2018 10th International Conference on Cyber Conflict
CyCon X: Maximising Effects
T. Minárik, R. Jakschis, L. Lindström (Eds.)
2018 © NATO CCD COE Publications, Tallinn

FeedRank: A Tamper-resistant Method for the Ranking of Cyber Threat Intelligence Feeds

Roland Meier
Department of Information Technology and Electrical Engineering
ETH Zürich
Zürich, Switzerland
meierrol@ethz.ch

Cornelia Scherrer
Department of Information Technology and Electrical Engineering
ETH Zürich
Zürich, Switzerland
cornelia.scherrer@alumni.ethz.ch

David Gugelmann
Exeon Analytics
Zürich, Switzerland
david.gugelmann@exeon.ch

Vincent Lenders
Science and Technology
armasuisse
Thun, Switzerland
vincent.lenders@armasuisse.ch

Laurent Vanbever
Department of Information Technology and Electrical Engineering
ETH Zürich
Zürich, Switzerland
lvanbever@ethz.ch

Abstract: Organizations increasingly rely on cyber threat intelligence feeds to protect their infrastructure from attacks. These feeds typically list IP addresses or domains associated with malicious activities such as spreading malware or participating in a botnet. Today, there is a rich ecosystem of commercial and free cyber threat intelligence feeds, making it difficult, yet essential, for network defenders to quantify the quality and to select the optimal set of feeds to follow. Selecting too many or low-quality feeds results in many false alerts, while considering too few feeds increases the risk of missing relevant threats. Naïve individual metrics like size and update rate

give a somewhat good overview about a feed, but they do not allow conclusions about its quality and they can easily be manipulated by feed providers.

In this paper, we present FeedRank, a novel ranking approach for cyber threat intelligence feeds. In contrast to individual metrics, FeedRank is robust against tampering attempts by feed providers. FeedRank's key insight is to rank feeds according to the originality of their content and the reuse of entries by other feeds. Such correlations between feeds are modelled in a graph, which allows FeedRank to find temporal and spatial correlations without requiring any ground truth or an operator's feedback.

We illustrate FeedRank's usefulness with two characteristic examples: *(i)* selecting the best feeds that together contain as many distinct entries as possible; and *(ii)* selecting the best feeds that list new entries before they appear on other feeds. We evaluate FeedRank based on a large set of real feeds. The evaluation shows that FeedRank identifies dishonest feeds as outliers and that dishonest feeds do not achieve a better FeedRank score than the top-rated real feeds.

Keywords: *cyber threat intelligence, intelligence feeds, cyber attacks, malware, botnets, situational awareness*

1. INTRODUCTION

States, organizations, companies and individuals are faced with ever-growing cyber threats. The most prominent among these threats include phishing or spam campaigns, malware distribution and DDoS attacks [1] [2]. To mitigate these threats, Cyber Threat Intelligence Feeds (CTIFs, also known as blacklists or block lists) are a major source of information for most network defenders [3]. The CTIF ecosystem is currently very large and complex [4] and for reliable protection, network defenders need to correlate data from multiple CTIFs [1].

However, while selecting the best set of CTIFs is crucial to maximizing efficiency, it is also difficult as there is no easy way to objectively compare CTIFs. In fact, network defenders often only evaluate feeds individually based on naïve metrics such as the feed's size. While these metrics allow for a rough assessment, they do not allow conclusions about the combination of multiple feeds and – as we will show in this paper – they are easy to manipulate for a dishonest CTIF provider in order to pretend a higher quality and thus increase its impact and revenue.

Problem statement. In this paper, we address the problem of finding an objective, tamper-resistant ranking algorithm that allows well-grounded selections of high quality CTIFs. We determine the quality of a feed by three key properties: *completeness, accuracy* and *speed*. That is, an ideal CTIF lists all malicious entities, does not list non-malicious entities and updates its entries promptly. In particular, we address the following research questions:

- How can the quality of a CTIF be estimated in a robust and scalable way? Achieving this is challenging because there is no ground truth to compare it with. Hence, one cannot rely on standard metrics such as precision and recall.
- How can the structure of the CTIF ecosystem be evaluated and how do existing CTIFs differ in terms of completeness, accuracy and speed? In particular, do CTIF providers cluster in groups or is there a large diversity regarding the reported threats among the different providers?
- Can we identify individual CTIFs that consistently outperform others and CTIFs that seem to lack behind or borrow information from other feeds? Specially, what are good metrics to identify outperformers and tampering-attempts by a subset of the feeds?

FeedRank. We present FeedRank, a novel metric for the ranking of CTIFs. The key idea behind FeedRank is to model the correlations between CTIFs as a graph and to obtain the ranking by applying algorithms to this graph. This way, FeedRank quantifies the relative performance among CTIFs and is able to evaluate the quality of feeds without a ground truth. At its core, FeedRank bears similarities with collective intelligence approaches or PageRank [5], an algorithm to rank websites that is used by Google.

The setting for ranking CTIFs bears similarities with the ranking of websites by search engines in the following aspects:

- Websites can contain arbitrary content (including dummy keywords to improve their ranking).
- Websites can contain links to any other website.
- There is no ground truth for the quality of websites.
- A website to which many other websites refer to is likely to be important.

Similar properties hold for CTIFs:

- CTIFs can contain arbitrary entries.
- CTIFs can copy entries from any other CTIF.
- There is no ground truth for the quality and validity of CTIF entries.
- A CTIF whose entries appear in other CTIFs is likely to be of high quality.

Despite these similarities, applying website ranking algorithms (such as PageRank) to CTIFs is challenging because of the particular semantic of the CTIF application domain. The key idea to apply website ranking algorithms to CTIFs is to model correlations between CTIFs in a graph. In particular, while PageRank uses the web graph (a graph that models the links between websites), FeedRank uses a correlation graph to model common entries in CTIFs and the time at which they appear in each of the considered CTIFs. The correlation graph provides us with the foundation to assess a CTIF's quality as we argue that a CTIF whose entries later appear on many other feeds has a high quality (like a website with many incoming links is assumed to be important). However, since the correlation graph alone does not allow conclusions about the completeness of a particular CTIF, FeedRank also performs an analysis of the contribution of each CTIF.

Contributions. The main contributions of this paper are:

- A tamper-resistant approach to rank CTIFs at scale (Section 3) based on:
  - correlations between CTIFs (Section 3B); and
  - the individual contribution of each CTIF (Section 3C).
- A comprehensive evaluation based on large sets of freely available CTIFs (Section 4).
- Two case-studies to demonstrate useful use-cases of FeedRank (Section 5).

2. EVALUATING CYBER THREAT INTELLIGENCE FEEDS

In this section, we provide an overview over Cyber Threat Intelligence Feeds (CTIFs) and identify key properties that characterize good CTIFs, sketch traditional evaluation metrics and identify strategies for how dishonest CTIF providers can tamper with them.

A. Cyber Threat Intelligence Feeds

In general, CTIFs are collections of *Indicators of Compromise* (IOC) that characterize malicious or non-malicious endpoints or activities. In this paper, we focus on feeds that list IP addresses associated with malicious activities (such as sending spam or hosting phishing sites). However, the obtained results are also applicable to other types of feeds.

CTIFs are available from a variety of commercial and non-commercial providers and can cover one or multiple types of threats (e.g. spam or phishing). The feeds are typically provided in real time; that is, the contents are updated continuously or with a certain frequency. New entries may be added when, for example, an endpoint is

found to behave maliciously and removed if the malicious activity has stopped. CTIFs obtain information about malicious endpoints in various ways. For example, malicious activity can be detected by email providers, honeypots, CERTs or by manual reports from users. CTIFs can also copy or fuse information from other CTIFs.

B. Properties of High Quality Feeds

An ideal CTIF is complete, accurate and fast. To be *complete*, the CTIF needs to contain all malicious endpoints at a given time. To be *accurate*, the CTIF must not list benign endpoints. To be fast, the completeness and accuracy property must hold at any given point in time, i.e., an endpoint must appear exactly during the time it behaves maliciously. This ideal state is obviously difficult to reach in practice, as there always exist malicious endpoints that have not yet been identified as such.

C. Individual Feed Metrics

Naïve metrics which evaluate each CTIF individually are easy to calculate and widely used. Examples of such individual metrics include the feed's size, the update frequency and the number of entries that are added or removed (cf. Table I). However, a major problem with individual metrics is that they provide little insight about the quality of a CTIF without a ground truth (i.e. a way to objectively verify the correctness of the feed's contents). Even worse, all the listed individual metrics can easily be manipulated by adding or removing entries to/from a CTIF (cf. Table I).

With FeedRank, we present an advanced and tamper-resistant ranking metric that does not require a ground truth. As we will describe in the following sections, analyzing the correlations of CTIFs allows reasoning about the feed's completeness, accuracy and speed.

TABLE I: EXAMPLES OF INDIVIDUAL FEED METRICS. THESE METRICS DO NOT ALLOW CONCLUSIONS ABOUT A FEED'S QUALITY AND CAN BE MANIPULATED BY THE FEED PROVIDER.

Metric	Description	Manipulation strategy
Size	Number of entries in the CTIF	Add random entries
Insertion rate	Number of entries that are added to CTIF per time unit	Add random entries
Removal rate	Number of entries that are removed from a CTIF per time unit	Remove random entries
Update rate	Rate at which entries are added or removed	Frequently replace random entries

3. FEEDRANK

In this section, we present the design goals and an overview of FeedRank. Further, we describe the two core components of FeedRank in more detail and explain why FeedRank is robust against tampering attempts.

A. Overview

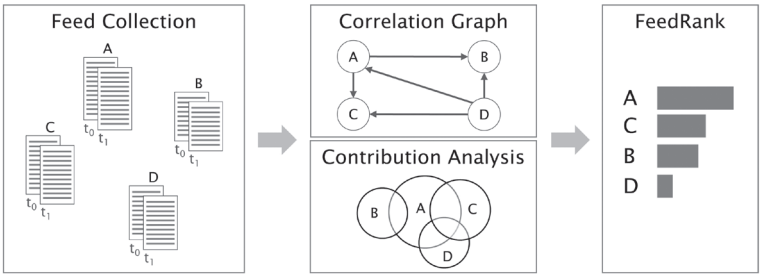
FeedRank allows us to identify high quality CTIFs, while at the same time being robust against tampering attempts from CTIF providers, by combining the contribution analysis and the correlation graph (see Table II).

TABLE II: KEY PROPERTIES OF HIGH QUALITY CTIFS AND HOW THEY ARE REPRESENTED IN FEEDRANK.

Property	Represented in
Completeness	Contribution analysis
Accuracy	Correlation graph (weighted edges model the entries that are confirmed by other CTIFs)
Speed	Correlation graph (directed edges between CTIFs represent the order in which common entries were listed)

FeedRank operates in three steps (see Figure 1). First, it collects snapshots of considered CTIFs; second it builds a feed correlation graph and performs a contribution analysis; and third, it computes a score for each considered CTIF.

FIGURE 1: FEEDRANK OPERATES IN THREE STEPS: IT COLLECTS SNAPSHOTS OF CTIFS, COMPUTES A CORRELATION GRAPH AND A CONTRIBUTION ANALYSIS AND OUTPUTS A RANKING.



I. Feed Collection. As an input, FeedRank requires at least two snapshots of each considered feed. A snapshot consists of the timestamp and all entries of a CTIF. For the most accurate results, the time between the two snapshots should be long enough such that all CTIFs provide an update of the entries. The set of considered feeds can

be specified by the operator who uses FeedRank. It should contain all the CTIFs that the operator considers using in their environment.

II. a) Correlation Graph. Based on the snapshots, FeedRank builds a correlation graph. The nodes in this graph correspond to the CTIFs and the (directed) edges represent correlations between them.

II. b) Contribution Analysis. For each CTIF, FeedRank computes a contribution metric that measures the CTIF's contribution to the total number of listed entries.

III. Feed Rating. FeedRank runs an algorithm similar to PageRank on the correlation graph. This, together with the results from the contribution analysis, assigns each feed a score and allows to rank them.

B. Correlation Graph

The correlation graph is used to model correlations between CTIFs. It is a directed graph where the vertices represent feeds and the weighted edges describe correlations between them. Two CTIFs (X and Y) are connected with a directed edge from X to Y if X contains entries that were contained in Y before they appeared in X. This means that X implicitly confirms the respective entries from Y. In other words: both feeds classify the entries as malicious, and Y was faster in listing them, which makes it more likely that Y is accurate with respect to these entries.

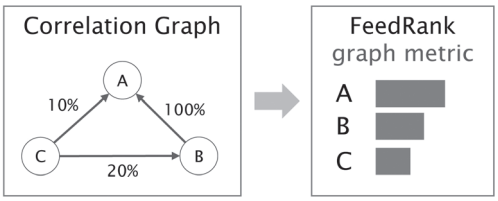
The weight of this edge is determined by the percentage of entries that appear first in Y and are later mentioned by X. For example, if Y contains 20 entries and 10 of them appear later on X, the weight of the edge would be 50% as this is the percentage of entries in Y that were confirmed by X.

Figure 2 illustrates an example of a correlation graph with 3 feeds with the following characteristics:

- B confirms 100% of the entries in A (i.e. every entry that appears in A later appears in B)
- C confirms 10% of the entries in A and 20% of the entries in B (i.e. 10% of the entries in A and 20% of the entries in B appear later in C)

In this example, feed A achieves the highest score according to the correlation graph and would thus be considered as the most valuable feed. The intuitive explanation for this is that all of A's entries are confirmed by B and no other feed is faster than A.

FIGURE 2: EXAMPLE OF THE FEEDRANK GRAPH METRIC. IT RANKS FEEDS ACCORDING TO THE AMOUNT OF ENTRIES THAT ARE CONFIRMED BY OTHER FEEDS (E.G. B CONFIRMS 100% OF THE ENTRIES IN A).



To determine a ranking of CTIFs in the correlation graph, we apply the PageRank algorithm [5]. PageRank is a ranking algorithm for websites (famously used by Google) and is based on a graph that models the hyperlinks between websites. Besides the web graph, PageRank requires two additional parameters: the damping factor and a convergence condition.

The damping factor d in PageRank describes the probability with which a user browsing at a certain website will click on any of the links to visit another website. For FeedRank, we calculate the damping factor depending on the average path length l (i.e., the average number of CTIFs that subsequently list an entry) of all entries that appear in at least two CTIFs within the analyzed dataset. From l, we calculate the probability that an entry "propagates" to another feed – along the lines of a user that clicks on a link to move to another website – as:

$$d = P(continue) = 1 - P(stop) = 1 - \frac{1}{l}$$

Being an iterative algorithm, PageRank further requires a convergence condition. The convergence condition in PageRank specifies the maximal delta between the graph score of all nodes (i.e. the precision of the result).

C. Contribution Analysis

The contribution metric is the result of the contribution analysis and measures the relative contribution of a single CTIF compared to the complete set of analyzed feeds. Therefore, it provides the foundation to select a subset of the analyzed feeds that together have a maximal contribution.

FeedRank's contribution analysis works as follows. First, it computes the complete set of all listed entries, i.e., the union of all entries listed in the considered feeds at any of the recorded snapshots. For each entry, it determines the feed that listed the entry first, and assigns the entry to that feed. In case multiple feeds add an entry at the same time, the entry is assigned to the biggest feed. The resulting contribution metric is then computed as the percentage of entries that each feed contributes to the complete set.

D. Tamper-resistance

In this section, we explain why FeedRank is robust against an unfair CTIF which tries to manipulate its rank. Since a CTIF provider can arbitrarily choose the contents of its feed, there are no guarantees about the validity of entries.

At a high level, we distinguish between the following tampering strategies:

- Adding entries that are not contained in the original CTIF.
- Removing of entries from the original CTIF.
- Replacing existing entries by other values (i.e. pretend updates).

For each of these strategies, the dishonest CTIF provider needs to choose the entries that will be added or removed. This can be done in at least the following ways:

- At random: New entries are generated randomly and randomly chosen entries are removed from the feed.
- Based on the contents of another CTIF: A dishonest CTIF can copy a subset or all entries from one or multiple other CTIFs and thus copy the behavior of these CTIFs.

The manipulation strategies mentioned above work well for individual metrics (as described in Section 2. C) but, as we explain in the following, they do not work with FeedRank.

A dishonest CTIF that tries to manipulate its FeedRank score by adding entries is not successful because: *(i)* if the added entries are chosen randomly, they will not be confirmed by other feeds with very high probability; *(ii)* if the added entries are copied from another CTIF, this is considered as if the dishonest feed confirms the other feed's entries and can therefore help the other feed, but not the dishonest feed. Obviously, a feed that copies entries from another feed is always slower in listing these entries. If a dishonest CTIF tries to improve its score by removing entries, each of the removed entries is either of high quality (i.e. it is confirmed by other feeds) or of low quality (it does not appear on other feeds). If a CTIF removes high quality entries, this lowers its ranking because a smaller percentage of its entries are confirmed. If it removes low quality entries, its overall quality increases and it (deservedly) obtains a better ranking.

A dishonest CTIF that both adds and removes entries faces the union of the limitations mentioned above. FeedRank does not measure the update frequency of CTIFs and therefore a higher update frequency does not help to improve the score. Instead, FeedRank is run with a certain frequency and based on the feed's contents at the time of execution. Therefore, as long as the update frequency of a CTIF is larger than or

equal to the execution frequency of FeedRank, increasing the update frequency does not change the FeedRank score.

While FeedRank is robust against a small percentage of dishonest CTIFs, it can be susceptible to manipulation attempts by many colluding CTIFs. However, this is hardly feasible in practice because it would require many CTIFs to become dishonest and it only works if the user considers all of them when running FeedRank (it is easy for a single entity to publish a large number of dishonest feeds, but it is unlikely that a user would consider all of them). Such a set of colluding CTIFs can be identified by doing basic cluster analysis based on the considered feeds and the correlation graph (we show this in Section 4B).

4. EVALUATION

In this section, we use real CTIFs to compare FeedRank with individual metrics and show its tamper-resistance. In the following subsections, we describe and visualize the dataset and show the evaluation results.

A. Dataset and Methodology

To evaluate FeedRank, we use both real CTIFs which we collected over a timespan of almost 12 days and synthetic CTIFs with which we simulated the impact of tampering strategies. Below, we provide more details about both types of feeds.

1) Collecting Real Feeds

For a comprehensive dataset, we fetched the feeds listed in Table III at regular intervals (60 min) during almost 12 days in 2017. In this way, we obtained 277 snapshots representing the activity of 27 feeds with a total of around 40 million entries. These snapshots allowed us to analyze correlations between feeds at a granularity of an hour. Some of the snapshots were incomplete because our collection infrastructure was unable to fetch them due to connectivity issues, database overload or rate limiting by the CTIF provider. The feed collection functionality was implemented in Python on top of the stix [6] and libtaxii [7] modules. The collected feeds were normalized and stored in an Elasticsearch database to facilitate analysis. We anonymized the feeds as it is not our goal to provide a ranking of particular feed providers, but to demonstrate the practicality of our algorithm.

TABLE III: EVALUATED FEEDS. WE ANALYZE 27 FREELY AVAILABLE CTIFS (LISTED IN ALPHABETICAL ORDER HERE).

Feed △	Source
AlienvaultReputationIP	reputation.alienvault.com
Autoshun	www.autoshun.org
BinaryDefense	www.binarydefense.com
CIArmyBadGuys	www.cinsscore.com
CymonBlacklist	www.cymon.io
CymonBotnet	www.cymon.io
CymonMaliciousActivity	www.cymon.io
CymonMalware	www.cymon.io
CymonPhishing	www.cymon.io
CymonSpam	www.cymon.io
Cymondnsbl	www.cymon.io
EmergingThreatsBlockRules	rules.emergingthreats.net
EmergingThreatsCompromised	rules.emergingthreats.net
FeodoIpBlocklist	feodotracker.abuse.ch
MalcodeIP	www.malc0de.com
MalwareDomainIp	mirror1.malwaredomains.com
NoThinkDNS	www.nothink.org
NoThinkHTTP	www.nothink.org
NoThinkMalwareIRC	www.nothink.org
NoThinkSNMPWeek	www.nothink.org
NoThinkSSH	www.nothink.org
NoThinkSSHWeek	www.nothink.org
NoThinkTelnetWeek	www.nothink.org
OpenBLBase	www.openbl.org
PhishTankJSON	data.phishtank.com
SSLIPBlacklist	sslbl.abuse.ch
ZeusTracker	zeustracker.abuse.ch

2) Generating Dishonest Feeds

To capture the effect of dishonest feeds, we considered two strategies: listing random entries and imitating high-ranked feeds.

I. Adding random entries: Adding random entries is a straightforward approach for a dishonest feed to improve its rank because it makes the feed appear larger and more up-to-date. Particularly because random entries are unlikely to be contained in other feeds, thus the tampering feed is the first to report them.

Adding random entries can be risky for a CTIF provider as it can increase the false positive rate, especially if an entry maps to a popular non-malicious service. However, by choosing unused (or rarely used) IP addresses or domains, a dishonest CTIF provider can cheat with a low risk of being detected.

We call a synthetic feed that follows such a strategy "RandomFeed" and generate it by choosing 50k IP addresses uniformly at random at each time unit (i.e. 1 hour).

II. Copying entries from high-reputation feeds: For this case, we generate "CopyFeed" by assuming that it copies all entries from the two best-ranked feeds with a delay of one time unit (i.e. 1 hour). By doing so, CopyFeed becomes the most complete feed but it lacks speed as it is never the first to announce any entry.

3) Parameters

PageRank, which is part of the graph ranking, requires the specification of a damping factor and a convergence condition. As we explained in Section 3B, we compute the damping factor as d=1-1/l where l is the average path length. For the evaluated dataset, the average path length is 2.87, which leads to a damping factor of 0.65. For the convergence condition, we choose a maximum delta (i.e. the precision of the results) of 10^{-6}.

B. FeedRank Dataset Baseline

In this section, we illustrate our dataset and the input of FeedRank with Figure 3 and Figure 4 after listing basic properties of each analyzed CTIF in Table IV.

TABLE IV: SIZE OF THE EVALUATED CTIFS.

	Number of entries				Number of entries		
Nr.	average ▽	max	min	**Nr.**	average ▽	max	min
F1	49125	51349	46931	F15	771	772	765
F2	22997	24397	21523	F16	764	1988	37
F3	16092	16591	15717	F17	500	500	500
F4	16085	16948	15247	F18	464	522	252
F5	15587	15826	15305	F19	444	444	444
F6	12719	18467	2362	F20	184	283	85
F7	8260	8861	7640	F21	127	133	121
F8	7956	11618	7	F22	127	133	123
F9	2556	2807	1165	F23	115	115	115
F10	2134	3491	6	F24	43	43	43
F11	1756	1761	1750	F25	40	43	37
F12	1277	1330	1213	F26	28	33	20
F13	1034	2695	27	F27	21	25	1
F14	1029	1033	498				

For a first insight in correlations in our dataset, we use Figure 3 to visualize a clustering of the evaluated feeds according to the number of common entries. That is, we run the Stoer-Wagner HCS (highly connected subgraphs) clustering algorithm [8] on a graph where the nodes represent feeds and the edges connect feeds with common entries

and are assigned a weight that equals the number of common entries. In Figure 3, we observe four clusters:

- One big cluster containing 7 (out of 27) feeds of different providers.
- Two smaller clusters consisting of 2 and 3 feeds from the same provider.
- One small cluster of two feeds (F11 and F15) where the number of common elements corresponds to the size of the smaller feed. This depicts an example of a feed (F15) that most likely contains a subset of the entries from another feed (F11).

Even though this clustering is not directly contained in the FeedRank algorithm, it shows that there are indeed correlations between the analyzed feeds.

In Figure 4, we show the correlation graph for the evaluated feeds. As explained in Section 3. B, this graph consists of nodes representing the feeds and directed, weighted edges that describe the percentage of confirmed entries from another feed.

FIGURE 3: EVALUATED FEEDS CLUSTERED BY THE NUMBER OF COMMON ELEMENTS. ABOUT 25% OF ALL FEEDS ARE CONTAINED IN ONE CLUSTER. FEEDS FROM THE SAME PROVIDERS ARE CONTAINED IN SMALLER CLUSTERS AND F15 DOES NOT LIST ELEMENTS THAT ARE NOT IN F11.

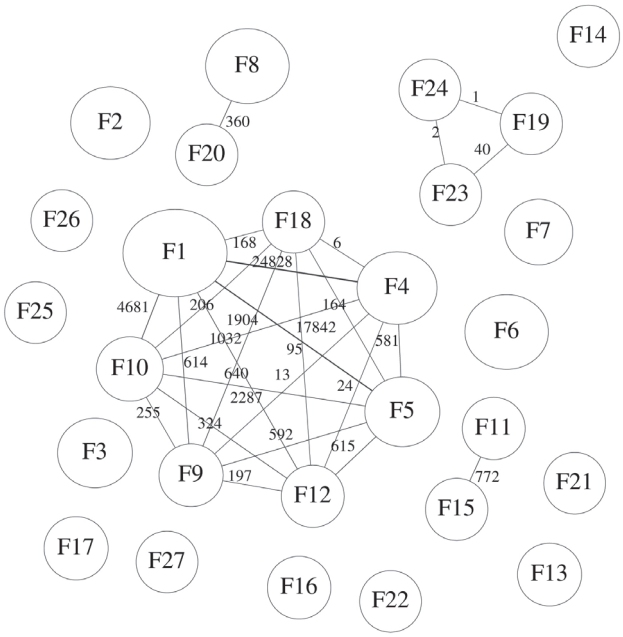


FIGURE 4: CORRELATION GRAPH FOR THE EVALUATED FEEDS. AS AN EXAMPLE, THE LARGEST FEED (F1) AND ALL ITS IN- AND OUTGOING EDGES ARE HIGHLIGHTED. THE EDGE LABEL DENOTES THE PERCENTAGE OF ENTRIES THAT A FEED CONFIRMS.

C. FeedRank vs. Individual Metrics

In this experiment, we compare the ranking obtained by individual metrics with the ranking according to FeedRank (see Table V). In Table VI, we show the ranking for all real feeds. The listed overall rank corresponds to the ranking according to the combination of all individual (or FeedRank) metrics. In this non-malicious case, we observe that the ranking according to the two metrics are strongly correlated (with a Spearman correlation coefficient of $\rho = 0.81$).

TABLE V: EVALUATED FEED METRICS.

Individual metrics	
Size	The average number of entries contained in the feed (more is better).
New	The average number of new entries per hour (more is better).
Removed	The average number of removed entries per hour (more is better).
FeedRank	
Contribution	A measure of how many additional entries a feed contributes.
Graph	The score obtained from the correlation graph.

TABLE VI: RANKING WITH INDIVIDUAL METRICS COMPARED WITH FEEDRANK. THE RANKINGS ARE STRONGLY CORRELATED ($\rho = 0.81$).

	Individual metrics				**FeedRank**		
Feed	Size	New	Removed	**Overall**	Contribution	Graph	**Overall** △
F1	1	5	7	3	1	10	1
F4	4	6	5	4	6	6	2
F13	13	4	4	6	9	5	3
F16	16	7	6	9	11	3	3
F8	8	2	2	2	2	14	5
F12	12	16	13	12	16	1	6
F9	9	15	21	17	10	8	7
F3	3	10	11	7	4	15	8
F10	10	3	3	5	13	7	9
F7	7	13	21	16	7	13	9
F2	2	9	21	11	3	18	11
F6	6	1	1	1	5	17	12
F5	5	11	9	8	14	9	13
F27	27	17	14	19	24	4	14
F21	21	20	16	18	18	11	15
F18	18	14	12	14	25	2	15
F17	17	8	8	10	8	21	15
F25	25	21	18	22	21	12	18
F11	11	18	15	14	12	22	19
F14	14	24	20	19	15	22	20
F20	20	12	10	13	19	20	21
F22	22	22	19	22	25	16	22
F19	19	25	21	25	17	24	23
F15	15	22	21	24	25	19	24
F23	23	25	21	26	20	24	25
F24	24	25	21	27	21	24	26
F26	26	19	17	21	23	24	27

D. Tamper-resistance

In this experiment, we evaluate the impact of RandomFeed and CopyFeed on the ranking. As the results in Table VII show, the dishonest feeds can obtain very good ranks (rank 1 for RandomFeed and rank 3 for CopyFeed) according to individual metrics, but not for FeedRank (rank 16 and 20).

TABLE VII: RANKING IN THE PRESENCE OF DISHONEST FEEDS. RANDOMFEED AND COPYFEED CAN TAMPER WITH WITH INDIVIDUAL METRICS, BUT NOT WITH FEEDRANK.

	Rank with individual metrics			**Rank with FeedRank**		
Feed	initial	+RandomFeed	+CopyFeed	initial △	+RandomFeed	+CopyFeed
F1	3	4	4	1	3	3
F4	4	5	5	2	2	2
F13	6	7	7	3	3	3
F16	9	10	10	3	3	5
F8	2	3	2	5	1	1
F12	12	13	13	6	7	7
F9	17	18	18	7	7	9
F3	7	8	8	8	11	11
F7	16	17	17	9	10	9
F10	5	6	5	9	6	5
F2	11	12	12	11	12	12
F6	1	2	1	12	7	8
F5	8	9	9	13	14	12
F27	19	20	20	14	18	17
F21	18	19	19	15	15	15
F17	10	11	11	15	13	14
F18	14	15	15	15	16	16
F25	22	23	23	18	19	19
F11	14	15	15	19	20	18
F14	19	20	20	20	21	20
F20	13	14	14	21	21	20
F22	22	23	23	22	23	24
F19	25	26	26	23	24	23
F15	24	25	25	24	25	27
F23	26	27	27	25	25	25
F24	27	28	28	26	27	27
F26	21	22	22	27	28	25
RandomFeed	n/a	1	n/a	n/a	16	n/a
CopyFeed	n/a	n/a	3	n/a	n/a	20

5. CASE-STUDY

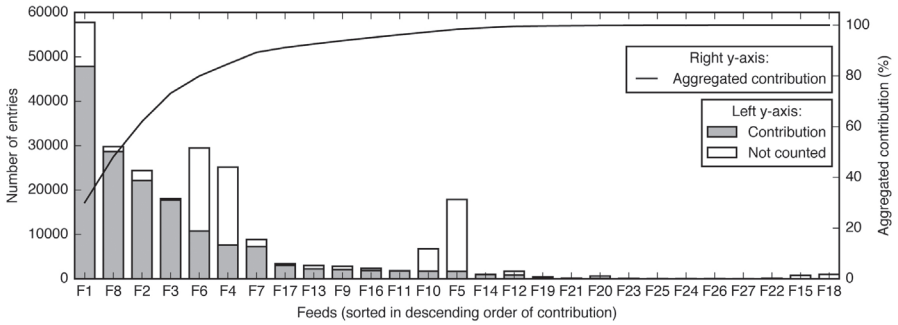
In this section, we come back to the two use-cases mentioned initially – network defenders that want to: *(i)* select the best feeds that together contain as many distinct entries as possible; and *(ii)* select the best feeds that list new entries before they appear on other feeds.

A. Prioritizing Completeness

To find a set of CTIFs that covers as many entries as possible (i.e. is as complete as possible) while not being susceptible to tampering attempts, FeedRank is used as follows. First, we compute the ranking solely according to the contribution. Since this ranking ignores the correlations, it is not tamper resistant and a CTIF that adds random entries can achieve a good rank. In a second step, we ensure tamper-resistance by excluding CTIFs whose graph score is below a user-defined percentile. Intuitively, the choice of this percentile reflects how many dishonest feeds the user expects. Here, we use the fifth percentile; that is, we assume that feeds whose graph metric is in the upper 95% are non-malicious.

In Figure 5, we plot the contribution of all collected CTIFs. As the figure shows, considering a small subset of all feeds is enough to cover a large percentage of all entries (e.g. the best 5 feeds together cover 80% of all entries). The figure also shows that by only looking at a feed's size it is not possible to derive the feed's contribution.

FIGURE 5: CONTRIBUTION OF ALL EVALUATED CTIFS. SELECTING 5 FEEDS IS ENOUGH TO COVER 80% OF ALL REPORTED IPS.



In Tables VIII and IX, we show the ranking in the presence of dishonest feeds.

RandomFeed has a high contribution score because its entries are most likely not listed on any other feed. However, because the vast majority of them are not confirmed by

any other feed, the graph score is very low. In particular, the graph score is below the 5th percentile, which is why the feed is not eligible to be used. CopyFeed has a poor contribution score because it is never the first feed to list any entry. However, because the copied entries originate from highly ranked feeds, CopyFeed deservedly achieves a good graph score.

TABLE VIII: RANKING ACCORDING TO THE CONTRIBUTION METRIC IN THE PRESENCE OF RANDOMFEED. THE GRAPH METRIC IS USED TO EXCLUDE POTENTIALLY DISHONEST FEEDS.

	FeedRank				
Feed	Contribution	Graph	Percentile	Rank △	Eligible
RandomFeed	28	1	0	1	No
F1	27	16	54	2	Yes
F8	26	21	71	3	Yes
F2	25	9	29	4	Yes
F3	24	11	36	5	Yes
F6	23	14	46	6	Yes
F4	22	22	75	7	Yes
F7	21	15	50	8	Yes
F17	20	13	43	9	Yes
F13	19	24	82	10	Yes
F9	18	19	64	11	Yes
F16	17	26	89	12	Yes
F11	16	7	21	13	Yes
F10	15	27	93	14	Yes
F5	14	17	57	15	Yes
F14	13	6	18	16	Yes
F12	12	25	86	17	Yes
F19	11	1	0	18	No
F21	10	20	68	19	Yes
F20	9	10	32	20	Yes
F23	8	1	0	21	No
F25	6	18	61	22	Yes
F24	6	1	0	22	No
F26	5	1	0	24	No
F27	4	23	79	25	Yes
F15	1	8	25	26	Yes
F18	1	28	96	26	Yes
F22	1	12	39	26	Yes

TABLE IX: RANKING ACCORDING TO THE CONTRIBUTION METRIC IN THE PRESENCE OF COPYFEED. THE GRAPH METRIC IS USED TO EXCLUDE POTENTIALLY DISHONEST FEEDS.

	FeedRank				
Feed	Contribution	Graph	Percentile	Rank △	Eligible
F1	28	16	54	1	Yes
F8	27	21	71	2	Yes
F2	26	8	25	3	Yes
F3	25	10	32	4	Yes
F6	24	13	43	5	Yes
F4	23	22	75	6	Yes
F7	22	14	46	7	Yes
F17	21	12	39	8	Yes
F13	20	24	82	9	Yes
F9	19	17	57	10	Yes
F16	18	25	86	11	Yes
F11	17	6	18	12	Yes
F10	16	27	93	13	Yes
F5	15	19	64	14	Yes
F14	14	5	14	15	Yes
F12	13	26	89	16	Yes
F19	12	1	0	17	No
F21	11	20	68	18	Yes
F20	10	9	29	19	Yes
F23	9	1	0	20	No
F25	7	15	50	21	Yes
F24	7	1	0	21	No
F26	6	4	11	23	Yes
F27	5	23	79	24	Yes
F15	1	7	21	25	Yes
F22	1	11	36	25	Yes
CopyFeed	1	18	61	25	Yes
F18	1	28	96	25	Yes

B. Prioritizing Speed

In this case study, a network defender wants to select CTIFs such that new entries are available as early as possible. For this, we rank the feeds according to the graph metric (that is, we ignore the contribution). FeedRank's correlation graph models the order in which entries appear in the feeds. Therefore, a feed that scores well in the

graph metric is one that is fast in including new entries. In contrast to computing the added entries for each feed individually, FeedRank ensures that it is impossible for a dishonest feed to tamper with the ranking.

In Table X, we show the resulting ranking with and without the dishonest feeds. RandomFeed appears at the very end of the ranking because its entries are not confirmed by other feeds. CopyFeed achieves a better rank because it copies the entries of highly ranked feeds with only a one-hour delay. By doing so, it is faster in listing these entries than other feeds that confirm the entries later.

TABLE X: RANKING ACCORDING TO THE GRAPH METRIC TO SELECT THE FASTEST FEEDS. THE DISHONEST FEEDS CANNOT ACHIEVE TOP RANKINGS.

	Rank with FeedRank		
Feed	initial △	+RandomFeed	+CopyFeed
F12	1	4	3
F18	2	1	1
F16	3	3	4
F27	4	6	6
F13	5	5	5
F4	6	7	7
F10	7	2	2
F9	8	10	12
F5	9	12	10
F1	10	13	13
F21	11	9	9
F25	12	11	14
F7	13	14	15
F8	14	8	8
F3	15	18	19
F22	16	17	18
F6	17	15	16
F2	18	20	21
F15	19	21	22
F20	20	19	20
F17	21	16	17
F14	22	23	24
F11	22	22	23
F19	24	24	26
F23	24	24	26
F24	24	24	26
F26	24	24	25
RandomFeed	n/a	24	n/a
CopyFeed	n/a	n/a	11

6. DISCUSSION

In this section, we first summarize the answers to the research questions, then discuss additional aspects of and choices that we made in the design of FeedRank.

A. Research Questions

Our research questions listed in Section 1 relate to the estimation of the quality of CTIFs, the CTIF ecosystem and the tamper-resistance of the evaluation metrics.

CTIF quality estimation. We use a graph-based correlation analysis together with a contribution analysis to measure correlations between CTIFs and the individual contribution of each CTIF. This allows us to estimate the relative quality of each CTIF with respect to all other analyzed CTIFs without requiring a ground truth.

CTIF ecosystem. Our correlation analysis allows finding clusters of highly correlated CTIFs and shows that most of the evaluated feeds are contained in the same cluster (i.e. most of the feeds overlap in terms of their entries but differ in terms of speed).

Tamper resistance. Our evaluation shows that correlation and contribution are tamper-resistant metrics for ranking CTIFs. While FeedRank produces a ranking that is strongly correlated with the ranking according to individual metrics in the absence of dishonest feeds, only FeedRank allows to identify dishonest feeds and to prevent them from achieving a good rank.

B. Speed of Dishonest Feeds vs. Execution Interval of FeedRank

For our evaluation, we use hourly snapshots and assume that the dishonest CopyFeed copies entries with a delay of one hour. If the delay were to be shorter than the snapshot interval, FeedRank could not determine whether CopyFeed and the two copied (legitimate) feeds listed the entries simultaneously or not. To prevent this inaccuracy, we envision the following mechanisms:

- The time between two snapshots can be decreased, which makes it more likely to be faster than dishonest feeds.
- CTIF providers can provide FeedRank with exclusive access to updates shortly before they are published.
- CTIF providers can add a few random (non-malicious and inactive) entries to their feeds to detect if another feed copies them (if these entries appear on another feed, it is highly likely that they were copied).

C. Evaluating CTIFs Instead of Evaluating Entries in CTIFs

With FeedRank, we assess the quality of CTIFs as a whole and not the quality of individual entries. With this, FeedRank provides the foundation to select CTIFs for deployment. The problem of evaluating the quality of particular entries is orthogonal to our work, but could be approached with a similar technique (e.g. by building a graph that models individual entries). From a network defender's point of view, the advantages of scoring threat intelligence at the level of CTIFs instead of individual entries are that: *(i)* reporting an IOC is delayed if an entry first needs to be verified by multiple feeds; and *(ii)* scoring CTIFs can be done once before deciding which feeds to use, which reduces operational and potential subscription costs.

D. Choice of the Evaluated Feeds

For our evaluation, we used a generic threat model and included a large set of freely available feeds covering different domains (e.g. generic, malware or phishing). Generally speaking, the set of considered CTIFs should contain all feeds that are suitable for the network defender's purpose. For example, a network defender that wants to select CTIFs for a spam filter should only consider feeds in this domain to get the most meaningful results.

Evaluating CTIFs for a more specific threat model or including commercial feeds is possible without modifying FeedRank but it is out of the scope of this paper.

7. RELATED WORK

To the best of our knowledge, we are the first to rank CTIFs based on their correlations and to consider potential manipulation strategies from CTIF providers. However, there has been previous work in the area of evaluating CTIFs and applying graph-based ranking algorithms in other domains.

A. Analysis and Evaluation of CTIFs

Sheng et al. study the effectiveness of phishing blacklists [9] and find that blacklists are ineffective when protecting users against phishing attacks because most phishing campaigns only last for a short time and blacklists are too slow in reacting.

Kührer and Holz describe a CTIF parser system [10] that records a large number of CTIFs and allows users to compute intersections between feeds and to query entries (e.g. domains). Entries that are contained in a large number of feeds are considered as being dishonest with high certainty. In later work [11], Kührer et al. propose a mechanism to identify parked domains and sinkholes (i.e. malicious domains that are identified and mitigated) in CTIFs.

Metcalf and Spring present an analysis of CTIFs over multiple years [12]. Similar to our approach, they analyze individual and combined features of CTIFs. However, they do not address the issue of dishonest CTIF providers that attempt to manipulate the rankings.

The Ponemon Institute found in a survey [3] that the application of threat intelligence is considered as very important for running secure systems but they did not investigate in the quality or the ecosystem of threat intelligence providers.

B. Graph-based Ranking

Page and Brin developed PageRank to rank websites [5]. They showed that the problem of ranking websites can be transferred to a graph problem and thus provided the foundation of transferring problems with several connected parties to graph problems.

Since then, concepts similar to PageRank have been applied to various problems, including to:

- Predict future relevance of scientific articles [13].
- Rank authors and publications [14].
- Rank correspondents according to their degree of expertise [15].
- Find influential users [16] and important content [17] in social networks.

8. CONCLUSION AND FUTURE WORK

The core concept of FeedRank is to model temporal correlations between feeds in a graph structure and to rank feeds based on this graph and the individual contribution of each feed. In contrast to traditional metrics that are applied to feeds individually, FeedRank is robust against tampering attempts by potentially dishonest feed providers.

In the evaluation and two case studies, we use data from 27 real feeds and show that FeedRank allows a reliable ranking even in the presence of dishonest feeds. For future work, we suggest using FeedRank to track the rankings of CTIFs over time. This will provide insights in the long-term behavior of CTIFs. Further, FeedRank could be extended by additional metrics and applied to related problems such as the evaluation of single entries in CTIFs.

REFERENCES

[1] D. Shackleford, “Who’s Using Cyberthreat Intelligence and How?,” SANS Survey, 2015.
[2] Symantec, “Internet Security Threat Report,” Bd. 22, 2017.
[3] “The Value of Threat Intelligence: The Second Annual Study of North American & United Kingdom Companies,” Ponemon Institute, 2017.
[4] H. Slatman, “awesome-threat-intelligence,” [Online]. Available: https://github.com/hslatman/awesome-threat-intelligence.
[5] L. Page, S. Brin, R. Motwani and T. Winograd, “The PageRank Citation Ranking: Bringing Order to the Web.,” Stanford InfoLab, 1999.
[6] “python-stix,” [Online]. Available: https://github.com/STIXProject/python-stix.
[7] “libtaxii,” [Online]. Available: https://github.com/TAXIIProject/libtaxii.
[8] M. Stoer and F. Wagner, “A Simple Min Cut Algorithm,” *Journal of the ACM (JACM)*, Bd. 44, Nr. 4, pp. 585-591, 1997.
[9] S. Sheng, B. Wardman, G. Warner, L. F. Cranor, J. Hong and C. Zhang, “An Empirical Analysis of Phishing Blacklists,” in *Proceedings of Sixth Conference on Email and Anti-Spam (CEAS)*, 2015.
[10] M. Kührer and T. Holz, “An Empirical Analysis of Malware Blacklists,” *PIK-Praxis der Informationsverarbeitung und Kommunikation 35.1*, pp. 11-16, 2012.
[11] M. Kührer, C. Rossow and T. Holz, “Paint It Black - Evaluating the Effectiveness of Malware Blacklists,” in *International Workshop on Recent Advances in Intrusion Detection*, 2014.
[12] L. Metcalf and J. M. Spring, “Blacklist ecosystem analysis: Spanning Jan 2012 to Jun 2014,” in *Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security*, 2015.
[13] H. Sayyadi and L. Getoor, “FutureRank: Ranking Scientific Articles by Predicting their Future PageRank,” in *Proceedings of the 2009 SIAM International Conference on Data Mining*, 2009.
[14] D. Zhou, S. A. Orshanskiy, H. Zha and C. L. Giles, “Co-ranking Authors and Documents in a Heterogeneous Network,” in *Seventh IEEE International Conference on Data Mining*, 2007.
[15] B. Dom, I. Eiron, A. Cozzi and Y. Zhang, “Graph-based ranking algorithms for e-mail expertise analysis,” in *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, 2003.
[16] Q. Wang, Y. Jin, S. Cheng and T. Yang, “ConformRank: A conformity-based rank for finding top-k influential users,” *Physica A: Statistical Mechanics and its Applications*, Bd. 474, pp. 39-48, 2017.
[17] E. Agichtein, C. Castillo, D. Donato, A. Gionis and G. Mishne, “Finding high-quality content in social media,” *Proceedings of the 2008 International Conference on Web Search and Data Mining*, 2008.