DDoSDB

*Collecting, Transforming, Applying, and
Disseminating DDoS Attack Knowledge*

1. Problem Definition and Challenges

Distributed Denial of Service (DDoS) attacks are one of the most severe and frequents cyber crimes of the decade. In 2015, small and medium size companies reported losing 53 thousand Dollars on average as a result of a DDoS attack, while big enterprises report 417 thousand Dollars of damage. Besides single organizations, DDoS attacks affect anyone connected to the Internet. For example, at the end of 2016, the attack against a company named Dyn strongly affected hundreds of other companies and millions of Internet users. Among the list of companies that reported outage because of the attack on Dyn are: Amazon.com, Twitter, Reddit, GitHub, BBC, CNN, Comcast, HBO, Netflix, PayPal, and Visa. The attack against Dyn is only one attack among hundred of others noticed on a daily basis. In the last eight years, the situation only got worse. The number of attacks doubled compared to the immediately previous year, the types of attacks and their power evolved together with the evolution of systems connected to the Internet (*e.g.,* Internet of Things devices).

DDoS attack is not a new problem, and there are efforts from all sides to mitigate the problem. For example, there are more than 33 thousand academic papers found via Google Scholar using the query “ddos attack”. In the last years, there was a blooming of network security companies worldwide. Governments have improved their laws against cyber crime. **The main problem of mitigation initiatives is that although these parties, *i.e.,* academia, private and public organizations, have improved the state-of-art on addressing the DDoS attack problem, they work isolated, barely together**. A collaborative approach would clearly benefit all these parties and substantially improve the mitigation of DDoS attacks. For example, with such approach academia would have access to actual and recent attack data, which would enable them to develop, test and improve solutions based on practical problems (not only theoretical). Private organizations would get the benefit of deploying successful academic solutions on their business. Governments would benefit with detailed information of misused devices and even information on attack perpetrators. Even Internet users would be able to collaborate by checking if their own devices are misused for performing attacks and receiving pieces of advice on how to protect their systems. Aiming at making this multidisciplinary and collaborative approach a reality we propose the DDoSDB project. Our main goal is the following.

Project goal: to develop a large-scale repository and an integrated set of tools for enabling the exchange of DDoS attack knowledge between Internet users, academia, and public & private organizations towards the mitigation of DDoS attacks.

In the context of DDoSDB project, the definition of *attack knowledge* is any type of information related to DDoS attacks. This knowledge can range from any type of data traces (packet, flow, and log-based) to a list of (academic) solutions for prevention, detection, and mitigation of attacks. Attack knowledge also ranges from best practices for preventing a particular type of attack to actual feedbacks from the network operators and the security community. To meet our main goal we identify four scientific challenges:

1. ***How to guarantee the privacy of our collaborators are willing to share attack data of themselves or/and of their customers (that suffered attacks)? and what approach best guarantees their privacy and still provides meaningful information to be further used?*** Although there are hundreds of academic works on data privacy, data sanitization, and anonymization, there also another set of papers that shows how to overcome these works. Organizations still have doubt what *de facto* would

guarantee their privacy and the privacy of their customers. Added to this mistrust there is a lack of solutions that filters normally from attack traffic, making even more challenging attack targets/victims to provide data free from potential privacy sensitive information.

2. ***How to summarize attack data in a unique, valuable, and smallest set of fields?*** Each DDoS attack contains a large amount of data. It is impractical to store and process everything from an attack trace. In one hand, by summarizing information, we run the risk of throwing away meaningful information. In another hand, if our summary is such able to uniquely identify an attack (*i.e.,* fingerprint) then it can be used for future detection purposes.
3. ***Which adaptations must be made to apply our attack fingerprints in attack/Intrusion Detection Systems (IDS)?*** Although the most used IDS (*e.g.,* SNORT, SURICATA, and BRO) were not conceptually designed for detecting DDoS attacks, to increase the change of usage and the impact of our attack fingerprints it is important for us to propose a suitable adaptation to IDS.
4. ***Which functionalities are the most beneficial to the expected and intended DDoSDB users?*** It is very challenging to please everyone. Each one of our intended user (*i.e.,* Internet users, academia, and public & private organizations) has their own requirements and some of these requirements conflict among them. Therefore we must be able to address the requirements of our intended uses and minimize conflict. Our end goal by addressing this challenge is to attend a wide range of users, which will provide more DDoS attack knowledge, and will lead to more successful approaches for preventing, detect and mitigate attacks.

2. Approach & Methodology

In this section each one of the four challenges previous described is addressed in a different step. The first, named **Collect**, aim at addresses the first challenges on privacy issues of our collaborators. Then, in the **Transform** step we aim at address the generation of attack fingerprints given different types of attack traces, *i.e.,* the second challenge. The third step, named **Apply**, aims at address the adaptation of existing IDS for working with our attack fingerprints. Finally, and intercalating all the other steps we have the step named **Disseminate**, which aim at address the requirements of our intended users. We connect, aggregate and made available all the outcome of these four steps in a unique big data infrastructure, called **DDoSDB**. Following we provide details on each step of our approach. Figure 1 depict our four steps.

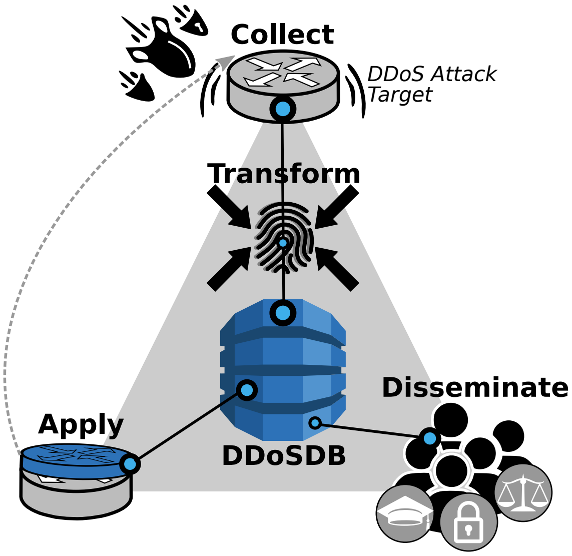


Figure 1: Project approach parts.

Collect

To address the privacy issues of our collaborators we intend to investigate, test and improve existing metrics for data sanitization (anonymization) towards find the most suitable applied for DDoS attack data sharing. Additionally, we intend to actively interact with Internet providers and big enterprises, which are the main expected collaborators, they are most frequently under attack and can provide the largest amount of attack data for our initiative. One of the main expected outcomes of this step is an open source tool that sanitizes attack data (packet, flow and log-based), which is intended for be used at the collaborators infrastructure. This last characteristic aim at provides additional guarantees to collaborator that intends to share attack data absent of privacy sensitive information.

Transform

To address the generation of attack fingerprints we intend to determine which network fields collected in different attack traces (packet-based, flow-based, and log-based) are sufficient for, first, classify a generic attack. Afterward, we analyze which remaining network field can be used for differentiate attacks in the same class. We then compare attacks in a same class to find unique identifiers. Finally, we investigate whether the remaining set of fields also meet the requirements of expected DDoSDB users. One of the main expected outcomes of this step is an open source tool that receives (input) any type of network data and transforms this data into a unique set of characteristics (*i.e.,* fingerprints), which can be further queried, compared, and used for detection and mitigate solutions;

Apply

To address the challenge of our attack fingerprint usage we will investigate how to adapt our fingerprints to be used in the most common IDS used in network operations (*e.g.,* BRO, Suricata, and Snort). In addition to that we will compare the performance of existing tools and propose improvements. Therefore, in this step of our approach, we expect to improve the state-of-the-art of signature-based DDoS attacks detection and allow users to use all attack fingerprints available at DDoSDB adapted to their IDSs (for further collection).

Disseminate

To address the challenge of attending our DDoSDB users requirements we will first perform a survey of functionalities expected by potential users (*i.e.,* Internet users, academia, and public & private organizations). Then, we will design and deploy an initial version of the database infrastructure. This infrastructure must to be adaptable and extensible to adjust modifications along the project. After each one of the previous described three steps of our approach we intend to re-visit our database infrastructure to better adapt the new functionalities with requirements of our users. The outcome of this step is a set of (visual) interfaces to intended DDoSDB users. Examples of interfaces are the following.

- For **any Internet user** we expect to have a public interface for querying and download attack traces. Example of queries is to check if an IP address was involved in an attack. If DDoSDB finds any attack record related to this hypothetical IP, then it can show a list of solutions for preventing the system behind this IP been misused in future attacks;
- For **academic researchers**, on top of enabling the (download and) usage of actual attack data, we aim at enabling researchers to advertise their published approaches & results to facilitate others to find solutions for specific attack types;
- For **network security groups** (*e.g.,* Computer Security Incident Response Team - CSIRT), we intend to enable feeds and alarms to notify every time that DDoSDB receives an attack involving devices managed by them. As a consequence of this feature we expect that infected and misused devices would be quicker isolated and solved by CSIRTs;

- For **law enforcement agencies**, we aim at facilitating find attack perpetrators based on the attacks in DDoSDB and support prosecution cases with pieces of evidence.

3. Relevance & Application of Knowledge

DDoS attack is an unquestionable relevant problem to our society. Reports show this problem will become even worse in the coming years. In this context DDoSDB is a unique approach that aims at *concentrating in one place all the knowledge on DDoS attack prevention, detection and mitigation* and facilitate interested people to work together improving the overall state-of-the-art.

There is a wide list on how DDoSDB is relevant for our society and how it strongly impacts the way to address attacks. First of all, DDoSDB enables *a remarkable step towards the cooperation between academia, public and private organizations*. Furthermore, the auxiliary tools developed along the project will be *open source* to incentive their usage and the evolution.

DDoSDB as a public infrastructure for knowledge exchange is extremely relevant for *facilitating the advertisement and usage of prevention, detection, and mitigation approaches* against DDoS attacks. From a scientific point of view the DDoSDB project will *answer open questions with strong validations and practical examples of usage*, such as how to sanitize attack data? and how to fingerprint DDoS attacks? In addition to this contribution, DDoSDB is a unique solution that integrated and *improves the state-of-the-art on attack data collection, sanitization and sharing*.

From a social point of view DDoSDB *facilitates anyone to actively participate on the mitigation of DDoS attacks* by either collaborating with attack traces or by receiving advices on how to prevent misused systems to be further used for attacks. From an educational point of view DDoSDB provides a unique *opportunity for new network security students and researchers to understand the actual and most recent characteristics of attacks* (and the already existing solutions).

By providing DDoSDB attack fingerprints, DDoSDB effectively contributes for *improving the early detection and mitigation of attacks*. Overtime, as more collaborators join DDoSDB as more (new) attacks will be fingerprinted and *higher becomes the chances to early detect existing types of attacks*. If any collaborator suffers a new attack then DDoSDB will spread the knowledge *avoiding anyone to be affected by such attack in the future*. Overall, based on the gain that collaborators would benefit, **we envision that DDoSDB will be the biggest public database with downloadable DDoS attack traces in the world**.

4. Expected and Preliminary Outcomes

In this section we list the expected outcomes related to each one of the four part of our approach (described in § 2).

- In the **Disseminate** part we expect to deliver the design and deployment of the DDoSDB infrastructure added to visual interfaces and functionalities to attend different types of users;
- In the **Collection** part we expect to deliver an open source tool for filter and sanitize attack data, *i.e.,* packet, flow and log-based;
- In the **Transform** step we expect to deliver an open source tool for attack fingerprints generation given attack network traces;
- In the **Apply** part we expect an open source tool for converting our attack fingerprints in signatures to be used in IDS systems;

In addition to these four outcomes, we expect to publish the scientific content of each outcome in high-profile academic venues in the fields of networking, network measurements and security, such as the Internet Measurement Conference (IMC), ACM Conference on Special Interest Gruoup on Data Communication (SIGCOMM), the Internet Society Network and Distributed System Security Symposium (NDSS) and the

Conference on Computer and Communications Security (CCS); and journals like Transaction on Networking and ACM Computer Communication Review.

We already started to (indirectly) work on DDoSDB. In 2016 we have helped a company called *Radically Open Security (ROS)* on the automated collection of DDoS attacks mechanism. Such mechanism is an outcome of a project named OPEN SOURCE ANTI-DDOS SOLUTION PROJECT (OSAS), granted by SIDN fonds (Call 1 - 2015). This method is already an preliminary step for the **Collection** part of our project. We intend to use the output of this mechanism as input of our filtering and sanitizing tool.

Last year, 2016, we also developed `http://ddosdb.org/` which is a Website with hundred DDoS attacks and an alfa prototype of our future DDoSDB. The attacks presented on this Website were automatically classified with a methodology developed for the *Nationale Beheersorganisatie Internet Providers (NBIP)* in the context of their project named PATROONHERKENNING DDOS EN BOTNETS. This project was also granted by SIDN fonds (Call 1 - 2015). This classification methodology is already an preliminary step for the **Transform** part of our project. We intend to use the output of this classification as initial step towards attack fingerprinting.

Overall, these two preliminary results highlight our successful collaboration with ROS and NBIP on a topic that already contributes to the DDoSDB project. Both organizations are also members of the DDoSDB project, described in the next section.

5. Membership Team & Key Publications

The team members of this research proposal are composed aiming to address the multidisciplinary characteristic of the project. Therefore, we count on support of specialists on DDoS attacks and on (big) data infrastructures. We also count on the practical experience of a network security company and the extensive experience of an organization that protects Internet providers against attacks. The members are described bellow followed by the list of publication that are related to the project.

- **M.Sc. José Jair C. de Santanna** (`http://jairsantanna.com`) is a Ph.D. researcher concluding his thesis in the first semester of 2017. He also acted as technical adviser of the Dutch National Cyber Security Center and the Scientific Research & Documentation Center and the Japanese Ministry of Internal Affairs and Communication, both on the topic of DDoS attacks. Among relevant papers for this project we highlight:
  - J.J. Santanna, R. van Rijswijk-Deij, A. Sperotto, R. Hofstede, M. Wierbosch, L. Granville, and A. Pras. *Booters - An Analysis of DDoS-as-a-Service Attacks* (Inproceeding) IFIP/IEEE International Symposium on Integrated Network Management (IM), 2015.
  - D. Douglas, J.J. Santanna, R. de O. Schmidt, L. Granville, A. Pras. *Booters: Can Anything Justify Distributed Denial-of-Service (DDoS) Attacks for Hire?* (Journal Article) Journal of Information, Communication & Ethics in Society (JICES), 15 (1), 2017.
  - J.J. Santanna, R. de O. Schmidt, D. Tuncer, J. de Vries, L. Granville, A. Pras. *Booter Blacklist: Unveiling DDoS-for-hire Websites* (Inproceeding)International Conference on Network and Service Management (CNSM). 2016.
- **Prof. Dr. Ir. Aiko Pras** (`http://wwwhome.cs.utwente.nl/~pras/`) is professor and research leader of Design and Analysis of Communication Systems group, which is one of the few research groups specialized in DDoS attacks in the Netherlands. Among relevant papers for this project we highlight:
  - A. Pras, J.J. Cardoso de Santanna, J. Steinberger, and A. Sperotto *DDoS 3.0 ? How Terrorists Bring Down the Internet*. Proceedings of the 18th International GI/ITG Conference on “Measurement, Modelling and Evaluation of Computing Systems” and “Dependability and Fault-Tolerance”, MMB & DFT 2016;
  - R. van Rijswijk-Deij, M. Jonker, A. Sperotto, and A. Pras. *The Internet of Names: A DNS Big Dataset*. Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication, SIGCOMM 2015;

- **Dr. Djoerd Hiemstra** (http://wwwhome.cs.utwente.nl/~hiemstra) is an associate professor at the DataBase group at the University of Twente, and chair of the IGS/CTIT Twente Data Science Center. These two organizations that have a large expertise in database infrastructure also promise to support our project on the development of DDoSDB infrastructure. Among relevant papers for this project we highlight:
  - D. Hiemstra. *De kracht van Big Data: Slimme modellen afgetroefd door eenvoudige modellen en heel veel data*. I/O Vivat, Inter-Actief 30(1), ISSN 1389-0468, pages 30-32, 2014;
  - D. Hiemstra. *Eenvoudige modellen en Big Data troeven slimme modellen af*. STAtOR 14(3-4), Vereniging voor Statistiek en Operationele Research, ISSN 1567-3383, pages 24-26, 2013;
  - P. Daas, B. Braaksma, R. Aly, Y. Engelhardt, D. Hiemstra, R. *Big Data Masterclass and DataCamp 2015*, Technical Report 2016-15, Statistics Netherlands (CBS), 2016.
- **Radically Open Security (ROS)** is the world's first not-for-profit computer security consultancy company. The contact person is its founder and CEO Dr. Melanie Rieback (https://radicallyopensecurity.com/team/MelanieRieback/), which was also an assistant professor in the Computer Systems group at the Vrije Universiteit in Amsterdam. Dr. Rieback promised to support DDoSDB project on the development of open source tools.
- **NationaleBeheersorganisatie Internet Providers (NBIP)** is a Dutch organization that provides the "Nationale anti-DDoS Wasstraat", which is a on-demand protection against DDoS attacks used to protect more than 50 small and medium ISP. The contact person at NBIP is its co-owner Gerald Schaapman (https://www.linkedin.com/in/geraldschaapman).

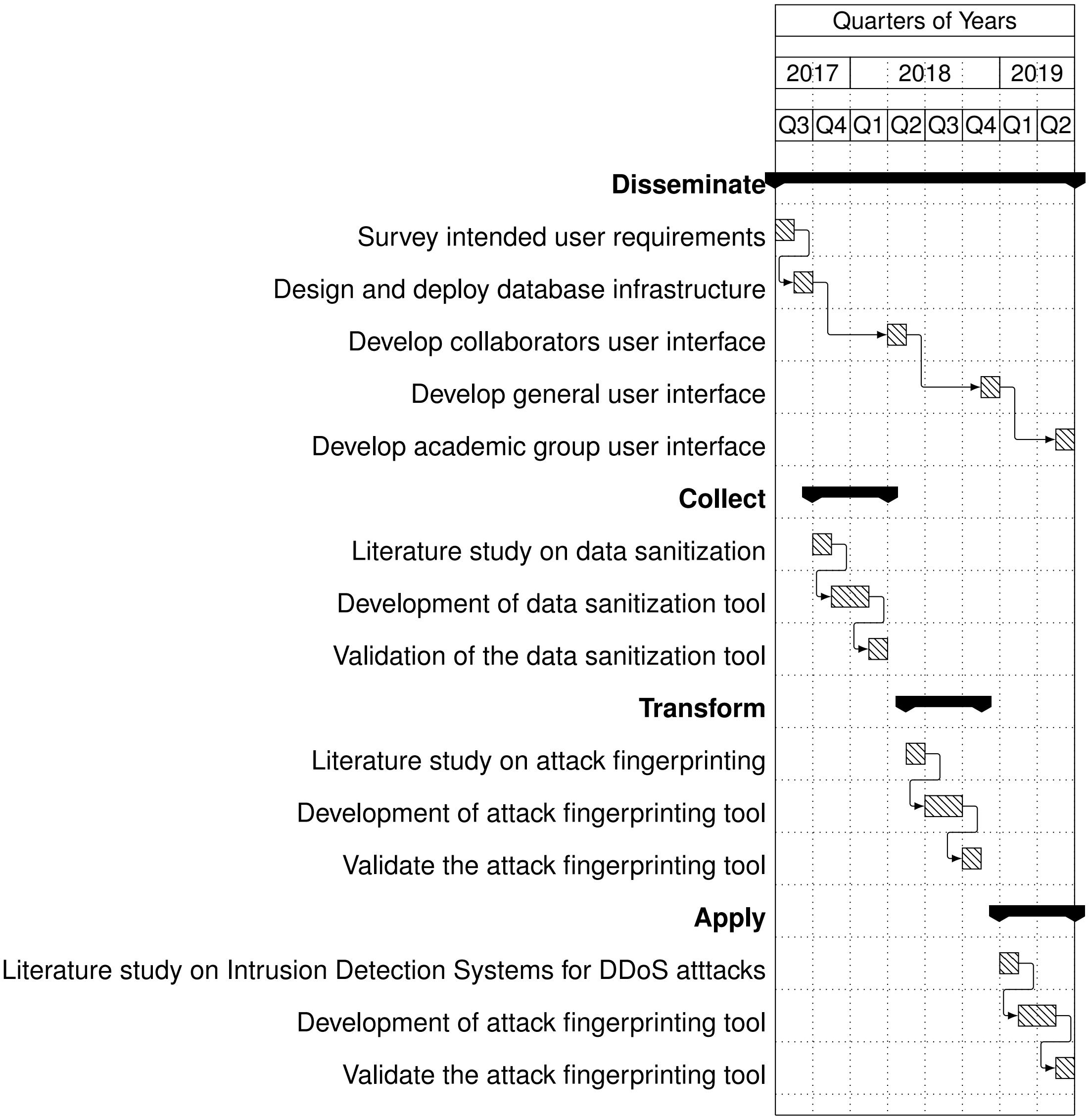
We added support letters of collaborators in the Appendix section. We also summarize in Table 1 the rule of each collaborator in the DDoSDB project.

Table 1: Team member rule.

	Collaborator	**Rule**
1	M.Sc. Santanna	Principal investigator
2	Prof. Dr. Ir. Pras	Project supervisor
3	Dr. Hiemstra	Support on the DDoSDB big data infrastructure
4	ROS	Support on DDoS expertize and development of tools
5	NBIP	Support on DDoS expertize and main attack data collaborator

6. Schedule

In this section we list the activities described in § 2 and add the time planned to address each task. Note that our time planning is divided in fours parts accordingly to our approach steps (*i.e.,* Disseminate, Collect, Transform, and Apply). Note, also, that this planning is merely meant as a guideline, and is not set in stone.



7. Final Remarks

We would like to clarify that the proponent of this project (Mr. Santanna) is finishing his Ph.D. thesis in the upcoming months. Therefore he would be glad to be founded by SIDN fonds to extend his research as a Postdoctoral researcher. Finally, there is an (still) open possibility of changing from the University of Twente to another Dutch University for the PostDoc. The current project supervisor (Prof. Dr. Pras) is aware of this remark and still supports this proposal (see Appendix). The definition of the actual university in which Mr. Santanna will do his PostDoc will be defined before the beginning of this project. Do not hesitate to contact us for any further information.