The Probability of Loss

How Threat Intelligence Quantifies Risk for the Business

by Levi Gundert and Bill Ladd, PhD

Recorded Future

Introduction

It can be argued that cyber threat intelligence (CTI) is most valuable to a business when it continuously informs a quantitative risk assessment model that contains specific probabilities for loss from a specific threat type.

In this paper, we'll use the techniques and formula articulated in *How to Measure Anything in Cybersecurity Risk* by Douglas W. Hubbard and Richard Seiersen[1], to estimate the probability of financial loss at Recorded Future (the company) based on multiple information security (INFOSEC) threats. We'll follow Hubbard and Seiersen's methodology for risk assessment, and illustrate how threat intelligence produced from Recorded Future (the product) helps the company make improved business decisions regarding internal security controls and the budget for those respective controls. The process we outline is designed to be replicated by other businesses searching for CTI benefits and/or an improved risk analysis model.

[1] Hubbard, Douglas W, and Seiersen, Richard. (2016). How to Measure Anything in Cybersecurity Risk. Hoboken, NJ: John Wiley & Sons.

The Problem

CTI is a necessary capability for the military, law enforcement, and intelligence agencies. Businesses have a different mandate though — to achieve long-term profitability. Thus, businesses are concerned with minimizing risk to avoid monetary loss. Operational risk specifically from cyber threats is a modern day primary concern. The problem is that CTI is a technical domain, but businesses typically speak and understand the language of risk, especially senior executives and board members.

CTI is a straightforward proposition — identify existing and evolving information threats and recommend corresponding controls to mitigate the threat's potential future impact to the business and parallel monetary loss. In the enterprise, the translation between threat and risk is often miscommunicated and thus misunderstood. Without a robust framework for measuring business benefit through risk, CTI is perceived solely as an expensive information security control.

Businesses also run the risk of conflating governance, risk, and compliance (GRC) programs with specific quantifiable risk and loss analysis. For example, NIST's Cyber Security Framework (CSF) is useful for auditing, understanding control gaps, and monitoring progress on improvements. Sections ID.RA-3 through ID.RA-6 even cover risk assessment, but this topic is often translated as a GRC requirement already completed through inadequate existing processes.

The current risk analysis equation most often used by enterprises today involves "likelihood of occurrence X impact." The inputs to the equation are usually subjective and rarely require any substantive thought. The outputs generally lead to a traffic light system (red/yellow/green) that doesn't communicate anything specific about the probability for business impact and associated loss. As Hubbard and Seiersen point out, this qualitative process leads to inadequate inputs, and therefore inadequate results, which senior leaders rely on to make risk management decisions.

Profitable insurance companies only underwrite policies subsequent to calculating comprehensive risk with the help of actuaries and/or data scientists. Chief information security officers (CISOs) need to do the same before making purchasing decisions that affect the business's bottom line.

The CTI industry is currently failing to materially benefit the business, which is untenable long term given the cost of building or outsourcing a comprehensive CTI program. To be successful, CTI must improve business decisions related to risk and loss, specifically the corresponding budget decisions for controls.

Definitions and Context

Threat intelligence is **the act of formulating an analysis** based on the identification, collection, and enrichment of relevant threat information.

CTI is principally delivered in two forms: strategic and operational (sometimes referred to as "tactical").

Operational threat intelligence is a programmatic process by which computers are configured to collect various indicators of compromise (hashes, IP addresses, etc.) and correlate them — usually in security incident and event management (SIEM) products — with internal telemetry (log) data to identify previously undetected compromises. The value of operational threat intelligence is directly proportional to the source(s) of proprietary or third-party data in use. The benefit of operational threat intelligence is that it acts as a security control itself to monitor and improve other existing controls. There are three primary operational value measurements:

1. Telemetry detection ratio (true positive alerts). This number should decrease over time as controls improve, unless new threat sources are being added.
2. Number of new control rules created. This is measured in regular intervals (monthly/quarterly/annually).
3. Future efficacy rate of new controls based on the previous bullet.

Operational threat intelligence is useful, but it produces ancillary value to the business.

Conversely, **strategic threat intelligence** requires human analysts to identify and deconstruct relevant threats to a business's employees, customers, infrastructure, applications, and vendors. Strategic threat intelligence produces primary value for a business when properly translated through a robust quantitative risk analysis framework. Business decision improvement (via strategic threat intelligence) metrics include:

- Number of security architecture improvements or changes.
- Number of new penetration testing or red team scenarios and associated control scores.
- Number of new internal threat hunting methodologies and corresponding future threat identification efficacy rates.
- Number of employees trained on specific information threat scenarios and prevention strategies.
- Number of risk management decisions affected (risk manager input required).
- **Annualized number of updated loss probabilities delivered to senior executives.**

A pithy definition of **risk**, as defined by Hubbard and Seiersen, is that "something bad could happen." Risk moves both ways, there is upside (missing an opportunity) and downside risk, but for the purposes of this exercise, we are constraining risk to the negative context.

The Process

The first step toward providing useful values in a risk assessment formula is becoming a trained estimator. For more information on the proven science of estimating, consult Hubbard and Seiersen's research. Humans are not good estimators of combined risk, but reasonable confidence levels around individual threats can be estimated with remarkable accuracy. Hubbard and Seiersen present exercises for training to become calibrated estimators equal to a 90% confidence interval, which means nine out of ten questions will be answered with ranges (a low and high value) where the answer falls somewhere in-between.

Hubbard and Seiersen point out, and our experience confirms, that when first estimating answers to random trivia questions, the tendency is toward overconfidence which leads to correctly answering far less than 90%. After successive attempts toward understanding when to mentally expand or contract a range of values, we became calibrated estimators. It is easy to discount the value of estimation training, but we found this process invaluable to identify our respective biases, and carefully estimate the ranges for our risk model.

The second step in creating a CTI risk analysis formula is listing cyber threat categories. While not comprehensive, the following is a sample threat taxonomy for this exercise:

Commodity phishing
Spear phishing
Pharming
Domain theft
Exploit kit
Watering hole and zero day
Unpatched RCE against application
Unpatched RCE against infrastructure
Web application exploit
Vendor intrusion
DDoS via botnet
DDoS via reflection/amplification
Ransomware
Recorded Future application compromise
Removable media
BGP hijacking
Malware on accounting workstation

After identifying cyber threats, we estimated a low and high bound for each of the following four questions given a specific threat:

1. What is the likelihood (in percentage terms) that a specific threat will impact the business this year?
2. What is the likelihood that, if the specific threat occurs, it will cause loss to the business's information confidentiality and/or integrity?
3. What is the likelihood that, if the specific threat occurs, it will cause loss to the business's service availability?
4. What is the range of expected loss (in currency) if the specific threat impacts the business?

Applying a 90% confidence threshold to estimating these ranges requires deep thought and often input from other individual contributors in the organization, specifically related to existing controls. Additional calibrated estimators contributing ranges to these questions creates higher-quality estimates overall as the process promotes discussion of cyber threats and corresponding controls. It may be difficult to articulate subconscious information, but practitioners know more than they think they know. Hubbard and Seiersen begin with removing the ridiculous, at which point 90% confidence becomes attainable, and that is the starting point for a useful quantitative model. A carefully estimated range from multiple calibrated estimators is always more valuable than time-pressured arbitrary values.

The following values were entered specifically for Recorded Future (the company), and we provide five specific threats as an example. A sample Excel spreadsheet and accompanying formulae are provided in Chapter 6 of H*ow to Measure Anything in Cybersecurity Risk*.

Event Name	Prob. Event Will Happen (Annual)	Type of Event If It Occurs		
		Only Conf/Int	Only Availability	Both
Targeted Spear Phishing	7%	98%	1%	1%
Web Application Exploit (SQLi, XSS, CSRF, etc.)	10%	50%	0%	50%
Botnet DDoS	2%	0%	100%	0%
Reflection/Amplification DDoS	13%	0%	100%	0%
Ransomware	5%	0%	0%	100%

In the above chart, we provide sample probability estimates from the five highest risk threats in our initial model. We estimate that the highest probability threat events for Recorded Future occurring in 2017 are (in no particular order) spear phishing, web application exploits, ransomware, and denial of service.

It is import to re-emphasize that these probability estimates are simply that, estimates. They are based on assessments of existing controls and industry-wide threat prevalence. They are also extremely transparent. As more people encounter the model's results, they may provide improved model estimates.

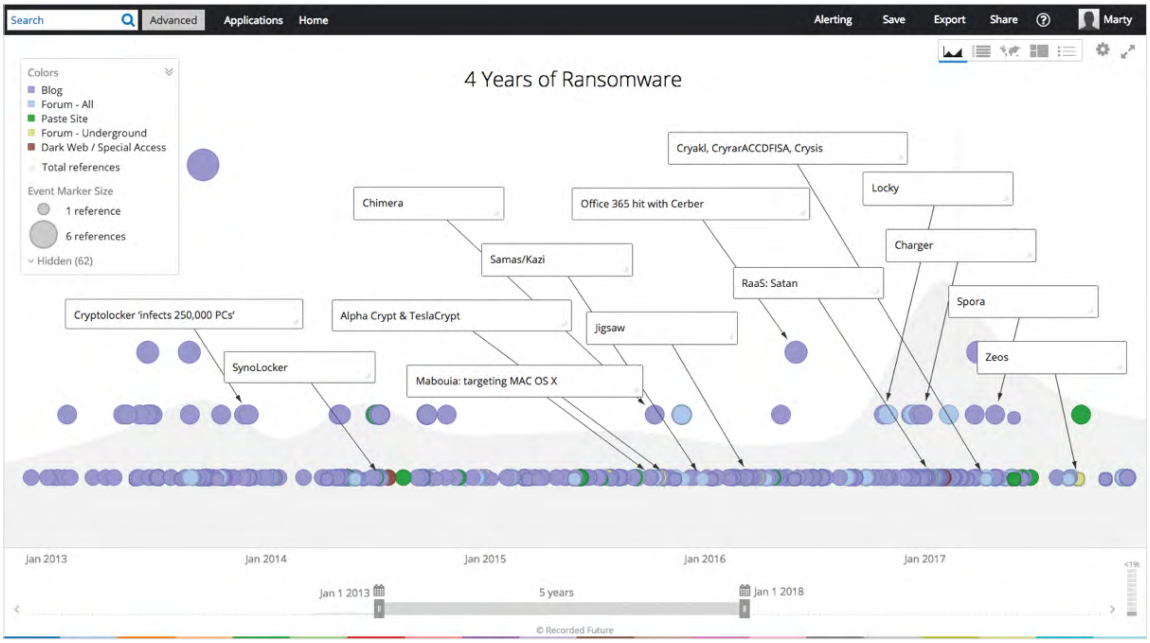
For example, the estimate for targeted spear phishing is based on knowledge of Recorded Future's current email controls and knowledge of prior successful spear phishing campaigns against other companies.

The estimate for a web application exploit similarly derives from current knowledge about the rate of new exploits, specifically for web application frameworks. Recorded Future's developers are well versed in secure coding practices, and internal security professionals monitor for new vulnerabilities, and administrate a bug bounty program. Yet the constant threat of previously undisclosed web application framework vulnerabilities leads to an estimate of 10%.

The third event category, botnet DDoS (distributed denial of service), is different — from a DDoS attack leveraging stateless protocols (UDP) to achieve amplification and reflection — because this type of DDoS attack has historically demonstrated lower impact capabilities. Recorded Future's current DDoS mitigation controls are likely to prevent a DDoS scenario initiated by a traditional botnet (e.g., BlackEnergy). Thus the estimated probability is 2%.

A DDoS attack initiated via vulnerable internet devices that cause amplification and reflection may be capable of overwhelming Recorded Future's controls in the most extreme cases. The recent Mirai attacks achieving 600 GBps are unstoppable without a coordinated response from tier-one internet service providers. The probability for this type of attack against Recorded Future is 13%. Additionally, Recorded Future publishes security research which factors into the estimate of future targeting by disgruntled actors.

The last event type, ransomware, is estimated based on the increasing number of families being created and propagated in the underground criminal economy. Initial infection of a Recorded Future computer is most likely via phishing or web drive-by (malvertising). The mix of employee workstation operating systems combined with application patching/updates does not remove all probability of ransomware infection via web drive-by, but it is minimal. However, there is always a small probability that commodity phishing could lead to ransomware infection via a typical mechanism like a Microsoft Office document employing a macro. Thus, the ransomware probability estimate is 5%.



Recorded Future timeline depicting the proliferation of new ransomware families.
RF Source: https://app.recordedfuture.com/live/sc/7v5xte11koeW

The next set of estimates involve dollar sample damage ranges triggered by the loss of confidential information/integrity and system availability, followed by estimates of system availability loss and associated dollar costs.

Event Name	90% Interval for Loss Due to Confidentiality/ Integrity		90% Confidence of Intervals for Loss Due to Availability			
	Lower Bound	Upper Bound	Duration of Outage (Hours)		Cost Per Hour ($)	
Targeted Spear Phishing	$500	$350,000	1	4	$300	$2,000
Web Application Exploit (SQLi, XSS, CSRF, etc.)	$250	$150,000	0.5	4	$100	$5,000
Botnet DDoS	$200	$2,000	0.5	24	$50	$10,000
Reflection/Amplification DDoS	$200	$50,000	0.5	72	$50	$10,000
Ransomware	$100	$10,000	0.5	8	$50	$500

Calibrated 90% confidence limits of losses.

The calibrated ranges for dollar loss are crucial to think through with as many stakeholders as possible.

The Results

After our variables are entered, we use random seeds to run the model numerous times. For example, in one of our sample runs using Hubbard and Seiersen's Excel formula, the outcome reveals a loss of $48,600 due to reflection/amplification DDoS attack(s). The majority of individual model simulations result in zero losses.

Next, to adequately summarize the outcomes from the model, we port the formula to R to run the model 250,000 times and summarize how often we observe losses at different levels.

Loss	Percent of Simulations With This Loss or Higher
$0.00	50%
$1,064.32	35%
$8,130.57	25%
$48,583.68	10%
$55,396.93	5%
$240,757.59	1%

Summarized loss results for Recorded Future in 2017.

The previous chart reveals the levels of loss Recorded Future is likely to experience from information threats. Sixty percent of the simulations had no loss, and 1% of the simulations had a loss of $240,000 or more.

The losses start at the 35% level, which means there is a 35% probability that Recorded Future could experience any information security-related loss in 2017. The potential losses increase through the 10% level where there is a 10% risk for a $48,000 loss.

Assessing which threats and corresponding control gaps are most likely to result in significant loss, is as interesting as the overall probability of losses.

Risk Factor	1% Frequency Loss
Spear Phishing	$107,481.82
Webapp.exploit	$73,183.46
Refl.Amp.DDoS	$48,583.68
Botnet.DDoS	$17,907.77
Ransomware	$3,753.98

Potential significant losses for selected risk factors. 1% of simulations had losses at these levels or higher.

These levels both illustrate important points about prevalence and level of loss. If we only have a 1% probability of losing about $18,000 to a botnet DDoS attack, do we invest in additional controls for further protection?

Following the results of our simulations, we discuss appropriate risk thresholds with internal risk managers, and implement additional controls where necessary. George E. P. Box, a famous statistician, once wrote "essentially, all models are wrong, but some are useful." When calculating risk, Box's motto applies. The advantage of a quantitative formula and simulations is that the outcome can be discussed and defended or changed due to the transparent assumptions, variables, and formula in use.

Engaging a risk manager or senior executive with specific probabilities about future loss accomplishes two important goals: one, it facilitates discussion about the model used to produce the probabilities, which may lead to improvement in the model, and two, once decision makers trust the model, the resulting specificity makes risk management much easier.

Consider the difference in these two statements:

> "The threat of denial of service to our business has changed from high to medium (red to yellow)."

vs.

> "Based on our best estimates, there is a 10% probability that our business will incur a loss of over $48,000 in 2017 caused by damage to availability of information via distributed denial of service."

Similar statements regarding the threat of ransomware:

> "The threat of ransomware to our business has changed from low to medium (green to yellow)."

vs.

> "There is a 5% probability that our business will incur a loss of $1,408.90 in 2017 due to ransomware."

The combined losses from multiple threat categories are also summarized with useful specificity. Based on the current simulations, there is a 20% probability that Recorded Future will incur losses of $48,583.68 or more based on multiple threats impacting the confidentiality, integrity, and/or availability of information. Granular figures improve trust and confidence in risk management, and facilitate more efficient use of company resources to lower risk where needed. Thus, it is the calibrated estimators producing ranges as variables in the model that are the crucial product of strategic threat intelligence.

Using Recorded Future to Continually Update Loss Probabilities

Recorded Future's patented machine learning and natural language processing (NLP) technology provides threat intelligence from open, closed, and technical sources in an intuitive analyst interface.

Three different Recorded Future strategic threat intelligence features facilitate real-time updates to improve calibrated estimator's knowledge and awareness:

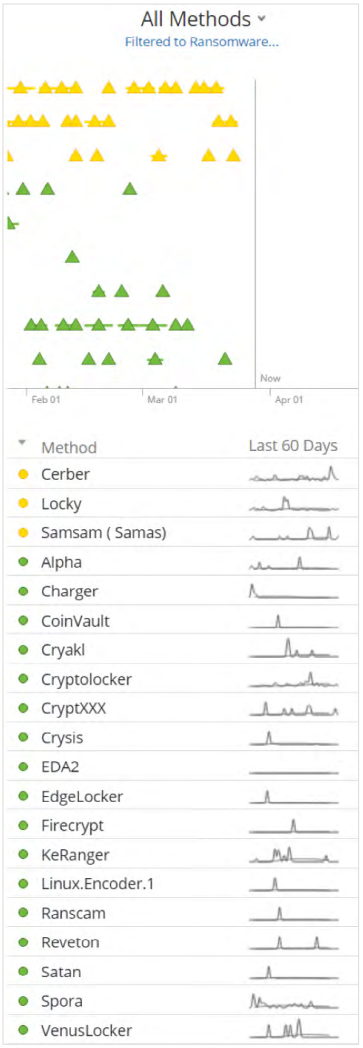
› Customizable dashboard

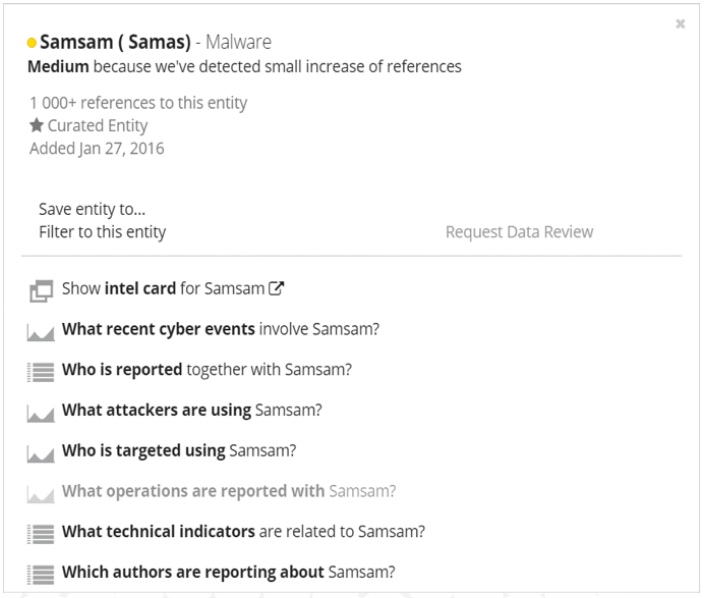
› Alerts

› API

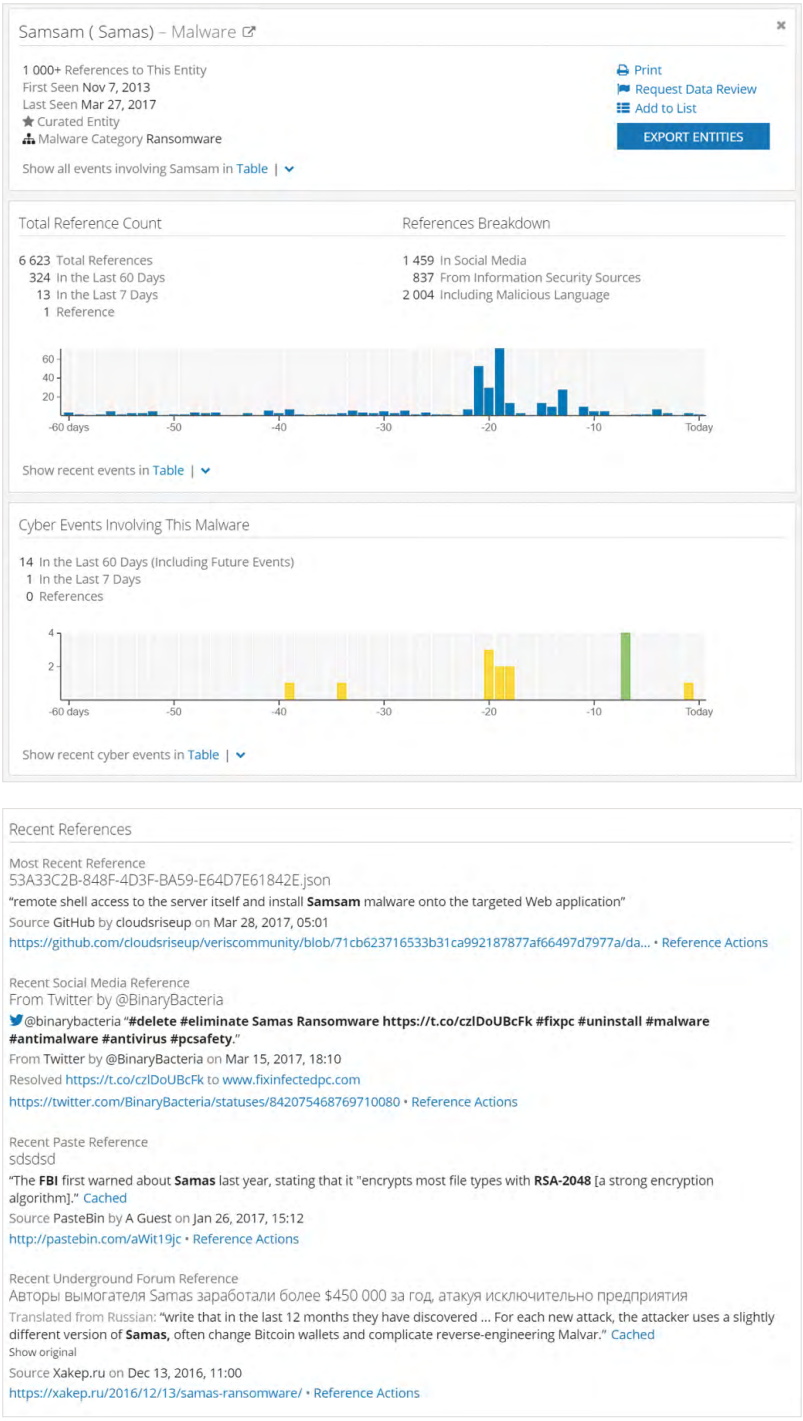
The customizable dashboard is a useful option for ongoing monitoring of a large threat category such as ransomware. In the below image, the trend line to the right of each ransomware family indicates increasing or decreasing mentions across thousands of web properties including code repositories, paste sites, security research blogs, criminal forums, and .onion (Tor accessible) forums. Similarly, the top summary timeline depicts trending and new ransomware families.

Clicking on a specific ransomware family produces an intelligence summary card which is the aggregation of large disparate data that empowers 10 times faster analysis.



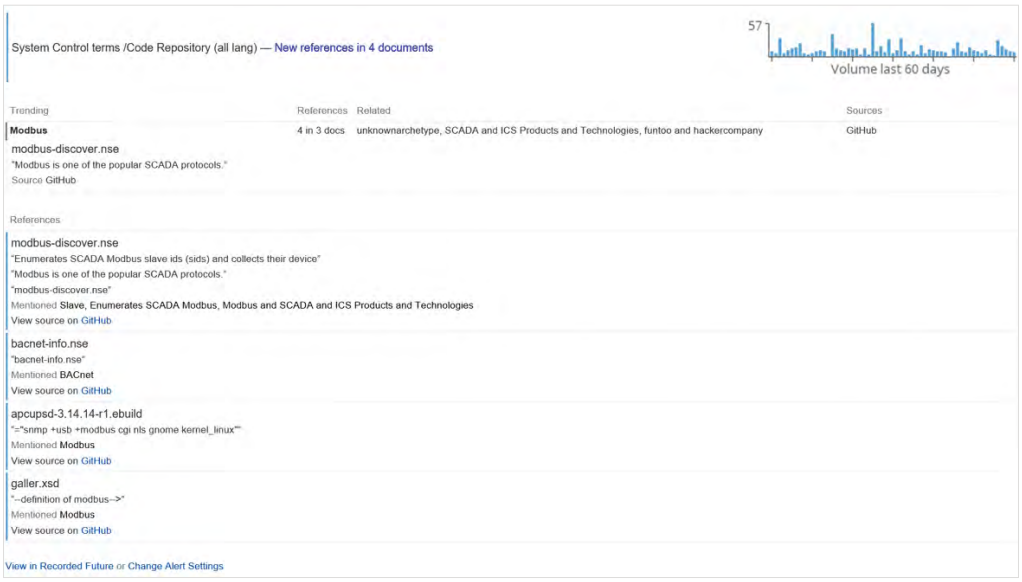


Click image to view larger version in browser.



The Intel Card for Samsam malware.
Click image to view larger version in browser.

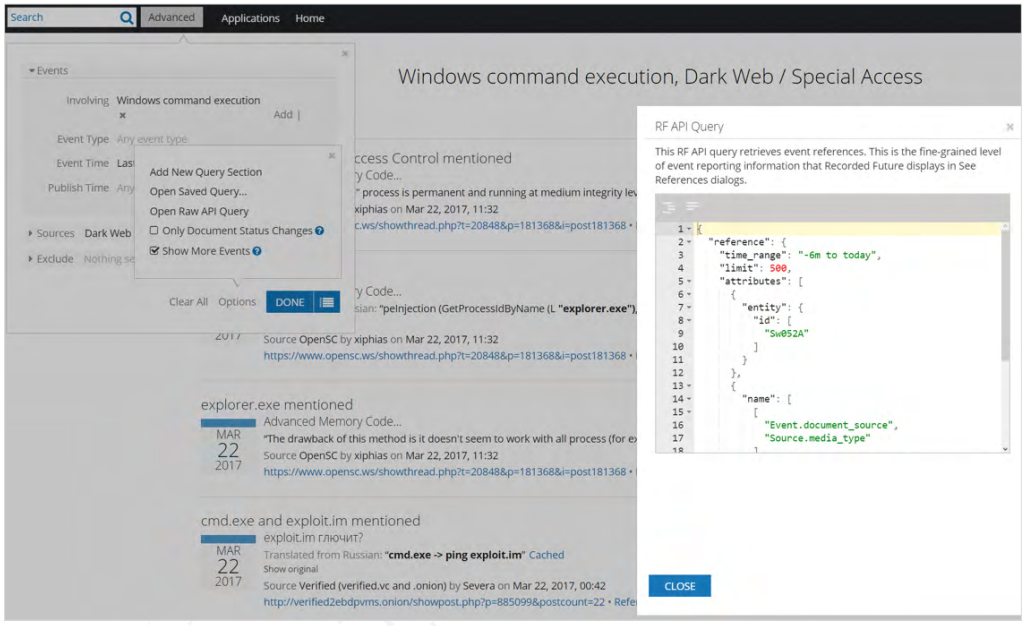
A second option is email alerting. The following daily alert is configured to surface new documents from code repositories that mention ModBus or DNP3, two control system protocols. For energy companies, this strategic awareness is directly beneficial to understanding the latest scanners and tools which may be used to target and attempt exploit of vulnerable control systems. Updated knowledge leads to improved calibrated estimates of loss.



Potential significant losses for selected risk factors. 1% of simulations had losses at these levels or higher.
Click image to view larger version in browser.

Last, Recorded Future's RESTful API is flexible for integration with third-party software for alerting or event processing. A query in Recorded Future is easily exportable as JSON to be used with a Python wrapper (or similar).

A Recorded Future query for a list of Windows command execution terms (wscript.exe, wmiprvse.exe, powershell.exe, etc.) for the past three months in dark web and special access forums produces the following JSON for export.



Exporting a Recorded Future query for API use.
Click image to view larger version in browser.

A Python program example using the exported JSON and Recorded Future API appears below.

```
from rfapi import ApiClient
from pprint import pprint
import json
API = ApiClient()


def query():
    return API.paged_query({
  “reference”: {
    “time_range”: “-3m to today”,
    “limit”: 500,
    “attributes”: [
      {
        “entity”: {
          “id”: [
            “Sw052A”
          ]
        }
      },
      {
        “name”: [
          [
            “Event.document_source”,
            “Source.media_type”
          ]
        ],
        “entity”: {
          “id”: [
            “OYHH7k”
          ]
        }
      }
    ]
  },
  “comment”: “306XrAYcHc”
}
  }
}, batch_size=10000, field=’instances’)

def main():
    instance = list(query())
    pprint(json.dumps(instance, indent=2))
if __name__ == ‘__main__’:
    main()
```

Conclusion

CTI is critical for a business when it is continuously updating a quantitative risk model with transparent assumptions, variables, and outcomes. The current industry default equation of "likelihood of occurrence x impact" is inadequate to inform decision makers about the future probability of loss to the business, because it begins with ambiguous variables and leads to ambiguous conclusions that cause confusion.

Businesses speak the language of risk, and financial loss due to cyber threats is a top concern. However, there is currently an inability to articulate specific probabilities of future loss based on information threats and current security controls. Clarity comes from specificity. Without specific loss probabilities, businesses are potentially wasting resources on unnecessary controls or lacking controls where they are most needed.

Becoming a calibrated estimator is a critical step to providing correct ranges in the Hubbard and Seiersen formula. Thus, calibrated estimators need near real-time CTI to improve their knowledge and awareness as threats evolve. Recorded Future is unique threat intelligence software with an unrivaled breadth of open, closed, and technical sources that is ideal for informing calibrated estimators, and by extension a robust risk analysis model.

For additional information on building a quantitative risk model refer to *How to Measure Anything in Cybersecurity Risk* by Douglas W. Hubbard and Richard Seiersen. For additional information on updating a quantitative risk model with relevant threat intelligence contact us at info@recordedfuture.com.

About Recorded Future

Recorded Future delivers threat intelligence powered by machine learning, arming you to significantly lower risk. We enable you to connect the dots to rapidly reveal unknown threats before they impact your business, and empower you to respond to security alerts 10 times faster. Our patented technology automatically collects and analyzes intelligence from technical, open, and dark web sources to deliver radically more context than ever before, updates in real time so intelligence stays relevant, and packages information ready for human analysis or instant integration with your existing security systems.