Towards Distributed Threat Intelligence in Real-Time

Philipp Meyer
HAW Hamburg
philipp.meyer@haw-hamburg.de

Raphael Hiesgen
HAW Hamburg
raphael.hiesgen@haw-hamburg.de

Thomas C. Schmidt
HAW Hamburg
t.schmidt@haw-hamburg.de

Marcin Nawrocki
FU Berlin
marcin.nawrocki@fu-berlin.de

Matthias Wählisch
FU Berlin
m.waehlisch@fu-berlin.de

ABSTRACT

In this demo, we address the problem of detecting anomalies on the Internet backbone in near real-time. Many of today's incidents may only become visible from inspecting multiple data sources and by considering multiple vantage points simultaneously. We present a setup based on the distributed forensic platform VAST that was extended to import various data streams from passive measurements and incident reporting at multiple locations, and perform an effective correlation analysis shortly after the data becomes exposed to our queries.

CCS CONCEPTS

• **Security and privacy** → *Intrusion detection systems*; *Denial-of-service attacks*; • **Applied computing** → *Network forensics*;

KEYWORDS

Internet security, threat detection, network forensic

ACM Reference format:
Philipp Meyer, Raphael Hiesgen, Thomas C. Schmidt, Marcin Nawrocki, and Matthias Wählisch. 2017. Towards Distributed Threat Intelligence in Real-Time. In *Proceedings of SIGCOMM Posters and Demos '17, Los Angeles, CA, USA, August 22–24, 2017,* 3 pages.
https://doi.org/10.1145/3123878.3123899

1 INTRODUCTION

The Internet has emerged to a critical infrastructure while at the same time attacks on virtually any of its resources have risen in frequency, diversity, and intensity. Massive attacks on the availability of Internet infrastructure (e.g., DDoS), its integrity (e.g., route forgery), the usability of its services (e.g., email spam) and many others are continuously observed today.

The identification and root-cause analysis of attacks may be intricate, and often can be only disclosed by correlating multiple observables or vantage points [2]. For example, email spamming quickly leads to identification and possible blacklisting of the originating IP space, but in reality this IP space was often hijacked either from dormant address pools or from inattentive operators. To reveal who is responsible for spamming and how to counteract, a BGP route analysis is needed. These analyses are complex, but often time critical: "Damage is done in seconds, minutes, and hours while discovery and containment are more often measured in days, weeks or even months" [4].

In this demo, we present a tool-chain for integrating various data sources into a distributed Threat Intelligence System that can process life data streams with high performance. We base our approach on VAST [8] as a distributed, though integrated and near real-time capable system. In contrast to traditional big data frameworks which evaluate regular expressions, VAST offers an explorative approach to data by enabling flexible searches on indexed heterogeneous data records.

We introduce use cases and our methodology in § 2, describe our demo setup in § 3, and conclude in § 4. Practical settings and requirements are explained in the appendix.

2 USE CASES AND METHODOLOGY

We motivate our demo with three use cases:

(1) Massive email spamming commonly leads to a quick reporting and blacklisting of the originating IP addresses. We retrieve several blacklists (e.g., blocklist, abuse, etc.) continuously. Once visible, we want to detect whether the spam is originating from hijacked IP space and query the BGP announcements that stream from monitors in near real-time. Queries can search for newly advertised prefixes that had been dormant, or test on recent multi origin AS (MOA) occurrences as an indicator of hijacks, or inquire on invalid route origin authorizations (ROAs), in cases where the resource public key infrastructure (RPKI) is deployed. These simple, possibly predefined query statements will automatically grant further insights into the root cause of the spamming incident.

(2) Blackholing on the upstream or an IXP is an effective countermeasure to massive DDoS attacks. However, being blackholed the victim is unable to observe continuation, modification, or termination of the attack. Querying DDoS traffic characteristics on SFLOW/IPFIX flow data may be used to (semi-) automatically adapt the blackholing configuration.

SIGCOMM Posters and Demos '17, August 22–24, 2017, Los Angeles, CA, USA
© 2017 Association for Computing Machinery.
ACM ISBN 978-1-4503-5057-0/17/08. . . $15.00
https://doi.org/10.1145/3123878.3123899

(3) We use honeypot data to gather unsolicited traffic. We see several options to combine these measurements with other data of our Threat Intelligence System (TIS). First, to verify in near real-time whether the source IP address is illegitimate in the global Internet, we inspect BGP dumps. If no covering prefix for the source IP address is announced, we assume a spoofed IP address. To trace the attacker back, we explore flow data from inter-domain vantage points. Note that flow sampling does not guarantee that the source is visible. However, the very fast processing capabilities of our TIS facilitate this attempt. Second, in case the address is routed, we assume that the host is infected by malware. As a significant amount of malware is distributed via email, a further step may include the analysis of mail server logs. This use case is more forward looking as it requires interaction of several threat intelligent systems. It is worth to explore this option in the future as the distribution of IP addresses of email clients is rather narrowed with respect to IP prefixes [9]. It may help to identify additional potential attackers early.

We start with VAST, a highly scalable software platform built on top of CAF, the C++ actor framework [3]. In previous work [7], VAST has proven to support scalable network forensic. We extend this work to cover a continuous real-time monitoring of distributed data sources. VAST can continuously import data streams and index its content records. Its distribution layer CAF enables a transparent deployment of multiple VAST instances at distant locations, without loosing the coherent data view.

In VAST, we can issue continuous queries, i.e query statements that permanently apply and return data once it appears in the index. As an example, the query

```
$ vast export ascii "&time > now - 1d && bro::\
blacklist.source.ip in bgp4mp::announcement.prefix"
```

asks for any subnet that had a reported spam IP within the last day and a BGP update, as well.

3 DEMO SETUP

The demo setup we present here reflects a real-world scenario that analyses data from the following sources: (a) *switches* provide data in the sFlow [6] format which contains samples of the flow data cut off after a fixed number of bytes, (b) *honeypots* accept incoming flows or connection attempts which are then recorded with the Bro network security monitor [5], (c) *BGP Routers* provide control plane data including prefix announcements and withdraws in the MRT format [1], and (d) *IntelMQ* aggregates data from incident reporting feeds—in our case blacklisted IP addresses.

Figure 1 shows the topology and data flow in this system. Sources can be located at different vantage points where their data is collected and cleaned up locally before being imported into a distributed VAST database. A continuous incident monitor scans new data and raises alerts depending on configureable queries. Also connected to the database is a user shell that allows real-time exploration through interactive queries as well as the export of dataset for further analysis.

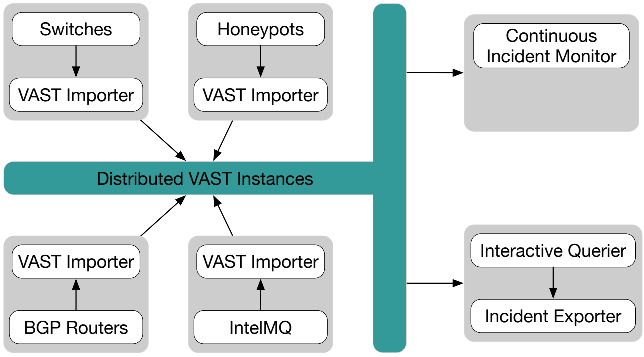

Figure 1: The system captures data from different vantage points for continuous monitoring and interactive analysis.

Our demo setup maps this distributed application onto a compact setup that omits the Internet-wide distribution to provide representative view into use cases. For this purpose, the database will run distributed across two local laptops that further deploy the continuous incident monitor and provide a shell for interactive queries.

Going forward, the incidents that are observed and validated in the control plane could be reported using *MISP* (Malware Information Sharing Platform and Threat Sharing) enriched with the related log excerpts.

4 CONCLUSION

We presented a demonstration system for distributed Threat Intelligence System that can perform correlation analysis in near real-time.

Future work in this project will lead in three directions. First, we will extend our data processing capabilities to include additional formats. Second, we will seek prototypic deployment at upstream providers or IXPs. Third, we will add an evaluation layer on top or the continuous querier that allows for complex, more delicate analyses on top of the (single statement) queries in VAST.

ACKNOWLEDGMENTS

This work was supported in parts by the German Federal Ministry of Research and Eduction within the project *X-Check*.

REFERENCES

[1] L. Blunk, M. Karir, and C. Labovitz. 2011. *Multi-Threaded Routing Toolkit (MRT) Routing Information Export Format*. RFC 6396. IETF.
[2] Sarah Brown, Joep Gommers, and Oscar Serrano. 2015. From Cyber Security Information Sharing to Threat Management. In *Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security (WISCS '15)*. ACM, New York, NY, USA, 43–49.
[3] Dominik Charousset, Thomas C. Schmidt, Raphael Hiesgen, and Matthias Wählisch. 2013. Native Actors – A Scalable Software Platform for Distributed, Heterogeneous Environments. In *Proc. of the 4rd ACM SIGPLAN Conference on Systems, Programming, and Applications (SPLASH '13), Workshop AGERE!* ACM, New York, NY, USA, 87–96.

[4] Jon C. Haass, Gail-Joon Ahn, and Frank Grimmelmann. 2015. ACTRA: A Case Study for Threat Information Sharing. In *Proceedings of the 2Nd ACM Workshop on Information Sharing and Collaborative Security (WISCS '15)*. ACM, New York, NY, USA, 23–26.

[5] Vern Paxson. 1999. Bro: a system for detecting network intruders in real-time. *Computer Networks* 31, 23–24 (1999), 2435–2463.

[6] P. Phaal, S. Panchen, and N. McKee. 2001. *InMon Corporation's sFlow: A Method for Monitoring Traffic in Switched and Routed Networks*. RFC 3176. IETF.

[7] Matthias Vallentin, Dominik Charousset, Thomas C. Schmidt, Vern Paxson, and Matthias Wählisch. 2014. Native Actors: How to Scale Network Forensics. In *Proc. of ACM SIGCOMM, Demo Session*. ACM, New York, 141–142.

[8] Matthias Vallentin, Vern Paxson, and Robin Sommer. 2016. VAST: A Unified Platform for Interactive Network Forensics. In *Proceedings of the USENIX Symposium on Networked Systems Design and Implementation (NSDI)*.

[9] Yinglian Xie, Fang Yu, Kannan Achan, Eliot Gillum, Moises Goldszmidt, and Ted Wobber. 2007. How Dynamic Are IP Addresses?. In *Proc. of ACM SIGCOMM*. ACM, New York, NY, USA, 301–312.