Threats of Tomorrow: Using Artificial Intelligence to Predict Malicious Infrastructure Activity

Staffan Truvé

Recorded Future
truve@recordedfuture.com

Summary

The ever-increasing scale and complexity of cyber threats is bringing us to a point where human threat analysts are approaching the limit of what they can handle. We believe the next-generation of cyber threats must be tackled by a combination of machines equipped with artificial intelligence (AI) and human analysts. One example of this is will be discussed: a new approach to forecasting malicious IP infrastructure by using machine learning.

1 Introduction

In recent years, humanity's ability to make accurate predictions has improved in many fields thanks to a combination of augmented sensor capabilities and new prediction algorithms.

As an example, today's weather forecasts benefit both from improved sensing by weather satellites and from new algorithms run on powerful parallel computers. In a similar way, web intelligence provides new sensing capabilities which can be combined with novel algorithms to predict future cyber threats.

Cyber defenders would benefit tremendously from being informed of threats ahead of time, instead of just reacting to identified, ongoing, or already completed attacks. Wouldn't it be good if we had a method for predicting tomorrow's threats? Such anticipatory intelligence can be produced by continuously training a machine-learning model with data sets derived from both technical intelligence (threat lists, etc.) and context from open source intelligence (OSINT). There has been some work on predicting vulnerable websites that might turn malicious after being infected [SoCh14] and on predicting malicious domains [XuSZ14], but predicting malicious infrastructure represented by their IP addresses has remained an open problem. The observations made in [XuSZ14] about re-use of valuable resources among different malicious domains is directly relevant for our work as well.

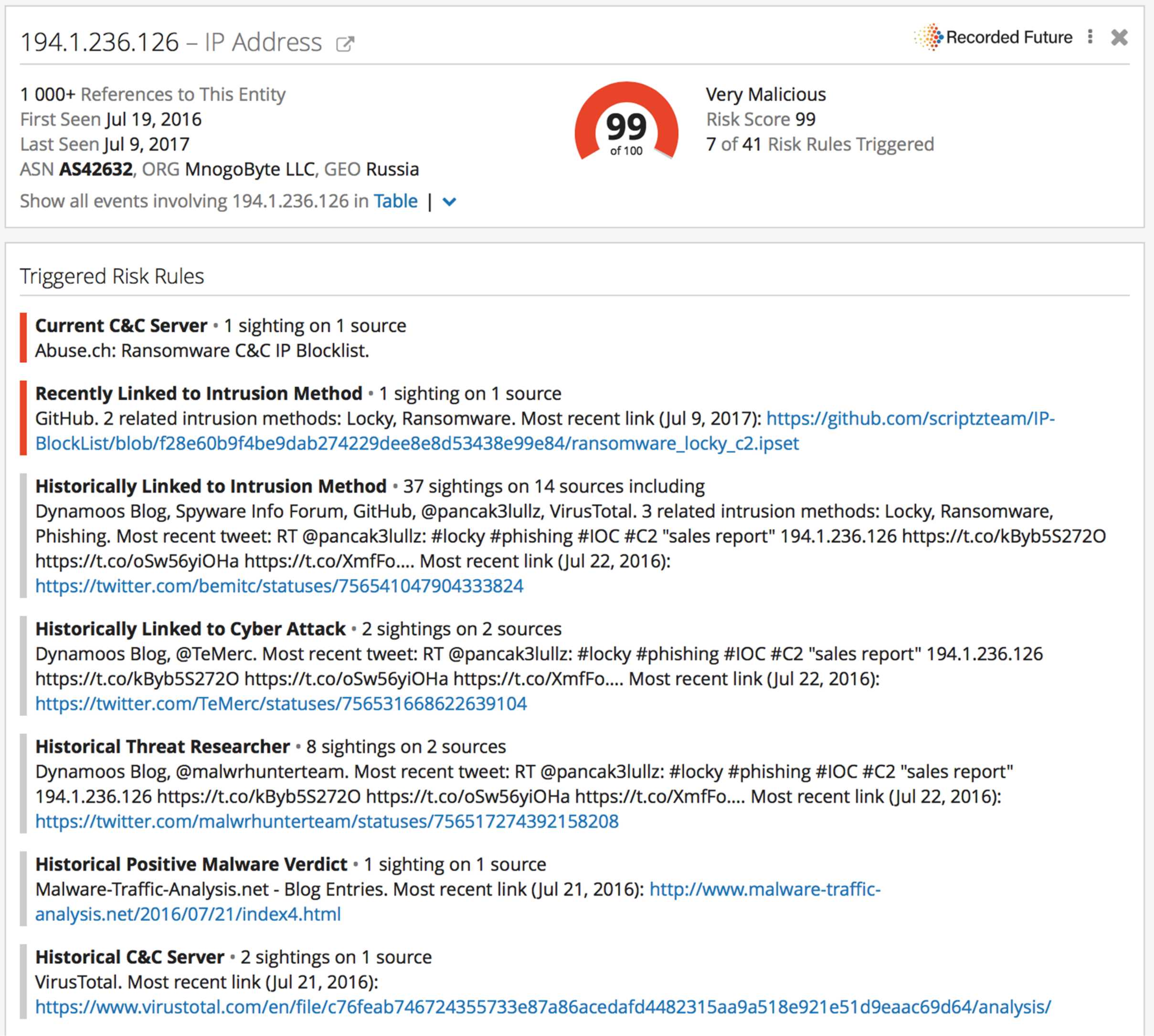
2 Risk Scoring

All threat actors, state-run intelligence agencies to cyber criminals, need to manage their resources and re-use assets (e.g., technical infrastructure and associated IP address ranges). This "business logic" is the underlying reason that we believe predicting future maliciousness is possible, based on large-scale observations of historic and current activity.

P. Schartner · A. Baumann (Hrsg.) · D•A•CH Security 2017 · syssec (2017) 204-212.

Today, defenders use threat lists (often also referred to as block lists or black lists) [Zelt17] to gain information about malicious IP addresses which might put their systems at risk.

Recorded Future computes Risk Scores for IP addresses based on threat list observations and natural language processing of open source text descriptions of threat. These risk scores are shown together with historic contextual information on an Intel Card [Todr16], to give threat analysts a quick overview of relevant information.



194.1.236.126 – IP Address

1 000+ References to This Entity
First Seen Jul 19, 2016
Last Seen Jul 9, 2017
ASN **AS42632**, ORG MnogoByte LLC, GEO Russia
Show all events involving 194.1.236.126 in Table |

Very Malicious
Risk Score 99
7 of 41 Risk Rules Triggered

Triggered Risk Rules

Current C&C Server • 1 sighting on 1 source
Abuse.ch: Ransomware C&C IP Blocklist.

Recently Linked to Intrusion Method • 1 sighting on 1 source
GitHub. 2 related intrusion methods: Locky, Ransomware. Most recent link (Jul 9, 2017): https://github.com/scriptzteam/IP-BlockList/blob/f28e60b9f4be9dab274229dee8e8d53438e99e84/ransomware_locky_c2.ipset

Historically Linked to Intrusion Method • 37 sightings on 14 sources including
Dynamoos Blog, Spyware Info Forum, GitHub, @pancak3lullz, VirusTotal. 3 related intrusion methods: Locky, Ransomware, Phishing. Most recent tweet: RT @pancak3lullz: #locky #phishing #IOC #C2 "sales report" 194.1.236.126 https://t.co/kByb5S272O https://t.co/oSw56yiOHa https://t.co/XmfFo.... Most recent link (Jul 22, 2016): https://twitter.com/bemitc/statuses/756541047904333824

Historically Linked to Cyber Attack • 2 sightings on 2 sources
Dynamoos Blog, @TeMerc. Most recent tweet: RT @pancak3lullz: #locky #phishing #IOC #C2 "sales report" 194.1.236.126 https://t.co/kByb5S272O https://t.co/oSw56yiOHa https://t.co/XmfFo.... Most recent link (Jul 22, 2016): https://twitter.com/TeMerc/statuses/756531668622639104

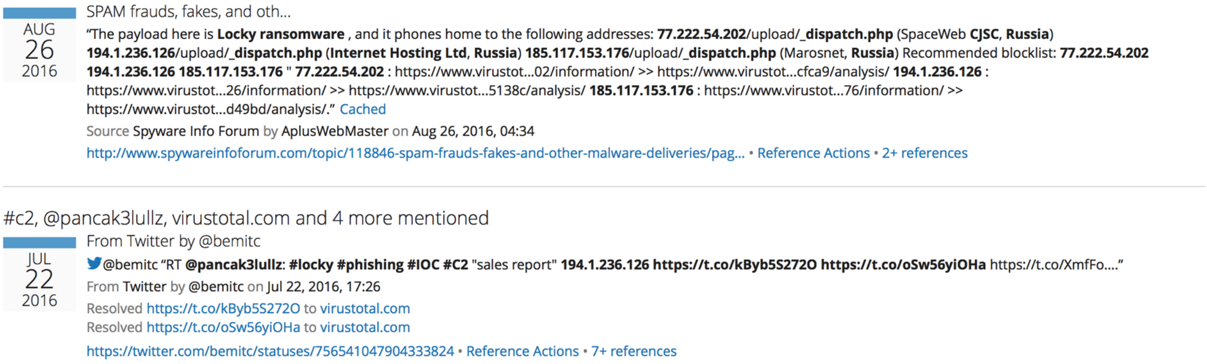
Historical Threat Researcher • 8 sightings on 2 sources
Dynamoos Blog, @malwrhunterteam. Most recent tweet: RT @pancak3lullz: #locky #phishing #IOC #C2 "sales report" 194.1.236.126 https://t.co/kByb5S272O https://t.co/oSw56yiOHa https://t.co/XmfFo.... Most recent link (Jul 22, 2016): https://twitter.com/malwrhunterteam/statuses/756517274392158208

Historical Positive Malware Verdict • 1 sighting on 1 source
Malware-Traffic-Analysis.net - Blog Entries. Most recent link (Jul 21, 2016): http://www.malware-traffic-analysis.net/2016/07/21/index4.html

Historical C&C Server • 2 sightings on 1 source
VirusTotal. Most recent link (Jul 21, 2016): https://www.virustotal.com/en/file/c76feab746724355733e87a86acedafd4482315aa9a518e921e51d9eaac69d64/analysis/

Fig. 1: A Recorded Future Intel Card shows the Risk Score of an IP address together with a description of which rules were used to compute the aggregated risk score.

To compute the risk score of an IP address, a large number of rules (currently more than 40) are used, and the result of them combined. As shown in the example above, these risk rules take into account both mentions on different threat lists and mentions in open source text, such as illustrated in Figure 2. The rules combine the content and context of mentions with metadata such as source and author to decide the risk score. In this example, the co-mention with the Locky ransomware triggers a rule which increases the risk score of the IP address 194.1.236.126



AUG 26 2016

SPAM frauds, fakes, and oth...

"The payload here is **Locky ransomware** , and it phones home to the following addresses: **77.222.54.202/upload/_dispatch.php** (SpaceWeb **CJSC**, **Russia**) **194.1.236.126**/upload/**_dispatch.php (Internet Hosting Ltd, Russia) 185.117.153.176**/upload/**_dispatch.php** (Marosnet, **Russia**) Recommended blocklist: **77.222.54.202 194.1.236.126 185.117.153.176** " **77.222.54.202** : https://www.virustot...02/information/ >> https://www.virustot...cfca9/analysis/ **194.1.236.126** : https://www.virustot...26/information/ >> https://www.virustot...5138c/analysis/ **185.117.153.176** : https://www.virustot...76/information/ >> https://www.virustot...d49bd/analysis/." Cached

Source Spyware Info Forum by AplusWebMaster on Aug 26, 2016, 04:34

http://www.spywareinfoforum.com/topic/118846-spam-frauds-fakes-and-other-malware-deliveries/pag... • Reference Actions • 2+ references

#c2, @pancak3lullz, virustotal.com and 4 more mentioned

JUL 22 2016

From Twitter by @bemitc

@bemitc "RT **@pancak3lullz: #locky #phishing #IOC #C2** "sales report" **194.1.236.126 https://t.co/kByb5S272O https://t.co/oSw56yiOHa** https://t.co/XmfFo...."

From Twitter by @bemitc on Jul 22, 2016, 17:26

Resolved https://t.co/kByb5S272O to virustotal.com

Resolved https://t.co/oSw56yiOHa to virustotal.com

https://twitter.com/bemitc/statuses/756541047904333824 • Reference Actions • 7+ references

Fig. 2: Two text references where an IP address is mentioned in a way that increases its risk score.

3 Visualizing the Malicious IP Address Space

Our intuition is that threat actors need to be efficient in how they manage their key resources, such as IP address ranges and hosting infrastructures. If this is true, then that should be reflected in the distribution of known malicious infrastructure across the IPv4 address space[1].

The IPv4 address space can be visualized using a Hilbert curve visualization [Ande09], as illustrated for 2006 in Figure 3 (left). The mapping keeps addresses close to each other in IPv4 address space close in the 2D visualization.

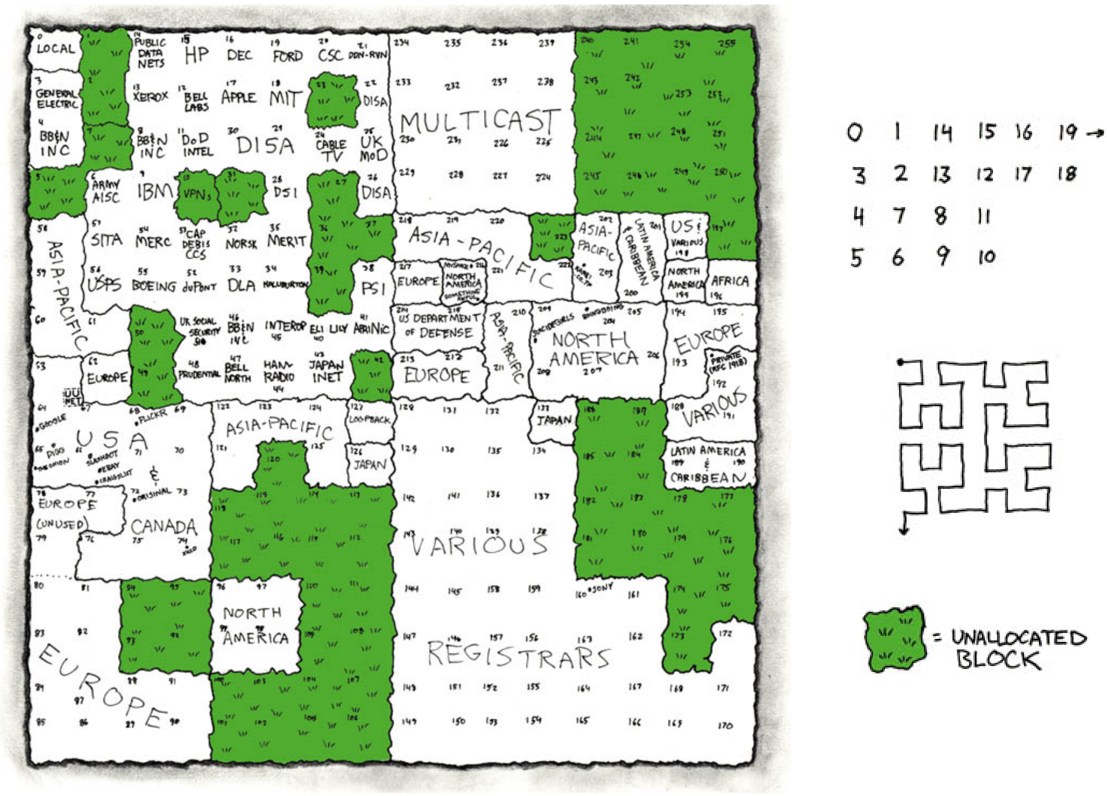


Fig. 3: Hilbert curve visualization of the IPv4 address space, courtesy of XKCD.com [Munr06].

[1] We restrict this study to the IPv4 address space, and not IPv6, since threat list data today contains very few IPv6 addresses.

According to [Munr06] the chart in Figure 3 shows the IP address space on a plane using a fractal mapping which preserves grouping – any consecutive string of IPs will translate to a single compact, contiguous region on the map. Each of the 256 numbered blocks represents one /8 subnet (containing all IPs that start with that number). The upper left section shows the blocks sold directly to corporations and governments in the 1990's before the RIRs (Regional Internet Registry) took over allocation.

In Figure 4, we show 2,856 IPv4 addresses with a Recorded Future Risk Score greater than 70. Note the clusters indicating IP address blocks representing infrastructure controlled by threat actors (see Figure 5 below for detail).

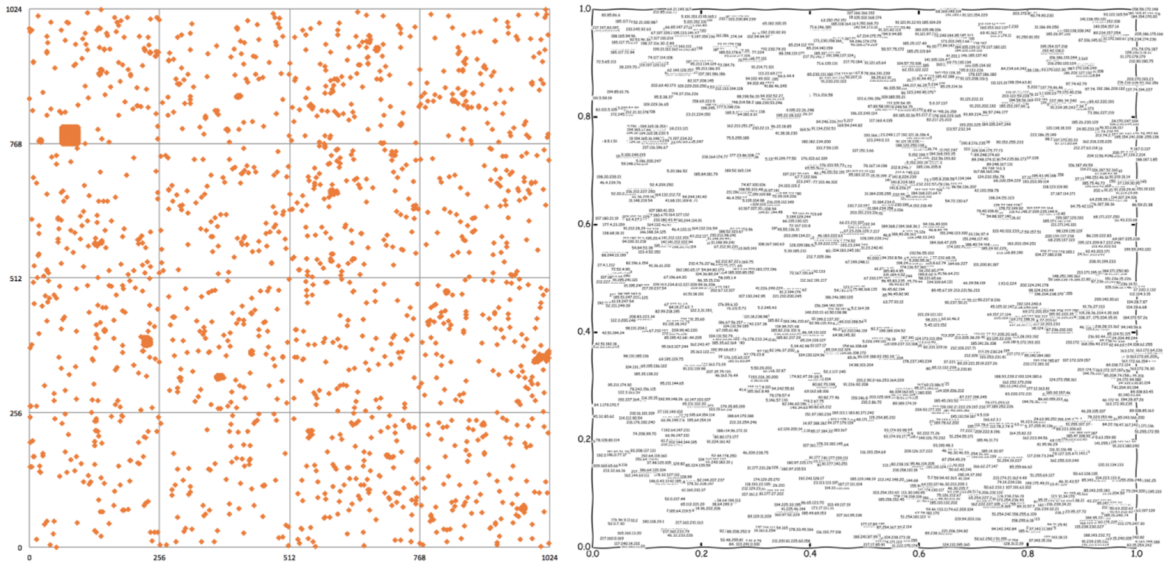


Fig. 4: Hilbert space plot of IPv4 addresses, with clusters of addresses with high risk score.

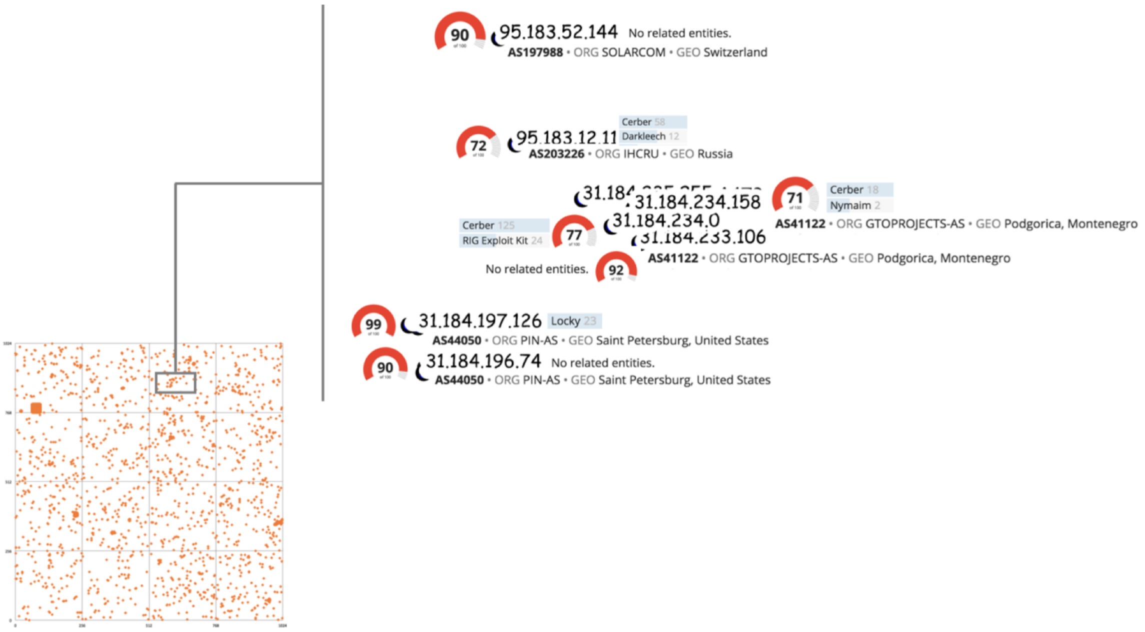


Fig. 5: A subset of the plot: IPv4 addresses with a high-risk score, ASN, and associated malware from open source mentions.

This visualization verifies our hypothesis that malicious IP addresses tend to cluster together in IP address space, and that combining closeness in this space with other attributes should be a relevant way of predicting future malicious IP addresses as well.

4 Predicting Malicious Infrastructure

Moving from risk scores based on historic observations to predicting future malicious behavior requires a predictive model that can use historic observations to forecast future ones.

First of all, there is of course some "inertia" on threat lists – our study shows that about 45% of the IP addresses on our selected threats lists remain on the lists from one week to the next.

To identify future malicious IP addresses, we trained a Support Vector Machine [CoVa95] model using historic threat lists combined with contextual OSINT data from the Recorded Future system. The resulting model takes into consideration not only CIDR (Classless Internet Domain Routing) neighborhoods [FLYV93], but also the context in which the neighbors of an IP address are being discussed. These contexts include e.g. IP addresses being identified as command and control servers, or associated with known threat actors, geographies, or malware – some 50 factors in total. In addition to this context that is derived from the text, we add metadata such as the source of the text – both explicit sources such as Twitter and Pastebin and source categories such as Social Media, Forums, Code Repositories, and Paste Sites.

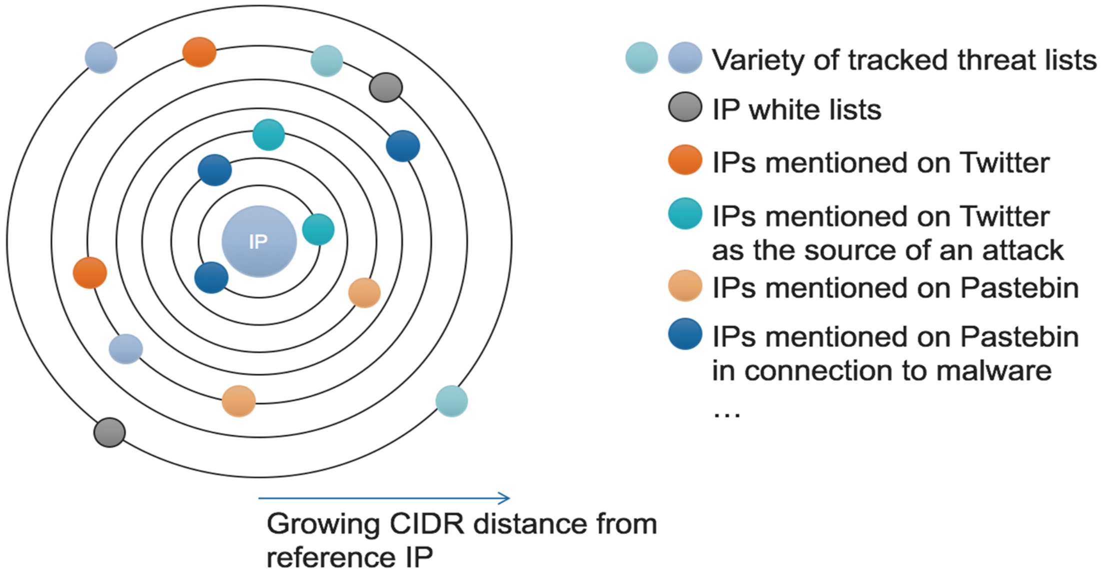
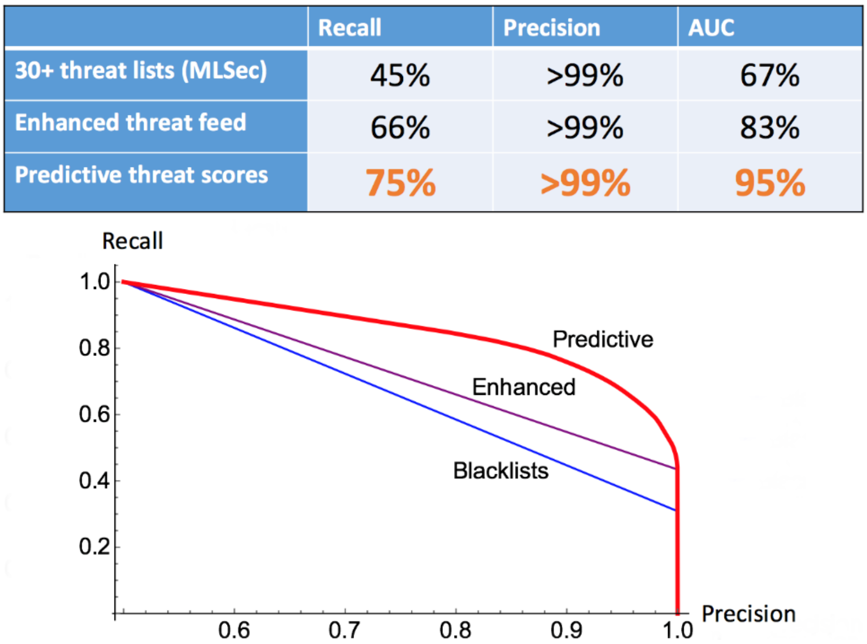

IP
Variety of tracked threat lists
IP white lists
IPs mentioned on Twitter
IPs mentioned on Twitter as the source of an attack
IPs mentioned on Pastebin
IPs mentioned on Pastebin in connection to malware
…
Growing CIDR distance from reference IP

Fig. 6: Different threat lists, different sources of OSINT mentions, and different contexts are used when training and applying the SVM model.

The output from the SVM model is used to assign a predictive risk score to IP addresses, and this score can then be used to alert security operations center (SOC) operators or threat analysts, or even to automatically block the address in firewalls and other network security systems.

In our study, 75% of next week's IP address threat list content was predicted by the SVM. Precision on a balanced test set was 99%, but on an unbalanced test set results are lower, since our models are really predicting the riskiness of a region defined by a CIDR block, not of individual addresses.

We compared the precision and recall of the set of threat lists used in the MLSec paper [PiMa14], an extended set of threat lists used by Recorded Future, and the predictive model described in this paper. The results are depicted in Figure 7 (where AUC denotes "area under the precision-recall curve"), and show that recall improves significantly while retaining precision when the predictive model is used.



	Recall	Precision	AUC
30+ threat lists (MLSec)	45%	>99%	67%
Enhanced threat feed	66%	>99%	83%
Predictive threat scores	75%	>99%	95%

Fig. 7: Precision and recall of threat lists vs. the predictive model described in this paper.

5 Experimental Results

The model was evaluated on historic data as well as out-of-sample data.

As a concrete example, IP address 88.249.184.71 was flagged by our predictive risk scoring on October 4, 2016 (using a 0.75 threshold), and not until October 14 did it first occur on any threat list, identified as a DarkComet RAT controller:

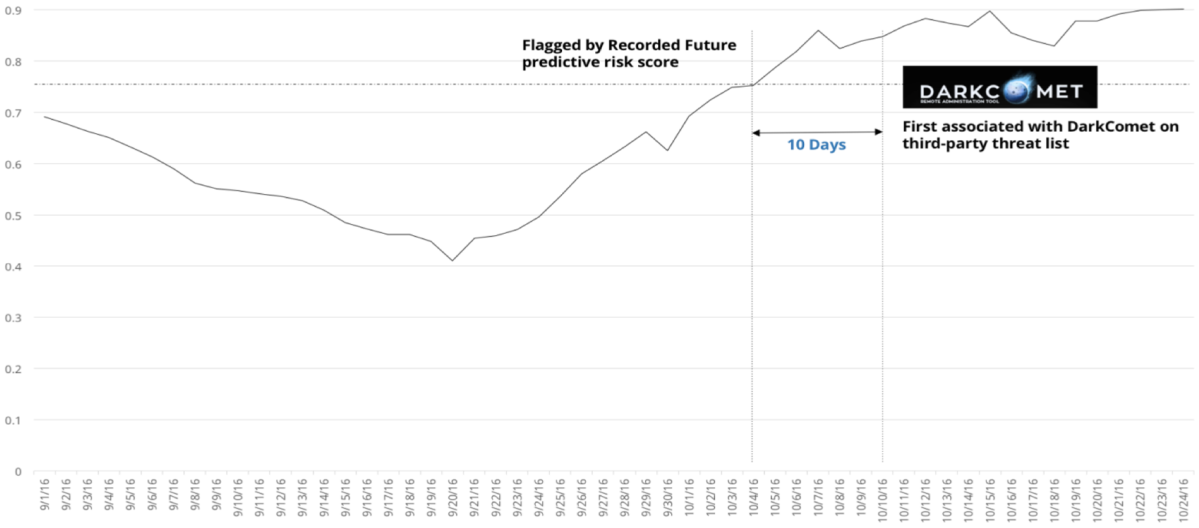


Fig. 8: Predictive risk score of an IP address over time.

In another example, IP address 189.218.148.73 reached a predictive risk score of 0.75 on March 8, 2016. The first malicious mention on a threat list (Socks24) occurred three days later, on March 11. Not until March 16 was the address mentioned as malicious in a source available as open source intelligence, in this case on Pastebin.

Once an address is published, the multiplicative effect of the Internet means it spreads quickly to multiple threat lists and open source resources – see Figure 9.

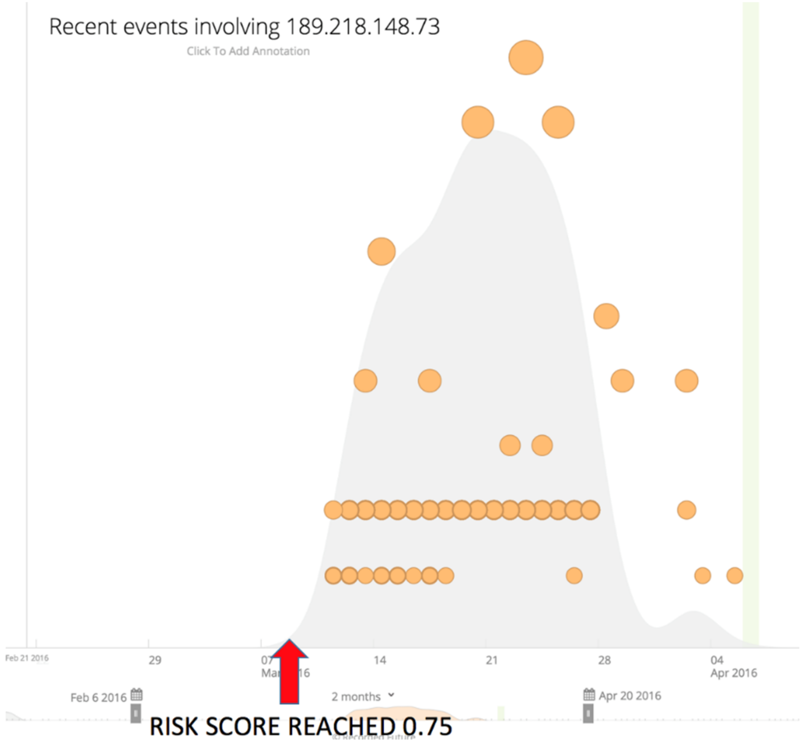


Fig. 9: Once an address is published on a threat list or in open source the multiplicative effect of the Internet allows it to spread rapidly.

We also performed a small quantitative study. In this, over 25% of 500 previously unseen IPs with a predictive risk score turned malicious (as reported by OSINT) within seven days of being flagged by our algorithms – see Figure 10.

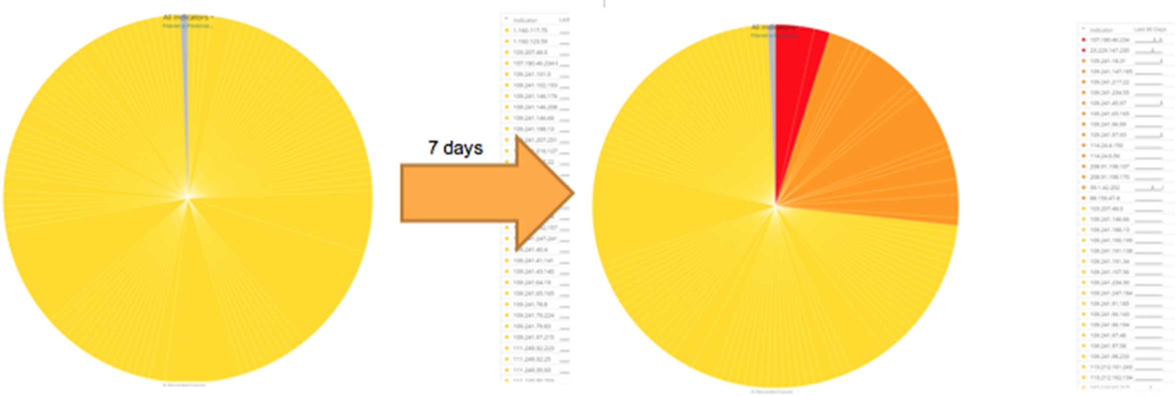


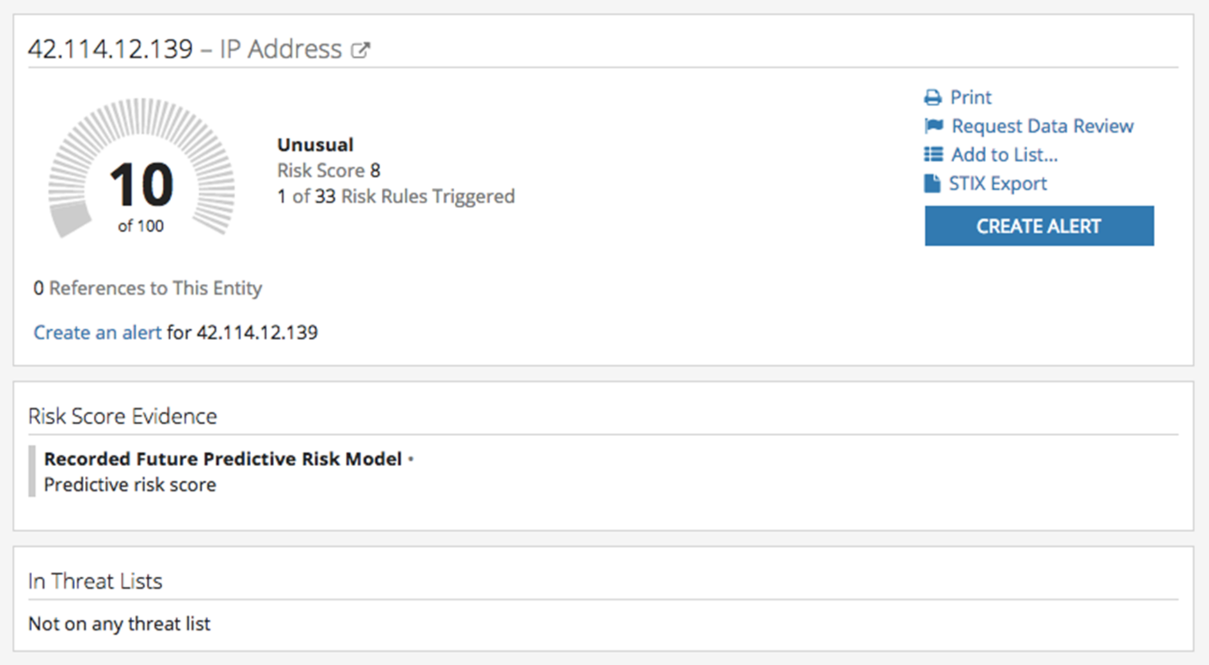
Fig. 10: Over seven days, more than 25% of IP addresses with a predictive risk score were flagged as malicious by OSINT.

In comparison, out of the entire IPv4 address space, typically less than 0.02% of all addresses at any given time have a risk score.

In a larger study, we trained an SVM model on historic data and applied it to the entire IPv4 address space. Of the IPs with a predictive risk score, 327,549 occurred on any external threat lists in the following seven days. Out of these, 185,209 had not appeared on any threat lists in the past three weeks (21 days total). Our model predicted 242,206 (74%) of future threat-listed IP addresses while maintaining greater than 99% precision on a balanced test set.

In general, IP addresses are being flagged as potentially malicious by our predictive risk score 3-5 days before actually occurring on threat lists — giving the defenders the heads-up needed to prepare their defenses ahead of time, instead of just being reactive.

Predictive risk scoring of IP addresses is available in Recorded Future Intel Cards, and through an API, allowing for easy integration with other tools and resources like Security Information and Event Management (SIEM) systems. This is the first of what we anticipate to be a series of novel ways of applying artificial intelligence to improve cyber security.


42.114.12.139 – IP Address

Unusual
Risk Score 8
1 of 33 Risk Rules Triggered

0 References to This Entity

Create an alert for 42.114.12.139

Risk Score Evidence

Recorded Future Predictive Risk Model •
Predictive risk score

In Threat Lists

Not on any threat list

Fig. 11: An intel card for an IP address with predictive risk score.

6 Conclusions and Future Work

We have shown how a machine learning model trained on both threat lists and open source based contextual data can help in identifying future malicious infrastructure, as represented by IP addresses. We believe this methodology gives defenders a unique way of moving from reactive to proactive network defense. The same methodology should also be applicable to predicting malicious Internet Domain Names, given that an appropriate distance / similarity measure can be identified (similar to the CIDR distance used in this study), and that sufficient training data in the form of threat lists and contextual data from open source intelligence can be identified.

References

[Ande09] S. Anders: “Visualization of genomic data with the Hilbert curve”, Bioinformatics, Vol. 25 (2009) pp. 1231-1235

[CoVa95] C. Cortes; V. Vapnik, (1995). “Support-vector networks”. Machine Learning. 20 (3): 273–297.

[FLYV93] V. Fuller, T. Li, J. Yu, K. Varadhan, “RFC 1519, Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy” (September 1993)

[Munr06] R. Munroe, “Map of the Internet”, https://xkcd.com/195/

[PiMa14] A. Pinto, K. Maxwell, “Measuring the IQ of your Threat Intelligence Feeds”, BSides Las Vegas, 2014

[SoCh14] K. Soska, N. Christin, “Automatically Detecting Vulnerable Websites Before They Turn Malicious”, Proceedings of the 23rd USENIX Security Symposium

[Todr16] M. Todros, “Anatomy of a Recorded Future Intel Card: Make Threat Analysis Fast”, https://www.recordedfuture.com/intel-cards-overview/

[XuSZ14] W. Xu, K. Sanders, Y. Zhang, “We Know it Before You Do: Predicting Malicious Domains”, Virus Bulletin Conference, September 2014

[Zelt17] L. Zeltser, “Blocklists of Suspected Malicious IPs and URLs”, https://zeltser.com/malicious-ip-blocklists/