrastructure components can be found in underground communities. These markets therefore are interesting avenues to investigate for this emergence. As has been noted, the data consists of unfiltered messages on the chat platform Discord which are related to booter services.

The growth of text on the internet has led to an increase in the field of Natural Language Processing (NLP) (Sun *et al.*, 2017). Natural language processing is the transformation of natural languages, in this case English, into something that the computer can work with (Lane *et al.*, 2019). This allows us to go through a vast amount of unstructured text data with the help provided a computer's processing power. In order to answer the research question such processing power is necessary. As there is a lot of potential with NLP, there still remains a lot of challenges associated on the syntax, semantic, and pragmatic level (Chowdhary, 2020). To increase the validity of the results this section will look at NLP usage in similar research to construct the methodology.

However, the use of chat communication data adds another layer of complexity different from other text analysis. The messages used in the context of Discord messages are often similar to text messages, such as short sentences and abbreviations. Tang *et al.* (2012) named these as the short docs and rampant abbreviations problems exacerbate the difficulties associated with synonymy and polysemy. Synonymy occurs when different words have the same meaning, whereas polysemy happens when the same words has different meanings depending on the context. As the phrase *"it is loose"* can refer to a hold not being tightened at the climbing wall, that there is a relaxed atmosphere or that a monster has escaped. As these messages are quite short, there is not enough information to be able to fully understand the context and the meaning of all the words.

Topic Modelling

To deal with these limitations, the research applies an algorithm for topic modelling on community channels rather than short messages. Angelov (2020) combines a Doc2Vec approach with UMAP feature reduction and HDBSCAN clustering. These techniques allow us to discover topics automatically compared to other topic modelling techniques such as K-Means, NMF, and LDA where you have to define the number of clusters. The steps will be discussed here which allows us to identify topics and clusters of channels in these communities. The steps are also seen in Figure 3.2.

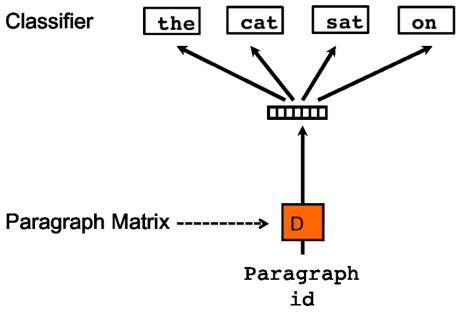


Figure 3.1: PV-DBOW Concept borrowed from Le and Mikolov (2014)

Firstly, a *semantic space* is created where the spatial representation of words are close to the documents they feature. Where a document is some text. While there are different ways to vectorize the text, translating text into a numerical representation a machine would understand, a Doc2Vec approach is used. To explain Doc2Vec, Word2Vec needs to be explained first (Rong, 2014). Word2Vec translates words into numerical representations but with relationships in tact. For example: Netherlands (01), Amsterdam (02), and the sun (03). Word2Vec is able to represent Amsterdam closer to the Netherlands than the sun. However, with how much the Dutch complain about the weather this could be a close call. Doc2Vec utilizes this thinking to apply it to documents rather than words (Le and Mikolov, 2014). While there are two ways to do this, such as the Distributed Memory version of Paragraph Vector (PV-DM) in an evaluation the other one performed better (Lau and Baldwin, 2016). Namely the Distributed Bag of Words version of Paragraph Vector (PV-DBOW), the concept is shown in Figure 3.1. It samples random words from a document which are then classified to determine if these words are part of a document during training. So by inputting a paragraph (document), it gives you randomly sampled words associated with it.

Secondly, because these vectorizations create many features to cluster documents on a dimension reduction technique will be applied. The Uniform Manifold Approximation and Projection (UMAP) for dimension reduction (McInnes *et al.*, 2020) as it preserves the local (each cluster group) and global (clusters in relation to each other) structures. However, cluster sizes and distances between clusters might mean nothing due to its reduction. It uses the number of approximate neighbors to construct a high-dimensional graphs but then it clumps points together in low-dimensional space with a minimum distance. A more detailed explanation can be found in this interactive notebook[1]. A Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) is applied after UMAP in order to find the dense areas of documents (Campello *et al.*, 2013). It works well by applying density based clustering to find similar documents and group them together, it deals with sparse noise by labelling them as such. The density of a cluster is decided by

[1]https://pair-code.github.io/understanding-umap/

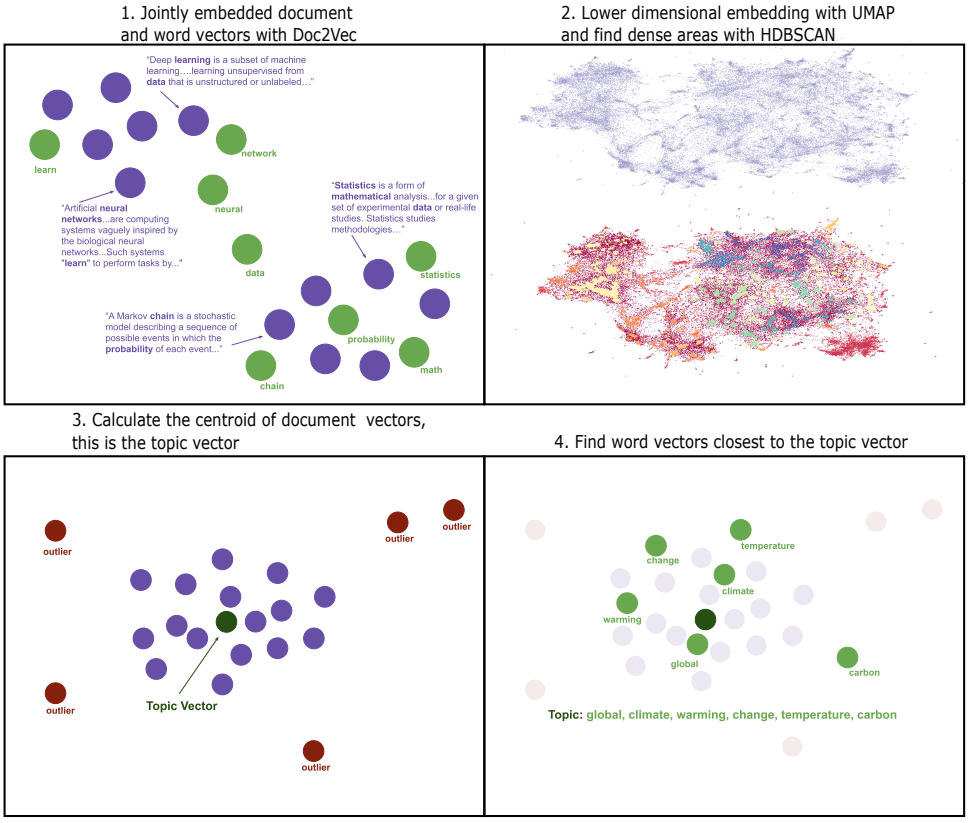


Figure 3.2: Topic modelling strategy borrowed from Angelov (2020)

the *core distance* which will be lower or higher respectively to density and sparsity. Due to the varying densities it is able to apply a cluster hierarchy, which shows if a cluster should be created in subclusters or not.

Thirdly, the topic vectors and words are calculated. The centroid is calculated by the centroid, the mean of all the document vectors in the same dense cluster. Fourthly, each point can then be labelled by its nearest word vectors.

Top2Vec means it is possible to automatically apply an unsupervised learning algorithm which discovers the number of topics and rather than a pre-defined number. This helps as it is unclear how many topics there are in these communities. Furthermore, compared to traditional methods such as Latent Dirichlet Allocation (LDA) and Probabilistic Latent Semantic Analysis (PLSA) the words are also more informative (Angelov, 2020). This is because those models are generative in nature therefore including general words used in sentences which do not give more information however.

3.2.2. Attack Characteristics

The second research question: "What are the differences in attack characteristics and victimization of amplification attacks and IoT-enabled attacks?" will be answered with the use of data visualization and statistical methods. Graphical techniques to describe data are often helpful as an exploratory technique and in conjunction with formal numerical techniques (D'Agostino, 1986).

A descriptive overview will be given regarding attack over time, protocol abuse, country comparisons, and attack duration. These detail the attack characteristics which help understand how they behave and affect victims in several ways.

The survival analysis or time to event analysis is done on the attack duration of DDoS attacks, where the event is the end of an attack. Survival analysis is commonly used to discuss how long it takes before a particular event occurs (Kleinbaum and Klein, 2010). In this case the Kaplan-Meier estimator describes the survival times of members of a group where the logrank test compares these survival times (Kaplan and Meier, 1958; Bland and Altman, 2004). This gives an indication of the attack duration and a statistical comparison between the two attack types.

3.2.3. Victimization

The third research question discusses the victimization pattern of these attacks. It answers who gets attacked and tries to explain why. The victims are classified based on their network type, i.e. what they are mostly used for. Essentially, these will be dissected into more detail to look why certain systems in each category gets attacked more than others. Using various statistical comparisons and models this question can be answered.

Statistical Comparisons

Firstly, statistical modelling will be used to compare the two attack datasets. Goodness of fit measures can be used to describe how well a sample conforms or differs from a hypothesized distribution (D'Agostino, 1986). Hereby, the null hypothesis (H0) is that the distribution of the IoT attacks should follow a similar distribution to the amplification attacks.

There are many different methods which look for goodness of fit in the data. A very clear method would be the use of a chi-squared test, which would look if there is statistical significance between the observed and the expected data in a contingency table. The formula is seen in in the equation 3.1 where O is the observed value and E the expected value. Grouping attack frequency with the type of each victim would give the chi-squared test the necessary ingredients to determine if there are discrepancies between the different victim types. These assumptions of the chi-squared test need to be met (McHugh, 2013)

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{3.1}$$

Furthermore, distributions can be compared using the Kolmogorov-Smirnov two-sample test, as it tests for when two samples come from the same distribution. Looking at equation 3.2 we can see that different empirical distribution functions defined as F in a supreme function. However, it remains sensitive against all possible types of differences. This will be useful to compare certain attack characteristics directly between the two datasets, such as the cumulative distribution function of attack durations or number of packets.

$$K_n = \sup_x |(F_n - F)(x)| \tag{3.2}$$

ISP Broadband Providers

One group which gets attacked a lot are individual users on an ISP network but also computing infrastructure not related to the hosting of websites necessarily. However, earlier classification should have clarified this difference which allows us to look into transit/access network types in more detail. Following the paper from Noroozian, Korczyński, *et al.* (2016) broadband providers can be analyzed in more detail when we would look with linear models. Looking at subscribers, ICT development index, and GDP PPP per capita will allow us to see if this model has changed contrasted with IoT data.

The negative binomial generalized linear model is a regression model based on a Poisson mixture distribution as its underlying probability distribution function (Hilbe, 2011). There are many different variants of this model which can be applied based on the shape of the data, including but not limited to: gamma, inverse Gaussian, or lognormal distributions. Meaning, that it allows for the data to have a different mean and variance unlike the Poisson PDF with □. An advantage of this is that the model fits better when the equi-dispersion assumption is not met.

To understand the effect of these independent variables on the dependent variable a generalized linear model will be used. The assumptions of a regular linear model are often not met which is the reason for a function that links the predictor variable, including (1) independence of residual errors, (2) normal distribution for residual errors, (3), homoscedasticity of residual errors, and (4) linearity. They differ slightly from the usual least squares model to deal with nonhomogenous variance in the dataset (Myers and Montgomery, 1997).

Hosting Providers

For this research there is a specific focus on The Netherlands rather than the whole world. While certain analysis can be applied globally, certain manual classification

actions would require time that this research does not have. Therefore, the analysis of hosting providers is limited to the Netherlands. The Netherlands has a very connected digital infrastructure and is ranked 4th on the Digital Economy and Society Index in the EU (Commission, 2017). Their location near transatlantic communication cables and capable data infrastructure with AMS-IX and NL-IX makes them very attractive for data centers and hosting providers.

Hosting providers are another popular victim type. The methodology for reviewing the incentives of the attacks depend on the type of the victim. Attacks which are linked to a domain, such as tudelft.nl, can provide us with more information on the specifics of the customers of hosting providers. As the attack data only includes IP addresses it is important to distinguish these two set of groups using passive DNS data. This means that historical DNS data is used to check if an IP address is linked to a (website) hostname. To make this as accurate as possible while being able to do this with limited computational resources is to check the DNS records for the attack happening in that month.

The victims of attacks targeting domains will be categorized into general categories such as enterprise, gaming, hosting, illegal activities, and education. By looking at the victims in more detail several hypotheses can be generated on why they are being attacked. However, one IP address can be linked to multiple domains due to a change in the infrastructure or shared IP hosting. Shared IP hosting refers to the ability where one IP address is able to refer to multiple hostnames as the hostname is part of the request accessing the website.

3.3. Data Availability and Collection

This section will discuss the data available to the project. There are in principle three main areas in which data is available, namely amplification, IoT, and underground communication data. Figure 3.3 shows that it is difficult to analyze if IoT botnets are also performing amplification attacks due to the prevalence of IP spoofing. These will therefore be treated as separate datasets, however, there might be overlap between them.

3.3.1. Underground Communication Data

The dataset of underground communication data contains information from messaging platforms such as Discord and Telegram which hosted booter services. These conversation happen on the chat platform Discord. This means that the data consists of real-time short messages in public group channels. The New York University has gracefully provided this information in order to find evidence of IoT-enabled attacks on these platforms. The underground market data is also sourced by Flashpoint[2] which gathers chat services data for the monitoring of threat-actor communities. The dataset contains unfiltered text messages being sent on public

[2]https://www.flashpoint-intel.com/

channels.

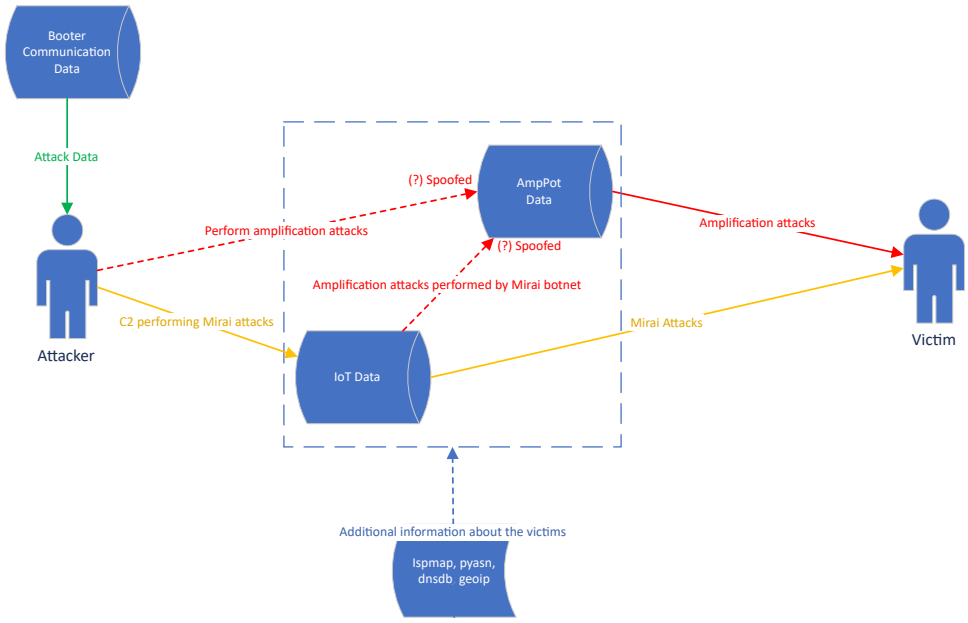


Figure 3.3: Database Information

3.3.2. Amplification Attack Data

DDoS attacks using amplification techniques are gathered using AmpPot based on the paper from (Krämer *et al.*, 2015). This contains information on attacks performed using amplification protocols, such as Memcached (11211), NTP (123), SNMP (161), QOTD (17), CHARGEN (19), SSDP (1900), and DNS (53). While there are 170,411,76 recorded attacks in the year 2019 in this dataset, it is therefore not a guarantee it can capture all the attacks. However, for the purpose of this research the broad sample taken should be enough. The dataset contains information on the victim's IP address, duration, number of packets, location, and the underlying autonomous system being attacked. Frequency of attacks in 2019 can be seen in Figure **??**. The data was gracefully provided by Yokohama national university.

AmpPot is a novel open-source honeypot designed to monitor amplification attacks, as they mimic vulnerable amplification services for attackers to use. Spitzner (2003) describes a general honeypot as: *"a decoy computer resource whose value lies in being probed, attacked, or compromised"*. The purpose here is to gather more information about the attacks as these honeypots capture data of the victim and attack characteristics. The first paper and design of these honeypots were created by Krämer *et al.* (2015), however, its scanners have been expanded with other capabilities since then Krupp *et al.* (2016), such as selective response.

Their definition of an attack means that sources have to send at least 100 con-

secutive requests to their honeypots. This helps to isolate the scans from the attacks. Furthermore, the number of honeypots used showed that there was convergence to measure all attacks. Their mixture of different honeypots allow them to capture a broad net of amplification attacks. They use emulated, proxied, and agnostic honeypots where the former two are protocol-specific the latter will respond regardless. Using this dataset it is quite clear that a significant portion of the amplification attacks will be found, including known and unknown protocols.

The dataset contains 11 sensors, which have captured 21564177 number of attacks across 7 different protocols abused for amplification. A heatmap of the attacks over the services is depicted in Figure 3.4. Notably there are some differences between the sensors, as *sensor018* identifies 0 attacks in DNS most-likely due to a configuration issue. All these sensors started and ended at the same time from 2019-03-01 till 2020-04-22.

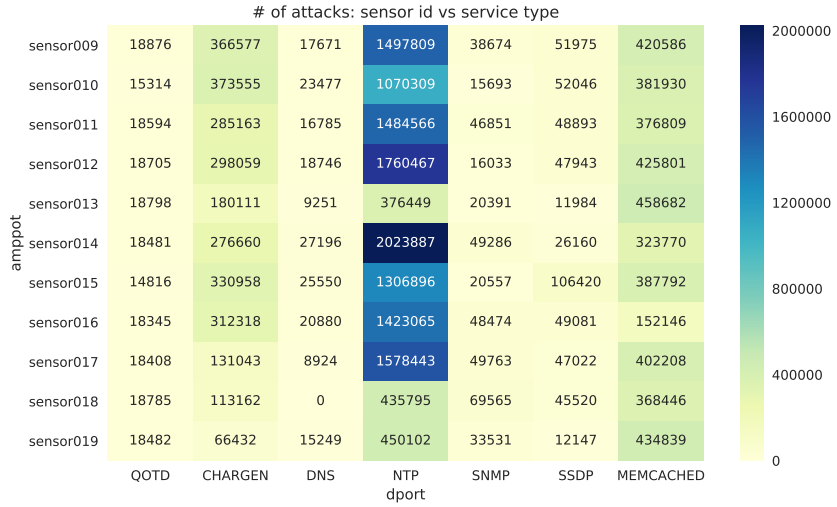


Figure 3.4: Heatmap of observed attacks per protocol and AmpPot sensor

The dataset will help give insight into the attack characteristics of amplification attacks, such as information on frequency of attacks, protocols, duration, location. In addition to that the dataset also reveals information about the victim, however, this might need to be supplemented with additional data from other sources to give a better view.

3.3.3. Internet-of-Things-enabled Attack Data

There are datasets available which contain information of attacks performed by IoT-devices. They are difficult to maintain because of the temporarily nature of Command and Control servers attackers use to launch an attack, which are used

to track these attacks. Netlab collects and extracts the C2 server from a significant number of Mirai samples. They publish their data on their website [3]. The data is very similar to the amplification attack data.

Liu and H. Wang (2018) describes how they have been able to track Mirai variants using over 32,000 Mirai samples. The Netlab 360 organization is a network security research lab that is tracking Mirai attacks through their command-and-control (C2) servers. These servers communicate with the infected bots, attack commands will be given by the server for the IoT devices to process. As they are able to automatically extract these servers from the samples they have a significant collection of C2 servers to track. This allowed them to connect to these servers by simulating an infected device to retrieve the attack commands. Recent blog posts from their lab show that they continue to track the newest Mirai variants too (H. Wang *et al.*, 2020).

Comparing the amplification attack and IoT-enabled attack datasets can show the similarities and differences of the attack. Keeping in mind that these datasets are a sample of the overall attack and that their could be overlap between the attacks, a comparison can give key insights in attack characteristics and victims.

3.3.4. Categorization of Victims

To analyze the victims of these attacks two classification datasets will be used, namely the CAIDA dataset and the ispmap dataset. Both datasets are chosen to give a more complete overview of the victim types, as CAIDA while having more classifications they are not as detailed as the ispmap dataset.

However, it remains a difficult task as networks are often not homogenous, meaning they do not offer a single service. The data used for this classification is a mix between ground-truth data (self-reported by the networks), machine learning, and manual classification.

CAIDA Dataset

The CAIDA dataset uses a machine-learning classifier to classify each AS to their business type: Transit/Access, Content, or Enterprise, see Table 3.1 (Applied Internet Data, 2020). They use a ground-truth dataset called PeeringDB where organizations are able to self-report their organization type for their algorithm to classify the unknowns. Their current claim is that their Positive Predictive Value (PPV) of the classifier is 70%, where $\text{PPV} = \frac{\text{Number of true positives}}{\text{Number of positive calls}}$.

To determine if this is true, a manual classification analysis was performed on 100 ASes. Out of these 100, the CAIDA and the manual classification differed 28 times. However, 18 of those differences were when CAIDA incorrectly labelled ASes as Transit/Access while they are either in Enterprise (14) or in Content (4). From

[3] https://data.netlab.360.com/mirai-c2/

Table 3.1: Overview of CAIDA classifications

Class	Description
Transit/Access	*ASes which was inferred to be either as a transit and/or access provider*
Content	*ASes which provide content hosting and distribution systems*
Enterprise	*Various organisations, universities and companies at the network edge that are mostly users, rather than providers of internet access, transit or content*

Enterprise 2 moved to Transit/Access and one Enterprise was mislabeled as Content. The rest (7) could not be identified through manual inspection so they are given NA. There seems to be an overrepresentation of Transit/Access types when CAIDA was unable to use the ground truth data from PeeringDB.

Ispmap Dataset

This dataset has been used in previous research by Asghari, M. J. van Eeten, *et al.* (2015), Tajalizadehkhoob *et al.* (2016), and Noroozian, Ciere, *et al.* (2017). Based on different data sources they assigned ASes to these five types: education, government, hosting, ISP-mobile, ISP-other, ISP-broadband, and corporate networks through manual categorization of 2050 ASes.

Furthermore, it adds additional information to isp-broadband ASes using (commercial) data from the TeleGeography Globalcomms dataset containing broadband subscriber numbers (TeleGeography, 2020).

3.3.5. Legal and Ethical Use of Data

The data used here comes from partnerships of the aforementioned universities associated with this project. To prevent legal and ethical complications that could arise from this data there are multiple safeguards in place.

Legal issues can arise when these sources are retrieved with illicit means. The data collection procedures to retrieve the attack data are focussed on doing the most minimal impact in the case of an amplification honeypot or removal of the attack bandwidth. The data collected in the underground communication data is in essence open to everyone in the chat. There was no breach needed to enter these communities.

Data privacy remains an important topic and should always be included when we are discussing personally identifiable information (PII). IP addresses in the attack

datasets can be classified as such. While the GDPR mentions that personal data can be used for scientific research in the public interest, the data should not be publicized and its process should be clear (D. R. Thomas *et al.*, 2017). This research will comply with these rules by being as transparent as possible and only publishing aggregated data.

Furthermore, there are several ethical issues which require consideration as mentioned by D. R. Thomas *et al.* (2017). For example, the identification of stakeholders and harms, safeguards, informed consent, justice, and public interest are important facets of this research. The stakeholders of relevance in this case are the attackers, victims, and booter communities. Secondary to that there are stakeholders who are involved but not relevant: Discord as the messaging platform, chat users who do not participate in illegal activities. It is important that these stakeholders will not be harmed by this research by not revealing any PII nor any closed information that is not publicly accessible. Informed consent and justice are not as relevant as they do not speak for this type of data. However, public interest for this research is limited by the lack of reproducibility due to stringent measures taken to safeguard the privacy.

3.4. Data Preparation

In order to use the data for its purpose it is important that the data is available and processed for this use.

3.4.1. Underground Chat Data

The underground chat data requires several preprocessing steps in order to extract it for analysis. Text preprocessing is an important role as it can significantly reduce the amount of text needed to process and increase the efficiency and effectiveness of the analysis.

The first step is to remove punctuation in the raw chat data based on the string module and its punctuation set. This keeps important symbols such as the currency signs and the question mark (offering services for money and the difference between giving or receiving services) Removing punctuation cleans the data up as our model will not support or benefit with the addition of it. While punctuation is important for the understanding of a message, the approach used here does not require the understanding of such subtleties. It also removes the added complexity of dealing with words attached to a punctuation element, e.g. *"word."* is not the same as *"word"*. The same problem occurs when there are words with different capitalizations. This problem is dealt with by putting everything into lowercase.

Secondly, stop words are removed using a 'canonical' pre-defined list from the Natural Language Toolkit (NLTK) (Lane *et al.*, 2019). The full list of the 178 stopwords is available in the Appendix A.2. While Kaur and Buttar (2018) argue that other methodologies based on frequency of words are also valid, the unqiue lan-

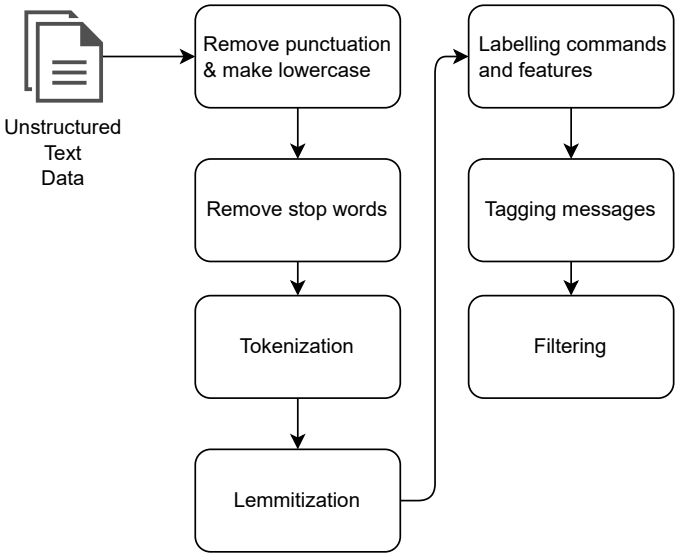


Figure 3.5: Text preparation of the underground chat data

guage occuring in chat data makes the elimination of certain words more difficult. Removing words based on high frequency could eliminate what we are looking for, as these communities are based on DDoS attacks, or low frequency which misses a certain subset of the data. A pre-defined list helps make it transparent what is exactly removed.

Note that the raw text is kept for validation purposes and that, if necessary, the text can be corrected.

Thirdly, tokenizing in this context means splitting the text up. There are different ways to do this but the robust option this research applies is based on a RegEx which tokenizes everything except spaces. Each "token" then becomes a sequence of characters which was grouped by the division of spaces. This will create tokens such that words like *"aren't"* are correct and not *"aren"* with *"'t"*. These tokens help identify key elements in a message and help NLP models to understand the text (Manning *et al.*, 2008).

Fourthly, lemmatizing is important to deal with the challenge of words having multiple variants. The other option here would be stemming. While stemming chops off the ends of words to create its most basic form, lemmatization uses a vocabulary and a morphological analysis to create a root form (Manning *et al.*, 2008). For example, *"saw"* would become *"s"* under stemming, however, with lemmitization this becomes *"see"*. For this use, the WordNetLemmatizer is applied from NLTK which is an online lexical database for online computing use (Miller, 1995).

Further preprocessing was done to reduce the feature space by replacing the lemmatized words which have a specific meaning in Discord or attacks. This was

similar to Hudic *et al.* (2014). They change a variety of values into a single label, for example urls were classified as 'tag_url' rather than the the plethora of diversified urls available. While the original remained still available, the vector space only looks at the tags. This was done for urls, IP addresses (with ports), discord usernames, discord mentions, discord commands, and money messages.

Messages will be identified and tagged accordingly. In order to find the right messages in the sea of messages it is important to find good examples of what you want to build a classification model. Based on these examples it is possible to create a dictionary with the right keywords to identify the messages which are of importance. These dictionaries allow us to 'tag' these messages based on several categories, such as marketplace, access, DDoS, power, and IoT. The full list of keywords used can be found in Appendix A.1. With the help of experts on this subject the list was narrowed down to these keywords. This way of working is based on previous text data analysis research from (Kontostathis *et al.*, 2010; Alami and Elbeqqali, 2015; An and Kim, 2018).

The last step is filtering for relevant channels and messages. This is an iterative process that determines the performance of the models. Several techniques are applied such as removal of words that are used a lot or words that are used below a certain percentage in the documents. Furthermore, removal of irrelevant messages on a higher scale is possible by removing irrelevant channels or channels which do not meet the threshold of the number of messages. Channels which have less than 100 messages were removed. However, it is important that this process is done carefully as to not remove relevant information.

3.4.2. Attack Data

Firstly, the attack data needs to be organized and additional information appended to be able to answer the main research questions. The Netlab and AmpPot datasets both have to include information from GeoIP databases of Maxmind, BGP routing data of Routeviews, CAIDA AS classification data, and ispmap data to be able to extract information on the victims. Based on the IP addresses they fill in the relevant information on the country of the victim, the autonomous system, and the classified type of the AS. See Figure 3.6 for more information to how they contributed.

Furthermore, it was important that the attack data and the information used to tag them were based on a similar time period. To account for this, the closest time period of each data source to the attack data was used. This is to prevent wrong tags based on differences in time, however, this solution is not perfect. The internet is a complex entity in which these data sources are not able to perfectly represent reality. The internet can be seen as an evolving interconnected system that continues to change (Shakkottai *et al.*, 2010).

The raw amplification attacks are separated into monthly data for 2019 from March till December, these files are cumulatively 4.8GB in size. While they also include GeoIP data on the victim, such as the AS organization these data fields are

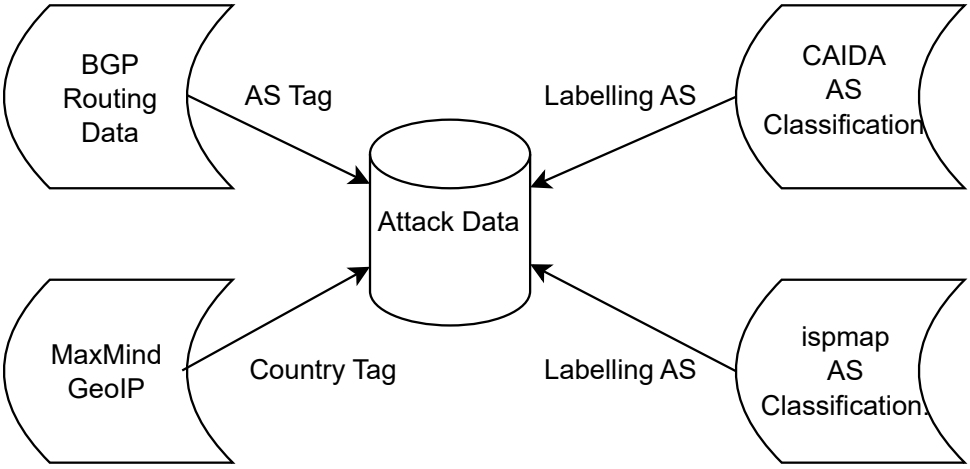


Figure 3.6: Data Structure

not fully sanitized. Meaning that names which include "," or other separator values will not be parsable by the $pandas.read_csv()$ function. Therefore, only the original attack data and columns are taken into account. See the following table of what is included:

The raw data from Netlab 360 is scraped directly from their website in daily intervals. They provide the following information, see Table.

As the size of the data created issues for analysis on the VM (8 Intel Xeon Skylake Processors with 16GB RAM) in a standard Jupyter Notebook environment, the data had to be restructured into two entities. The UML diagrams (Appendix B) shows that it is split into attack and victim data with its necessary information. Atomizing the data in such a way helps with the speed and performance of the analysis. Other helpful tools to deal with the technical constraints which were used are: Modin [4] and Dask [5] to improve performance.

Lastly, manual classification is necessary for certain aspects of this research. Such as the classification of the top autonomous systems and the domains attacked. As accurate classifications for these type of entities using machine learning is still quite difficult, the research decides to manually classify the top categories. For example, the CAIDA dataset only has a positive predictive value of 70%. With margins like these, it becomes difficult to trust the dataset for more detailed analysis. The AS are classified by entering the AS number in the Hurricane Electric BGP Report [6]), find the relevant prefixes, and an internet search on these company descriptions to find out where they belong to.

[4] https://github.com/modin-project/modin
[5] https://github.com/dask/dask
[6] bgp.he.net - Hurricane Electric is a global internet service provider

Table 3.2: Metadata Amppot

Column	Description
id	Attack Unique ID
proto	UDP(17)
src	Victim's IP Address
dst	Honeypot's IP Address
dport	Destination Port, Abused Port
mode	Agnostic / Proxied
starrtime	Time when the attack packet addressed to the victim was observed
endtime	Time when the attack event ended
duration	Attack duration
totalpacket	The number of packets which honeypot received from the attack
avepps	Average packets per seconds in the attack event
maxpps	Max packet per seconds in the attack event

Table 3.3: Metadata IoT Data

Column	Description
Time	Registered time of the attack
C2	Which C2 servers sent the command out
Attack Type	Attack vector, this field maybe null. If attack_type be null means anomaly traffic spike be detected however unknown attack type refers to things such as SYN and FLOOD
Target	IP Address of the victim
Target Port	Which port was attacked during the attack
Duration	Duration of the attack

3.5. Conclusion

This chapter has shown the manners in which the main research objective will be achieved. The choices made in methodology can influence the results heavily which is why it is important that the assumptions and thought processes have been outlined. The triangulation approach combining qualitative and quantitative data reveals that each sub-question needs different tools to answer it. Visualization and statistical modelling are going to be used to answer comparative questions related to amplification and IoT differences. Natural Language Processing will help deal with the large unstructured text data to investigate underground forums. Furthermore, this research relies a lot on data. The capturing methodologies of amplification (AmpPot) and IoT (Netlab) data were shown and how each data source contributed to more information about the victims. Preparing the data to be usable for this study turned out to be a challenging task as the real world data had to be processed for

suitable use.

4

IoT in Underground Markets

This chapter discusses the activity of the chat platform Discord in relation to DDOS attacks. A collection of communities related to booter services on Discord are processed and analyzed.

4.1. General Overview

Table 4.1 gives an overview of the 41 communities in the dataset. There are in total 2916091 messages recorded from 37673 channel-specific unique users. Interestingly enough, the temporary aspect of these communities can be seen as there are communities who do not last for very long. Even new versions of the marketplaces are also recorded such as *"VSB Marketplace 2.0"* and *Mirai Variant & VPN V2.0*. This is even more accurately seen in Figure 4.1 as the majority of the communities do not exists for more than a year with the longest one being *Hardchats 2018* recorded for 963 active days. Despite that, they are not the most active community measured by messages and users. That goes to the community with a very original name: *Hacking*.

Figure 4.2 shows the number of messages on a given day per community. The lines on top represent the start of a new community being recorded in the data. The number of messages ramps up in 2019 and again at the end of 2019 leading into 2020. The gap between 2019 and 2020 is most-likely an issue with the logging service rather than a winter break.

Figure 4.3 takes all these considerations into account to determine the most popular community in this dataset. These are min-maxed normalized scores and it shows that the popular communities are quite centralized at the top. This might suggest network effects at play where bigger communities continue to draw more popularity. Aside from the top communities there are a majority of smaller com-

Table 4.1: Overview Discord Communities related to DDoS

	Community Name	Channels	Date Covered	Active Days	Messages	Msgs/Day	Users	Msgs/User
0	Hardchats 2018	28	2017-07-17 till 2020-03-07	963	52790	54.818276	999	52.842843
1	ThugCrowd HQ	57	2018-12-01 till 2019-04-20	140	274132	1958.085714	2439	112.395244
2	OofLand	13	2018-06-22 till 2020-02-21	608	11490	18.898026	294	39.081633
3	Crossfade	8	2018-10-06 till 2019-01-20	105	32692	311.352381	306	106.836601
4	Official Hackintosh	11	2018-12-01 till 2020-02-23	449	1987	4.425390	181	10.977901
5	ComfyChoo! ☐ (´„•ω•„) ☐	62	2018-12-01 till 2019-08-08	250	88269	353.076000	2539	34.765262
6	Cyber Terrorists	27	2018-12-01 till 2019-02-04	65	66004	1015.446154	824	80.101942
7	☐ Hacking ☐	33	2018-12-01 till 2019-04-19	139	662706	4767.669065	7420	89.313477
8	We Hack For Hentai	18	2018-08-27 till 2020-02-10	531	270126	508.711864	1077	250.813370
9	Nija	24	2018-11-23 till 2019-09-09	289	8896	30.782007	331	26.876133
10	ghost@kirin(キリン)	18	2018-05-10 till 2019-02-28	294	21500	73.129252	723	29.737206
11	CrossFade - AntiSec	7	2018-12-06 till 2019-11-21	350	15120	43.200000	140	108.000000
12	Shadoh's Big Hack Lounge	4	2018-10-03 till 2019-07-29	299	9468	31.665552	200	47.340000
13	VSB Marketplace 2.0	17	2018-12-14 till 2019-02-08	55	8658	157.418182	1194	7.251256
14	Supreme Security Services	9	2018-08-07 till 2019-02-01	178	12458	69.988764	400	31.145000
15	Big Hekks	5	2018-12-14 till 2019-07-29	226	1789	7.915929	59	30.322034
16	hakka shit	6	2018-12-28 till 2019-07-10	194	1738	8.958763	77	22.571429
17	digitalgangster.com	4	2018-12-01 till 2020-03-07	462	699163	1513.339827	1839	380.186514
18	OmitVPN | Support Chat	7	2018-12-01 till 2020-03-05	459	619	1.348584	79	7.835443
19	7 Market	8	2019-01-30 till 2019-03-22	50	16058	321.160000	301	53.348837
20	UvU Kingdom	21	2019-01-17 till 2019-07-29	193	15335	79.455959	780	19.660256
21	Stresser.WTF Community	8	2019-01-04 till 2019-04-21	106	1412	13.320755	125	11.296000
22	TOP-STRESSER'S	8	2019-02-19 till 2019-05-24	93	27801	298.935484	579	48.015544
23	Usernames.org (Delayed)	19	2019-02-27 till 2019-06-05	97	181154	1867.567010	4280	42.325701
24	Illegal Community	13	2019-03-04 till 2019-07-29	146	8001	54.801370	764	10.472513
25	UN5T48L3 Cyber Security	14	2019-04-07 till 2019-05-30	53	9710	183.207547	279	34.802867
26	Console Prospect	43	2019-05-30 till 2020-03-07	282	5632	19.971631	464	12.137931
27	D N L 4	21	2019-05-16 till 2019-07-23	67	7420	110.746269	205	36.195122
28	x0rz.co	20	2019-06-22 till 2019-09-03	72	32521	451.680556	373	87.187668
29	Stress.gg	38	2019-10-30 till 2020-01-17	79	18230	230.759494	1260	14.468254
30	CyberHackers.eu Community	28	2019-10-23 till 2020-03-07	135	76073	563.503704	2557	29.750880
31	NSSALES	8	2019-09-05 till 2020-02-18	166	364	2.192771	85	4.282353
32	Volkoz Booter Discord	7	2019-09-18 till 2019-10-01	12	213	17.750000	34	6.264706
33	Running the System	13	2019-11-28 till 2020-03-06	99	20402	206.080808	553	36.893309
34	CTR 10.0	16	2019-10-23 till 2019-12-30	68	216840	3188.823529	1745	124.263610
35	☐MIRAI VARIANT / VPN☐	13	2019-07-13 till 2019-08-28	46	20008	434.956522	879	22.762230
36	☐Mirai Variant & VPN v2.0☐	10	2019-10-23 till 2019-11-02	9	585	65.000000	91	6.428571
37	MushroomClub	21	2019-07-03 till 2020-02-12	224	4006	17.883929	395	10.141772
38	OVERFLOW.LTD	7	2019-07-26 till 2020-02-28	217	188	0.866359	41	4.585366
39	BangStresser.com	10	2020-01-25 till 2020-02-12	18	6039	335.500000	362	16.682320
40	CYBER NET	16	2020-02-04 till 2020-02-23	18	8483	471.277778	395	21.475949
41	" Legit Shop's "	4	2020-01-15 till 2020-02-21	37	11	0.297297	5	2.200000

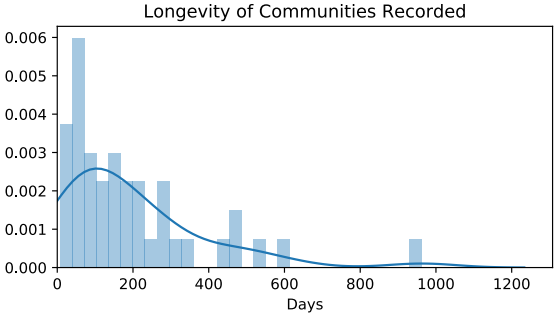


Figure 4.1: Distribution plot showing the longevity of communities in days

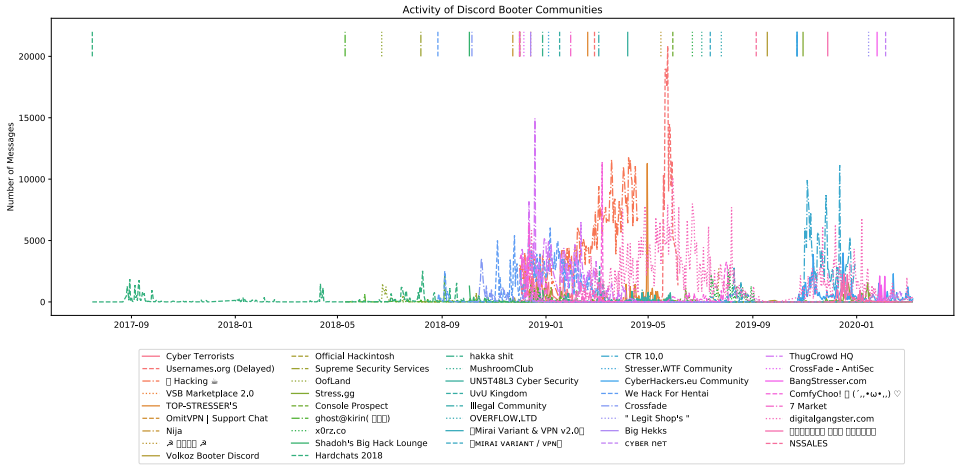


Figure 4.2: Number of messages posted per Discord community

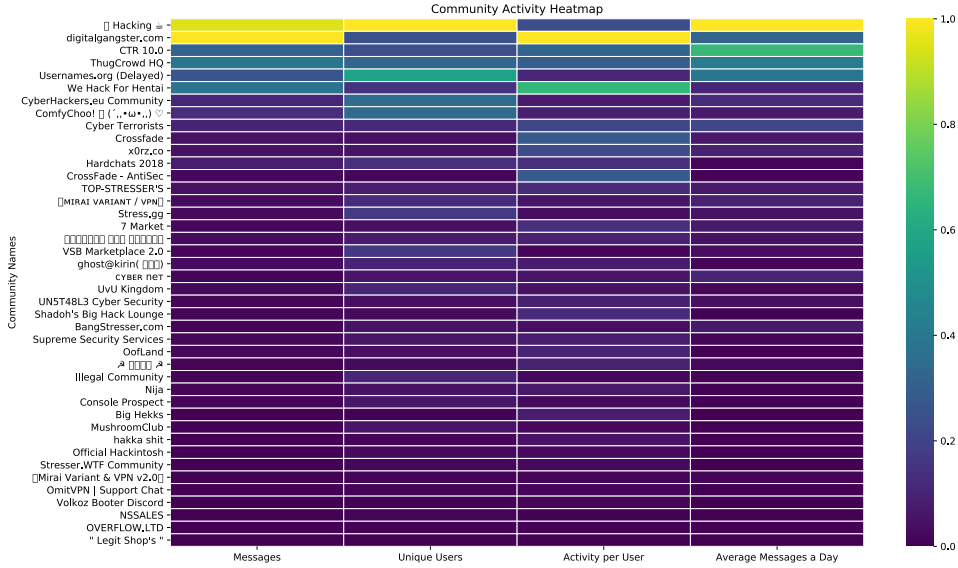


Figure 4.3: Popularity of communities measured across several dimensions

Figure 4.4: Visualizing wordclouds of topics found in the chat data

munities who do not record as much activity.

4.2. Channels

Channels can be seen as digital rooms on Discord where the community interacts with each other. For organizational purposes, channels are often intended for one specific conversational topic. For example, a *memes* channel in Discord is often used by their users to share images with text that often makes a humorous political or social commentary. Figure 4.4 reveals seven different topics which was found during topic modelling.

With the help of the algorithm provided by Angelov (2020) it reveals seven categorizations in the hundreds of channels from the 41 different communities. Namely, *General Channels (112), Community Channels (100), IoT Channels (64), Bot Channels (61), Marketplace Channels (52), DDoS Channels (50), and Hacking Channels (34)*. A full list of all the channels are given in the Appendix C. The topic given to them by the classification is shown in combination with the cosine score. The parameters for the vectorization, dimensionality reduction, and clustering are found in Appendix D.

The Figure 4.5 reveals the documents (channels) that are tagged to their specific topic. UMAP is a dimensionality reduction technique often used to reduce high dimensional data into 2D. Each document featured 300 vectors which are now mapped onto this 2D graph. The relative position of each point is given to the other. There are certainly points which do not belong to a certain cluster, such as the few points on the right and in the centre. This shows the difficulty associated with automated topic modelling as certain channels do not have a single purpose. The majority of the points are in clusters which look to be visually distinct from each other which seems promising for the reliability of the algorithm.

An important mention is that is much more an exploratory tool at the moment as we lack accurate labels for full classification. Nonetheless, it helps scope the relevant channels for this research. Manual inspection of the IoT channels reveals

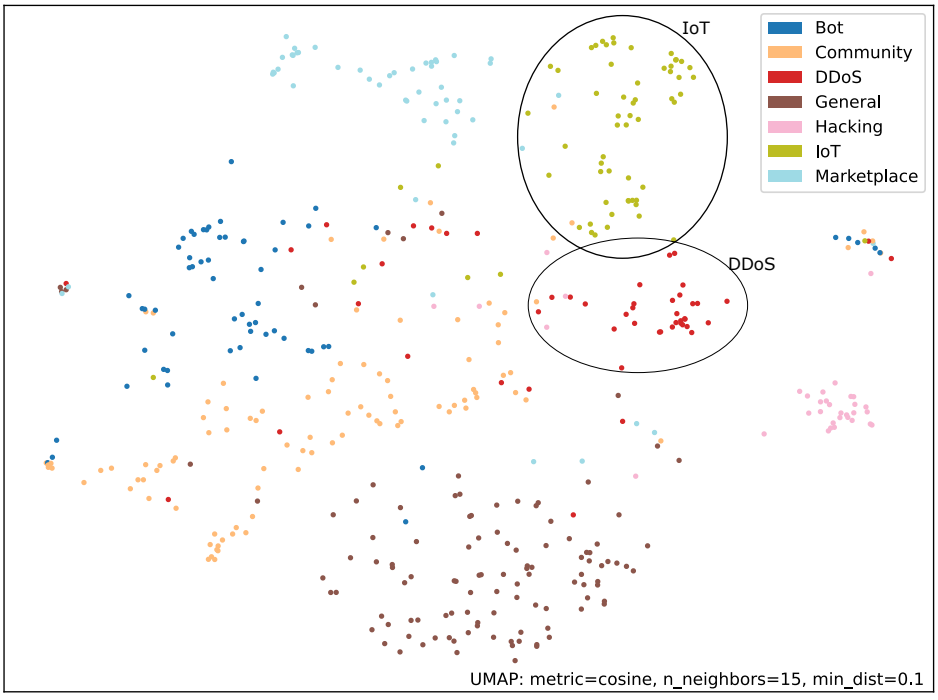


Figure 4.5: Dimension reduction with UMAP to show the clusters

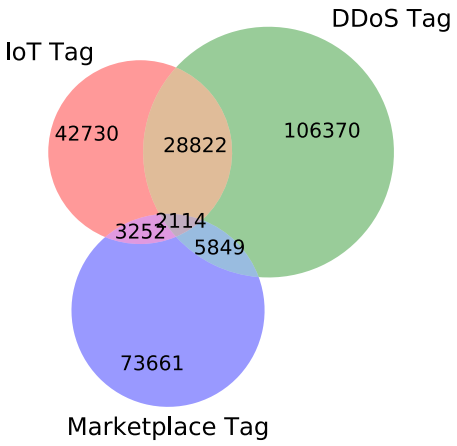


Figure 4.6: Venn Diagram of messages tagged with either IoT, DDoS, or Marketplace keywords

that this form of topic modelling with a lot of features seem to work relatively well compared to other possibilities (LDA, NMF, and LSI). Those tools did not give very interpretable results, they are viewable in the Appendix E.

To verify, the ten channels with the highest cosine score in the IoT cluster were looked at manually. It remains difficult to fully separate channels dedicated to marketplace and DDoS with specific IoT channels. While out of those 8 channels can be seen with messages dedicated towards advertising IoT attacks they also featured a lot of unrelated channels. However, it should be noted that certain channels with Top2Vec also had very low cosine scores related to the topic. It remains a challenge to assign channels which cannot be categorized with a certain topic. It could also be that these channels feature a mix of topics. This model, however, shows promise in its ability to look commonly used words together and the ability to find channels with keywords.

Out of the full 2916091 messages in the dataset, 246839 messages are tagged in one form or another including *Access and Power*. They result in 246839 messages (8.47%) that are tagged accordingly. There remains difficulty in assigning relevancy to all these messages, as a mention of these words does not mean they are immediately of interest for this research. Some keywords can be using in normal conversation rather than actually advertising DDoS attacks. To deal with the high volume of noise this is a good first step.

The Venn Diagram 4.6 gives an indication of the number of messages which are tagged as *IoT, DDoS, or Marketplace*. It adds even more evidence that IoT attacks are being discussed, sold, and bought. The Venn diagram gives us a vague proportion between IoT and DDoS talk on these marketplaces. While IoT attacks are still smaller, they are not that far off from DDoS attacks. The discussions of IoT in these communities are also plotted in Figure 4.7, showcasing discussions

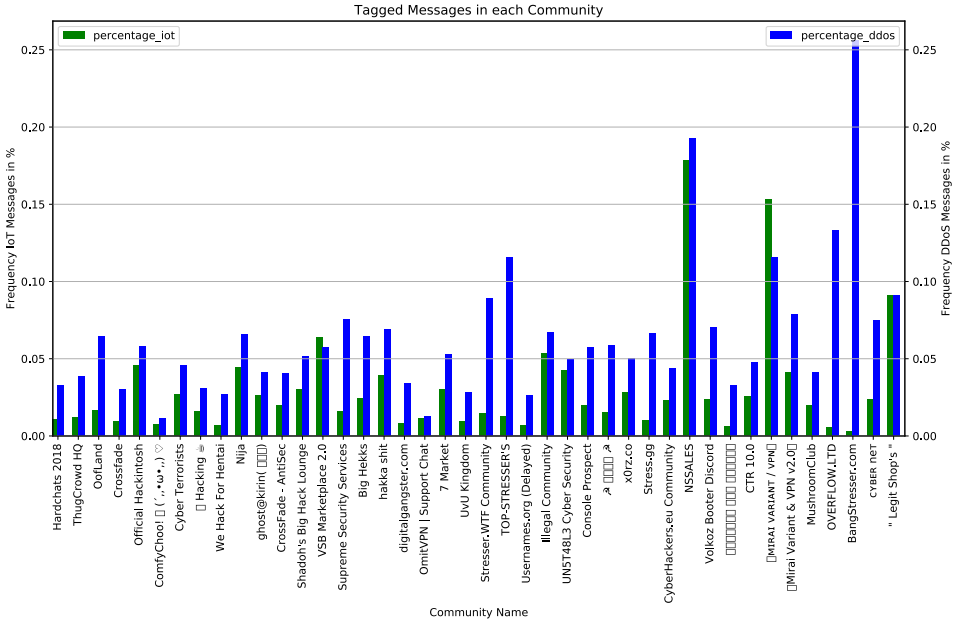


Figure 4.7: Proportion of messages in each community discussing IoT or DDoS

surrounding IoT compared to DDoS in general.

4.3. Conclusion

Concluding, it showed that these underground marketplaces are already commercializing IoT attacks. With the help of clustering algorithms it became clear which channels were used for IoT advertisements. Activity of these underground marketplaces are also concentrated on a select few, most-likely due to network effects.

This means that there is evidence of IoT-based attacks being commoditized by DDoS-as-a-Service groups. The commoditization allows the scene to become more professionalized as there is a monetary incentive to keep adapting. This will increase the risk of IoT-based attacks. We can also see IoT talk in proportion with DDoS in general, where it suggests certain communities are more focussed on it than others. This could suggest specializations of communities developing new technologies in their own niches. Furthermore, it also showed certain characteristics of these Discord cybercrime communities. Their short duration and the popularity of a select few could have implications for law enforcement. Taking down communities could be less beneficial as there already seems to be fast dynamics at play. Communication for new communities could happen on other platforms such as the websites of booters, forums or social media. Popular communities could be identified and checked in order track the mainstream approaches at the moment. Lastly, these short messages occurring in Discord chats complicates finding rele-

vant messages with high entropy for researchers and practitioners alike. The vast amount of unstructured text data makes it more difficult but breakthroughs in NLP will help considerably. The automatic clustering and topic modelling here shows it is possible to filter out the noise and focus on the relevant topics at hand. Now it is of importance what for implications this will bring to DDoS attacks.

5

DDoS Attack Analysis

This Chapter looks at the different attack characteristics and victimization of amplification attacks and IoT-enabled attacks.

The Table 5.1 gives a quick glance of the scale of the attack datasets. The amplification and IoT attack datasets are both collected over approximately one year. However, the difference in attacks collected by each of them is significantly different by two orders of magnitude for the global scope. Given this variation in the dataset succeeding analysis will look at the distribution and pattern of the attack characteristics rather than absolute numbers.

It is difficult to determine what the cause of this size difference could mean. There are several measurement challenges in capturing all the traffic data available, for instance AmpPot and Netlab both use different techniques in order to capture attack traffic. However, it is possible for the Netherlands at least to see that even when the total attacks differ by two orders of magnitude, it is only one when we look at the unique IPs being attacked, and none for unique ASes. This means that information of totally new victims a new attack marginally can add becomes smaller. Suggesting that these attacks are concentrated in relatively few victims. These ASes are often governed by large organizations controlling a sizeable part of the internet, which constraints the total number of them.

Table 5.1: Overview Attack Data

	AmpPot Data	AmpPot Data NL	Netlab Data	Netlab Data NL
Collection Date	2019-03-01 till 2020-03-04		2020-01-09 till 2020-12-23	
Total Attacks	21564177	532221	749523	34084
Victims (Unique IPs)	1065907	20173	145042	6169
Victims (Unique ASes)	15019	351	6000	202

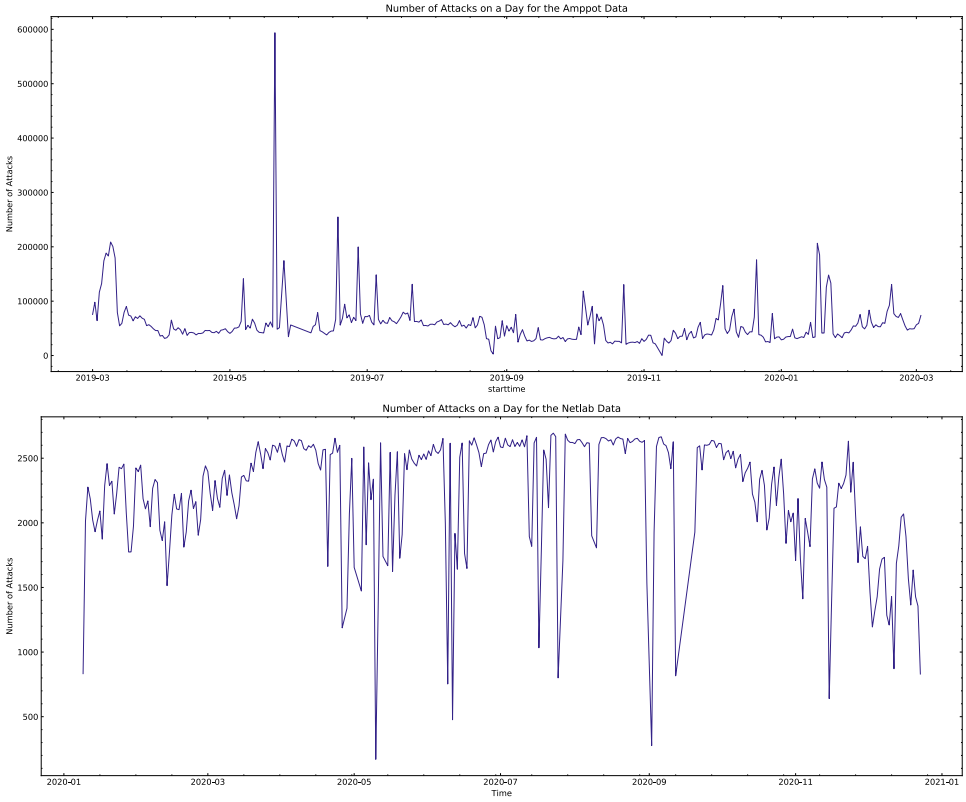


Figure 5.1: Number of attacks over time

5.1. Timeline and Protocol Use

While the data for AmpPot extends into 2020-04-22 there were issues with the honeypot data between March 4th to April 11th. Therefore, everything from March 4th has been cut from the analysis.

The graphs mapping out the attack pattern over their collected time in Figure 5.1 gives two very different patterns. While the amplification patterns show peaks, the IoT pattern reveals drops. This can be explained in two ways. The first possibility shows that amplification attacks are much more stable compared to IoT attacks. There is no significant period in time where they drop below a stable attack count. This possibility would explain why they are so prevalent in DDoS-as-a-Service as they are a robust technique for attacks. The other possibility is that because of the way these IoT variants operate, it becomes more difficult to track them. IoT variants purposely update their Command-and-Control servers to avoid being captured, which is the exact methodology Netlab uses to track them. Therefore, there is a slight delay before they are able to track them again.

The Figure 5.2 shows the frequency of each protocol used for each attack. The

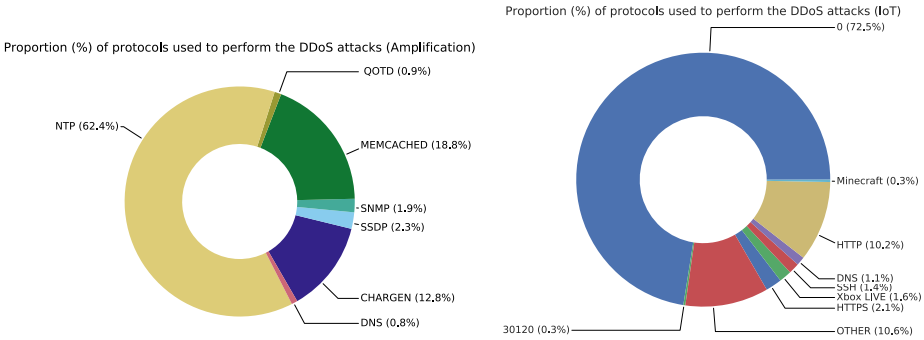


Figure 5.2: Proportion of attacks using these protocols

amplification dataset only lists seven protocols used for amplification, where the most popular one is NTP followed up by Memcached and CHARGEN. The IoT dataset contains a lot of attacks targeting ports not associated with a specific use, that is why only the top ten ports are visible. Whereas amplification attacks make use of protocols viable for amplification IoT attacks are not limited by it. IoT attacks are able to utilize their multi-vector approach by targeting all sorts of ports.

Interestingly enough, the popularity of protocols used for amplification attacks seem to differ from different industry reports such as Nexusguard (2019) and Netscout (2019). First, the variety in attack patterns show the difficulty of capturing this phenomenon in a consistent manner. Differences can be caused by different capture methodologies, however, each analysis shows a different piece of the puzzle. Secondly, this Figure only shows the number of attacks in proportion to the total attacks but not the total volume of traffic sent.

5.2. Country Comparisons

The Figure 5.3 gives a quick overview which countries are over- or underrepresented in unique DDoS victims by population in a log log plot. The United States, Saudi-Arabia, China, France, and the United Kingdom make up the top five for the amplification attacks. Compared to the United States, United Kingdom, Germany, Canada, and Netherlands for IoT attacks. The Netherlands (bolded) is a popular country for both amplification as IoT attacks due to their strong digital infrastructure and hosting capabilities. Appendix F shows these figures without the country labels for a better trend visualization.

Furthermore, Figure 5.4 shows that amplification attacks attack the continents Asia, Europe, and North-America quite evenly. IoT attacks seem to favour North-America with Europe and Asia following quite far behind. The fact that IoT attacks have not been attacking Asia is interesting and could suggest some bias at play.

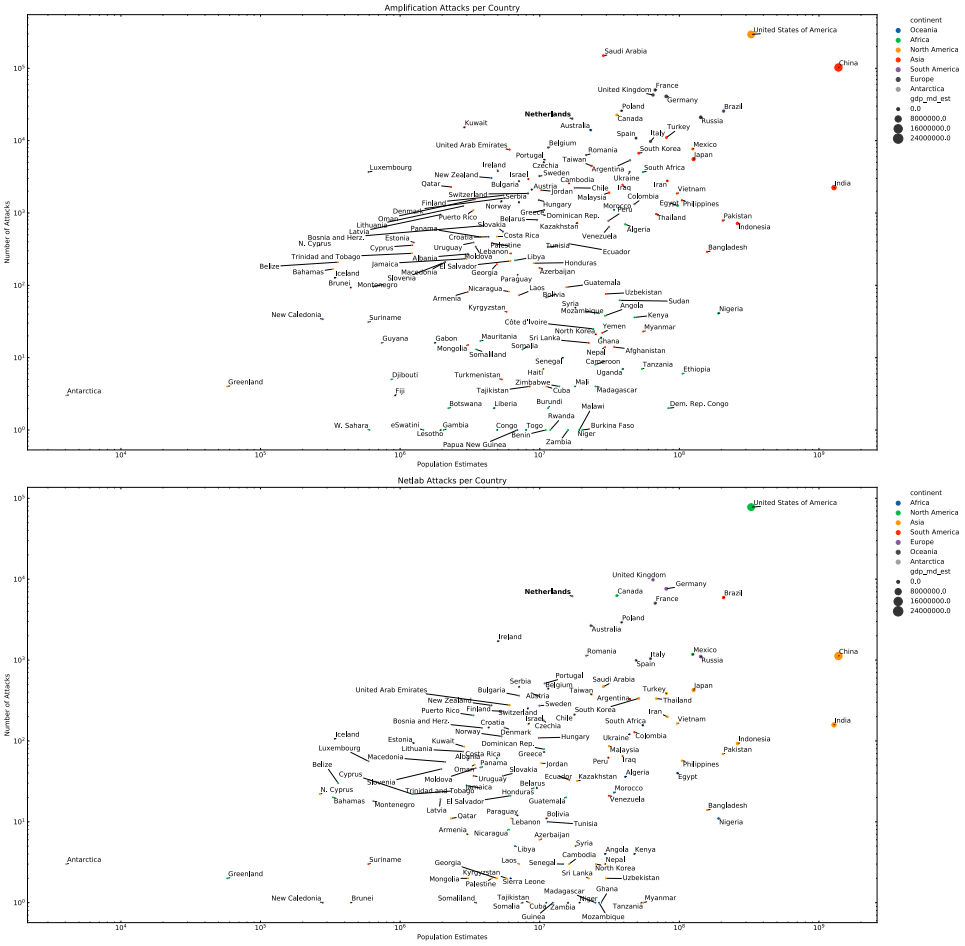


Figure 5.3: Country Comparison of number of unique victims (IP) attacked

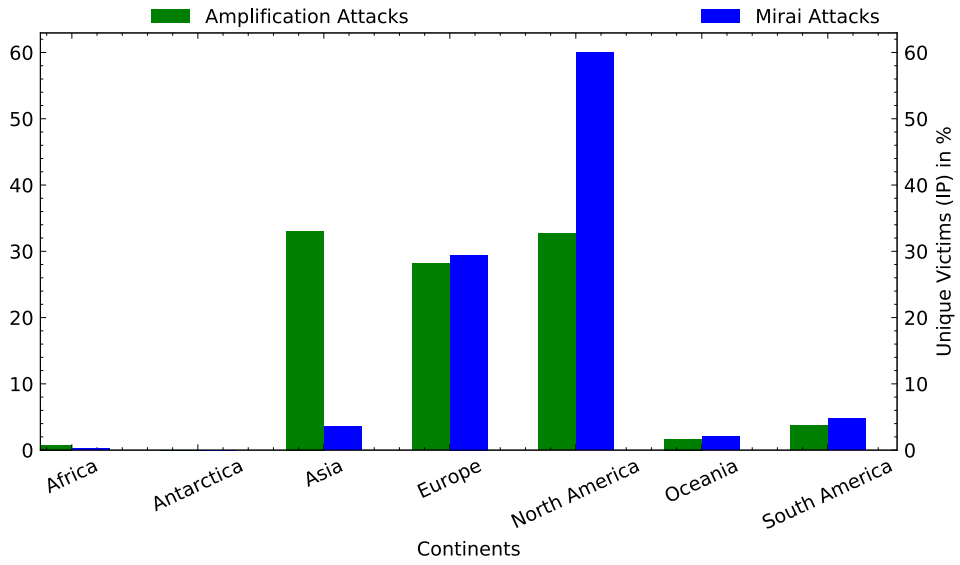


Figure 5.4: Comparison of the proportion of attacks on unique IP victims occurring in each continent

5.3. Attack Duration

The attack duration graphs in Figure 5.5 compare amplification and IoT attack durations in seconds. The attacks tend to be of short duration between 10 and 1000 seconds. The density functions feature multiple spikes and there are slight differences in the number of spikes. The amplification attack around 1.8 shows a very big spike for amplification attacks.

Furthermore, the survival analysis using the logrank test gives a significant difference between the two with 2590.31 as the test statistic. As the survival distributions of the two populations are compared they are significantly different from each other.

5.4. Conclusion

Comparing amplification and IoT attacks there are some differences in their attacks. There are varying attack patterns which could be attributed to the workings of IoT or the capturing methodology. Differences of the geographical space of the victims were found as IoT attacks seemingly do not target Asia as much compared to amplification attacks. Furthermore, survival analysis was performed on the attack duration showcasing differences between these two attack types.

The differences between amplification and IoT-based on technical characteristics mean IoT require different defences. As shown in the previous chapter the commoditization of IoT is happening. While it seems that amplification attacks re-

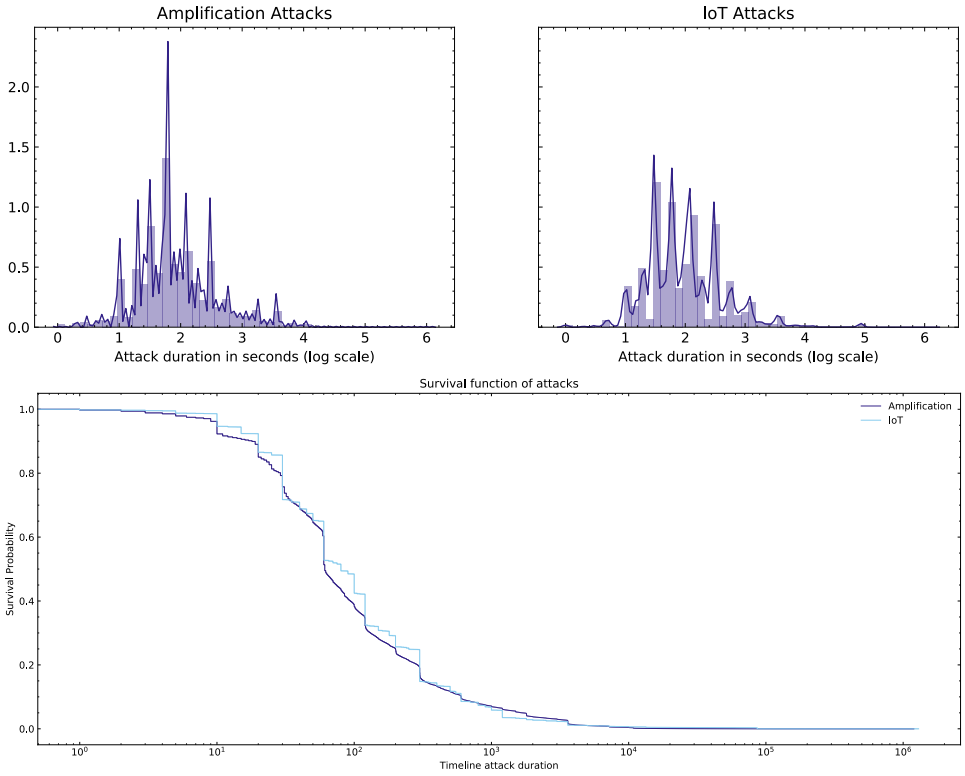


Figure 5.5: Attack duration and its survival function

main the dominant form at the moment, the growth of IoT devices might bring a change to this. What this means is that there should be a bigger focus on mitigating and safeguarding vulnerable IoT devices. Furthermore, it shows a different attack pattern when we take a look at where the victims are placed. This will be analyzed in further detail in the next chapter.

6

Victimization in Detail

This section will look into victimization in more detail. The previous section showed that there are large group of victims in Transit/Access and Content (CAIDA) and isp-broadband and hosting (ispmap). Going through broadband providers and hosting providers it is possible to give more information about the motivation of an attack.

6.1. Victimization Comparison

This section will compare the attack characteristics and victimization patterns of amplification and IoT-enabled attacks (Mirai).

6.1.1. Unique AS and IPs Victim Types

The Figures 6.1 show the CAIDA victim types for amplification and Mirai attacks respectively. The graph represents the victims through unique ASes and IPs which were identified in the dataset. An attack is able to affect the individual machine but also the network. An AS is a collection of IP prefixes under the control of a single entity with a defined routing policy to communicate with the internet (Hawkinson and Bates, 1996). It is the AS which contain the designated IP address, however, as each AS is operated by a single entity labelling these makes it easier to know what each attack is targeting.

The distribution of the attacks between amplification and IoT are roughly similar, where a lot of attacks are located in Transit/Access, with Content and Enterprise who are less preferable as targets. A chi-squared test between unique AS and the unique IPs result in p-values of $2.59e-36$ and 0, as they have a significant relation between the groups. The CAIDA classification shows a rough equal proportion of victims and attacks occurring between unique ASes and unique IPs.

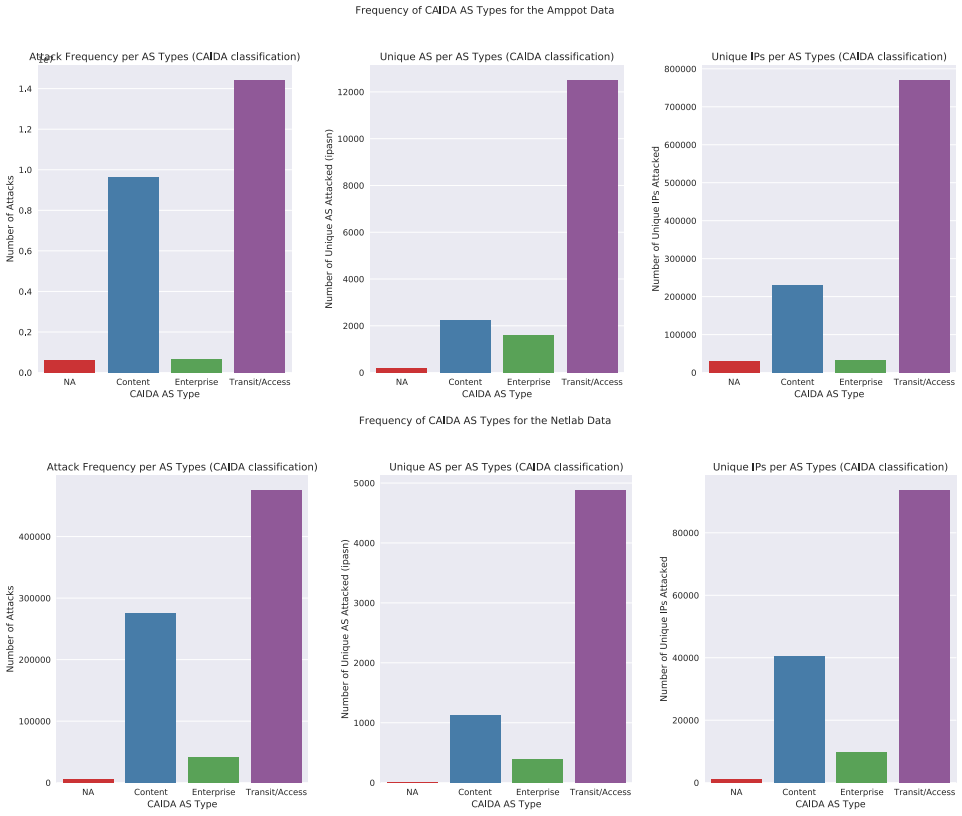


Figure 6.1: CAIDA classification victims

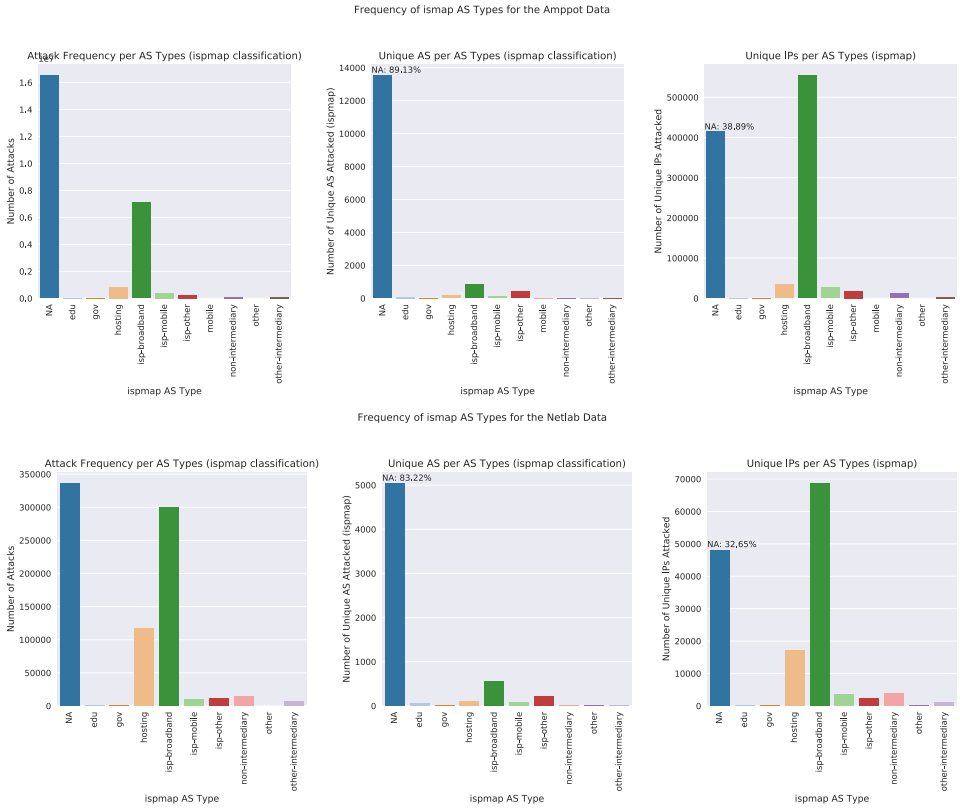


Figure 6.2: Ispmap classification victims

The CAIDA classification algorithm prefers to label an AS as Transit/Access quite easily, as a stratified manual of 100 ASes show. Of these there were 28 inconsistencies which is roughly similar to their aforementioned PPV. Of these inconsistencies 18 of them were classified as Transit/Access while they were not, respectively 14 and 4 for Enterprise and Content. Only 2 were mislabelled as Enterprise and should have been Transit/Access, and 1 AS changed from Content to Enterprise. It was not possible to manually classify the other ASes due to lack of information. For the distribution of CAIDA and ispmap AS types of the given sample, see Appendix G.

Figure 6.2 classify these ASes using the ispmap dataset. It should be noted that there are a lot of unknowns in this dataset, for amplification attacks 89.13% of the unique ASes and 38.89% of the unique IPs cannot be correctly labelled. Similar ratios appear for IoT data with 83.22% and 32.65% unknowns respectively. To understand this better, the manual classification of 2050 ASes provide high reliability in the data, however, misses significant portions of the victims. These unknowns form a significant portion of the data which will be explained in the next section.

From this particular subset, it shows that a small number of unique ASes harbor a

significant proportion of the victims under isp-broadbands. Broadband ISPs usually would have a large pool of IP addresses which can be targeted as each customer has either a static or dynamic IP.

With finer granularity differences become slightly more visible. These graphs also show that IoT attacks have a relative higher tendency to attack hosting compared to amplification data. While attack frequency and unique IP victims for hosting score relatively high, the number of unique hosting ASes attackers are not that significant.

Proportions between amplification and IoT attacks are again roughly similar, with significant p-values for the chi-squared test. Respective p-scores of $1.89e-28$ and 0 comparing ASes and unique IPs.

6.1.2. Missing ispmap data

The high number of missing classifications from the ispmap data is surprising, as the dataset has been used in previous research by Noroozian, Korczyński, *et al.* (2016) where they managed to classify most of the ASes found in the amplification data. The missing data limits certain interpretation of the data, however, given that they classify the same 2700 ASes the results can discuss the attack differences on these high-profile ASes.

Further investigation into the missing and non-missing ispmap data shows that there are some differences between amplification and IoT attacks. One way is to compare the CAIDA classifications of the missing and non-missing ispmap data. If this is similar it is possible to conclude that on a higher level the ispmap classifications contain a representative set of the population. If they are not, they also help reveal shortcomings of the dataset. The available Figures are in Appendix G.

For Transit/Access ispmap covers the same proportion as the non-classified ASes. But for both Content and Enterprise they are underrepresented as the missing classifications have a higher proportion in both. Looking at both attack types it becomes clear that the attack frequency and unique IP victims of Content victims are not clearly labelled with ispmap, for the IoT attacks there is also a high degree of Enterprise victims not being taken into account.

This is why it is important to include both sets of classifications. While CAIDA is able to label a very high percentage of victims they are often not as accurate and contain less details but they do give a big picture overview. The ispmap classification labels a smaller (selective) portion but gives higher granularity and accuracy.

6.2. ISP Broadband Providers

Broadband providers can provide multiple services to their customers, such as hosting capabilities. Previous research on broadband providers showed that attacks are on IP addresses that did not have any domains associated for 95% of the cases.

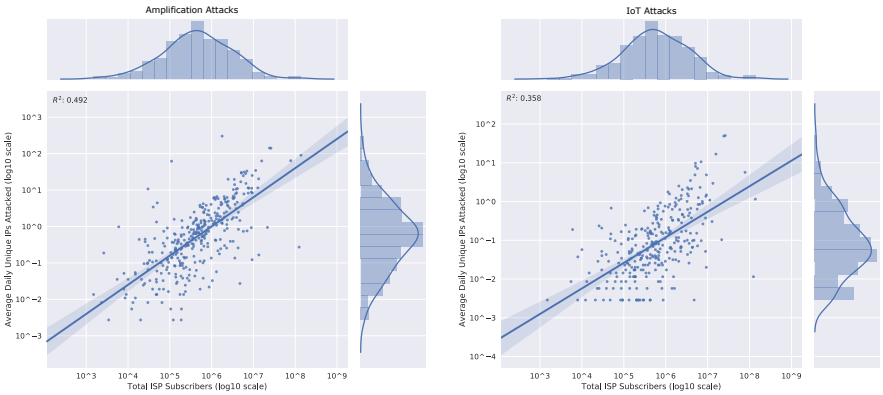


Figure 6.3: Correlation unique IP Victims with total ISP subscribers

(Noroozian, Korczyński, *et al.*, 2016). As most IP addresses are not associated with a domain it makes sense to conclude that the transit customers of an ISP are being attacked. These are often regular individuals making use of the internet.

6.2.1. Total ISP Subscribers

To track victimization on an ISP it is important to take the IP churn rate into account (Asghari, M. J. van Eeten, *et al.*, 2015). As ISPs often apply dynamic IP address allocation an attack on different IPs can still mean it is the same target leading to an overestimation of the number of victims. To combat this dynamic churn rate the metric average daily unique IPs attacked is calculated. The DHCP churn varies significantly among ISP networks meaning the 24 hour window used might not be perfect (Moura *et al.*, 2015). However, it reduces its effect significantly (Noroozian, Korczyński, *et al.*, 2016). An attack on each unique IP victim of an ISP is counted daily. These are then summarized and then divided by the number of days of the captured data to create the average of the daily unique IPs attacked. See Equation 6.1 where k is the number of days.

$$AverageDailyUniqueIPsAttacked = \frac{\sum_{i=1}^{k} DailyUniqueIPsAttacked}{NumberofDaysinAttackDataset} \quad (6.1)$$

The Figure 6.3 shows that the number of ISP subscribers and the victimization of an ISP are correlated with each other. The high coefficient of determination (R^2) explained in both amplification (0.490) and IoT (0.358) attacks adds more evidence that individual users are being targeted. This OLS regression is also significant for both with p being 0. The R^2 for both attack types is lower than previously reported by Noroozian, Korczyński, *et al.* (2016). In recent years the number of subscribers does not explain the majority variance in the victimization of ISPs. It could suggest

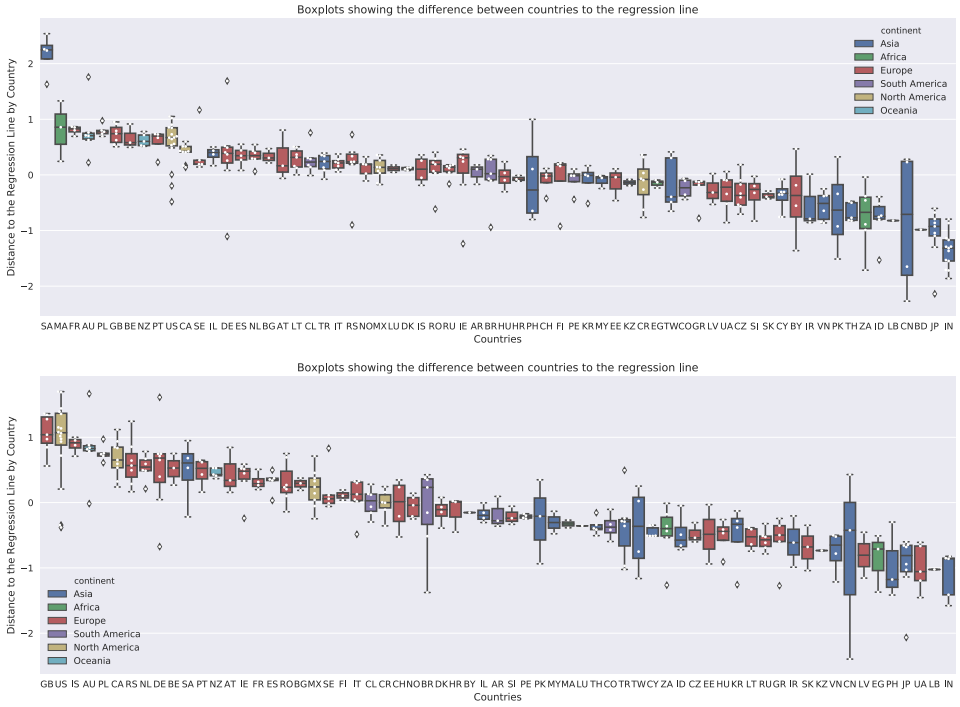


Figure 6.4: Broadband ISPs attacked grouped by country relative to the regression line

a shift to other factors playing a more important role. The same conclusion can be said towards the Netlab data as the coefficient of determination is relatively lower.

While the log scale helps reveal this correlation it also hides significant differences between ISPs. There still remains quite a big spread between certain ISPs which will be investigated further in the following sections.

6.2.2. Country Effects

Noroozian, Korczyński, *et al.* (2016) suggested country-level effects at play as there are differences in the economic status and infrastructure of each country. This analysis will be replicated here as well.

Figure 6.4 show the specific ISPs in each country and their distance to the regression line in Figure 6.3. This is calculated by $ActualValue - ExpectedValue$ if this is positive it means that they get relatively more victims than predicted with the subscriber data and vice versa.

This graph shows a global trend where broadband providers in (West) Europe, North America, and Oceania are usually targeted unevenly more given the number of customers (with the exception of Morocco). While Asian and Eastern-European countries are attacked less. It remains uncertain why these continental differences

Table 6.1: Negative binomial GLM regression models - Amplification

	Dependent variable:		
	# Victims per ISP		
	1	2	3
Subscribers (log10)	1.5779*** (0.095)	1.5130*** (0.095)	1.5608*** (0.099)
ICT Dev. Index		0.2624*** (0.048)	
GDP PPP per Capita			4.505e-05*** (3.59e-06)
Constant	-8.1307*** (0.571)	-9.6010*** (0.663)	-9.8724*** (0.638)
Residuals	335	334	334
Observations	337	337	337
Log Likelihood	-712.64	-702.83	-657.72
Note: *p<0.1; **p<0.05; ***p<0.01			

exist but they seem to often underperform given their ISP size. Factors related to culture, wealth, ISP services, but also gaming habits should be investigated to find the true cause.

The figure also shows interesting individual country differences. Notably, amplification attacks have been attacking Saudi-Arabia disproportionately. There might have been frequent attacks against the citizens of Saudi-Arabia, however, no notable campaign has been reported. As expected when the victims are most-likely individuals rather than high-profile organizations.

Furthermore, Taiwan, China, and the Philippines show some big differences inside the countries. The ISPs of other countries are clustered together or if there is variation it does not occur in both datasets. These variations in the country might suggest their broadband providers serve different demographics, or that for these specific countries the country-level effects are not at all that viable due to specific circumstances. While this remains uncertain due to lack of data of these ISPs, the other countries do seem to have country effects with the clustering.

6.2.3. Institutional Country Effects

These Generalized Linear Models (GLM) with ISP subscribers in log_10 , the ICT Development Index of 2017, and the GDP PPP per Capita 2019 in International $ are used to predict the number of victims per ISP in log_10. Therefore, it is important that the coefficients are read correctly with the use of log. For example, an increase of one unit in the ICT development index increases the number of victims by $e^{0}.2624 = 1.3$. For each dataset, three models are made. One with only ISP subscribers and the other two add either ICT Development Index or the GDP PPP per Capita. These are not added together as there is a strong correlation

Table 6.2: Negative binomial GLM regression models - IoT

	Dependent variable:		
	# Victims per ISP		
	1	2	3
Subscribers (log10)	2.3898*** (0.192)	2.1763*** (0.194)	2.1526*** (0.200)
ICT Dev. Index		0.6524*** (0.114)	
GDP PPP per Capita			4.887e-05*** (6.98e-06)
Constant	-15.4643*** (1.245)	-19.0303*** (1.622)	-16.1626*** (1.423)
Residuals	293	292	292
Observations	295	295	295
Log Likelihood	-232.31	-214.47	-206.57
Note: *p<0.1; **p<0.05; ***p<0.01			

between these two institutional factors.

The ICT development index looks at ICT access, use, and skills capturing a nation's ICT development. That they are a significant factor means that not only access (measured by number of subscribers) but also the general technical knowledge of the population is important. One way this reasoning makes sense is that the tech-savyness of a population makes victims more attractive. As they utilize and use ICT skills these victims might be more prevalent on the internet therefore increasing their victimization rate. Another explanation can be found in personal motivation for DDoS attacks as seen in 2.2, then it does make sense to incorporate reasoning why a higher ICT development index also makes them more capable of DDoS attacks: (1) higher awareness of DDoS attacks, (2) increase capabilities to utilize DDoS attacks or find commoditized ways to do, and (3) create motivation for attacks. But it remains important to acknowledge that the model only looks at increase in victims and not increase in attacks.

The GDP PPP per Capita is an economic factor of a country which can contribute to higher victims in different ways. Several explanations an be found, such as (1) higher time consumption in gaming and (2) increase in economic capabilities to purchase DDoS attacks from booter services. The gaming explanation is often used as most of the DDoS attacks on individuals occur because of gaming see 2.2.

The assumptions of these models are shown in Appendix G. There seems to be violation of several assumptions, such as independent and identically distributed resiual errors, normal distribution, and homoscedasticity. These violations seem to suggest that there are either missing explanatory variables which are not considered in the model or multicolinearity of factors. This would suggest more research should be done to take into consideration more variables to explain the variance in DDoS attacks on broadband ISPs.

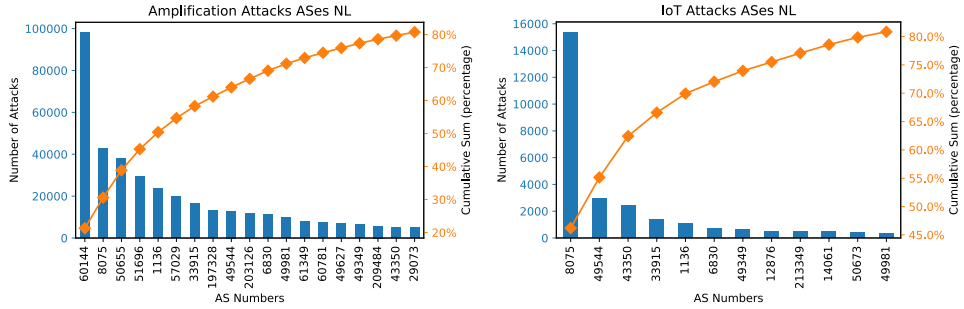


Figure 6.5: Hosting ASes attacked with the cumulative sum

As these factors are significant in the models, it becomes clear that the number of subscribers, ICT development index, and the GDP PPP per capita contribute to a higher victimization of a broadband provider. This seems to correspond with Figure 6.4.

6.3. Hosting Providers

This section discusses DDoS attacks happening on domains. Compared to the last section they provide more information on the type of victim and gives insight into the type of organizations which are attacked. This section also differs from other sections because it only discusses the Netherlands. This is because of the high manual labour required to look into this. The Netherlands is chosen because of the author's familiarity with the country and its strong digital infrastructure in Europe. The implications of this, however, is that the analysis of hosting providers cannot be applied to the entire dataset. These results are limited to the Netherlands which has their own unique characteristics. There are around 1200 registrars and 800 providers in the Netherlands alone (NBIP, 2020).

6.3.1. Victimization Pattern Netherlands

Firstly, the victimization pattern of the Netherlands will be quickly discussed again. As briefly discussed in Table 5.1 there are 351 and 202 unique ASes for AmpPot and Netlab data respectively. However, 80% of the attacks occur on a select few ASes. Figure 6.5 shows that 80% of amplification attacks are targeted on 19 ASes and it is 12 ASes for IoT attacks in the Netherlands. Respectively 11 and 4 of these ASes were added to the existing ispmap classification. See the Appendix G for these classifications.

The victimization pattern for the Netherlands is visualized in Figure 6.6. Despite only classifying 5% of the ASes in the country 80% of the attacks are classified. There are already some differences between the Dutch victimization and the global victimization patterns. While the global trend shows that the majority of victims

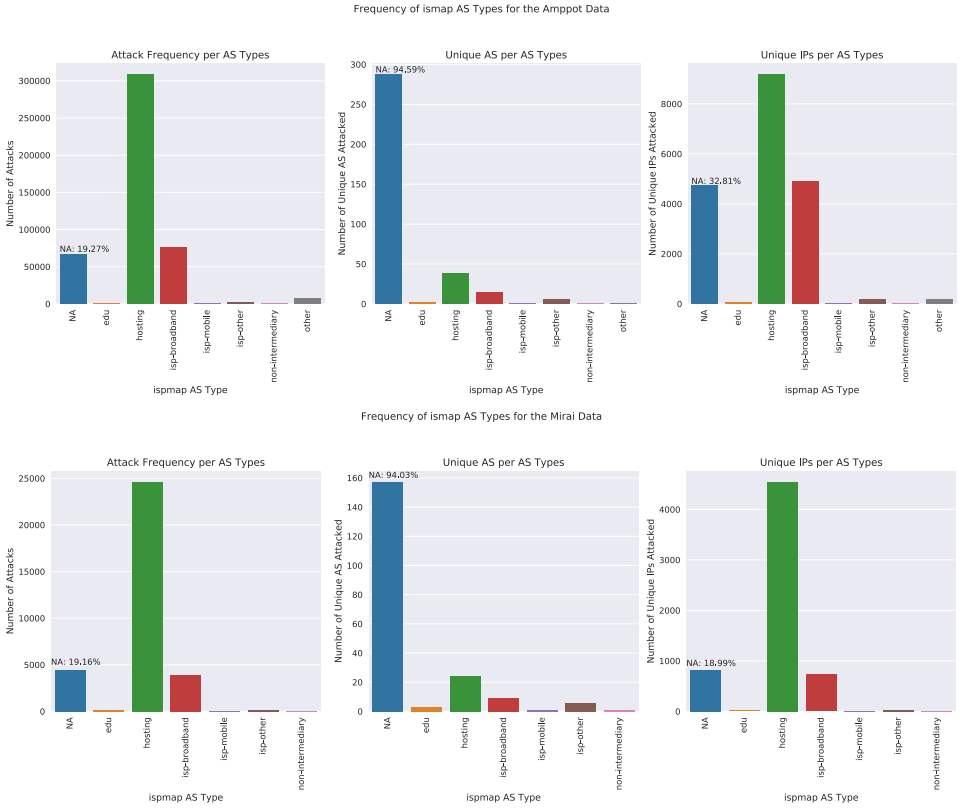


Figure 6.6: Ispmap classification victims in the Netherlands

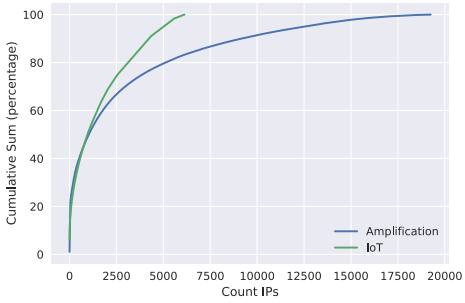


Figure 6.7: Cumulative sum over IPs attacked in the Netherlands

are end-users of ISPs who are attacked, the Netherlands seem to deal with attacks on hosting providers relatively more. Suggesting that DDoS attacks are targeting services rather than users in the Netherlands. This adds more evidence to the country-level differences of DDoS attacks and provides a good basis to analyze the hosted domains.

The Figure 6.7 shows the cumulative sum of the attack frequency per IP, it is comparable to the ASes in Figure 6.5. This trend shows that the concentration of attacks is not as bad compared to the ASes but it is still not equally distributed. For example, it might be interesting to notice if these attacks follow the Pareto principle. The attack types differ quite significantly from each other. For amplification 80% of the attacks can be explained by 5116 IPs (26.58%). IoT attacks need 3116 IPs (50.85%) to explain 80% of their attacks. Although one IP from a hosting provider is already responsible for 5% of the attacks. Statistically we can show this as the Kolmogorov–Smirnov two-sample test show that the p-value of the test is $4.85e-05$ and the test score is 0.20, rejecting the null hypothesis that these distributions are similar. While amplification attacks remain concentrated on a minority of the targets, IoT attacks spread their attacks more.

The Netherlands features a higher percentage of ISP broadband victims who own domains. Whereas previous research showed that only 5% of the (unique) IPs attacked have a domain, this research shows that 28.15% (amplification) and 16.69% (IoT) of unique IPs classified as isp-broadband host domains out of all the unique isp-broadband IPs. Respectively their attack frequency accounts for 32.14% and 25.48% of total attacks on ISP broadband victims.

6.3.2. Domains in the Netherlands

Figure 6.8 shows the frequency of the number of domains associated with an IP address. It is clear that IP addresses classified as ISP broadband are often for single (personal) domains while non-ISP providers are also able to use a single IP address for multiple domains. Usually these are large hosting providers who can map up to

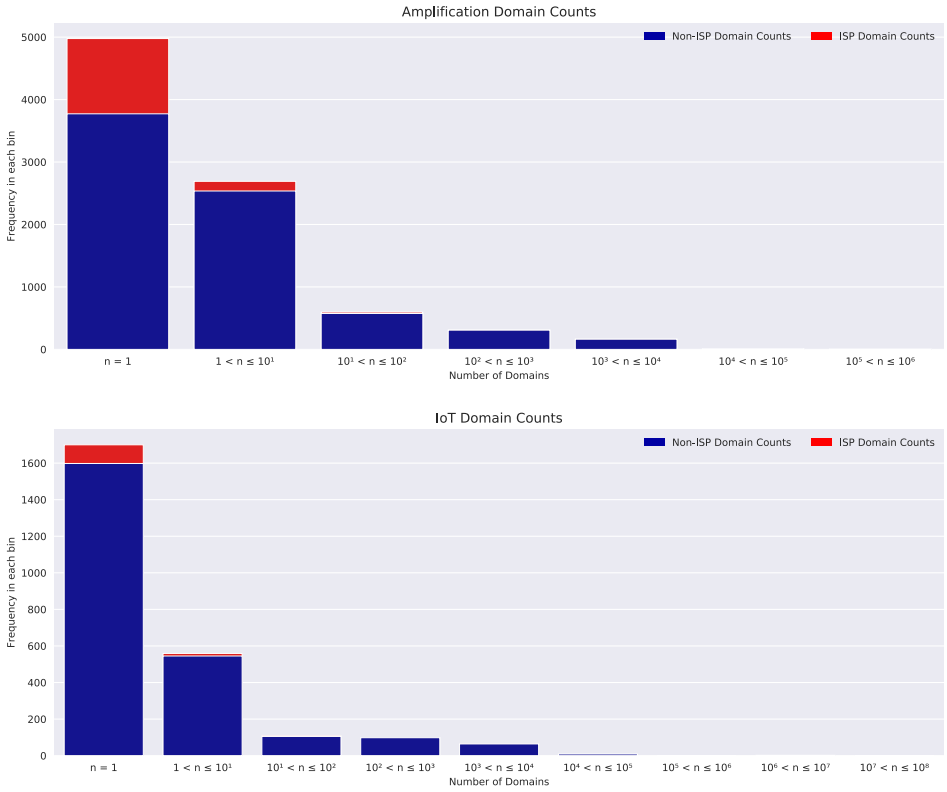


Figure 6.8: Number of domains associated with an IP address

47366607 domains for one IP in a single year. [1] While the majority of IP addresses for hosting providers are also single domains, 43.17% and 33.20% are some form of shared hosting for the amplification and IoT victims respectively.

For the next step, the manual classification of the top 100 attacked IPs for each attack set will be shown. Hosting takes up a significant majority of these attacks with 77 and 61 IPs of these belonging to an AS classified as hosting respectively for AmpPot and Netlab victims. However, there are only 61 IPs that are associated with a domain for amplification victims but 64 for IoT victims. Out of these, only 28 and 21 IP addresses have single domains respectively. The rest are shared domains where it becomes difficult to discuss which domain is being attacked.

Manual classification remains quite difficult as many of these websites currently lead to nowhere. Interestingly enough, a lot of these attacked domains are of short duration. The dataset does not necessarily reveal that established or big organizations face the majority of the attack, see Appendix G. Unfortunately the data only includes attack frequency on IP and not on domain. It still could be that they are attacked due to the size of the organization the attacks are split across their services. For example, victims of amplification attacks to domains of ICT infrastructure support and hosting providers are all split across numerous IPs in the top 100. Surprisingly there were also attacks against the domain of a popular ISP in the Netherlands rather than its user. The examples in the Mirai dataset feature even less of big organizations, they seem to mainly target small discord servers, illegal services (such as booter websites), high-school education, tourism, cloud servers, and political websites.

6.4. Conclusion

Looking into victimization into more detail these country-level effects become more clear. This section looked at the two victim types which are attacked a lot, namely broadband ISPs and with the growth of IoT attacks also hosting providers. The fact that individual users are being targeted remains of interest as it shows the workings of booter services. Furthermore, attacks on hosting providers can be explained by a few ASes, although Mirai attacks do spread their attacks more over IPs compared to amplification attacks. This pattern is only applicable for the Netherlands, as other countries were excluded in the hosting provider analysis. The Netherlands does differ from the global trend, as attacks on hosting providers are more common.

These patterns of victimization showcase how important it is to understand who the victims are of DDoS attacks. By comparing amplification and IoT-based attacks this study shows that the method of attack can influence who gets attacked. However, its causal influence is still difficult to determine as it could also be influenced by attack capture methodology. One plausible theory is that the investment of IoT-based attacks is still higher than amplification attacks at the moment, which is why they would prefer to spend their resources to attack high profile targets in hosting

[1] This one was mapped to a technology website in the Netherlands and their subsequent subdomains.

and western countries. Despite that, it is interesting that these attacks do differ suggesting that new organizations will be under the threat of DDoS attacks.

7

Discussion and Recommendations

This section discusses the limitations and the implications of the previous results. This research applied data science tools and techniques to investigate the impact of IoT on the commoditization of DDoS attacks. Furthermore, it highlighted the difference between amplification and IoT attacks in attack characteristics and victimization.

7.1. Limitations

This research sets out to explore the impact of IoT devices on the DDoS ecosystem. To understand the full impact socio-technical concepts were explored and investigated. However, quantification and measurement of these concepts remain a difficult area of research in the economics of cybersecurity. Several of these limitations will be discussed.

7.1.1. Data

The reliability and trust in the data used for this research is important for the validity of the results.

Firstly, datasets are heterogeneous in nature. DDoS attack data vary based on capture methodology, meaning that different data sources can give different conclusions of the DDoS problem. This is caused by the difficulty in capturing these attacks, as shown by the two different methodologies in this research. Amplification data was captured using honeypots, while the IoT data is captured by extracting C&C commands. Evaluation of the scope of their captured data and its representa-

tiveness for the whole is a difficult question to answer. For example, amplification attacks are only capturing a certain set of protocols while new protocols can be used for amplification as well. Given that the attacks change continuously, new measurement abilities will have to be created to capture new attacks. Previous research already established that industry reports and scientific articles come up with different conclusions due to different incentives (Noroozian, Korczyński, *et al.*, 2016). Despite that, to this researcher's knowledge this is the first comparison of the victimization of amplification and IoT attacks. The data is noisy and selective in what it captures, yet it is also exhaustive and should provide an approximation to the ground truth. It remains uncertain if the results from these datasets are conclusive for all DDoS attacks, However, it does contribute additional knowledge in order to have insight in the bigger picture.

Furthermore, the datasets used capture two different time periods. The comparisons are performed with a time delay of a year. It remains uncertain if certain patterns have changed because of the characteristics of IoT attacks or because the whole field has evolved in that year. Lack of research in this field makes it difficult to estimate the evolution of DDoS attacks as well. While industry reports publish yearly DDoS reports they often limit themselves to specific industries. Industry reports do mention an evolution in new attack vectors, related to new amplification protocols and vulnerabilities in IoT devices (Link11, 2020; KPN, 2021; Nexusguard, 2020). A further complication is that 2020 is the year where a global pandemic occurred due to *SARS-Cov-2*. A large population started working from home and the reliance on digital infrastructure only became bigger. These uncertainties in the evolution of DDoS with the changing digital landscape due to the pandemic creates a doubt if these comparisons are due to IoT or time. Nonetheless, these adaptations are likely to become part of its growth and are important to study. The DDoS landscape is very dynamic and it is quickly able to adapt to new attacks and defences.

Lastly, the data is appended with information from other data sources in order to add valuable knowledge about the victims. The multitude of different data sources on geolocation, AS classification, broadband subscribers, and hosting domains were important to be able to answer the research questions. Despite that, they are not without faults as there is no absolute mapping and classification of the internet that is up to date all the time. When possible this research appended missing information with manual classification. Information can therefore be out of date or wrong, however, the datasets are trusted by security practitioners and academics alike. For this research it is only possible to use the data available to it.

7.1.2. Research Approach

The lack of ground truth data, missing or unreliable information makes it even more important to discuss how the research is performed and its limitations.

Firstly, large scale data on the motivation of DDoS attacks does not exist. While limited studies have explored this space by interviewing the attackers, most studies

including this one have to infer the motivation through the attacks. The multi-causality issue as described by de Bruijne *et al.* (2017) shows that a wide variety of factors can affect DDoS attacks. It remains impossible to know the true reason for the motivation of an attack. Yet, quantitative data in this space can contribute to insights previously unseen. Victimization remains an important aspect to combat DDoS attacks on a large scale.

Secondly, topic modelling of the underground chat data is still a very novel and exploratory tool. While NLP has existed for a very long time, the extra computational power has seen a resurgence in new tools and methodologies previously unseen. The use of Doc2Vec, UMAP, and HBDSCAN comes with their own caveats. UMAP for example suffers from the the curse of high dimenstionality (Zimek *et al.*, 2012; Schubert and Gertz, 2017). The lack of reproducibility in NLP research makes it difficult to evaluate the given methodologies for different use cases (Lau and Baldwin, 2016; A. Rogers, 2021). Another issue is that evaluation metrics in topic modelling often do not take into account the human interpretation of these topics. While they can say things about coherence and overlap of topics, it is hard to evaluate topics based on interpretability.

Thirdly, as computational power and storage remains an issue while dealing with big data the research had to be scoped down to the Netherlands for domain victims. This means that a country perspective was given rather than the global one. As seen in the Results section there are country specific factors at play which means these results cannot be representative for the entire attack set.

7.2. Implications for Policy Recommendations

The limitations in the data and the research approach looked at the results with a critical perspective. These boundaries need to be taken into account for the policy recommendations. These results are important for stakeholders who are invested in creating a reliable internet, such as policymakers, telecom, digital infrastructure companies, law enforcement but also enterprises and victims in general. This section will discuss the results for the sub-question: *"What policy measures can be taken to reduce the impact of DDoS attacks?"*.

This research is very exploratory in nature due to the limitations of the data and methodology. Yet, new key insights are developed which can help make another step in the right direction. The impact of IoT on DDoS attacks is not only relevant for now but will increasingly become more important with its growth. Continued research and development is necessary to create a safe and reliable internet as more and more services depend on it.

7.2.1. Underground Market

This is one of the first few research papers on the impact of IoT but also on its impact in the criminal underworld. Given access to the open communication of DDoS-as-

a-Service communities shows that IoT is used in combination with amplification attacks.

This research provides evidence of buying and selling of IoT attacks in DDoS-as-a-Service chat data. The implication of this is that IoT attacks are already commoditized. This means that it is viable for criminals to develop, sell or rent out their infrastructure. While also showing that they have not fully taken over the market as there remains a strong preference for the stability and ease of amplification booter services. However, his commoditization could mean that its development and growth will continue to grow as there is a viable profit incentive attached to it. Policy should focus on creating safety standards and personal device hygiene awareness in order to prevent a growing number of IoT devices to be infected.

These underground markets are an important avenue for law enforcement and cybersecurity experts to keep track of new developments. Not only do they show the buying and selling of infrastructure and tools, they also gives an unique opportunity to find out how cybercriminals think and operate. The problem remains that there is a lot of noise creating difficulties in finding high entropy information. This research shows that most of these communities exist in short duration and there is a concentration of activity around a few communities. While law enforcement could take these popular communities down, the ease of setting up new Discord communities would make this irrelevant. This would also suggest that the freedom Discord gives its user could lead to malicious and unwarranted behaviour on their platform. As Discord is only one form of communication it would reason that heavy censoring from Discord would make these communities flee to other applications, such as Telegram. There is most-likely more value in tracking and understanding these communities instead of taking them down. Policies based on communicating that Discord is willing to share personal information in case of malicious behaviour might act as a scare tactic but research often shows that people are not aware they are committing a crime.

Furthermore, this research showed the power of natural language processing in identifying relevant information. Machine learning is able to go identify relevant information for law enforcement and researchers alike. Creating tools and frameworks focussing in this domain can benefit our understanding of the world of crime.

7.2.2. Attack Characteristics

Regarding the attack characteristics of amplification and IoT attacks, this will have implications for the technical defensive measures. Previous research showed quite a few differences in single and multi-attack vector attacks, degree of automation, and source address validity. These differences were already quite well-established when IoT attacks became popular. However, empirical validation comparing amplification attacks with IoT attacks is to this author's knowledge new information.

Firstly, IoT attack methods seems to either (1) deal with more unreliability and/or (2) is more difficult to track due to the constant evolution of the variant. The first

probability actually reveals a weakness of IoT and it could explain why previous research on booter sites have not seen IoT attacks as a popular choice if at all. Booter sites would need reliable services in order to serve their customers, as IoT botnets are frequently in a turf war with each other and remediation of devices can affect their attack power making them less reliable in consistent output. Remediation is becoming a key interest of ISPs in order to prevent the spread (Bouwmeester, 2020). New variants pop up creating new methodologies for infecting old and new devices replacing other variants. The second implication is that the tracking of IoT attacks need to be improved in a way that new variants will be tracked as well. This means that there remains a barrier to commoditization of IoT attacks, however, the quick growth and creation of new variants could reduce this barrier in the future. It will be important to develop methodologies tracking IoT attacks accurately.

Secondly, the development of an accurate representation of attacks is also important for amplification attacks. This research also shows a discrepancy between this attack set and other industry reports. This fragmentation is essentially worrying, as policymakers and financial budgets could focus on the wrong issue at hand. Every report should include the level of uncertainty in their analysis due to their methodology.

Thirdly, it is interesting that there are some differences between amplification and IoT attacks: (1) there is a change in attack distribution on a continental level where North America and Europe are targeted relatively more, (2) attack durations differ significantly from each other as IoT attacks last slightly longer, and (3) while broadband users remain the majority victim IoT also favours attacking hosting providers. As discussed before, it is difficult to pinpoint if this is due to a methodology issue or due to an actual change in victim pattern. Nonetheless, these trends are important to consider. An explanation could be that commoditizing IoT attacks remains more difficult than amplification attacks resulting that their victims are preferable of similar statue to the amount of effort they have put in. As it remains, IoT attacks can also be in the hands of more advanced attackers who prefer higher profile targets. This means that with an increase in IoT attacks a different set of victims could become the target who might be unprepared for it. Preparing hosting providers for IoT might become a higher priority, similar to how certain countries should prepare more than others (or it could be reasoned differently that countries not dealing with IoT should look at the current countries and learn from them).

7.2.3. Victimization in Detail

Victimization of DDoS attacks in amplification and IoT attacks remains an underrepresented area of research. To be able to stay ahead in this arms race it is important to understand who are being attacked and how best policy can fit into that.

The paper of Noroozian, Korczyński, *et al.* (2016) revealed victimization patterns where the majority of the users attacked are in broadband ISPs. This research can confirm that this is still the case, meaning its conclusion about the democratization

of victimhood due to commoditization still plays an important issue. The cheap offerings of booter services allow attackers to go after anyone as there is no need for a return on investment. A key area for policy would be to include measures to protect the individual user as well. Putting more effort into protecting the individual IP addresses could be a key message, such as individual protection through VPNs or obfuscation by applications.

The results show clear country level differences even when corrected for ISP size. A GLM regression model show that certain institutional factors such as the ICT development index and GDP PPP Per Capita can explain some of this but they are not very strong. One hypothesis is that attacker and victims are geographically concentrated near each other as attacks are often becoming more personal or driven by gaming where you usually play on servers near you (Noroozian, Korczyński, *et al.*, 2016). What we do see are continental differences where the Asian continent sits under the expected attacks given their ISP size (with the exception of some Middle East countries). Despite DDoS attacks being a globalized problem the effects are very different on a country level. These differences have to be taken into account when sharing information with each other as policies would have to be individually applied.

This difference can be seen as the Netherlands deals with attacks on hosting relatively more than victims in ISP broadband suggesting a shift from the global perspective. An interesting difference between amplification and IoT attacks is the diversity of attacks. Amplification attacks seem to concentrate their attacks on a smaller portion while IoT targets various victims. This can open the attack frequency up to new organizations who did not have to deal with this problem before. A lot of organizations have already included or outsourced measures against DDoS attacks as it is important for hosting providers to be reliable. This trend shows that the risk could increase in the future and adjustments could be made on organizational level.

7.2.4. Key Takeaways for Policy

The implications for each results chapter have been discussed in the previous subsections. This will further discuss the implications in a holistic manner.

Firstly, as the commoditization of IoT-based DDoS attacks professionalize this tool more the policy measures will have to adapt. The pool of vulnerabilities includes more stakeholders who need to be included for society's mitigation. The basis of this is that there a lot of vulnerable IoT devices who do not have the necessary protection. And this will increase as IoT devices are now not only are they being taken over, their protocols are also abused for amplification (Cimpanu, 2020). Underground markets form an effective way to track these developments to create early mitigation. But more importantly, this is just another part of the ongoing arms race that have existed since the internet was created. While there is money to be made, it will continue to develop.

Policy measures which tracks the methods used or disrupt communication platforms are important but ultimately short-term solutions. While it is IoT and Discord now, the field contains much more and they have shown to be quite adaptable. Policy recommendations based on disrupting the market, such as removing the demand for DDoS or removing the financial incentives for the supplier would be much more effective in the long term. While previous research showed that blocking PayPal transactions, disrupting their infrastructure, and awareness campaigns were effective (Karami, Park, *et al.*, 2016; Santanna *et al.*, 2017; Collier, D. R. Thomas, *et al.*, 2019). The financial transactions and the back-end infrastructure has changed as IoT attacks do not require the standard booter infrastructure topology as seen in Figure 2.6. Disrupting operators in underground markets is much more difficult as they are able to hide their details until a private message is made, such as their financial accounts and infrastructure. This does mean that they cannot scale very effectively and only reach consumers who put in a little bit more effort. Creating trust and getting consumers is difficult for small marketplaces.

While this research does not show a displacement of 'standard' booter services commoditization of IoT can change the technical supply of the market. Which is why measures targeting consumers might be more robust, as they are the constant factor. However, the consumers persistence thoughts that it is not a serious crime, there is low levels of harm, and is legal (Collier, D. R. Thomas, *et al.*, 2019) makes this a difficult avenue.

Unfortunately, there is still a lot that we do not know about. For example, how is the infrastructure of commoditized IoT-based DDoS attack setup precisely? What would the differences be? What are the most effective ways to change the persistence thoughts of the consumers? But what we do know is that become more and more relevant in the battle against DDoS attacks.

Secondly, the victimization patterns does seem to show that to be the case where individual users get attacked more, meaning it still seems very localized. Additionally, this can also mean that the victims already have a high enough security defence which is why purchasers do not want to spend money on something they cannot take down. Hypothetically, while you could attack a big website that simply stays up the instant gratification of seeing a victim disconnect in a game is worth the money more. It can be off-putting when they do not manage to take down those enterprise or hosting providers. However, small victims might still suffer. Leaving the takedowns of big services to the professionals and not commonly purchasable. We would also see more disrupted web services if this were the case.

What this means is that policy measures might not need to be focussed on hosting and enterprises as their incentive already pushes them to find the measures necessary for their defence. However, this might be made more effective with coordination and shared defensive capabilities. Yet, a group overrepresented as victims who do not seem to have these capabilities are the individual users being attacked in Transit/Access networks. While this might seem to have low consequences, it allows booter services to keep operating and continue to spend time on improving

their attack vectors. Eliminating or improving the defensive capability of individuals users getting attacked might be beneficial in disrupting the market. However, it also proves a much more difficult task as they have less control or incentive to prepare for this. Targeting applications to prevent individual users from sniffing out the IP addresses of others would be one step.

Thirdly, the shift in victimization pattern is noticeable that IoT-based attacks can be seen attacking hosting providers more, also spreading their attacks over more victims. One hypothesis offered is that the expertise required for IoT-based attacks are higher therefore they want victims with bigger consequences. Another would be that these victims are not capable of dealing with the different attack characteristics giving IoT-based DDoS attacks instant gratification. Much is uncertain, however, that there are different victims being attacked is likely.

Yet, the limitations in the data prevent us from emboldening this claim. The current industry and research datasets are very much scattered and are selective in what they look at. For evidence-based decisions to be made you need to have validity and trust in your data. There is still much that is not clear. It is important that there is a shared repository of attack data available. Not only is this a very difficult task to implement due to the issues in capturing attacks it also requires the work of multiple parties together. You need data from the victims as well as the attackers. These conclusions can only be made about the data used and not for the population.

Despite that, the implications are significant. It means that the avenue used for an attack can be correlated with a type of victim. Fragmenting the notion that DDoS attacks are uniform and need to be researched independently of each other. Not only that, victims can require different needs and help. It might become a policy question where resources should be spent to protect whom, for example, should IoT vulnerabilities or amplification vulnerabilities be targeted more? While this research cannot give that answer, there needs to be more research on this issue.

7.3. Conclusion

This section discussed the limitations and the implications of the results found in the previous chapter. First, limitations in the data used creates uncertainty in the validity of the results. Some of these were due to issues related to how DDoS attacks are tracked and captured, where it might not be the most flexible solution in this dynamic landscape. The difficulty in having ground truth data related to the victims remains a challenge. Inferring what has happened based on proxies could lead to wrong conclusions. Furthermore, limitations in computational power and the lack of evaluation for topic modelling outside of its interpretability makes it difficult to process and analyze data as a whole. Second, these results have a lot of implications for practitioners as well as academics. Reliable and accurate representations of the DDoS attacks is an important step in achieving and validating warning signs of new

emerging attack threats. Continental differences and increase in attacks on hosting providers due to IoT attacks remain worrying. While top hosting providers already know how to deal with DDoS attacks it remains unclear if everyone is capable of this when the attacks are more spread out. The commoditization of IoT attacks should be a warning that it will continue to develop and grow as there is a profit incentive.

8

Conclusion

The objective of this study is to understand how vulnerable IoT devices affect DDoS attacks. Along the way it dived into the underground markets, attack characteristics, victimization patterns, and its policy implications. This section will give concluding remarks on the main research question **"What patterns of commoditization and victimization can we observe with IoT-based DDoS attacks compared to amplification attacks?"**. The study processed an enormous amount of data to answer these questions with the help of data science.

This chapter will give a recap of each research question in detail and summarize the answer to the main objective. Furthermore, the scientific and societal contribution of this research will be discussed. Lastly, recommendations will be given for future work.

8.1. Research Questions

To answer the main research objective the research is split into four different sub-questions which answer the key aspects. This research has looked at both the technical as well as the social aspects of the DDoS attacks.

In order to answer the main research question a thorough literature review set the context and definition of core concepts used for the rest of the research. First, it explained what DDoS attacks are and how difficult they are to capture through the myriad of complexities and variety they are able to show. In short, there are fundamental differences in their operation, such as their attack vectors and application by DDoS-as-a-Service, which is why it is important to see the effects of these differences. Secondly, the motivation for DDoS attacks are extensively discussed. Frameworks of criminology in cybercrime, previous research on DDoS attack motivation, and the danger of commoditization of cybercrime shows that it is

important to understand the workings of attackers. To mitigate the threat against our ICT infrastructure it is not only necessary to implement technical defensive measures but also how to stop them from happening in the first place.

Firstly, the research question *"Are IoT-based attacks being commoditized by DDoS-as-a-Service groups?"* gives an overview of the impact of IoT as a commoditized tool in the underground market. While Discord remains a platform for many legitimate communities, underground markets have become quite popular on the chat platform. The convenience of creating and monitoring a community has lowered the bar for underground markets to be made there. They often exist for a short duration (shorter than a year) and popular ones are seem to include strong network effects drawing most activity towards them. As their size might make it difficult for law enforcement and researchers to go through them, the use of natural language processing and machine learning can be helpful tools in finding relevant information. Clustering based on topics show that IoT attacks are commoditized in these communities. The buying and selling of IoT attacks mean that its development will be similar to that of a business where growth is key. Messages tagged as IoT proportional to DDoS show that they are not far behind in terms of discussion.

Secondly, the research question *"What are the differences in attack characteristics of amplification and IoT-based DDoS attacks?"* explores the technical dimensions of these attacks. Giving an overview of the attacks occurring, the protocols used, country comparisons, and attack duration. It shows that there are differences in amplification and IoT attacks. However, it remains difficult to pinpoint these differences due to technical shifts or the fact that the data only collects samples of the real world. Changes in attacks however remain a reliable threat and it shows how important accurate representations of the issue is. These differences can lead to a shift in structural defence measures.

Thirdly, the research question *"What is the pattern of victimization of amplification and IoT-based DDoS attacks?"* looks at the impact of the emerging threat on the social side. Its implication shows that victims focussed in Transit/Access, and more specifically ISP broadband, are dealing with DDoS attacks more. Suggesting that individuals are targeted a lot by DDoS attacks, which is one of the consequences of the commoditization of DDoS attacks as in agreement with the previous research by Noroozian, Korczyński, *et al.* (2016). The motives of an attacker seem to be aimed more at individual grievances, or as commonly hypothesized creating a competitive advantage when gaming. However, it is clear that IoT attacks show relatively more eagerness to attack content-related or hosting victims as well. Despite that, it is also visible that country-level effects are at play. Certain ISPs in continents such as (West) Europe and North America are affected more by DDoS attacks on their users even when corrected for size. Giving reason to suggest that policies should be adjusted on a country basis. One of these countries that break the global trend seem to be the Netherlands where hosting victims get relatively more attacked than individuals. Individual analysis on the victims of hosting providers has only been run for the Netherlands due to computing limitations. These attacks are mostly concentrated on popular hosting solutions, as a few ASes them can already explain

80% of the attacks. Looking at IPs and domains it seems that the spread seem to differ between these two attacking types. Whereas amplification attacks are still roughly concentrated, IoT attacks spread their attacks across several IPs. These IPs are also significantly related to shared hosting where one IP can cover many multiple domains. However, manual investigation of the attack domains revealed that it was mostly small websites being attacked.

Fourthly, the research question *"What policy measures can be taken to reduce the impact of DDoS attacks?"* is answered in section 7.2 where the implications of the previous results are discussed. The fast growth of vulnerable IoT devices coupled with different attack characteristics and victimization pattern can mean that previous underrepresented aspects should come more to the forefront. There needs to be more development in creating accurate representations of the DDoS landscape, one preferably not biased towards industry which can keep track of the fast developments of DDoS attacks. Concretely, this will help preparation and policymaking for the future. IoT attacks do have differing attack and victimization patterns compared to amplification attacks. Remaining vigilant of these developments will be a key issue in protecting country specific assets. This research shows that keeping track of development in the underground markets can help in discovering potential new attack vectors.

These results show the bigger implications for the field. Firstly, commoditization of IoT-based DDoS attacks is just part of the arms race between attackers and defenders. Previous interventions such as disrupting their infrastructure and money flow need to be adopted to these different methods. But more importantly, focussing on a constant factor such as the demand for these attacks might be more beneficial. As we know the supply side can and will adopt as long as there is a financial incentive. Secondly, individual users in Transit/Access remain a majority of the victims. Attackers targeting these victims help support the booter services and the underground markets. While it is much more difficult to protect these individuals there could be more gain for the removal of the financial motive. Thirdly, IoT-based DDoS attacks do give a different victimization pattern. Yet, it proves quite difficult to generalize the claims of the results to the general population. Data collection needs to become a top priority if you want evidence-based policy making. But the implications of this are that DDoS attack techniques are independent of each other how they operate and who they attack. Focussing on the policy question, it can help prioritize how resources should be spent on mitigating a DDoS attack.

Lastly, utilizing a holistic view from the knowledge gained by the sub-questions the main research question *"How has the misuse of vulnerable IoT devices towards launching DDoS attacks impacted victims?"* can be answered within the scope of this research. There is definitely a trend in the commoditization of IoT attacks, as they are popular topics in the underground communities related to DDoS-as-a-Service. This will have consequences for the type of DDoS attacks society will have to deal with as they differ from the previously highly commoditized amplification attacks. Vulnerable IoT devices are currently already being commoditized and an increase of them in the future could lead to more incentive for abuse.

8.2. Academic and Societal Contribution

This research adds to the growing evidence that cybersecurity is inherently a human problem as well as a technical problem. By combining these views into a socio-technical perspective this paper can contribute to the current state of knowledge in both academic as well as societal worlds.

Firstly, research into IoT-enabled DDoS attacks have namely focussed on the technical implementations rather than the societal impact. While other research have also looked at the supply chain of IoT vendors to mitigate vulnerabilities, a comparative study on the impact of IoT attacks with amplification attacks remains novel to this author's knowledge. It added new insight into IoT attacks and showcased areas of improvement for both academics as policymakers to reign in the dangers of vulnerable IoT devices.

Secondly, this is also the first paper to this author's knowledge that uses Discord chat data from booter services to determine how prevalent IoT is. Underground markets remain a high potential source of information for the workings and operation of cybercriminals. This research showcased that the use of data science can benefit researchers and practitioners alike to utilize this potential.

Thirdly, this research adds a holistic view of a problem that is often separated by domain expertise. Understanding the effects of emerging technology in IoT, in relation to the commoditized world of cybercrime where the technical and societal impact of it is being researched plays a key role in better understanding of complex problems. Hopefully this research showed its strength in providing a broad perspective of the problem which should help the academic and society at large.

8.3. Recommendations for Future Work

This study is aware of its faults as described in the limitations (section 7.1) and will therefore recommend pathways which can be used for future work.

First, the importance of studies investigating the development cannot be understated in a quickly changing field. The results that are relevant now might change in the future. Depending on the methodologies and the data used new insights can be created which are relevant then. This study discussed the issues with heterogeneous datasets and how we had to work with small pieces of the puzzle. Researching more robust and adaptable gathering attack data methodologies will improve our understanding of the total landscape.

Second, there are answers that are still unanswered due to the lack of time or data when we look at the victimization patterns of these attacks. The country-level differences do not seem to be fully explained by the institutional factors mentioned in this research. Other factors to include, such as gaming, knowledge of DDoS attacks, or internet culture could be interesting. The lack of ground truth data remains a problem where we can only infer from other variables that need to be

discovered.

Third, it is important to find out how commoditized IoT attacks are setup. Scraping relevant messages to discover the price and power of IoT attacks for example could help quantify the problem. Unfortunately, most of these transactions occur in private chatrooms which are unavailable to third parties. So questions also remain in how often these advertisements get called upon on. Much is still unclear on their exact operation, such as how access is given or its reliability issues cleared.

9

Bibliography

Abhishta, A., Junger, M., Joosten, R., and Nieuwenhuis, L. J. M. (2020). "A Note on Analysing the Attacker Aims Behind DDoS Attacks". In: *Intelligent Distributed Computing XIII*. Ed. by I. Kotenko, C. Badica, V. Desnitsky, D. El Baz, and M. Ivanovic. Studies in Computational Intelligence. Cham: Springer International Publishing, pp. 255–265. ISBN: 978-3-030-32258-8. DOI: 10.1007/978-3-030-32258-8_30.

Akamai (2017). *Akamai' s [State of the Internet] / Security. Q1 2017 Executive Summary*. URL: https://www.akamai.com/us/en/multimedia/documents/state-of-the-internet/q1-2017-state-of-the-internet-security-executive-summary.pdf (visited on 08/17/2020).

Alami, S. and Elbeqqali, O. (Oct. 2015). "Cybercrime Profiling: Text Mining Techniques to Detect and Predict Criminal Activities in Microblog Posts". In: *2015 10th International Conference on Intelligent Systems: Theories and Applications (SITA)*. 2015 10th International Conference on Intelligent Systems: Theories and Applications (SITA), pp. 1–5. DOI: 10.1109/SITA.2015.7358435.

An, J. and Kim, H. (2018). "A Data Analytics Approach to the Cybercrime Underground Economy". In: *IEEE Access* 6, pp. 26636–26652. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2018.2831667.

Anderson, R. (2001). "Why Information Security Is Hard-an Economic Perspective". In: *Seventeenth Annual Computer Security Applications Conference*. IEEE, pp. 358–365. ISBN: 0-7695-1405-7.

Angelov, D. (Aug. 19, 2020). *Top2Vec: Distributed Representations of Topics*. arXiv: 2008.09470 [cs, stat]. URL: http://arxiv.org/abs/2008.09470 (visited on 01/06/2021).

Antonakakis, M. *et al.* (2017). "Understanding the Mirai Botnet". In: p. 18.

Applied Internet Data, C. C. for (2020). *The CAIDA UCSD AS Classification Dataset*. CAIDA. URL: https://www.caida.org/data/as-classification/index.xml (visited on 12/20/2020).

Asghari, H., van Eeten, M., and Bauer, J. M. (2016). "Economics of Cybersecurity". In: *Handbook on the Economics of the Internet*. Edward Elgar Publishing.

Asghari, H., van Eeten, M. J., and Bauer, J. M. (Sept. 2015). "Economics of Fighting Botnets: Lessons from a Decade of Mitigation". In: *IEEE Security Privacy* 13.5, pp. 16–23. ISSN: 1558-4046. DOI: 10.1109/MSP.2015.110.

Ashton, K. (2009). "That 'Internet of Things' Thing". In: *RFID journal* 22.7, pp. 97–114.

Atzori, L., Iera, A., and Morabito, G. (Oct. 28, 2010). "The Internet of Things: A Survey". In: *Computer Networks* 54.15, pp. 2787–2805. ISSN: 1389-1286. DOI: 10.1016/j.comnet.2010.05.010.

AWS Shield (2020). *Threat Landscape Report –Q1 2020*. URL: https://aws-shield-tlr.s3.amazonaws.com/2020-Q1_AWS_Shield_TLR.pdf (visited on 08/17/2020).

Bland, J. M. and Altman, D. G. (2004). "The Logrank Test". In: *Bmj* 328.7447, p. 1073.

Bouwmeester, B. J. (2020). "A Visit to the Crime Scene: Monitoring End-Users during the Remediation Process of Mirai Infected Internet of Things Devices". In: URL: https://repository.tudelft.nl/islandora/object/uuid%3Ae4226202-2e1b-4d5f-8c71-8092dde562dc (visited on 01/07/2021).

Brewster, T. (Jan. 29, 2019). *Discord: The $2 Billion Gamer's Paradise Coming To Terms With Data Thieves, Child Groomers And FBI Investigators*. Forbes. URL: https://www.forbes.com/sites/thomasbrewster/2019/01/29/discord-the-2-billion-gamers-paradise-coming-to-terms-with-data-thieves-child-groomers-and-fbi-investigators/ (visited on 01/05/2021).

Broadband Commission (2019). "State of Broadband Report 2019". In: *Geneva: International Telecommunication Union and United Nations Educational, Scientific and Cultural Organization*, p. 148.

Büscher, A. and Holz, T. (2012). "Tracking DDoS Attacks: Insights into the Business of Disrupting the Web". In: *5th {USENIX Workshop on Large-Scale Exploits and Emergent Threats ({LEET 12)*.

Campello, R. J., Moulavi, D., and Sander, J. (2013). "Density-Based Clustering Based on Hierarchical Density Estimates". In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, pp. 160–172.

Cayubit, R. F. O., Rebolledo, K. M., Kintanar, R. G. A., Pastores, A. G., Santiago, A. J. A., and Valles, P. B. V. (Dec. 1, 2017). "A Cyber Phenomenon: A Q-Analysis on the Motivation of Computer Hackers". In: *Psychological Studies* 62.4, pp. 386–394. ISSN: 0974-9861. DOI: 10.1007/s12646-017-0423-9.

Cetin, O., Ganan, C., Altena, L., Kasama, T., Inoue, D., Tamiya, K., Tie, Y., Yoshioka, K., and van Eeten, M. (2019). "Cleaning Up the Internet of Evil Things: Real-World Evidence on ISP and Consumer Efforts to Remove Mirai". In: *Proceedings 2019 Network and Distributed System Security Symposium*. Network and Distributed System Security Symposium. San Diego, CA: Internet Society. ISBN: 978-1-891562-55-6. DOI: 10.14722/ndss.2019.23438.

Cheung, R. (2017). "Targeting Financial Organisations with DDoS: A Multi-Sided Perspective: Comparing Patterns in AmpPot Data to Experts View on Target Selection in the Financial Sector". In: URL: https://repository.tudelft.nl/islandora/object/uuid%3Acd5fd5ea-ec4e-4cf4-bb8c-2fc4c84f583a (visited on 12/09/2019).

Chowdhary, K. R. (2020). "Natural Language Processing". In: *Fundamentals of Artificial Intelligence*. Springer, pp. 603–649.

Chromik, J. J., Santanna, J. J., Sperotto, A., and Pras, A. (2015). "Booter Websites Characterization: Towards a List of Threats". In: *Brazilian Symposium on Computer Networks and Distributed Systems (SBRC)*.

Cimpanu, C. (2020). *FBI Warns of New DDoS Attack Vectors: CoAP, WS-DD, ARMS, and Jenkins*. ZDNet. URL: https://www.zdnet.com/article/fbi-warns-of-new-ddos-attack-vectors-coap-ws-dd-arms-and-jenkins/ (visited on 02/15/2021).

Cloudflare (2021). *Famous DDoS Attacks | Biggest DDoS Attacks*. Cloudflare. URL: https://www.cloudflare.com/learning/ddos/famous-ddos-attacks/ (visited on 01/12/2021).

Cohen, L. E. and Felson, M. (1979). "Social Change and Crime Rate Trends: A Routine Activity Approach". In: *American sociological review*, pp. 588–608.

Collier, B., Clayton, R., Hutchings, A., and Thomas, D. R. (2020). "Cybercrime Is (Often) Boring: Maintaining the Infrastructure of Cybercrime Economies". In:

Collier, B., Thomas, D. R., Clayton, R., and Hutchings, A. (2019). "Booting the Booters: Evaluating the Effects of Police Interventions in the Market for Denial-of-Service Attacks". In: *Proceedings of the Internet Measurement Conference on - IMC '19*. Amsterdam, Netherlands: ACM Press, pp. 50–64. ISBN: 978-1-4503-6948-0. DOI: 10.1145/3355369.3355592.

Commission, E. (Apr. 27, 2017). *Europe's Digital Progress Report 2017*. Text. URL: https://ec.europa.eu/digital-single-market/en/news/europes-digital-progress-report-2017 (visited on 12/15/2020).

Czyz, J., Kallitsis, M., Gharaibeh, M., Papadopoulos, C., Bailey, M., and Karir, M. (2014). "Taming the 800 Pound Gorilla: The Rise and Decline of NTP DDoS Attacks". In: *Proceedings of the 2014 Conference on Internet Measurement Conference*. ACM, pp. 435–448. ISBN: 1-4503-3213-7.

D'Agostino, R. B. (1986). *Goodness-of-Fit-Techniques*. Vol. 68. CRC press. ISBN: 0-8247-7487-6.

DDoS Quick Guide (2020). US Cert, p. 5. URL: https://us-cert.cisa.gov/sites/default/files/publications/DDoS%20Quick%20Guide.pdf (visited on 12/08/2020).

De Donno, M., Dragoni, N., Giaretta, A., and Mazzara, M. (2016). "AntibIoTic: Protecting IoT Devices against DDoS Attacks". In: *International Conference in Software Engineering for Defence Applications*. Springer, pp. 59–72.

De Donno, M., Dragoni, N., Giaretta, A., and Spognardi, A. (Sept. 2017). "Analysis of DDoS-Capable IoT Malwares". In: *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*. 2017 Federated Conference on

Computer Science and Information Systems (FedCSIS), pp. 807–816. DOI: 10.15439/2017F288.

De Bruijne, M., van Eeten, M., Gañán, C. H., and Pieters, W. (2017). "Towards a New Cyber Threat Actor Typology". In: p. 72.

Denzin, N. K. (1978). *The Research Act: A Theoretical Introduction to Sociological Methods*. Transaction Publishers. 384 pp. ISBN: 978-0-202-36859-7. Google Books: UjcpxFE0T4cC.

De Santanna, J. J. C. (2017). "DDoS-as-a-Service: Investigating Booter Websites". In:

Doshi, R., Apthorpe, N., and Feamster, N. (May 2018). "Machine Learning DDoS Detection for Consumer Internet of Things Devices". In: *2018 IEEE Security and Privacy Workshops (SPW)*, pp. 29–35. arXiv: 1804.04159. DOI: 10.1109/SPW.2018.00013.

Douligeris, C. and Mitrokotsa, A. (Apr. 5, 2004). "DDoS Attacks and Defense Mechanisms: Classification and State-of-the-Art". In: *Computer Networks* 44.5, pp. 643–666. ISSN: 1389-1286. DOI: 10.1016/j.comnet.2003.10.003.

Dwivedi, Y. K. *et al.* (Dec. 1, 2020). "Impact of COVID-19 Pandemic on Information Management Research and Practice: Transforming Education, Work and Life". In: *International Journal of Information Management*. Impact of COVID-19 Pandemic on Information Management Research and Practice: Editorial Perspectives 55, p. 102211. ISSN: 0268-4012. DOI: 10.1016/j.ijinfomgt.2020.102211.

Framingham, M. (2019). *The Growth in Connected IoT Devices Is Expected to Generate 79.4ZB of Data in 2025, According to a New IDC Forecast*. IDC: The premier global market intelligence company. URL: https://www.idc.com/getdoc.jsp?containerId=prUS45213219 (visited on 11/22/2019).

Gandhi, R., Sharma, A., Mahoney, W., Sousan, W., Zhu, Q., and Laplante, P. (Spr. 2011). "Dimensions of Cyber-Attacks: Cultural, Social, Economic, and Political". In: *IEEE Technology and Society Magazine* 30.1, pp. 28–38. ISSN: 1937-416X. DOI: 10.1109/MTS.2011.940293.

Greenberg, D. F. (2017). "The Contested Place of Motivation in Criminological Theory". In: *Challenging Criminological Theory: The Legacy of Ruth Rosner Kornhauser*.

Hawkinson, J. and Bates, T. (1996). *Guidelines for Creation, Selection, and Registration of an Autonomous System (AS)*. URL: https://tools.ietf.org/html/rfc1930#section-3 (visited on 09/10/2020).

Herley, C. and Florêncio, D. (2010). "Nobody Sells Gold for the Price of Silver: Dishonesty, Uncertainty and the Underground Economy". In: *Economics of Information Security and Privacy*. Springer, pp. 33–53.

Hilbe, J. M. (2011). *Negative Binomial Regression*. Cambridge University Press. ISBN: 0-521-19815-1.

Hilt, S., Kropotov, V., Mercês, F., Rosario, M., and Sancho, D. (2019). *The Internet of Things in the Cybercrime Underground*.

Hilt, S., Mercês, F., Rosario, M., and Sancho, D. (2020). "Worm War: The Botnet Battle for IoT Territory". In: p. 30.

Hirschi, T. and Gottfredson, M. R. (2008). "Critiquing the Critics: The Authors Respond". In: *Out of control: Assessing the general theory of crime*, pp. 217–231.
Holl, P. (2015). "Exploring DDoS Defense Mechanisms". In: *Network* 25.
Holt, T. J. and Bossler, A. M. (2015). *Cybercrime in Progress: Theory and Prevention of Technology-Enabled Offenses*. Routledge. ISBN: 1-317-68899-6.
Holt, T. J., Leukfeldt, R., and Weijer, S. van de (Jan. 19, 2020). "An Examination of Motivation and Routine Activity Theory to Account for Cyberattacks Against Dutch Web Sites:" in: *Criminal Justice and Behavior*. DOI: 10.1177/0093854819900322.
Hoque, N., Bhattacharyya, D. K., and Kalita, J. K. (2015). "Botnet in DDoS Attacks: Trends and Challenges". In: *IEEE Communications Surveys Tutorials* 17.4, pp. 2242–2270. ISSN: 1553-877X, 2373-745X. DOI: 10.1109/COMST.2015.2457491.
Hudic, A., Krombholz, K., Otterbein, T., Platzer, C., and Weippl, E. (2014). "Automated Analysis of Underground Marketplaces". In: *Advances in Digital Forensics X*. Ed. by G. Peterson and S. Shenoi. IFIP Advances in Information and Communication Technology. Berlin, Heidelberg: Springer, pp. 31–42. ISBN: 978-3-662-44952-3. DOI: 10.1007/978-3-662-44952-3_3.
Hutchings, A. and Clayton, R. (Oct. 2, 2016). "Exploring the Provision of Online Booter Services". In: *Deviant Behavior* 37.10, pp. 1163–1178. ISSN: 0163-9625, 1521-0456. DOI: 10.1080/01639625.2016.1169829.
Imperva (2014). *The Top 10 DDoS Attack Trends*. Imperva. URL: https://www.imperva.com/docs/DS_Incapsula_The_Top_10_DDoS_Attack_Trends_ebook.pdf (visited on 12/08/2020).
Javaid, U., Siang, A. K., Aman, M. N., and Sikdar, B. (2018). "Mitigating IoT Device Based DDoS Attacks Using Blockchain". In: *Proceedings of the 1st Workshop on Cryptocurrencies and Blockchains for Distributed Systems*, pp. 71–76.
Jick, T. D. (Dec. 1979). "Mixing Qualitative and Quantitative Methods: Triangulation in Action". In: *Administrative Science Quarterly* 24.4, p. 602. ISSN: 00018392. JSTOR: 2392366. DOI: 10.2307/2392366.
Kambourakis, G., Kolias, C., and Stavrou, A. (Oct. 2017). "The Mirai Botnet and the IoT Zombie Armies". In: *MILCOM 2017 - 2017 IEEE Military Communications Conference (MILCOM)*. MILCOM 2017 - 2017 IEEE Military Communications Conference (MILCOM), pp. 267–272. DOI: 10.1109/MILCOM.2017.8170867.
Kamps, J. and Kleinberg, B. (2018). "To the Moon: Defining and Detecting Cryptocurrency Pump-and-Dumps". In: *Crime Science* 7.1, p. 18.
Kaplan, E. L. and Meier, P. (1958). "Nonparametric Estimation from Incomplete Observations". In: *Journal of the American statistical association* 53.282, pp. 457–481.
Karami, M. (2016). "Understanding and Undermining the Business of DDoS Booter Services". George Mason University.
Karami, M. and McCoy, D. (2013). "Understanding the Emerging Threat of Ddos-as-a-Service". In: *Presented as Part of the 6th {USENIX Workshop on Large-Scale Exploits and Emergent Threats*.
Karami, M., Park, Y., and McCoy, D. (2016). "Stress Testing the Booters: Understanding and Undermining the Business of DDoS Services". In: *Proceedings*

of the 25th International Conference on World Wide Web - WWW '16. The 25th International Conference. Montréal, Québec, Canada: ACM Press, pp. 1033–1043. ISBN: 978-1-4503-4143-1. DOI: 10.1145/2872427.2883004.

Kaur, J. and Buttar, P. K. (2018). "A Systematic Review on Stopword Removal Algorithms". In: *Int. J. Futur. Revolut. Comput. Sci. Commun. Eng* 4.4.

Kleinbaum, D. G. and Klein, M. (2010). *Survival Analysis*. Springer. ISBN: 1-4419-6645-5.

Kolias, C., Kambourakis, G., Stavrou, A., and Voas, J. (2017). "DDoS in the IoT: Mirai and Other Botnets". In: *Computer* 50.7, pp. 80–84. ISSN: 1558-0814. DOI: 10.1109/MC.2017.201.

Kontostathis, A., Edwards, L., and Leatherman, A. (Mar. 4, 2010). "Text Mining and Cybercrime". In: *Text Mining*. Ed. by M. W. Berry and J. Kogan. Chichester, UK: John Wiley & Sons, Ltd, pp. 149–164. ISBN: 978-0-470-68964-6 978-0-470-74982-1. DOI: 10.1002/9780470689646.ch8.

KPN (2021). *DDoS-aanvallen in 2020: dit moet u weten*. URL: https://www.kpn.com/zakelijk/blog/ddos-aanvallen-in-2020-dit-moet-u-weten.htm (visited on 01/07/2021).

Krämer, L., Krupp, J., Makita, D., Nishizoe, T., Koide, T., Yoshioka, K., and Rossow, C. (2015). "Amppot: Monitoring and Defending against Amplification Ddos Attacks". In: *International Symposium on Recent Advances in Intrusion Detection*. Springer, pp. 615–636.

Krupp, J., Backes, M., and Rossow, C. (2016). "Identifying the Scan and Attack Infrastructures behind Amplification DDoS Attacks". In: *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1426–1437.

Kumar, S. and Carley, K. M. (Sept. 2016). "Understanding DDoS Cyber-Attacks Using Social Media Analytics". In: *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*. 2016 IEEE Conference on Intelligence and Security Informatics (ISI), pp. 231–236. DOI: 10.1109/ISI.2016.7745480.

Lane, H., Howard, C., and Hapke, H. M. (2019). *Natural Language Processing in Action*. Manning.

Lau, J. H. and Baldwin, T. (July 18, 2016). *An Empirical Evaluation of Doc2vec with Practical Insights into Document Embedding Generation*. arXiv: 1607.05368 [cs]. URL: http://arxiv.org/abs/1607.05368 (visited on 01/07/2021).

Le, Q. and Mikolov, T. (2014). "Distributed Representations of Sentences and Documents". In: *International Conference on Machine Learning*, pp. 1188–1196.

Link11 (June 8, 2020). *Distributed Denial of Service Report: First Half of 2020*. URL: https://www.link11.com/en/downloads/ddos-report-1st-half-year-2020/.

Lipson, H. F. (Nov. 1, 2002). *Tracking and Tracing Cyber-Attacks: Technical Challenges and Global Policy Issues*. CARNEGIE-MELLON UNIV PITTSBURGH PA SOFTWARE ENGINEERING INST. URL: https://apps.dtic.mil/sti/citations/ADA408853 (visited on 09/15/2020).

Liu, Y. and Wang, H. (2018). "Tracking Mirai Variants". In: p. 18.

Lohachab, A. and Karambir, B. (Sept. 1, 2018). "Critical Analysis of DDoS—An Emerging Security Threat over IoT Networks". In: *Journal of Communications and Information Networks* 3.3, pp. 57–78. ISSN: 2509-3312. DOI: 10.1007/s41650-018-0022-5.

Lueth, K. L. (2020). *State of the IoT 2020: 12 Billion IoT Connections*. URL: https://iot-analytics.com/state-of-the-iot-2020-12-billion-iot-connections-surpassing-non-iot-for-the-first-time/ (visited on 12/19/2020).

Luo, T., Xu, Z., Jin, X., Jia, Y., and Ouyang, X. (2017). "IoTCandyJar: Towards an Intelligent-Interaction Honeypot for IoT Devices". In: *Black Hat*, p. 11.

Lyu, M., Sherratt, D., Sivanathan, A., Gharakheili, H. H., Radford, A., and Sivaraman, V. (July 18, 2017). "Quantifying the Reflective DDoS Attack Capability of Household IoT Devices". In: *Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks*. WiSec '17. New York, NY, USA: Association for Computing Machinery, pp. 46–51. ISBN: 978-1-4503-5084-6. DOI: 10.1145/3098243.3098264.

Mahmoud, R., Yousuf, T., Aloul, F., and Zualkernan, I. (Dec. 2015). "Internet of Things (IoT) Security: Current Status, Challenges and Prospective Measures". In: *2015 10th International Conference for Internet Technology and Secured Transactions (ICITST)*. 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST). London, United Kingdom: IEEE, pp. 336–341. ISBN: 978-1-908320-52-0. DOI: 10.1109/ICITST.2015.7412116.

Mahoney, J. and Goertz, G. (2006). "A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research". In: *Political Analysis* 14.3, pp. 227–249. ISSN: 1047-1987, 1476-4989. DOI: 10.1093/pan/mpj017.

Majkowski, M. (2020). *Reflections on Reflection (Attacks)*. The Cloudflare Blog. URL: https://blog.cloudflare.com/reflections-on-reflections/ (visited on 12/08/2020).

Mandelcorn, S. M. (2013). "An Explanatory Model of Motivation for Cyber-Attacks Drawn from Criminological Theories".

Manning, C. D., Schütze, H., and Raghavan, P. (2008). *Introduction to Information Retrieval*. Cambridge university press. ISBN: 0-521-86571-9.

Marzano, A. *et al.* (June 2018). "The Evolution of Bashlite and Mirai IoT Botnets". In: *2018 IEEE Symposium on Computers and Communications (ISCC)*. 2018 IEEE Symposium on Computers and Communications (ISCC), pp. 00813–00818. DOI: 10.1109/ISCC.2018.8538636.

McHugh, M. L. (June 15, 2013). "The Chi-Square Test of Independence". In: *Biochemia Medica* 23.2, pp. 143–149. ISSN: 1330-0962. pmid: 23894860. DOI: 10.11613/BM.2013.018.

McInnes, L., Healy, J., and Melville, J. (Sept. 17, 2020). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. arXiv: 1802.03426 [cs, stat]. URL: http://arxiv.org/abs/1802.03426 (visited on 01/07/2021).

Miller, G. A. (1995). "WordNet: A Lexical Database for English". In: *Communications of the ACM* 38.11, pp. 39–41.

Mirkovic, J. and Reiher, P. (Apr. 1, 2004). "A Taxonomy of DDoS Attack and DDoS Defense Mechanisms". In: *ACM SIGCOMM Computer Communication Review* 34.2, pp. 39–53. ISSN: 0146-4833. DOI: 10.1145/997150.997156.

Moore, T. (Dec. 1, 2010). "The Economics of Cybersecurity: Principles and Policy Options". In: *International Journal of Critical Infrastructure Protection* 3.3, pp. 103–117. ISSN: 1874-5482. DOI: 10.1016/j.ijcip.2010.10.002.

Motoyama, M., McCoy, D., Levchenko, K., Savage, S., and Voelker, G. M. (Nov. 2, 2011). "An Analysis of Underground Forums". In: *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*. IMC '11. Berlin, Germany: Association for Computing Machinery, pp. 71–80. ISBN: 978-1-4503-1013-0. DOI: 10.1145/2068816.2068824.

Moura, G. C. M., Ganan, C., Lone, Q., Poursaied, P., Asghari, H., and van Eeten, M. (May 2015). "How Dynamic Is the ISPs Address Space? Towards Internet-Wide DHCP Churn Estimation". In: *2015 IFIP Networking Conference (IFIP Networking)*. 2015 IFIP Networking Conference (IFIP Networking). Toulouse, France: IEEE, pp. 1–9. ISBN: 978-3-901882-68-5. DOI: 10.1109/IFIPNetworking.2015.7145335.

Myers, R. H. and Montgomery, D. C. (July 1, 1997). "A Tutorial on Generalized Linear Models". In: *Journal of Quality Technology* 29.3, pp. 274–291. ISSN: 0022-4065. DOI: 10.1080/00224065.1997.11979769.

Nasanbuyn (Feb. 20, 2016). *DDoS Attack*. URL: https://commons.wikimedia.org/wiki/File:Ddos-attack-ex.png#filelinks (visited on 12/05/2019).

Nawrocki, M., Wahlisch, M., Schmidt, T. C., Keil, C., and Schonfelder, J. (2016). "A Survey on Honeypot Software and Data Analysis". In: p. 38.

Nazario, J. (July 1, 2008). "DDoS Attack Evolution". In: *Network Security* 2008.7, pp. 7–10. ISSN: 1353-4858. DOI: 10.1016/S1353-4858(08)70086-2.

NBIP (2020). *R&D projecten*. NBIP. URL: https://www.nbip.nl/kenniscentrum/projecten/ (visited on 01/08/2021).

Netscout (2019). *NETSCOUT THREAT INTELLIGENCE REPORT*. URL: https://www.netscout.com/sites/default/files/2020-02/SECR_001_EN-2001_Web.pdf (visited on 12/27/2020).

Nexusguard (2019). *DDoS Threat Report 2019 Q4*. URL: https://blog.nexusguard.com/threat-report/ddos-threat-report-2019-q4 (visited on 12/27/2020).

— (2020). *DDoS Threat Report 2020 Q2*. URL: https://blog.nexusguard.com/threat-report/ddos-threat-report-2020-q2 (visited on 01/07/2021).

Noroozian, A., Ciere, M., Korczynski, M., Tajalizadehkhoob, S., and Van Eeten, M. (2017). "Inferring the Security Performance of Providers from Noisy and Heterogenous Abuse Datasets". In: *16th Workshop on the Economics of Information Security. http://weis2017. Econinfosec. Org/Wp-Content/Uploads/Sites/3/2017/05/WEIS_2 Pdf*.

Noroozian, A., Korczyński, M., Gañan, C. H., Makita, D., Yoshioka, K., and van Eeten, M. (2016). "Who Gets the Boot? Analyzing Victimization by DDoS-as-a-Service". In: *Research in Attacks, Intrusions, and Defenses*. Ed. by F. Monrose,

M. Dacier, G. Blanc, and J. Garcia-Alfaro. Vol. 9854. Cham: Springer International Publishing, pp. 368–389. ISBN: 978-3-319-45718-5 978-3-319-45719-2. DOI: 10.1007/978-3-319-45719-2_17.

Özçelik, M., Chalabianloo, N., and Gür, G. (2017). "Software-Defined Edge Defense against IoT-Based DDoS". In: *2017 IEEE International Conference on Computer and Information Technology (CIT)*. IEEE, pp. 308–313. ISBN: 1-5386-0958-4.

Paxson, V. (2001). "An Analysis of Using Reflectors for Distributed Denial-of-Service Attacks". In: *ACM SIGCOMM Computer Communication Review* 31.3, pp. 38–47.

Peraković, D., Periša, M., and Cvitić, I. (2015). "Analysis of the IoT Impact on Volume of DDoS Attacks". In: p. 10.

Provos, N. (2004). "A Virtual Honeypot Framework." In: *USENIX Security Symposium*. Vol. 173, pp. 1–14.

Rishi, R. and Saluja, R. (2019). *Future of IoT*. India: Ernst & Young.

Rogers, A. (2021). *How Can We Improve Peer Review in NLP?* The Gradient. URL: https://thegradient.pub/how-can-we-improve-peer-review-in-nlp/ (visited on 01/07/2021).

Rogers, M. K. (2006). "A Two-Dimensional Circumplex Approach to the Development of a Hacker Taxonomy". In: *Digital investigation* 3.2, pp. 97–102.

Rong, X. (2014). *Word2vec Parameter Learning Explained*. arXiv: 1411.2738.

Rossow, C. (2014). "Amplification Hell: Revisiting Network Protocols for DDoS Abuse." In: *NDSS*.

Ryan, P. and Watson, R. (Mar. 14, 2017). "Research Challenges for the Internet of Things: What Role Can OR Play?" In: *Systems* 5.1, p. 24. ISSN: 2079-8954. DOI: 10.3390/systems5010024.

Ryba, F. J., Orlinski, M., Wählisch, M., Rossow, C., and Schmidt, T. C. (May 17, 2016). *Amplification and DRDoS Attack Defense – A Survey and New Perspectives*. arXiv: 1505.07892 [cs]. URL: http://arxiv.org/abs/1505.07892 (visited on 03/16/2020).

Safaei Pour, M., Mangino, A., Friday, K., Rathbun, M., Bou-Harb, E., Iqbal, F., Samtani, S., Crichigno, J., and Ghani, N. (Apr. 1, 2020). "On Data-Driven Curation, Learning, and Analysis for Inferring Evolving Internet-of-Things (IoT) Botnets in the Wild". In: *Computers & Security* 91, p. 101707. ISSN: 0167-4048. DOI: 10.1016/j.cose.2019.101707.

Salim, M. M., Rathore, S., and Park, J. H. (July 2020). "Distributed Denial of Service Attacks and Its Defenses in IoT: A Survey". In: *The Journal of Supercomputing* 76.7, pp. 5320–5363. ISSN: 0920-8542, 1573-0484. DOI: 10.1007/s11227-019-02945-z.

Santanna, J. J., Schmidt, R. d. O., Tuncer, D., Sperotto, A., Granville, L. Z., and Pras, A. (July 2017). "Quiet Dogs Can Bite: Which Booters Should We Go After, and What Are Our Mitigation Options?" In: *IEEE Communications Magazine* 55.7, pp. 50–56. ISSN: 1558-1896. DOI: 10.1109/MCOM.2017.1600992.

Sara Boddy, Justin Shattuck, Debbie Walkowski, and David Warburton (2019). *The Hunt for IoT: Multi-Purpose Attack Thingbots Threaten Internet Stability and Human Life*. URL: https://www.f5.com/labs/articles/threat-

intelligence/the-hunt-for-iot--multi-purpose-attack-thingbots-threaten-intern.html (visited on 01/10/2020).

Schubert, E. and Gertz, M. (2017). "Intrinsic T-Stochastic Neighbor Embedding for Visualization and Outlier Detection". In: *Similarity Search and Applications*. Ed. by C. Beecks, F. Borutta, P. Kröger, and T. Seidl. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 188–203. ISBN: 978-3-319-68474-1. DOI: 10.1007/978-3-319-68474-1_13.

Shakkottai, S., Fomenkov, M., Koga, R., Krioukov, D., and Claffy, K. C. (Mar. 2010). "Evolution of the Internet AS-Level Ecosystem". In: *The European Physical Journal B* 74.2, pp. 271–278. ISSN: 1434-6028, 1434-6036. DOI: 10.1140/epjb/e2010-00057-x.

Silverio-Fernández, M., Renukappa, S., and Suresh, S. (2018). "What Is a Smart Device?-A Conceptualisation within the Paradigm of the Internet of Things". In: *Visualization in Engineering* 6.1, p. 3.

Sinanović, H. and Mrdovic, S. (2017). "Analysis of Mirai Malicious Software". In: *2017 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. IEEE, pp. 1–5. ISBN: 953-290-078-0.

Sood, A. K. and Enbody, R. J. (Mar. 1, 2013). "Crimeware-as-a-Service—A Survey of Commoditized Crimeware in the Underground Market". In: *International Journal of Critical Infrastructure Protection* 6.1, pp. 28–38. ISSN: 1874-5482. DOI: 10.1016/j.ijcip.2013.01.002.

Spitzner, L. (2003). "The Honeynet Project: Trapping the Hackers". In: *IEEE Security & Privacy* 1.2, pp. 15–23.

Sun, S., Luo, C., and Chen, J. (July 1, 2017). "A Review of Natural Language Processing Techniques for Opinion Mining Systems". In: *Information Fusion* 36, pp. 10–25. ISSN: 1566-2535. DOI: 10.1016/j.inffus.2016.10.004.

Tajalizadehkhoob, S., Korczyński, M., Noroozian, A., Gañán, C., and van Eeten, M. (Apr. 2016). "Apples, Oranges and Hosting Providers: Heterogeneity and Security in the Hosting Market". In: *NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium*. NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium, pp. 289–297. DOI: 10.1109/NOMS.2016.7502824.

Tan, L. and Wang, N. (Aug. 2010). "Future Internet: The Internet of Things". In: *2010 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE)*. 2010 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE). Vol. 5, pp. V5-376-V5–380. DOI: 10.1109/ICACTE.2010.5579543.

Tang, J., Wang, X., Gao, H., Hu, X., and Liu, H. (Feb. 1, 2012). "Enriching Short Text Representation in Microblog for Clustering". In: *Frontiers of Computer Science* 6.1, pp. 88–101. ISSN: 1673-7466. DOI: 10.1007/s11704-011-1167-7.

TeleGeography (2020). *TeleGeography's GlobalComms Database Service*. URL: https://www2.telegeography.com/globalcomms-database-service (visited on 09/10/2020).

Thomas, D. R., Pastrana, S., Hutchings, A., Clayton, R., and Beresford, A. R. (Nov. 2017). "Ethical Issues in Research Using Datasets of Illicit Origin". In: *Proceed-*

ings of the 2017 Internet Measurement Conference. IMC '17: Internet Measurement Conference. London United Kingdom: ACM, pp. 445–462. ISBN: 978-1-4503-5118-8. DOI: 10.1145/3131365.3131389.

Thomas, K. *et al.* (2015). "Framing Dependencies Introduced by Underground Commoditization". In:

Trend Micro (2020). *Worm War: The Botnet Battle for IoT Territory*. URL: https://www.trendmicro.com/vinfo/us/security/news/internet-of-things/caught-in-the-crossfire-defending-devices-from-battling-botnets (visited on 08/17/2020).

United Nations (June 2017). *Global Indicator Framework for the Sustainable Development Goalsand Targets of the 2030 Agenda for Sustainable Development*. URL: https://unstats.un.org/sdgs/indicators/indicators-list/ (visited on 01/31/2020).

Verstegen, S. (2019). "Understanding the Role of IoT End Users in Miria-Like Bot Remediation". In:

Vishwakarma, R. and Jain, A. K. (Jan. 1, 2020). "A Survey of DDoS Attacking Techniques and Defence Mechanisms in the IoT Network". In: *Telecommunication Systems* 73.1, pp. 3–25. ISSN: 1572-9451. DOI: 10.1007/s11235-019-00599-z.

Vlajic, N. and Zhou, D. (July 2018). "IoT as a Land of Opportunity for DDoS Hackers". In: *Computer* 51.7, pp. 26–34. ISSN: 1558-0814. DOI: 10.1109/MC.2018.3011046.

Wall, D. (2007). *Cybercrime: The Transformation of Crime in the Information Age*. Vol. 4. Polity. ISBN: 0-7456-2736-6.

Wang, H., Turing, A., Liu, Y., and Genshen, Y. (2020). *The Botnet Cluster on the 185.244.25.0/24*. 360 Netlab Blog - Network Security Research Lab at 360. URL: https://blog.netlab.360.com/the-botnet-cluster-on-185-244-25-0-24-en/ (visited on 12/12/2020).

Wegberg, R. van, Tajalizadehkhoob, S., Soska, K., Akyazi, U., Ganan, C. H., Klievink, B., Christin, N., and Eeten, M. van (2018). "Plug and Prey? Measuring the Commoditization of Cybercrime via Online Anonymous Markets". In: 27th {USENIX Security Symposium ({USENIX Security 18), pp. 1009–1026. ISBN: 978-1-939133-04-5. URL: https://www.usenix.org/conference/usenixsecurity18/presentation/van-wegberg (visited on 08/12/2020).

What Is a Distributed Denial-of-Service (DDoS) Attack? (2020). Cloudflare. URL: https://www.cloudflare.com/learning/ddos/what-is-a-ddos-attack/ (visited on 12/08/2020).

What Is an ACK Flood DDoS Attack? (2020). *What Is an ACK Flood DDoS Attack? | Types of DDoS Attacks*. Cloudflare. URL: https://www.cloudflare.com/learning/ddos/what-is-an-ack-flood/ (visited on 12/08/2020).

Yar, M. (2005). "The Novelty of 'Cybercrime' An Assessment in Light of Routine Activity Theory". In: *European Journal of Criminology* 2.4, pp. 407–427.

Zand, A., Modelo-Howard, G., Tongaonkar, A., Lee, S.-J., Kruegel, C., and Vigna, G. (July 2017). "Demystifying DDoS as a Service". In: *IEEE Communications*

Magazine 55.7, pp. 14–21. ISSN: 1558-1896. DOI: 10.1109/MCOM.2017.1600980.

Zargar, S. T., Joshi, J., and Tipper, D. (2013). "A Survey of Defense Mechanisms Against Distributed Denial of Service (DDoS) Flooding Attacks". In: *IEEE Communications Surveys Tutorials* 15.4, pp. 2046–2069. ISSN: 1553-877X. DOI: 10.1109/SURV.2013.031413.00127.

Zhang, Z.-K., Cho, M. C. Y., Wang, C.-W., Hsu, C.-W., Chen, C.-K., and Shieh, S. (Nov. 2014). "IoT Security: Ongoing Challenges and Research Opportunities". In: *2014 IEEE 7th International Conference on Service-Oriented Computing and Applications*. 2014 IEEE 7th International Conference on Service-Oriented Computing and Applications, pp. 230–234. DOI: 10.1109/SOCA.2014.58.

Zimek, A., Schubert, E., and Kriegel, H.-P. (2012). "A Survey on Unsupervised Outlier Detection in High□dimensional Numerical Data". In: *Statistical Analysis and Data Mining: The ASA Data Science Journal* 5.5, pp. 363–387.

Cover image from National Institutes of Health (U.S.). Medical Arts and Photography Branch (1969). Public domain. Retrieved from: https://collections.nlm.nih.gov/catalog/nlm:nlmuid-101454190-img. Chosen as it depicts triangulation where many different perspectives help uncover the pattern within.

A

Natural Language Processing Words

A.1. Keywords for Tagging

	Marketplace	Access	DDoS	IoT	Attack Power
0	$	authentication	booters	device	gbps
1	£	auth	bot	bot	mbps
2	€	login	botnet	botnet	fast
3	€	access	box	bot	seconds!
4	bitcoin	NaN	ddos	mirai	mb
5	bitcoin cash	NaN	ddosed	spot	gb
6	btc	NaN	ddosing	botnets	ping flood
7	btc	NaN	hit	qbot	flood
8	buy	NaN	hit	iot	null
9	buying	NaN	hitting	device	null attack
10	cash	NaN	stress	botnet?	snmp
11	coin	NaN	stresser	NaN	ntp
12	credit	NaN	stresser.wtf	NaN	ssdp
13	dash	NaN	NaN	NaN	http
14	ecoin	NaN	NaN	NaN	NaN
15	ether	NaN	NaN	NaN	NaN
16	ethereum	NaN	NaN	NaN	NaN
17	eur	NaN	NaN	NaN	NaN
18	euro	NaN	NaN	NaN	NaN

Continued on next page

	Marketplace	Access	DDoS	IoT	Attack Power
19	interkassa	NaN	NaN	NaN	NaN
20	lite coin	NaN	NaN	NaN	NaN
21	litecoin	NaN	NaN	NaN	NaN
22	market	NaN	NaN	NaN	NaN
23	monero	NaN	NaN	NaN	NaN
24	money	NaN	NaN	NaN	NaN
25	omise	NaN	NaN	NaN	NaN
26	paid	NaN	NaN	NaN	NaN
27	pay	NaN	NaN	NaN	NaN
28	payment	NaN	NaN	NaN	NaN
29	paypal	NaN	NaN	NaN	NaN
30	perfect money	NaN	NaN	NaN	NaN
31	perfectmoney	NaN	NaN	NaN	NaN
32	purchase	NaN	NaN	NaN	NaN
33	qiwi	NaN	NaN	NaN	NaN
34	ripple	NaN	NaN	NaN	NaN
35	sell	NaN	NaN	NaN	NaN
36	seller	NaN	NaN	NaN	NaN
37	selling	NaN	NaN	NaN	NaN
38	selly	NaN	NaN	NaN	NaN
39	store	NaN	NaN	NaN	NaN
40	trade	NaN	NaN	NaN	NaN
41	usd	NaN	NaN	NaN	NaN
42	web money	NaN	NaN	NaN	NaN
43	webmoney	NaN	NaN	NaN	NaN
44	zcash	NaN	NaN	NaN	NaN

A.2. Stopwords

['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves', 'what', 'which', 'who', 'whom', 'this', 'that', "that'll", 'these', 'those', 'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through', 'during', 'before', 'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under', 'again', 'further', 'then', 'once', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both', 'each', 'few', 'more', 'most', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than', 'too', 'very', 's', 't', 'can', 'will', 'just', 'don', "don't", 'should', "should've", 'now', 'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't", 'couldn', "couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn't", 'hasn', "hasn't", 'haven', "haven't", 'isn', "isn't", 'ma', 'mightn', "mightn't", 'mustn', "mustn't", 'needn', "needn't", 'shan', "shan't",

'shouldn', "shouldn't", 'wasn', "wasn't", 'weren', "weren't", 'won', "won't", 'wouldn', "wouldn't"]

B

Data

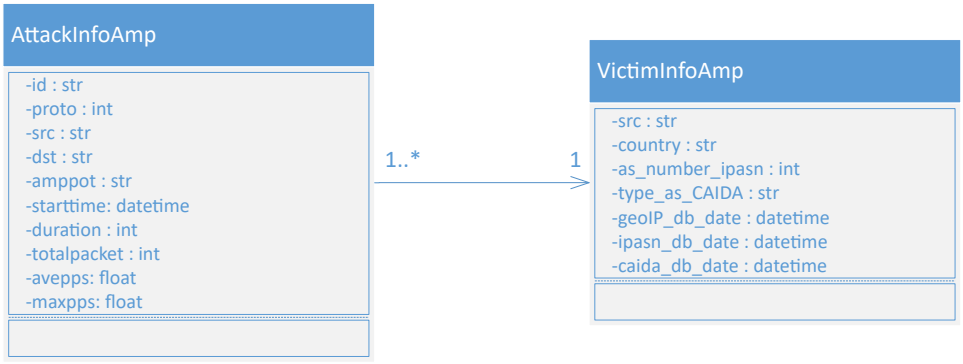


Figure B.1: Amppot UML Diagram

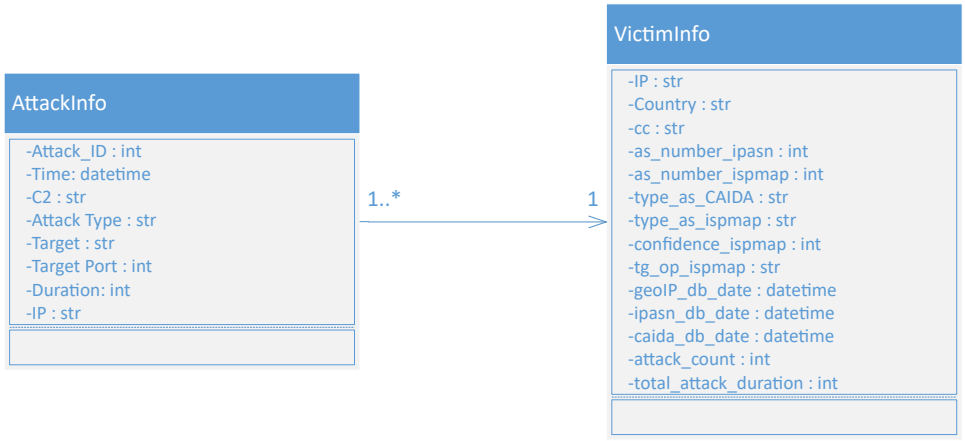


Figure B.2: Netlab UML Diagram

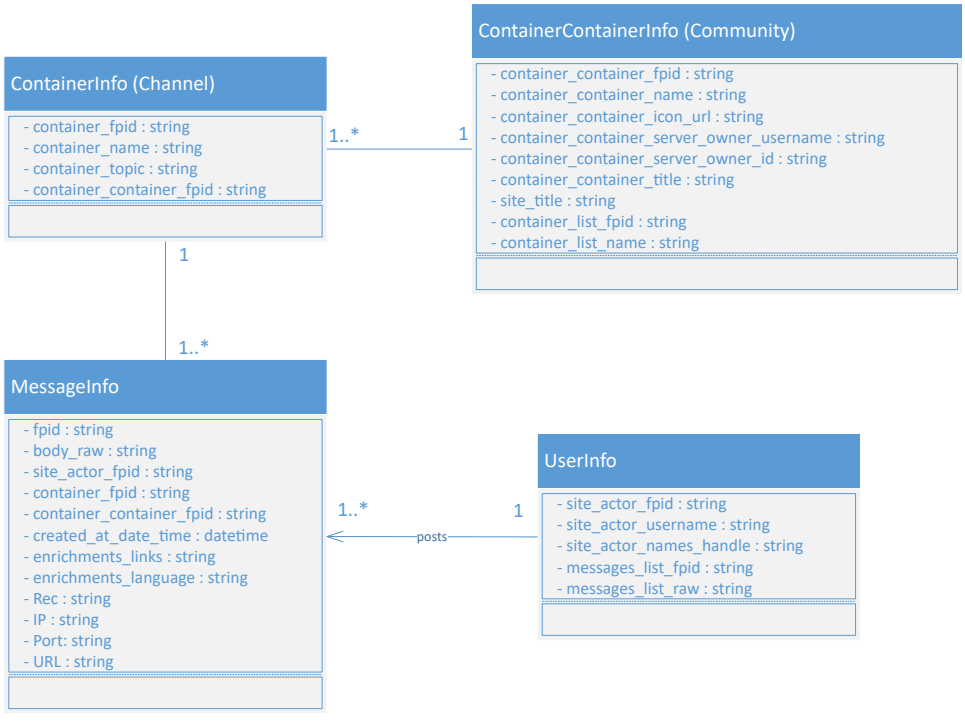


Figure B.3: Underground Market UML Diagram

C

Community and Channels

Community	Channel	Topic	Messages	Score
” Legit Shop’s ”	chat	IoT	7	0.420255
	leave-join	N/A	2	0.000000
	scammers	N/A	1	0.000000
	unverified-market	N/A	1	0.000000
7 Market	beef	Hacking	3411	0.126197
	bots	DDoS	3345	0.088440
	gamble	N/A	4	0.000000
	partners	Community	593	0.559520
	☐-general-☐	Community	8592	0.173569
	☐-buying-☐	N/A	3	0.000000
	☐-selling-☐	Marketplace	109	0.442196
	☐promote☐	N/A	1	0.000000
BangStresser.com	welcome	N/A	12	0.000000
	☐ 🄵 bump-server	DDoS	1516	0.263906
	☐ 🄵 partnerships	IoT	8	0.249692
	☐ 🄵 discussion	DDoS	4407	0.238892
	☐ 🄵 recommendations	N/A	2	0.000000
	☐ 🄵 updates	N/A	3	0.000000
	☐ 🄵 socials	N/A	2	0.000000
	☐ 🄵 announcements	N/A	4	0.000000
	☐ 🄵 report-bugs	General	85	0.175789
Big Hekks	advertise	N/A	2	0.000000
	general	IoT	1527	0.290202
	leaks	N/A	2	0.000000

Continued on next page

Community	Channel	Topic	Messages	Score
	nfsw	Community	257	0.182836
	rules	N/A	1	0.000000
CTR 10.0	anime	Bot	56	0.158925
	announcements	Community	74	0.142953
	beef	General	18	0.178056
	bot-commands	Bot	1325	0.395218
	childrens-corner	DDoS	134	0.176485
	debate	DDoS	193	0.116034
	general	IoT	200482	0.263184
	help	IoT	8030	0.392405
	memes	DDoS	1031	0.095782
	nsfw	Community	4894	0.114827
	rules	N/A	2	0.000000
	scammers	N/A	16	0.000000
	shitposting	Community	320	0.168610
	spam	Bot	262	0.228473
	verify	N/A	3	0.000000
ComfyChoo! □ (´„•ω•„) □	announcements	Community	52	0.519129
	bot-commands	N/A	9	0.000000
	click-me	Community	15489	0.098253
	complain-here	Community	200	0.371360
	here-you-ask-for-a-role	N/A	3	0.000000
	memes-and-chaos	N/A	34	0.000000
	music-chat□	N/A	19	0.000000
	where-you-type	Community	392	0.117689
	□□□□ crayons ~ □□	N/A	2	0.000000
	kawaii market	Community	3	0.363850
	□□□□ mediasu ~ □□	Community	35	0.260163
	□□□□ patsu ~ □□	N/A	50	0.000000
	□□□□ pets ~ □□	N/A	4	0.000000
	□□□ colors ~ □	N/A	1	0.000000
	□□□ qotd ~ □	N/A	2	0.000000
	□□□ shybunnies ~ □	N/A	15	0.000000
	affiliates	N/A	3	0.000000
	□□□□🇫🇫 info ~ □	Community	4	0.541996
	□advertising	Community	339	0.365630
	□aesthetics	Community	45	0.256701
	□art	Community	35	0.264890
	□bot-commands	Community	2188	0.168616
	□casino	Community	1099	0.123168
	□counting	Community	298	0.267222

Continued on next page

Community	Channel	Topic	Messages	Score
	☐general	Community	31723	0.189418
	☐images	Community	57	0.252051
	☐introduce-yourself	Community	212	0.340298
	☐math	Community	7988	0.151249
	☐memes	Community	522	0.311552
	☐music	Community	108	0.263611
	☐role-assignment	Community	352	0.198535
	☐selfies	Community	81	0.248871
	☐spam	Community	431	0.194039
	☐suggestions	Community	44	0.243044
	☐vip	Community	103	0.246915
	☐voice-text	N/A	39	0.000000
	☐announcements	Community	26	0.390455
	☐join-leave-logs	Community	2573	0.172830
	☐partnerships	N/A	2	0.000000
	☐polls	N/A	3	0.000000
	🄵 ☐☐ 🄵 emote-spam ~ ☐	N/A	28	0.000000
	🄵 ☐☐ 🄵 oneword ~ ☐	N/A	27	0.000000
	🄵 ☐☐ 🄵 pokecord ~ ☐	Bot	139	0.497156
	🄵 ☐☐ 🄵 supportsu ~ ☐	N/A	2	0.000000
	🄵 ☐☐ 🄵 botchu ~ ☐	Bot	10075	0.143446
	🄵 ☐☐ 🄵 bumpchu ~ ☐	DDoS	116	0.141524
	🄵 ☐☐ 🄵 chatchu ~ ☐	Community	10726	0.299529
	🄵 ☐☐ 🄵 drawings ~ ☐	Community	39	0.223840
	🄵 ☐☐ 🄵 giggles ~ ☐	N/A	15	0.000000
	🄵 ☐☐ 🄵 introductions ~ ☐	Community	80	0.355933
	🄵 ☐☐ 🄵 kawaii ~ ☐	N/A	6	0.000000
	🄵 ☐☐ 🄵 moosic ~ ☐	Bot	11	0.273065
	🄵 ☐☐ 🄵 pingpong ~ ☐	Community	10	0.470820
	🄵 ☐☐ 🄵 roles ~ ☐	Community	20	0.347925
	🄵 ☐☐ 🄵 staffies ~ ☐	Community	21	0.370002
	🄵 ☐☐ 🄵 userphone ~ ☐	Community	2165	0.157701
	☐☐°☐🄵☐selfies ~ ☐	Community	60	0.252588
	☐☐ media ~ ☐🄵🄵	N/A	2	0.000000
	☐☐ partners ~ ☐🄵🄵	Community	94	0.566422
	☐☐ rules ~ ☐🄵🄵	Community	15	0.473469
	☐☐ server info ~ ☐🄵🄵	Community	33	0.465223
Console	announcements	Hacking	24	1.286471
Prospect				
		N/A	5	0.000000
	bot-spam	N/A	17	0.000000
	invites	Hacking	84	0.660550

Continued on next page

Community	Channel	Topic	Messages	Score
		N/A	4	0.000000
	lounge	Community	59	0.353923
		Hacking	3905	10.503192
		N/A	22	0.000000
	memes	DDoS	54	0.248190
		N/A	31	0.000000
	new-member	Hacking	108	0.446961
		N/A	1	0.000000
	rules	Hacking	3	0.606855
	suggestions	Hacking	117	1.311273
	support	Hacking	1182	1.950193
		N/A	14	0.000000
	tutorial	N/A	2	0.000000
CrossFade - AntiSec	advertisement	N/A	23	0.000000
	discussion	N/A	32	0.000000
	general	IoT	13978	0.121659
	how-to-enable-video	DDoS	1	0.191145
	music-general	Bot	854	0.540647
	role-assignment	Bot	232	0.125159
Crossfade	advertise	Community	27	0.321845
	cris-rat-paradise-lole	Bot	77	0.217445
	dinos	Bot	87	0.139949
	file-drop	N/A	18	0.000000
	general	Bot	27916	0.101779
	music-requests	Bot	3405	0.709404
	nsfw	Bot	579	0.183928
	self-assign-roles	Bot	583	0.237074
Cyber Terrorists	0w0	IoT	96	0.237706
	about	N/A	2	0.000000
	accounts	N/A	18	0.000000
	amps-cat	N/A	3	0.000000
	announcements	IoT	51	0.310258
	application-channel	Community	225	0.313673
	application-form	N/A	2	0.000000
	application-rules	N/A	2	0.000000
	application-updates	N/A	4	0.000000
	beef	IoT	964	0.248583
	botnet	IoT	2880	1.019155
	clovers-dog	N/A	44	0.000000
	file-drop	IoT	423	0.295663
	files	IoT	74	0.515205

Continued on next page

Community	Channel	Topic	Messages	Score
	general	IoT	56880	0.345158
	help	IoT	1907	0.411980
	memes	IoT	103	0.266705
	music	Bot	1257	0.451922
	ranks	Community	2	0.215884
		N/A	4	0.000000
	react-to-message	N/A	1	0.000000
	rules	Community	2	0.384055
	scammers	IoT	50	0.284444
	servers	IoT	30	0.468277
	show-off	IoT	980	0.300588
CyberHackers.eu Community	alts	Marketplace	3519	0.164254
	c	General	61	0.279799
	challenge-channel	N/A	2	0.000000
	challenge-publications	General	54	0.265825
	coding-channel	General	414	0.267805
	google-ctf□	Community	146	0.057719
	hacker101-ctf□	Hacking	365	0.158379
	hidra-multitool	IoT	56	0.175412
	htb-ctf□	General	40	0.197227
	join-left	N/A	4	0.000000
	memes	DDoS	152	0.094644
	nsfw	Marketplace	516	0.097517
	phone-related	General	91	0.225936
	pico-ctf□	N/A	37	0.000000
	pre-orders	N/A	3	0.000000
	python	General	153	0.223803
	staff-reporting	N/A	2	0.000000
	stream-channel	N/A	3	0.000000
	web	General	70	0.252552
	□general	General	68542	0.090775
	□vps-list	N/A	3	0.000000
	□discord-links	N/A	5	0.000000
	□giveaway	N/A	2	0.000000
	□partners	N/A	1	0.000000
	□projects	N/A	1	0.000000
	□rules	Community	15	0.425820
	□tutorials	N/A	1	0.000000
	□help	General	1815	0.255233
Hardchats 2018	anime_nsfw	Community	476	0.434916
	announcements	Community	13	0.411842

Continued on next page

Community	Channel	Topic	Messages	Score
	binary-exploitation	N/A	4	0.000000
	bot-commands	Community	482	0.100080
	cute-animal-shit	Bot	100	0.438958
	dumps	Marketplace	186	0.225269
	facts	N/A	1	0.000000
	feet	N/A	14	0.000000
	female_nsfw	Community	563	0.336392
	gaming	Community	11	0.335229
	general_nsfw	Community	985	0.216522
	hands	N/A	12	0.000000
	hard_chats	General	45932	0.121754
	male_nsfw	N/A	4	0.000000
	marketplace	Marketplace	964	0.380141
	mathematics	N/A	19	0.000000
	memes	Community	733	0.101990
	music	Bot	1516	0.462893
	partnered-servers	Community	4	0.565651
	pentesting	N/A	6	0.000000
	physics	N/A	15	0.000000
	programming	General	72	0.275591
	reverse-engineering	N/A	6	0.000000
	rules	N/A	4	0.000000
	selfie	N/A	17	0.000000
	spam-shit	Community	158	0.290056
	welcome	Community	493	0.182943
Illegal Community	⌈☐⌋ scammers	N/A	14	0.000000
	⌈☐⌋ scammers	N/A	2	0.000000
	⌈☐⌋ welcome	Bot	1057	0.156138
	⌈☐⌋ power-proof	N/A	3	0.000000
	⌈☐⌋ main-chat	Marketplace	4791	0.264237
	⌈☐⌋ file-drop	N/A	14	0.000000
	⌈☐⌋ unverified-market	Marketplace	579	0.424658
	⌈☐⌋ verified-market	IoT	31	0.499638
	⌈☐⌋ sell	Marketplace	69	0.439884
	⌈☐⌋ trade	Marketplace	11	0.651261
	⌈☐⌋ announcements	N/A	9	0.000000
	⌈☐⌋ photo-chat	Marketplace	782	0.188450
	⌈☐⌋ bot-commands	Bot	639	0.492034
MushroomClub	apply-for-raiding	N/A	7	0.000000
	aussie-adapt	N/A	2	0.000000
	blair	N/A	3	0.000000

Continued on next page

Community	Channel	Topic	Messages	Score
	bot-configuration	N/A	10	0.000000
	bump	Bot	92	0.199515
	give-aways	N/A	28	0.000000
	giveaway-rewards	N/A	4	0.000000
	hydra_songrequests	N/A	1	0.000000
	staff-applications	N/A	43	0.000000
	unverified-shop	Marketplace	119	0.368515
	vouches	N/A	14	0.000000
	□description□	N/A	1	0.000000
	□rules□	Community	1	0.489612
	□general□	Marketplace	3100	0.189389
	□self-role□	N/A	3	0.000000
	□arrivals□	N/A	11	0.000000
	□hack-chat□	General	160	0.237256
	□official-shop□	Marketplace	6	0.425605
	□announcments□	Community	23	0.392522
	□memes□	Community	378	0.324462
NSSALES	announcements	N/A	2	0.000000
	config-request	N/A	5	0.000000
	configs	N/A	4	0.000000
	general	Marketplace	18	0.506521
	methods	N/A	6	0.000000
	vouches	N/A	8	0.000000
	wtb	N/A	8	0.000000
	wts	Marketplace	313	0.410288
Nija	announcements	IoT	28	0.407362
	bot-commands	N/A	8	0.000000
	c	N/A	7	0.000000
	css	N/A	1	0.000000
	dot-net	N/A	7	0.000000
	general	IoT	7798	0.533368
	html	N/A	12	0.000000
	java	N/A	5	0.000000
	market	IoT	134	0.426864
	memes	N/A	8	0.000000
	nsfw	N/A	36	0.000000
	offtopic	N/A	4	0.000000
	php	N/A	1	0.000000
	python	N/A	10	0.000000
	rules	N/A	1	0.000000
	scammers	N/A	1	0.000000
	server-info	N/A	1	0.000000

Continued on next page

Community	Channel	Topic	Messages	Score
	spam	N/A	29	0.000000
	sql	N/A	5	0.000000
	support	IoT	220	0.344855
	useful-archives	N/A	4	0.000000
	vouches	N/A	22	0.000000
	welcome	Bot	554	0.110086
OVERFLOW.LTD	chat	DDoS	116	0.309921
		N/A	4	0.000000
	Customer Chat	N/A	2	0.000000
	Main	DDoS	48	0.493613
	Stresser Info	N/A	2	0.000000
	TOS	DDoS	2	0.324784
	Stresser News	DDoS	14	0.517859
Official Hackintosh	bot-commands	N/A	13	0.000000
	reports	N/A	2	0.000000
	spam	N/A	12	0.000000
	¤botnet¤	IoT	129	0.395124
	¤downloads¤	N/A	4	0.000000
	General	Marketplace	1325	0.278404
	Memes	N/A	8	0.000000
	NSFW	N/A	60	0.000000
	Marketplace□	Marketplace	13	0.499224
	Welcome	Bot	404	0.324784
	Underground	N/A	17	0.000000
OmitVPN | Support Chat	bot-spam	N/A	16	0.000000
	check-invites	N/A	17	0.000000
	create-ticket	Hacking	58	0.363710
	en-lounge	Community	131	0.259977
	logs	Community	374	0.277718
	off-topic	N/A	12	0.000000
	suggestions	N/A	11	0.000000
OofLand	bot-spam	Bot	89	0.604418
	botcommands	Bot	616	0.400657
	count-to-8-thousand	Bot	859	0.376714
	general	Bot	8953	0.219855
	leave-logs	N/A	8	0.000000
	media	Bot	31	0.411794
	memes	N/A	10	0.000000
	nsfw	Community	358	0.271283
	pokemon-duel	Bot	481	0.265973

Continued on next page

Community	Channel	Topic	Messages	Score
	role-vote	N/A	14	0.000000
	rules	Community	2	0.455913
	welcome	N/A	10	0.000000
	xbox-ps-pc	Bot	59	0.334382
Shadoh's Big Hack Lounge	big-hacker-hangout	IoT	8427	0.277516
	file-drop	IoT	928	0.267278
	sales	IoT	105	0.475982
	welcome	N/A	8	0.000000
Stress.gg	announcements	DDoS	94	1.206614
	bot-commands	Bot	839	0.762392
	change-logs	N/A	8	0.000000
	conversations	DDoS	14969	4.715581
	dstats	N/A	20	0.000000
	giveaways	Bot	122	1.193784
		N/A	41	0.000000
	information	N/A	12	0.000000
	report-bugs	DDoS	133	0.459832
		N/A	35	0.000000
	rules	Community	1	0.525232
	spam	DDoS	403	0.773818
		N/A	48	0.000000
	suggestions	DDoS	409	1.398314
		N/A	21	0.000000
	support	DDoS	657	0.563130
		N/A	15	0.000000
	tickets	DDoS	283	1.013374
	tos	N/A	3	0.000000
	verification	DDoS	117	0.286590
Stresser.WTF Community	chat	DDoS	1326	0.401820
	news	N/A	19	0.000000
	power-proof	N/A	17	0.000000
	pseudo-hackers-ips	N/A	3	0.000000
	purchase	N/A	2	0.000000
	shop_link	N/A	6	0.000000
	stresser-wtf-community	DDoS	39	0.515161
Supreme Security Services	general-chat	IoT	11028	0.316128
	info	Community	9	0.354081
	invite-rewards	DDoS	147	0.167538
	nonclient-support	DDoS	192	0.334738

Continued on next page

Community	Channel	Topic	Messages	Score
	nsfw	Community	322	0.101071
	power-proofs	N/A	4	0.000000
	readme	Community	8	0.296302
	supreme-market	IoT	728	0.420567
	updates-changelog	DDoS	20	0.316232
TOP-STRESSER'S	advertisements	N/A	2	0.000000
	bump	Bot	151	0.330864
	general	DDoS	20086	0.277799
	moments	N/A	7	0.000000
	news	N/A	4	0.000000
	request-a-hit	DDoS	7539	0.199521
	scammer-stressers	N/A	3	0.000000
	top-10	N/A	9	0.000000
ThugCrowd HQ	48hours-forensics	General	59	0.368206
	affiliates	General	118	0.250335
	announcements	General	13	0.270850
	art	General	195	0.260312
	binaries-and-reversing	General	126	0.302277
	bot-fuzzing	General	18821	0.215748
	bot-testing	General	5999	0.237688
	bug-bounty	N/A	22	0.000000
	cat-pics	General	101	0.220916
	certs-and-education	General	83	0.420460
	cloud-and-devops	General	511	0.390766
	code-shitposting	General	154	0.267808
	coding-questions	General	1255	0.309263
	cons	General	565	0.264484
	cryptocurrency	General	90	0.271454
	ctf-and-challenges	General	337	0.334976
	demoscene	General	17	0.274129
	general	General	212401	0.352916
	gta-irl	General	219	0.323490
	guide2thugcrowd	General	8	0.391067
	hardware	General	518	0.356308
	hire-a-hacker	General	314	0.306243
	holiday-family-pics	General	436	0.192452
	honeypot	N/A	12	0.000000
	ideas	General	318	0.355050
	johnathon-4-mp	General	33	0.225142
	log	General	45	0.226679
	manifesto	General	41	0.277370

Continued on next page

Community	Channel	Topic	Messages	Score
	mentalhealth	General	2232	0.329902
	mobile	General	182	0.333613
	music	General	434	0.265026
	networking	General	208	0.328894
	news-and-exploits	General	2680	0.407969
	occult	General	47	0.252818
	osint	General	91	0.342813
	pdfs-and-books	General	262	0.374897
	people-skills	General	31	0.365730
	physical-security	N/A	14	0.000000
	ricing	General	283	0.217857
	riot	General	318	0.274488
	rules-suggestions-temp	General	146	0.290175
	security	General	3395	0.469874
	shitposting	General	9320	0.298190
	show-notes	N/A	15	0.000000
	spaceforce	General	1028	0.331352
	square_up	Community	124	0.140618
	street-knowledge	General	248	0.351799
	swap-meet	N/A	4	0.000000
	tech-job-resources	General	52	0.332835
	tech-support	General	433	0.312566
	threat-intel	General	748	0.271346
	tools	N/A	3	0.000000
	video-games	General	172	0.237561
	voiceless-voice	General	8196	0.313408
	windows-world	General	88	0.308164
	wireless-and-sdr	General	565	0.289807
	日本語	N/A	2	0.000000
UN5T48L3	bot-commands	N/A	1	0.000000
Cyber Security				
	general	IoT	6876	0.265354
	help	IoT	126	0.372346
	join-leave	Bot	1250	0.127424
	memes	DDoS	28	0.126225
	music-commands	N/A	13	0.000000
	nsfw	Bot	1301	0.133014
	official-b0tn3t-store	IoT	25	0.560302
	red-team	N/A	10	0.000000
	scammers	IoT	16	0.144974
	special-free-drops	N/A	14	0.000000
	updates	IoT	13	0.438722

Continued on next page

Community	Channel	Topic	Messages	Score
	vouch	IoT	36	0.274410
	□welcome□□	N/A	1	0.000000
Usernames.org (Delayed)	announcements	Community	43	0.253418
	appraisals	Marketplace	571	0.293652
	bot-spam	Bot	967	0.399613
	bumps	N/A	5	0.000000
	engagement	Marketplace	115	0.279343
	fortnite	Marketplace	693	0.516992
	gamertags	Marketplace	6921	0.217871
	general-discussion	Marketplace	157635	0.210110
	giveaways	Community	27	0.306011
	instagram	Marketplace	4320	0.319711
	marketplace-general	Marketplace	5355	0.484870
	music	Bot	641	0.536459
	other	Marketplace	1292	0.476053
	pictures	Marketplace	39	0.500582
	playstations	Marketplace	437	0.359185
	questions-suggestions	Marketplace	303	0.224913
	sm-general	Marketplace	283	0.240236
	spam-jokes	Marketplace	655	0.290763
	twitter	Marketplace	852	0.311524
UvU Kingdom	botnet-proof	N/A	17	0.000000
	crypto-currency-news	N/A	2	0.000000
	funny-videos	N/A	2	0.000000
	memes	N/A	3	0.000000
	minecraft	N/A	6	0.000000
	news	Community	17	0.374792
	pokecord	Bot	372	0.381166
	promotion	Community	7	0.495352
	questions	N/A	1	0.000000
	quotes-by-uvu	N/A	4	0.000000
	roblox	N/A	6	0.000000
	share-your-music	Bot	235	0.487674
	steam-games	N/A	3	0.000000
	SINGLES	Bot	137	0.242972
	NSFW	Community	880	0.325569
	RULES	Community	7	0.526157
	BOTS	Bot	594	0.502719
	HOT-OR-NOT	Bot	452	0.216882
	DEBATE	Bot	94	0.455846
	MAIN	Bot	12382	0.372226

Continued on next page

Community	Channel	Topic	Messages	Score
	MEMERY	Community	114	0.292943
VSB Marketplace 2.0	☐verified-buying	Marketplace	26	0.548678
	☐verified-middleman	Marketplace	1	0.217632
	☐verified-selling	Marketplace	168	0.438761
	☐unverified-buying	Marketplace	487	0.566871
	☐unverified-selling	Marketplace	2649	0.558690
	☐free-methods	Marketplace	28	0.405771
	☐reports	Marketplace	641	0.348468
	☐general	Marketplace	4282	0.470101
	☐money-trading	Marketplace	42	0.617922
	☐accounts	Marketplace	67	0.553548
	☐usernames-selling	Marketplace	110	0.449183
	☐call-middleman	N/A	1	0.000000
	☐memes-pictures	N/A	4	0.000000
	☐sqrewymarket	Marketplace	20	0.534555
	☐partners	N/A	12	0.000000
	☐vouch	Marketplace	113	0.358137
	☐information	Community	7	0.302676
Volkoz Booter Discord	bot-commands	N/A	4	0.000000
	chat	Community	128	0.164092
	goodbye	N/A	3	0.000000
	ips-only	General	35	0.206231
	market	N/A	1	0.000000
	rank	N/A	16	0.000000
	spam	N/A	26	0.000000
We Hack For Hentai	affiliates	N/A	1	0.000000
	announcements	General	83	0.271618
	citrusec	N/A	3	0.000000
	coding-bunker	General	3515	0.168495
	cooking	Bot	477	0.226887
	ctf-event-general	General	2583	0.214049
	events	General	18	0.262622
	general	General	206123	0.283203
	hell	Community	288	0.237135
	hentai	Community	906	0.277150
	lincox-suggestions	General	375	0.226474
	meme-bin	General	942	0.123163
	politics	General	127	0.224572
	porn	Community	1673	0.139571

Continued on next page

Community	Channel	Topic	Messages	Score
	robo-faggots	General	53005	0.159926
	rules	N/A	1	0.000000
	smorgasbord	N/A	6	0.000000
digitalgangster.com	music	General	1261	0.213984
	neals-gaming-matrix	General	266	0.175600
	vamos-a-la-playa	General	693409	0.205562
	whyworkmargorp	General	4227	0.250247
ghost@kirin(キリン)	accounts	Marketplace	34	0.450342
	anime-pics	Community	44	0.232070
	beef	N/A	18	0.000000
	bot-commands	Bot	515	0.228100
	botnet	IoT	2385	0.252962
	downloads	IoT	418	0.224307
	follow-them-instas	Marketplace	129	0.178190
	general	IoT	15155	0.195590
	join-these-dank-shit	Community	184	0.216859
	kirin-new-stuff-to-add	N/A	26	0.000000
	market	IoT	1719	0.310395
		N/A	9	0.000000
	nice	N/A	25	0.000000
	porn	Community	660	0.136213
	rules	N/A	17	0.000000
	scammer-list	IoT	61	0.160950
	welcome	Bot	56	0.156604
		Community	45	0.345369
hakka shit	announcements	IoT	39	0.418551
	bot	N/A	1	0.000000
	file-drop	N/A	3	0.000000
	general	IoT	1688	0.394581
	help	N/A	6	0.000000
	sales	N/A	1	0.000000
x0rz.co	announcements	IoT	47	0.350641
	bot	Bot	736	1.018559
	current-affairs	N/A	1	0.000000
	dedicated-servers	IoT	8	0.446491
	donations	N/A	2	0.000000
	faq	IoT	1	0.341198
	how-this-works	N/A	2	0.000000
	lobby	IoT	20011	0.255184
	lounge	IoT	11104	0.293753
	memes	Hacking	105	0.144768

Continued on next page

Community	Channel	Topic	Messages	Score
	music	Bot	196	0.390252
	news	IoT	50	0.369770
	nsfw	Community	83	0.220678
	prices	IoT	4	0.415081
	read	N/A	1	0.000000
	reviews	IoT	56	0.350567
	suggestions	IoT	24	0.202359
	talk-with-ai	Community	54	0.135498
	voice-chat	Bot	36	0.521034
CYBER NET	for-kids-with-no-mic	Hacking	313	0.200091
	free-ip-stressers	N/A	1	0.000000
	ip-tools	N/A	1	0.000000
	verify	DDoS	89	0.299333
	welcome	Hacking	100	0.242066
	©credits©	N/A	1	0.000000
	□s̊ůp̊p̊o̊r̊t̊□	IoT	422	0.172992
	□general□	Hacking	6454	0.211940
	□□prices□□	DDoS	12	0.363338
	□tech-talk□	Community	305	0.133571
	□forums□	N/A	1	0.000000
	□announcements□	Community	30	0.349744
	□linked-server	N/A	7	0.000000
	□off-topic-chat□	Community	35	0.123109
	□BOT-COMMANDS□	Bot	586	0.460958
	□□MEME-CHAT□□	Marketplace	126	0.190351
DNL4	Scammers	N/A	1	0.000000
	Support	N/A	4	0.000000
	Music Commands	Bot	330	0.719476
	Vouches	N/A	3	0.000000
	Den	DDoS	4233	0.169177
		General	1998	0.322984
	Requests	N/A	17	0.000000
	Admin	N/A	11	0.000000
	Market	IoT	55	0.818623
	ANNC	N/A	5	0.000000
	Memes	General	327	0.136373
		Marketplace	263	0.296071
	Bots	Bot	116	0.103595
		Community	48	0.111826
		N/A	9	0.000000
Running the system	meme	Bot	49	0.131340

Continued on next page

Community	Channel	Topic	Messages	Score
	pryzraky_network	DDoS	286	0.204522
	christmas2k19	DDoS	10	0.258113
	Power Proof	N/A	17	0.000000
	Welcome	N/A	2	0.000000
	Announcements	Community	51	0.203316
	Cyberspace	DDoS	2128	0.093398
		General	4349	0.289574
	□□memes	Community	147	0.099172
	□□suggestions	Community	62	0.125235
	General Chat	DDoS	13290	0.090259
	Sales	Marketplace	11	0.288112
□Mirai Variant & VPN v2.0□	□botnetinfo□	N/A	6	0.000000
	□mainchat□	IoT	401	0.166592
	□welcome□	Marketplace	47	0.211752
	□announcements□	IoT	11	0.396010
	□reviews-vouches□	N/A	2	0.000000
	□-payments-□	Marketplace	79	0.201199
	□prices□	N/A	29	0.000000
	□nsfw□	N/A	8	0.000000
	□dox-releases□	N/A	2	0.000000
□MIRAI VARIANT / VPN□	paypal-cashapp-bitcoin	N/A	4	0.000000
	rep	N/A	1	0.000000
	newfagz	Community	1248	0.156237
	WEBSITE	N/A	2	0.000000
	□□□□□□□	IoT	56	0.402908
	□MAIN-CHAT□	IoT	16923	0.214313
	□□□□□□□□□□	N/A	1	0.000000
	□GIVEAWAYS□	Community	30	0.265417
	□BYE-BITCH□	Community	950	0.111998
	□-ANNOUNCEMENTS□	IoT	79	0.457411
	□PRICES□	N/A	1	0.000000
	□WELCOME□	N/A	1	0.000000
	□self-promo□	IoT	712	0.279597
□ Hacking □	announcements	General	57	0.303901
	app-hacking	General	1053	0.346727
	beginners	General	18943	0.338163
	blue-team	General	857	0.311185
	bot-commands	Bot	44511	0.059928
	certifications	General	468	0.378217

Continued on next page

Community	Channel	Topic	Messages	Score
	cryptography-math	General	2686	0.290569
	ctf-challenges	General	4900	0.279934
	ctf-info-scoreboard	General	1809	0.066053
	ctf-solves	General	2514	0.087652
	general	General	419321	0.299645
	hacking	General	15577	0.426145
	information-rules	General	8	0.312084
	music	General	1399	0.218694
	music-commands	Bot	4397	0.468043
	networking	General	4295	0.326630
	news	General	597	0.377479
	news-meta	General	1002	0.352401
	operating-systems	General	7310	0.370521
	osint	General	1202	0.400083
	other-tech	General	3098	0.372853
	partners	N/A	2	0.000000
	privacy	General	278	0.324662
	programming	General	18541	0.352372
	reverse-engineering	General	2010	0.333786
	sansholidayhack	General	628	0.293792
	social-engineering	General	914	0.310905
	spam-memes	General	41155	0.173522
	suggestions-issues	General	1778	0.410460
	the-void	General	32830	0.174040
	voice-text	Bot	8189	0.109226
	web-hacking	General	3445	0.349039
	welcome	General	16932	0.289164

D

Parameters Clustering

```
doc2vec_args = {"vector_size": 300,
"min_count": 50,
"window": 15,
"sample": 1e-5,
"negative": 0,
"hs": 1,
"epochs": 400,
"dm": 0,
"dbow_words": 1}

umap_model = umap.UMAP(n_neighbors=15,
                       n_components=5,
                       metric='cosine').
                       fit(self._get_document_vectors(norm=False))

# find dense areas of document vectors
cluster = hdbscan.HDBSCAN(min_cluster_size=15,
                          metric='euclidean',
                          cluster_selection_method='eom').
                          fit(umap_model.embedding_)
```

E

Other Topic Models

```
LDA Model:
Topic 0:
[('tag_discord_mention', 13775.668495753298), ('server',
    ↪ 9824.292984513719), ('lol', 7779.529042596922), ('
    ↪ joined', 6486.029527839725), ('shit',
    ↪ 5562.108771747763), ('account', 5036.787521440183),
    ↪  ('tag_url', 4382.364884554793), ('selling',
    ↪ 4274.89280931989), ('dm', 4172.840430558485), ('
    ↪ discord', 4030.106827446759)]
Topic 1:
[('tag_discord_mention', 64707.845204789424), ('lol',
    ↪ 17707.21131842165), ('tag_url', 14641.028192233342)
    ↪ , ('like', 12714.523231857407), ('im',
    ↪ 10585.95621209476), ('know', 8566.277497133071), ('
    ↪ dont', 8349.405681660808), ('yes',
    ↪ 8221.771743211337), ('shit', 7569.074299915103), ('
    ↪ fuck', 6886.958377232008)]
Topic 2:
[('like', 27668.369558165912), ('tag_discord_mention',
    ↪ 20187.13546071013), ('tag_url', 17228.723139076104)
    ↪ , ('lol', 16354.462878937142), ('yeah',
    ↪ 15523.705111978052), ('know', 15236.058204559879),
    ↪ ('want', 11880.249380104973), ('good',
    ↪ 11687.56900814656), ('im', 11588.882533388713), ('
    ↪ need', 11169.125656713857)]
Topic 3:
[('pls', 18294.88412259671), ('tag_discord_command',
    ↪ 12204.2842717167), ('tag_discord_mention',
```

```
    ↪ 9330.411619601791), ('ttag_discord_command',
    ↪ 7565.543416330745), ('work', 7320.3294127820045),
    ↪ ('meme', 5259.670363466513), ('score',
    ↪ 4330.507722776875), ('type', 3882.195570937133), ('
    ↪ tag_url', 3622.4109836992434), ('searching',
    ↪ 3534.5165152944983)]
Topic 4:
[('point', 44570.64488192066), ('welcome',
    ↪ 7818.696625332437), ('hacking', 7482.472277920165),
    ↪  ('work', 7211.550909541374), ('tag_discord_command
    ↪ ', 6592.054007885685), ('channel',
    ↪ 6514.65987611412), ('access', 6495.612275859732),
    ↪ ('coffee', 5975.913070063833), ('read',
    ↪ 5650.611629911337), ('server', 5331.767328602746)]
Topic 5:
[('gay', 54884.77189816795), ('like', 24984.69719872236),
    ↪  ('lol', 23028.941434151096), ('tag_url',
    ↪ 21902.854423470777), ('im', 18015.25025520096), ('
    ↪ tag_discord_mention', 17003.83257830591), ('shit',
    ↪ 16055.872315363227), ('know', 12622.638461861874),
    ↪ ('fuck', 12012.215876928942), ('ur',
    ↪ 11868.562166092823)]
Topic 6:
[('tag_url', 58486.39696483381), ('join',
    ↪ 19280.911036810652), ('leave', 15389.655209009932),
    ↪  ('file', 5396.339573272438), ('stresser',
    ↪ 3814.4990275031905), ('__', 3068.3544743749785), ('
    ↪ invited', 3047.1589121451993), ('plan',
    ↪ 2922.827610323238), ('server', 2835.8544356343796),
    ↪  ('best', 2760.7526084789374)]

====================

NMF Model:
Topic 0:
[('gay', 199.0864584331368), ('like', 84.73613131763835),
    ↪  ('lol', 69.6702995283417), ('tag_url',
    ↪ 68.59075108286345), ('im', 56.69649552337537), ('
    ↪ tag_discord_mention', 54.319232241137804), ('shit',
    ↪  51.56401757340244), ('know', 43.187741471898526),
    ↪ ('fuck', 35.10938158313921), ('got',
    ↪ 34.07488828467149)]
Topic 1:
[('point', 210.81448532171834), ('ctf', 8.45291832348087)
    ↪ , ('hacking', 8.441770170171022), ('coffee',
```

```
    ↪ 8.424204920874285), ('50', 8.418950564700998), ('
    ↪ tag_url', 8.149084538452255), ('20',
    ↪ 5.27609781702803), ('svelte', 4.69070709062453),
    ↪ ('45', 4.569131533093522), ('read',
    ↪ 4.494910140408933)]
Topic 2:
[('tag_discord_mention', 115.23124133079817), ('like',
    ↪ 52.13757607836196), ('lol', 42.8009355506594), ('
    ↪ know', 28.574425083521835), ('im',
    ↪ 27.976011928609243), ('yeah', 26.37064119572827),
    ↪ ('oh', 26.346156728026653), ('good',
    ↪ 23.52746259339257), ('shit', 21.244557453730774),
    ↪ ('yes', 20.715482153667867)]
Topic 3:
[('tag_url', 167.37865643173967), ('stresser',
    ↪ 13.54749085870771), ('raided', 13.477967095166429),
    ↪  ('plan', 12.085995653396889), ('server',
    ↪ 10.709171766731574), ('using', 10.390532160526725),
    ↪  ('want', 10.195561679339626), ('ill',
    ↪ 10.182977843140822), ('dm', 10.088567056612021),
    ↪ ('300', 9.718147193133364)]
Topic 4:
[('pls', 91.5447783901983), ('work', 44.4226017872503),
    ↪ ('ttag_discord_command', 43.13456449733872), ('
    ↪ tag_discord_command', 28.632933906676083), ('
    ↪ tag_discord_mention', 22.78501865409287), ('meme',
    ↪ 22.11090427219783), ('score', 21.67034398325419),
    ↪ ('total', 20.98548451478641), ('type',
    ↪ 20.9334148344561), ('coin', 20.891583578545585)]
Topic 5:
[('join', 107.30978884006204), ('tag_url',
    ↪ 106.89195769788714), ('leave', 106.08930423799823),
    ↪  ('tag_discord_mention', 1.5946828076989983), ('afk
    ↪ ', 0.6958550556646923), ('poor',
    ↪ 0.6499115048310555), ('free', 0.5557362803889617),
    ↪ ('insane', 0.4762393712744276), ('nitro',
    ↪ 0.4721679428752906), ('power', 0.46057056301691773)
    ↪ ]
Topic 6:
[('welcome', 40.63892489675287), ('access',
    ↪ 31.350995632411866), ('channel',
    ↪ 30.532792901024813), ('server', 27.75339479038825),
    ↪  ('tag_discord_command', 26.532499251295956), ('
    ↪ hacking', 24.110521837499732), ('stop',
    ↪ 21.092140088027044), ('information',
```

```
    ↪ 20.34625971282512), ('read', 20.23115099742832), ('
    ↪ rest', 19.873721282612014)]

====================

LSI Model:
Topic 0:
[('gay', 0.5020560083477508), ('tag_discord_mention',
    ↪ 0.2904125728835589), ('like', 0.2853078228633416),
    ↪ ('tag_url', 0.25980657001752655), ('lol',
    ↪ 0.2333897094481212), ('im', 0.18091404847978185),
    ↪ ('shit', 0.15907502669701942), ('know',
    ↪ 0.14794836931722075), ('got', 0.1145393388838604),
    ↪ ('fuck', 0.11284592582750551)]
Topic 1:
[('point', 0.9772910229127585), ('tag_url',
    ↪ 0.04378394728874014), ('hacking',
    ↪ 0.04279912858496331), ('ctf', 0.042302529846693394)
    ↪ , ('coffee', 0.04058250718939045), ('50',
    ↪ 0.039333911872266106), ('tag_discord_mention',
    ↪ 0.028949180575059002), ('pls',
    ↪ 0.024470987595167425), ('20', 0.023983650008046052)
    ↪ , ('read', 0.02261287631471683)]
Topic 2:
[('tag_discord_mention', 0.5747843368575517), ('
    ↪ tag_discord_command', 0.13132536836726108), ('oh',
    ↪ 0.10902352180124668), ('want', 0.08797034464020093)
    ↪ , ('use', 0.087498987407439), ('lmao',
    ↪ 0.07591439163190979), ('yeah', 0.0750081849317589),
    ↪  ('server', 0.07464976790684337), ('work',
    ↪ 0.06703841349696982), ('like', 0.06546405748897524)
    ↪ ]
Topic 3:
[('tag_url', 0.8050127563858998), ('join',
    ↪ 0.379810573104202), ('leave', 0.3544924180524418),
    ↪ ('stresser', 0.041540519173352886), ('raided',
    ↪ 0.0412880979112052), ('plan', 0.036910021493047236)
    ↪ , ('stressers', 0.02939909064377314), ('300',
    ↪ 0.028625579966693532), ('dm', 0.02617096963559717),
    ↪  ('boot', 0.025482537514259412)]
Topic 4:
[('pls', 0.625016383522316), ('ttag_discord_command',
    ↪ 0.2987852017214009), ('work', 0.2783587755716036),
    ↪ ('tag_discord_command', 0.21563893272498824), ('
    ↪ type', 0.16274883392650716), ('score',
```

```
    ↪ 0.14975331136128872), ('coin', 0.14464490975054542)
    ↪ , ('total', 0.1444053821394942), ('meme',
    ↪ 0.14329902755482413), ('gay', 0.11642100631717307)]
Topic 5:
[('leave', 0.582515101676924), ('join',
    ↪ 0.548621760588176), ('tag_discord_mention',
    ↪ 0.19260619019227365), ('gay', 0.10249386025898112),
    ↪  ('point', 0.024814611258466075), ('yes',
    ↪ 0.023633031664524965), ('lol',
    ↪ 0.022942479878896055), ('ur', 0.016722495526676658)
    ↪ , ('nigga', 0.01630358576181321), ('im',
    ↪ 0.01593584017428806)]
Topic 6:
[('welcome', 0.31817442715819416), ('access',
    ↪ 0.24422529098992252), ('channel',
    ↪ 0.2344310490268881), ('server', 0.1855492173127769)
    ↪ , ('hacking', 0.18310670175578708), ('read',
    ↪ 0.15816610700940661), ('rest', 0.1577434311621947),
    ↪  ('stop', 0.15513926435552275), ('information',
    ↪ 0.15434792122081786), ('rules',
    ↪ 0.15207665367896508)]
====================
```

F

DDoS Attacks Country Scatter

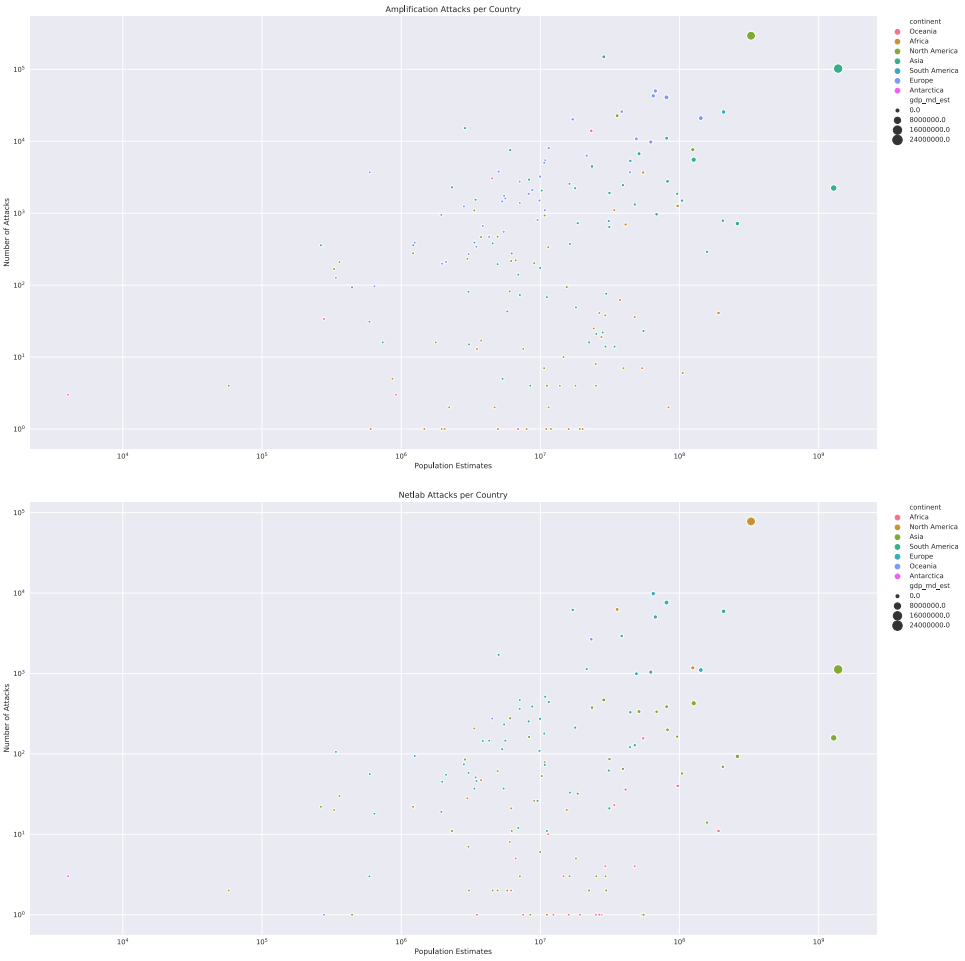


Figure F.1: Country Comparison of number of unique victims (IP) attacked

G

Victimization Supplemental Figures

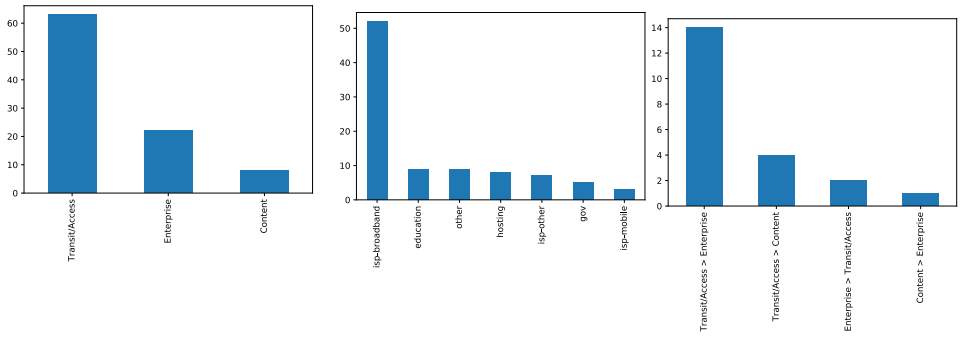


Figure G.1: Statistics for manual sampling

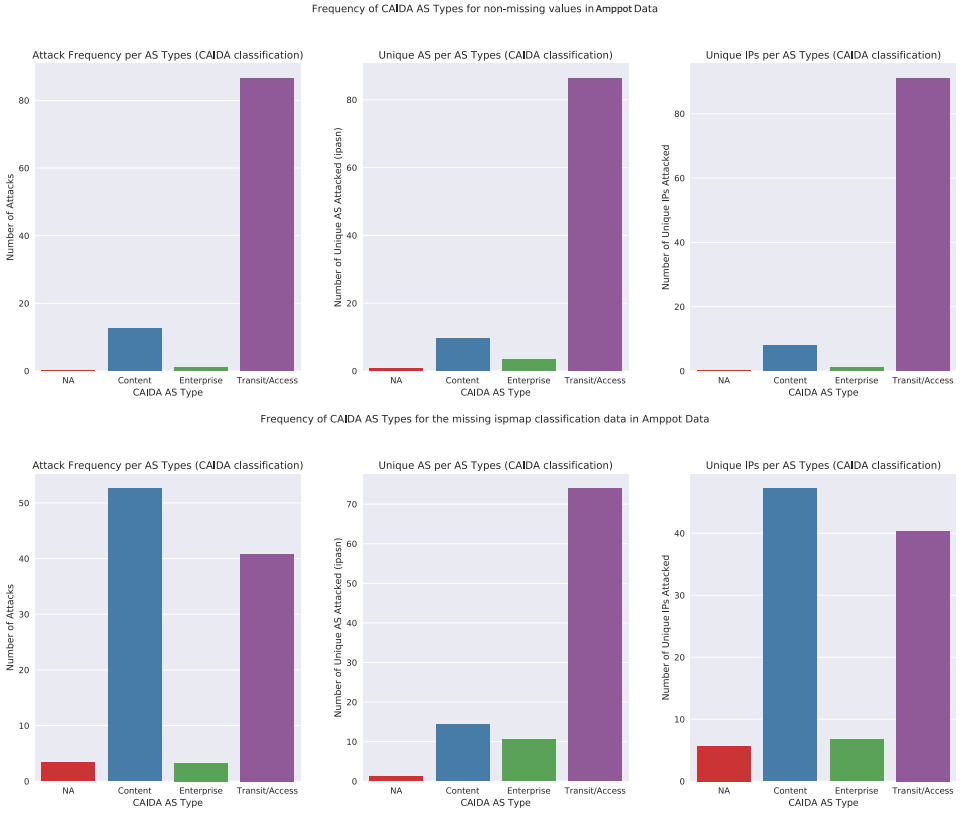


Figure G.2: Ispmap classification victims comparing non-missing and missing classification for amplification attacks

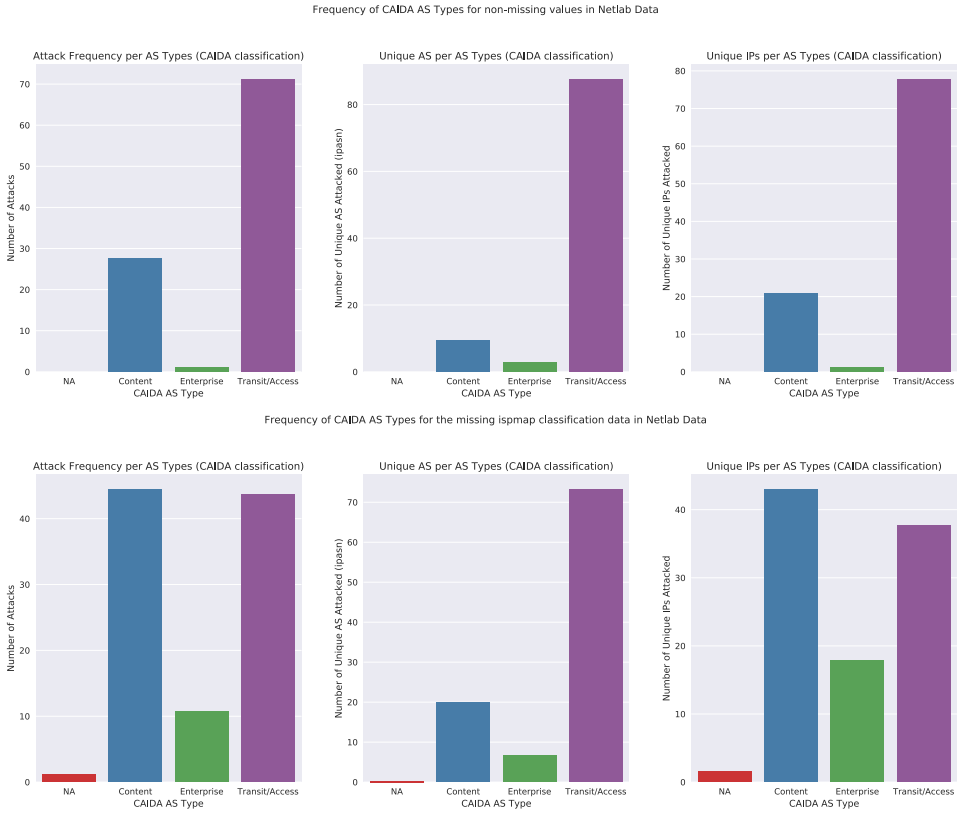


Figure G.3: Ispmap classification victims comparing non-missing and missing classification for IoT attacks

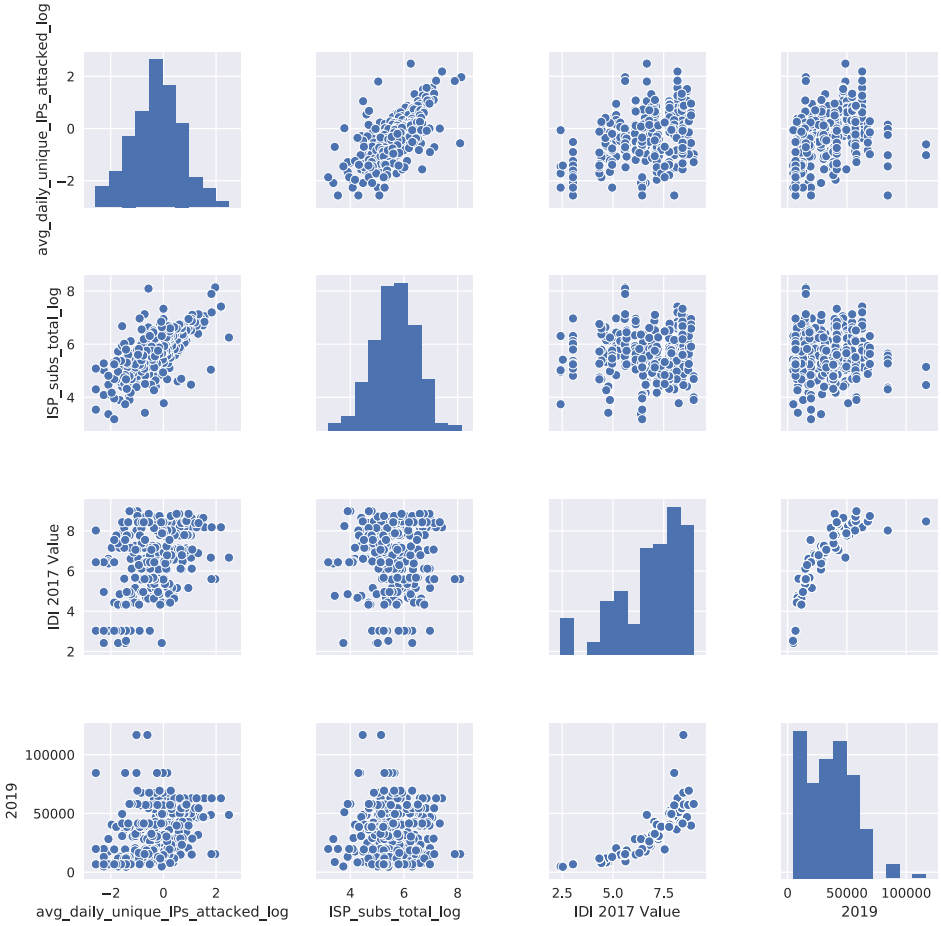


Figure G.4: Pairplot GLM Factors Amplification Attacks

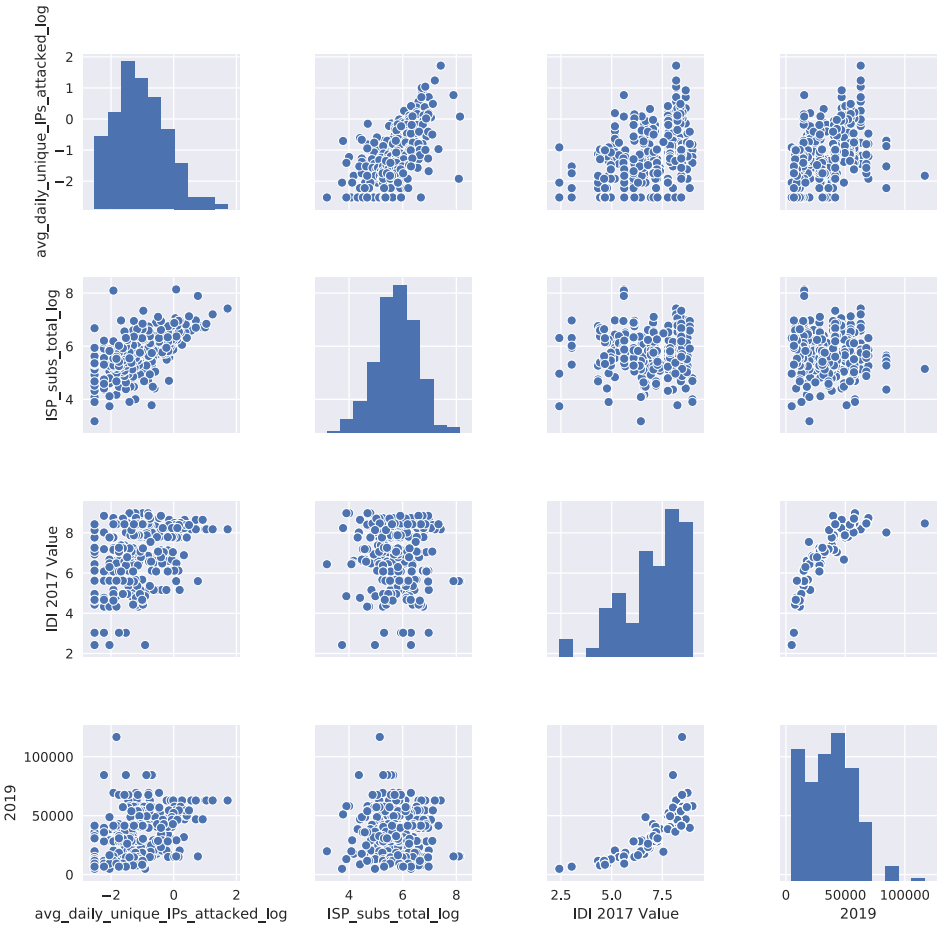


Figure G.5: Pairplot GLM Factors IoT Attacks

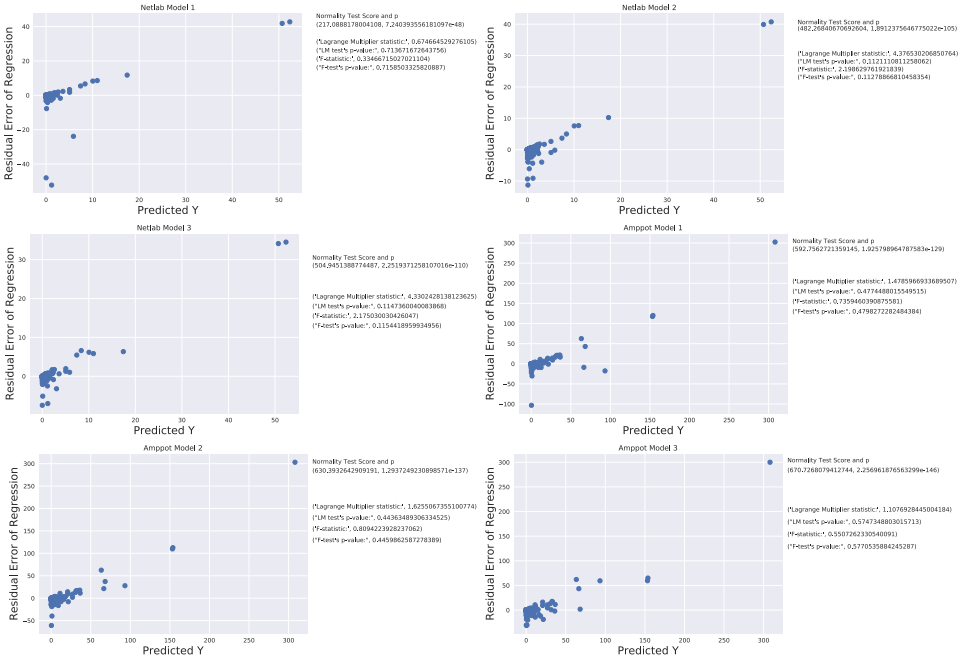


Figure G.6: GLM Model Assumptions

	AmpPot Domains	Mirai Domains
0	['00000000000000000000000000...	['0000000.in.', '000:\\032not\\032found-2.reas...
1	['0000.fo', '02567367823536273.com', '10server...	['0000.fo.', '000.xram.cc.', '002.xram.cc.', '...
2	['023789465298359283685922632...', '0dayexplo...	['00534bdns4p5.ts3.ba.', '08te9ozo.idea-rks.co...
3	['0btemoslab.com', '0vk.com', '0x10.pro', '0x3...	['013.0a9.myftpupload.com.', '027.779.myftpupl...
4	['0dayexploits.net', '1337tools.in', '5v5.eu',...	['01.bucharest.blazingfast.io.', '102.blazingf...
5	['100gbit.com', '100gbit.eu', '100gbit.info', ...	['0anel.crowland.ro.', '1069ef8f941d860c4075cb...
6	['16mb.com', '96.lt', 'a7design.al', 'abdalade...	['1000freespins.nl.', 'app.zrit.nedbase.net.',...
7	['27mc-radio.nl', '27mcradio.nl', 'analytix.nl...	['145-130-203-104.fixed.kpn.net.']
8	['7295632.com', '7295652.com', '7295672.com', ...	['1986kf88.xyz.', '8bdeaa8e5846e48ef916b1718c0...
9	['7teen8.nl', 'advensa.nl', 'bedrijfskeuringe...	['217-123-14-192.cable.dynamic.v4.ziggo.nl.']
10	['8test.nl', '9test.nl', 'floorhosting.nl', 'f...	['45-61-142-158.ams.priv.octovpn.net.', 'ninja...
11	['aag-ey.com', 'bxjsipdwvzvrrbsqgrzb.online', ...	['45-61-142-171.ams.priv.octovpn.net.']
12	['angelosreef.nl', 'beckerscms.nl', 'brict.it'...	['45-61-142-177.ams.priv.octovpn.net.', 'octos...
13	['anovuurwerk.nl', 'hurricanewerkkleding.nl', ...	['6hgu.bytebx.com.', 'nbg83.zbigz.com.', 'nfo....
14	['asuscomm.com']	['82-72-111-56.cable.dynamic.v4.ziggo.nl.']
15	['banxu78.com', 'danyu98.com', 'korimscdn.com'...	['83-86-123-229.cable.dynamic.v4.ziggo.nl.']
16	['blazingfast.io', 'evobattle.com']	['83-86-18-46.cable.dynamic.v4.ziggo.nl.']
17	['boatrep.xyz', 'botnetrep.xyz']	['acnh.world.', 'acosamc.com.', 'activatelearn...
18	['chaorenmimi.com']	['admin.dis.cool.', 'euams1.vpn.courvix.com.'...
19	['cycry.com', 'geldbild.com', 'investigoo.com'...	['a.heavy-r.com.', 'devtestnew.heavy-r.com.', ...

Continued on next page

	AmpPot Domains	Mirai Domains
20	['ddns.me']	['aifirstfit-dev-raise-memory.unicorn-platform...
21	['dyndns.org', 'xboot.pro']	['amazon.fr.18951704607820738521...
22	['elitehosting.nl', 'kvsolutions.nl']	['amsterdam-1.octovpn.net.']
23	['hibbohotel.nl', 'servitus.nl', 'writhingdoor...	['api.badmasti.us.', 'file.badmasti.com.', 'fi...
24	['inspectors.nl']	['api.tejiaoyun.com.', 'rianarlkmk11.r-e.kr.',...
25	['kbitholding.nl']	['bele-onjo.com.', 'ceb.gbl-europe.com.', 'cle...
26	['kbitholding.nl', 'lifehosting.net', 'writhin...	['bobins.xyz.', 'daddy-cool11.dyndns.org.', 'hi...
27	['korimscdn.com', 'otherse.com']	['brazil1610.discord.gg.', 'brazil18.discord.g...
28	['korimscdn.com', 'otherse.com']	['brazil2186.discord.gg.', 'brazil673.discord....
29	['korimscdn.com', 'otherse.com']	['brazil286.discord.gg.']
30	['korimscdn.com', 'otherse.com']	['brazil288.discord.gg.']
31	['korimscdn.com', 'otherse.com']	['brazil28.discord.gg.', 'brazil6686.discord.g...
32	['korimscdn.com']	['brazil306.discord.gg.']
33	['kvsolutions.nl']	['brazil389.discord.gg.', 'brazil5271.discord....
34	['kvsolutions.nl', 'lighthappy.net']	['brazil4285.discord.gg.', 'brazil44.discord.g...
35	['kvsolutions.nl']	['brazil704.discord.gg.', 'brazil849.discord.g...
36	['kvsolutions.nl']	['bst-1b34a97d-7d7c-4c3d-8999-97f770fcb3d2.bas...
37	['lifehosting.eu', 'lifehosting.nl', 'unfolded...	['bullstresser.to.', 'dienstleistungen-rehfuss...
38	['lighthappy.net']	['c60.gobust.net.']
39	['lighthappy.net']	['cdn2.mobideck.net.']
40	['lighthappy.net']	['content.axc.nl.', 'production.axc.nl.', 'ver...
41	['lydie-en-tom.com']	['dc-337ba681140e.livingmine.net.', 'prod-live...
42	['mezy.wtf', 'sausagecasserole.eu']	['dns0905d67d-04dc-4017-a95c-402142ce8290-azur...
43	['mezy.wtf', 'subby.xyz']	['dnsb7dc5329-beba-4f1b-a399-222f0acaba84-azur...

Continued on next page

	AmpPot Domains	Mirai Domains
44	['mtwaverleypreschool.org.au', 'sjifri-days.tk']	['eu-ams.vpn.courvix.com.']
45	['noobhotel.nl']	['eu-ams.vpn.courvix.com.']
46	['notaion.com']	['eu-central8162.discord.gg.']
47	['ripe.net']	['geertwilders.be.', 'geertwilders.eu.', 'geer...
48	['ripe.net']	['halt.nl.', 'www.halt.nl.']
49	['ripe.net']	['ict.sgboz.nl.', 'info.mollercollegestb.nl.',...
50	['ripe.net']	['idb-mbx.deonderwijsspecialisten.nl.', 'mail....
51	['santa3.ru', 'santa4.ru']	['iotnetsjokerbotnet.xyz.']
52	['stressthem.to']	['jhasdjahsdjasfkdaskdfasbot.nigg...
53	['writhingdoor.com']	['lists.smartschool.be.', 'mail.smartschool.be...
54	['writhingdoor.com']	['mail.dentalsupportuk.com.', 'mail.shillam.me...
55	['writhingdoor.com']	['mx2.ravelijnstb.nl.', 'mx2.sgboz.nl.', 'mx2....
56	['writhingdoor.com']	['nieuw-holland.rhmc.nl.', 'udp-method.xyz.', '...
57	['ziggo.nl']	['sbz061e64e2ec05.westeurope.cloud...
58	['ziggo.nl']	['sbz101071f581c4.westeurope.cloud...
59	['ziggo.nl']	['sbz366ba33ee657.westeurope.cloud...
60	['ziggo.nl']	['sd-119491.dedibox.fr.']
61		['styralphlaurenoutletuk.com.', 'www.styralphl...
62		['us-central781.discord.gg.']
63		['www.iq.hdl.990720-hdl-hc-dev.dev-aws.hanaclo...