STRIKING SECURITY GOLD

Uncovering hidden insights in a decade's worth of RSA Conference abstracts.

As the premier security conference in the world, RSA Conference offers an excellent lens through which to study the topics and trends within our industry. The Conference's slogan of "Where the World Talks Security" shows that's not just an accident; it's the goal.

But what exactly do we talk about when we talk "security?" That's the question we seek to answer in this report, which has its roots in a similar question asked by an eight-year-old daughter two and a half years ago: "What's the RSA Conference about, Daddy?" That root sprouted into a four-part blog series and a panel discussion a year later where we analyzed 25 years of session titles in honor of the 25th anniversary of RSA Conference.

To really study the question, however, titles provide limited value. They're often created to grab attention rather than impart information. Call for Paper (CFP) submissions, by comparison, are a veritable goldmine of details and insight about the sessions just waiting to mined. Once again, RSA Conference was kind enough to supply the ore for our digital pickaxes. Did we strike gold and unearth valuable nuggets of insight about our industry? You'll have to read on to find out.

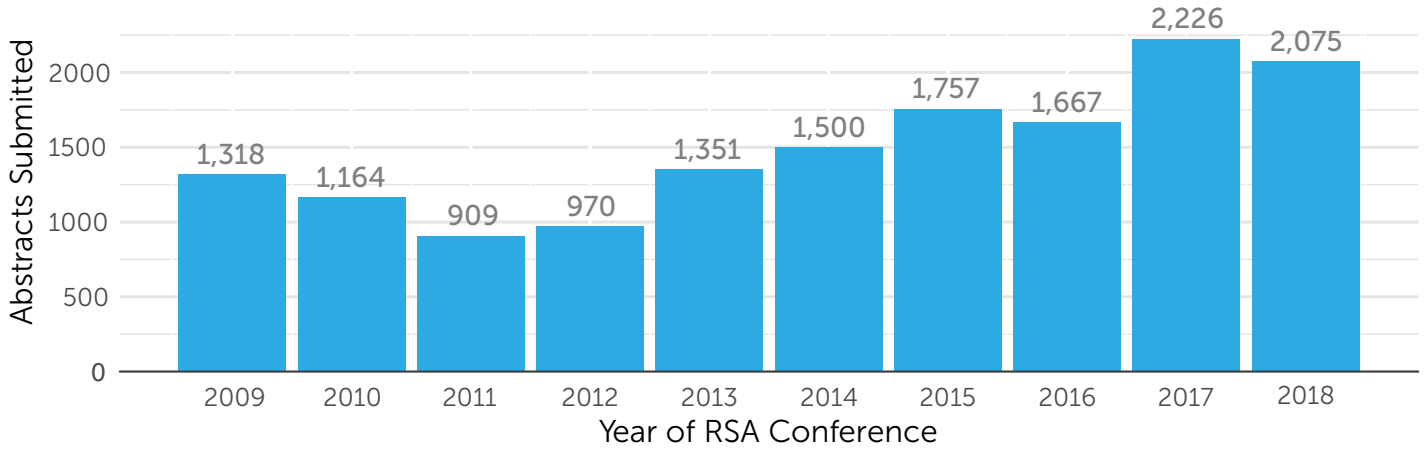
RSA Conference | Where the world talks security

CYENTIA INSTITUTE

This report was produced by the Cyentia Institute, a research firm that seeks to advance cybersecurity knowledge and practice through data-driven analysis. We curate knowledge for the community, partner with vendors to create compelling research, and help enterprises gain insight from their data. Find out more: www.cyentia.com.

The Corpus

All told, the RSA Conference committee provided just shy of 15,000 CFP submissions over a 10-year period from 2009 to 2018. Those familiar with the process know CFPs include information like the session title, abstracts of varying length, objectives, target audience, etc. This report focuses solely on the text contained within the Long Session Abstract (2009-2012) or Session Details (2013-2018) of the CFP, as it is the fullest description of topics covered in the proposed session.



Source: Cyentia Institute with data from RSA Conference

Figure 1: Annual RSA Conference Submissions

"Proposed session" is an important distinction here because only a subset of these CFPs was accepted and presented at the Conference. CFP proposals are thoughtfully selected to achieve reasonable balance across sessions of many different interests, categories, and target audiences. In that sense, we believe the CFPs give a more unfiltered measure of the cybersecurity community's level of interest across various topics. At the same time, this approach will naturally amplify hot topics to a greater degree than colder ones. We're okay with that; in fact, it's one of the reasons this analysis is so interesting.

As you may suspect, the corpus of text for this analysis is quite large. The 15,000 CFP abstracts, which are limited to 2,500 characters, contain 46,000 unique words. That number may seem small, but consider the fact that 25,000 unique words comprise the works of Shakespeare,[1] and you get the sense that we collectively possess a pretty strong vocabulary.

The techniques we leverage against this corpus fall under two broad categories. The first is Natural Language Processing (NLP), specifically topic modeling and clustering. The second is a classification system developed for the Cyentia Research Library, which contains hundreds of industry reports from cybersecurity vendors and other organizations. And much of what we present in this report is a blend of both techniques.

[1] https://www.twinword.com/blog/how-many-words-does-the-average-person-know/ Also, the typical U.S. adult has a vocabulary between 20,000 and 30,000 words, but uses only about 5,000 in everyday speech.

Topical Analysis

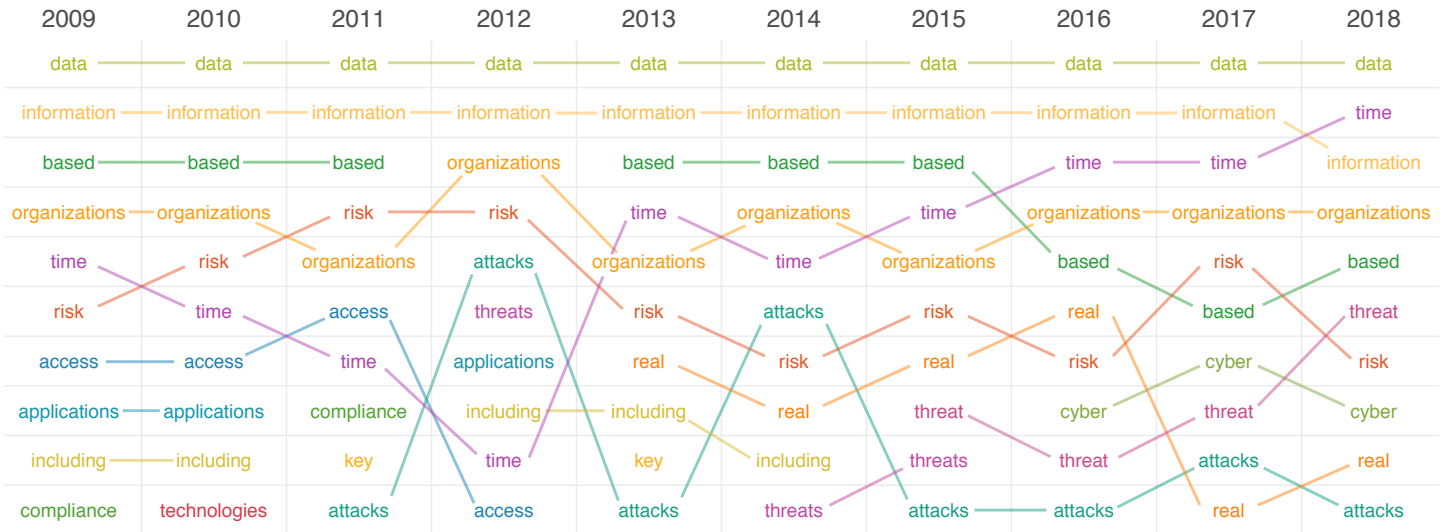
The goal of this section is to identify common topics covered by RSA Conference abstracts and study how those topics change over time. We employ several different approaches in pursuit of that goal, ranging from simple to complex. We'll start with simple.

Taking It From the Top

The absolute simplest way to begin exploring topics within a corpus of documents is by counting words. Figure 2 shows words that appear in the largest number of abstracts each year...and it's not very informative (though it does have a lot of "data" and "information"). In fact, the only reason we've included it here is to show the limits of simple word frequencies (you know, word clouds) for topical analysis. Let's try another approach.

To Cyber or Not to Cyber?

Though most of us have gotten over the "You said 'cyber'—drink!" silliness of several years ago, the question of how to reference our field is still unsettled in many circles. Rather than offer a personal contention in favor of "information security," "cybersecurity," or something else, we'll appeal to the corpus and accept its ruling on this case.



Source: Cyentia Institute with data from RSA Conference

Figure 2: Most Common Words in RSA Conference Abstracts Each Year

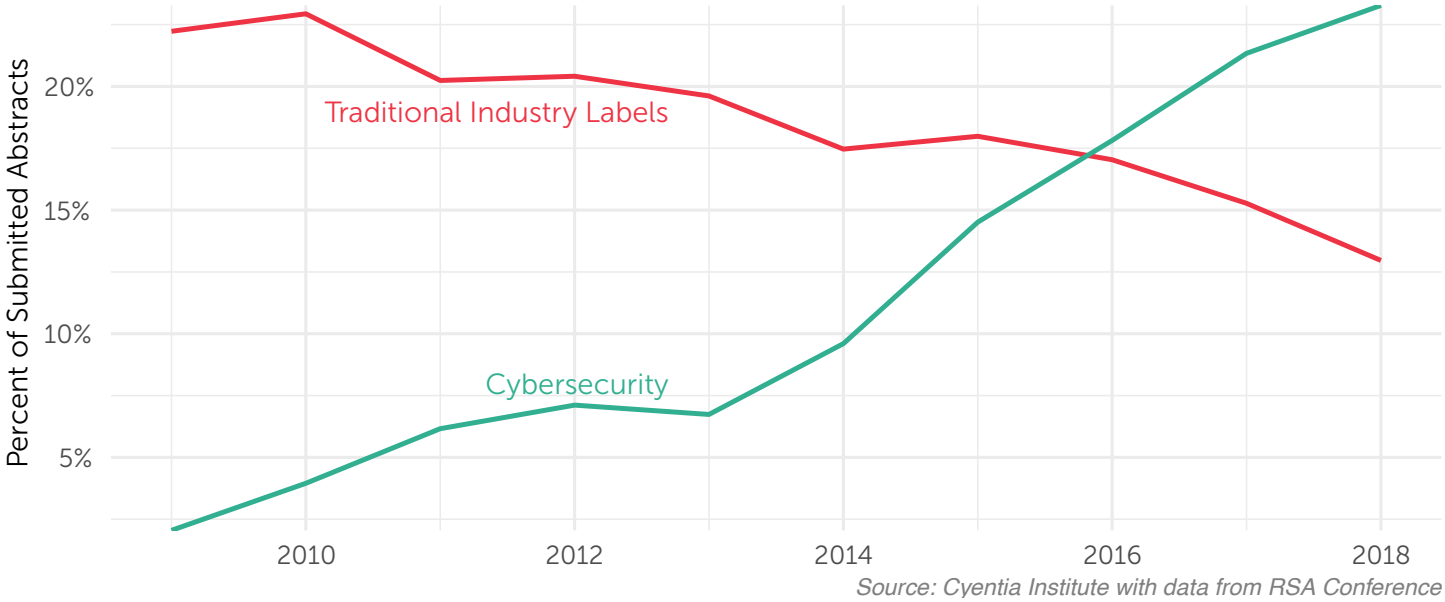


Figure 3: Cybersecurity vs Traditional Industry Labels[2] in RSA Conference Abstracts

To those still fighting the "cyber" war, the corpus definitively declares resistance is futile. The battle was lost in 2015. Remember that these are submissions from your peers, so yield to assimilation and join the fold. We offer this as an olive branch: you just might find those outside the industry better understand what you mean when you tell them you're in "cybersecurity" rather than whatever it is you've been using. To those still unconvinced who want to keep fighting the good fight, we wish you a heartfelt "good luck storming the cyber-castle."

Oh, I Remember That Year!

Perhaps a look back at words and phrases indicative of each year of the RSA Conference would be a good way to settle in to the corpus. For this, we use a statistic that measures the relative importance of terms among submissions of a given year compared to other years ("tf-idf"). We let the math do its thing for Figure 4 and didn't impose any rules or guidance on what it identified other than restricting it to single or paired words.

[2]Traditional industry labels include information security, infosec, network security, data security, and enterprise security.

2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
web 2.0	web 2.0	data	real–world	BYOD	APT	BYOD	IoT	IoT	IoT
network access control	anti–virus	risk	real–time	tablet	BYOD	IoT	threat actors	ransomware	ransomware
anti–virus	social networking	organizations	cloud–based	APT	security analytics	security analytics	BYOD	devops	GDPR
cross–site scripting	cross–site scripting	access	third–party	anti–virus	mobile apps	threat actors	security analytics	threat actors	iot devices
PCI–DSS	PCI–DSS	compliance	in–depth	MDM	software–defined	home depot	kill chain	kill chain	devops
sarbanes oxley	myspace	key	high–profile	iOS	MDM	snowden	devops	GDPR	blockchain
service–oriented	conficker	attacks	real–life	stuxnet	iOS	software–defined	OPM	blockchain	equifax
unified communications	VOIP	applications	Epsilon	Flame	stuxnet	data science	software–defined	cyber insurance	wannacry
javascript	payment card	control	end–user	mobile apps	tablets	devops	NIST CSF	security analytics	threat hunting
management strategy	ratio	process	enterprise–wide	advanced malware	prism	heartbleed	iot security	NIST CSF	bitcoin
PDAs	security standard	enterprise	zero–day	kill chain	advanced malware	kill chain	anthem	dark web	deep learning
email security	data security	environment	cost–effective	software–defined	dropbox	ransomware	dark web	bitcoin	devsecops

Source: Cyentia Institute with data from RSA Conference

Figure 4: Most Important/Special Words in RSA Conference Abstracts Each Year

"What a difference a decade makes!" That's our initial reaction to Figure 4, and perhaps yours as well. If that eight-year-old mentioned above were sitting here now, she'd ask "What's a PDA, Daddy?" Anti-virus was apparently a standout topic a decade ago, but its relative importance among Conference submissions (and enterprise security programs) has waned of late. Oh, and 2010 called; it wants its Myspace account reactivated.

Looking over the terms associated with each year is actually quite fascinating. We sure wanted to talk about data and risk concepts in 2011, but must have found that too restrictive and decided to just "keep it real" in 2012. BYOD wins 2013 hands down, while 2014 goes to Mandiant. (APT1 dropped a week before RSA Conference 2013) 2015 is a tough one; too bad we can't have a battle royale between Snowden, Home Depot, data science, and Heartbleed. We're giving the belt to data science for simply making that matchup possible. We'll cut this short and just award 2016 to IoT, 2017 to ransomware, and 2018 to GDPR.

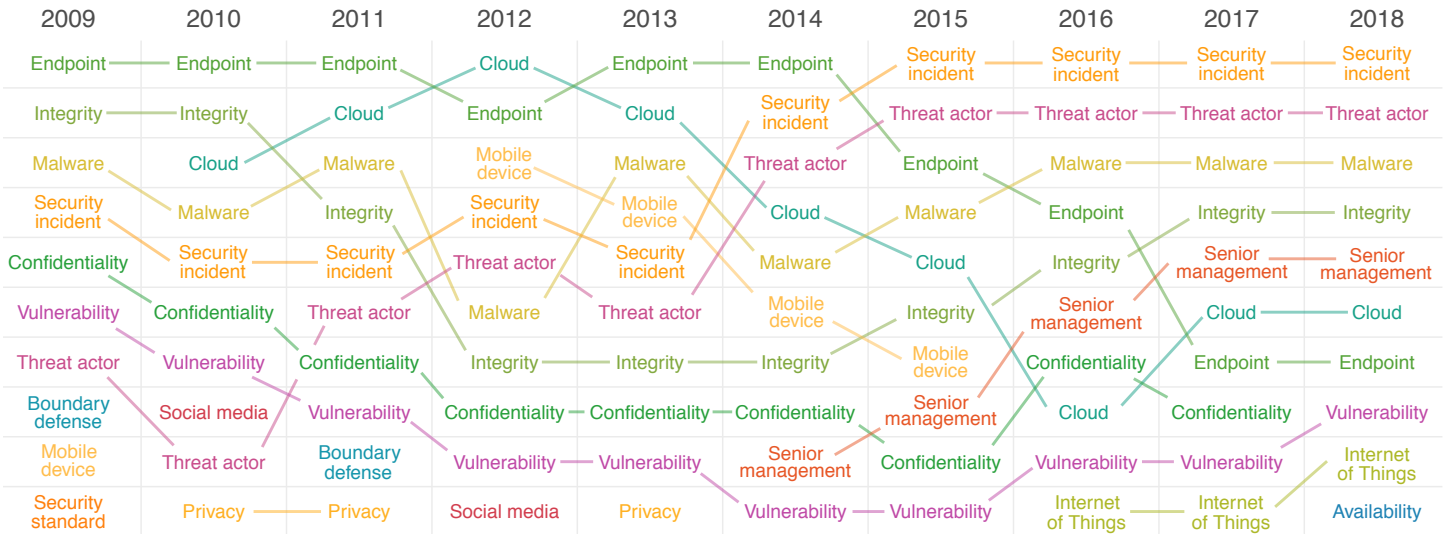
By now, you've undoubtedly noticed that Figure 4 lists some odd or related terms. This is a good example of where the unguided algorithm struggles to derive context, meaning, and associations that trained human eyes spot easily. In the next sections, we blend the strengths of man and machine.

Playing Tag With Algorithms

In tasks like the one before us, which seek to draw meaningful themes from a large corpus filled with complex concepts, guiding the analytical process with domain expertise can make a big difference. The classification system developed to tag security industry reports contained in the Cyentia Research Library is an example of such a guided algorithm. In it, multiple variations on a term of interest are manually mapped to a common tag.[3] The tagging system can then be used to train the algorithm—and vice versa, so the classification system gets smarter with age and experience.

Applying this classification system to the corpus yields Figure 5, a more meaningful rendition of Figure 2's simple word frequencies. It ranks the most common tags among abstracts for each year and also traces how that ranking changes over time. Keep in mind that each abstract will have multiple tags.

[3] i.e., the *Internet of Things* tag not only covers "Internet of Things," but also "IoT," "Internet of Everything," "Industrial Internet," etc.



Source: Cyentia Institute with data from RSA Conference

Figure 5: Most Common Tags in RSA Conference Abstracts Each Year

One observation from Figure 5 is the remarkable consistency it shows across the years. Certain tags enter and exit the stage, but seven of the top 10 stay for the whole show. We won't discuss these movements in detail here (That's coming next.), but we'd be remiss not to express our delight to see *Senior management* climbing the ladder. We're interpreting that as a sign the old barriers between security and business are breaking down.

What's Hot and What's Not

The "top 10" format of Figure 5 is not conducive to studying the broader array of topics and trends across RSA Conference submissions. For that, we need more tags and more tag-centered trending. Those criteria led to the creation of Figure 6, which is admittedly somewhat of a doozy. Give it a moment, though; you might find this is exactly the dataviz you're looking for. It may help to grab this scalable version.

The ordering of sub-charts in Figure 6 is based on the total number of abstracts flagged with each tag. So, more submissions were tagged with *Security incident* than anything else. Though shown last here, *Deep/Dark web* isn't the least common of all of the tags, because we've trimmed the list to fit on the page. There's no way we could possibly discuss or even pinpoint everything of interest in Figure 6, so we'll stick to some "color" commentary around topics that show ascending (green charts), descending (red), and flat (purple) trendlines.[4] Beyond colors, many of the topics in the figure can be conveniently viewed under 3 T's: threats, techs, and trends. Let's start with threats.

[4]We're using a linear regression line to establish the trending direction. If the slope is significant and positive, we're calling it ascending. If significant and negative, descending. A flat (~0 slope) regression line is tricky because it may indicate a static trend (e.g., Vulnerability) or a peak/valley in the middle (e.g., Mobile devices). For the latter, we trend based on the last six years to push it into the ascending or descending category, and shade them slightly differently to denote that push.

We're at peak threat heading into RSA Conference 2018 for *Ransomware, Extortion, Financial gain* (actor motive), and *Availability* (as in "loss of"). Though separate tags, those are all obviously related to the ransomware epidemic. *Threat actor, Insider, Stolen creds*, and *Security incident* are all one year off-peak, but still near it. Though trending up for the decade overall, it's a bit surprising to see some threat-related topics like *Threat intel, Intel sharing*, and *Kill Chain* declining in recent years. It doesn't seem long ago that those were white-hot, and they still are in many circles. Remember all this is relative.

This gets away from "green" trends, but we feel compelled to interrupt this program for a special announcement: The "APT-is-all-that-matters" era has officially drawn to a close! But don't take our word for it; look for yourself. *APT*, *Targeted attacks, Espionage*, and *Cyberwar* are well off-peak and falling. *State actor* is still in the green, but that bottom line is likely boosted by the Russian rather than Chinese variety we grew so familiar with several years ago. A moment of silence, please...

Source: Cyentia Institute with data from RSA Conference

Figure 6: Common Tags and Trends Among RSA Conference Abstracts. Full-size version here.

...and we're back. Let's move to tech-related topics, where we find *Machine learning, Internet of Things, Cyber-physical, Deep/Dark web,* and *CVE* all at-peak in 2018. A few others, like *Control systems*, look to be within the margin of error. *CVE*, due to its age, may be the only surprise there.

Some might be surprised to see declines among tags related to tech trends we once thought would trigger the Digital Apocalypse, such as *Cloud, Virtualization, Mobile device,* and *BYOD*. It seems the more we discuss things, the more they move from "Run for the hills!" to "We got this." And that's one of the main values of venues like the RSA Conference to our industry, isn't it?

A couple more semi-random observations to close out the techs. First, *Biometrics* is steadily climbing out of its low point of several years back. (2018's dip may just be a quick bio break.) Second, *Big data* isn't looking like such a big deal anymore.

Now on to some broader industry trends. To the comments in the previous section about the rise of *Senior management*, add *CISO* and *Board of Directors*. There's definitely a growing vibe at RSA Conference around security leadership. You could argue that *Cyber insurance* belongs in that grouping, as well (and Figure 10 might agree).

Apart from the soaring *GDPR* and *NIST (CSF)*, security and compliance standards like *FISMA, PCI-DSS*, and *ISO/IEC* aren't faring well. Even the broader tags like *Security standard* and *GRC* trend steadily downward. We suspect this is the natural fate of topics in an industry that often fixates on the new and/or scary.

Blocking a Bit on Bitcoin and Blockchain

In our highlights of hot techs and trends, some may have noted the absence of references to *Bitcoin* and *Blockchain*. The raw algorithm caught both in Figure 4, but we missed them in our classification system, and it's a good example of how any classification system—even one guided by domain expertise—can come up short. In fact, it might even lead astray, as we will show in this short sidebar.

First, though, let's rectify the lack of any view of abstracts dealing with *Cryptocurrency* (inclusive of Bitcoin, Ethereum, and several others) and *Blockchain*. Figure 7 shows both entering Conference submissions in 2015, dipping in 2016, then surging the last two years. But in what context are we discussing these topics?

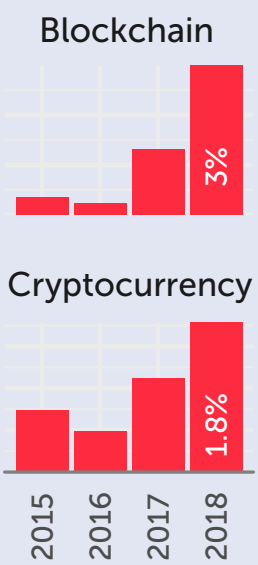
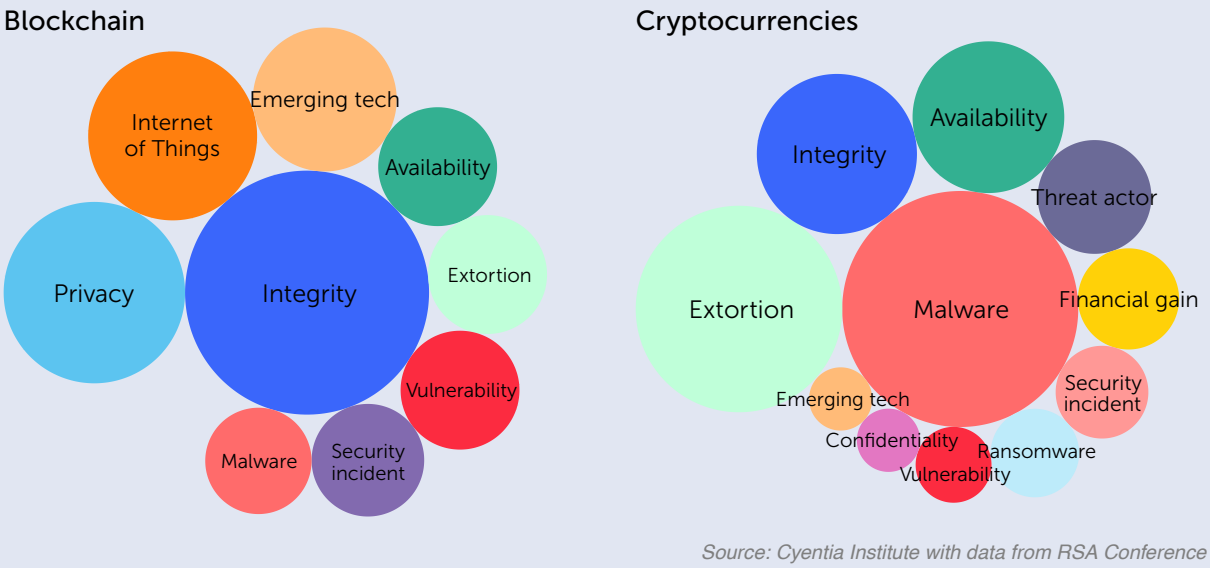


Figure 7: Number of RSA Conference Abstracts Tagged With Cryptocurrency and Blockchain

Figure 8 shows other topics (tags) that are most strongly associated with abstracts related to *Cryptocurrency* and *Blockchain*. In other words, "If a submission mentions one of these terms, what else does it mention?" The result is interesting and reveals the context in which security professionals tend to view these topics.

When security people discuss *Cryptocurrency*, it's often in the context of threats, namely ransomware payments. We suspect that would be dramatically different at a FinTech conference. *Blockchain* is more associated with protection and accountability, which is probably more in line with connotations outside our field. More generally, this exercise illustrates how algorithms sometimes need to guide experts to water.

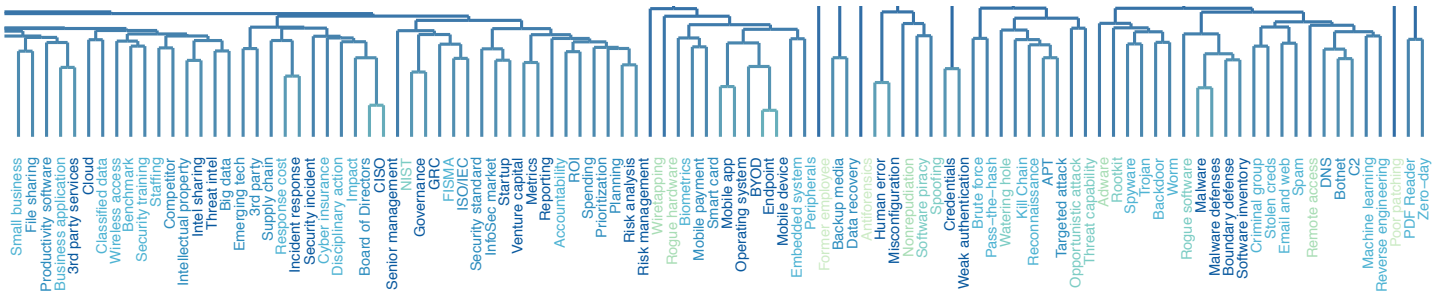


Source: Cyentia Institute with data from RSA Conference

Figure 8: Other Tags Associated With RSA Conference Abstracts Tagged With Cryptocurrency and Blockchain

What an Absolute Cluster...

Another area where algorithms can greatly assist expert-guided classifications systems is discovering latent associations among topics within a large corpus. Figure 9's dendrogram is one way of looking at closely-related terms. It may not be the easiest or clearest way of viewing relationships, mind you, but we've included it to help demonstrate the process of exploring relationships among topics.



Source: Cyentia Institute with data from RSA Conference

Figure 9: Truncated Dendrogram of Tag Associations Among RSA Conference Abstracts. Full-size version here.

If you scan across Figure 9 and find yourself thinking, "I can see why those terms would be on the same branch," then it's doing its job. (It also works for spotting odd associations for deeper review). For instance, find *Smart card* and *Mobile payment* on the same branch in the middle of the list. Going one level up (These plots are read from the bottom up.) connects those with *Biometrics*, and up once more ties in another series of branched tags related to mobile devices.

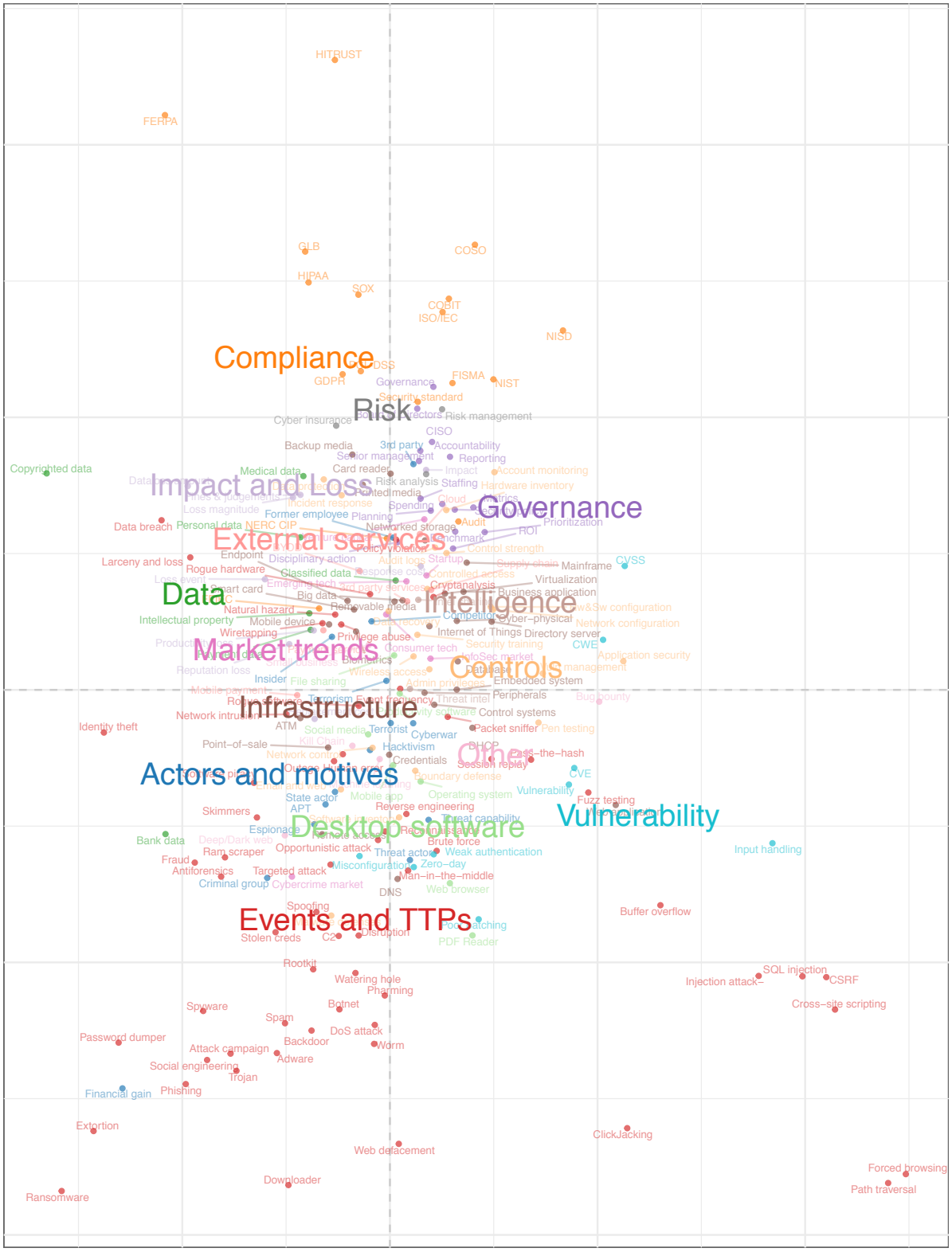
It's not hard to imagine a Conference session hitting on all of those topics. *Zero-day* and *PDF reader* sit at the right end of the figure, and we'll leave you to ruminate on why those terms are so strongly related. Scary how an untrained algorithm knows our weaknesses, isn't it?

Figure 10 may offer a more palatable (though still complex) view of associated topics. It uses a clustering technique to plot the strength of correlation among tags on a coordinate plane. Tags in close proximity often occur together in an abstract, while those farther apart rarely do. For instance, many sessions cover both *ransomware* and *extortion*, but you are very unlikely to attend an RSA Conference talk on the *Family Educational Rights and Privacy Act (FERPA)* and *forced browsing*.

Try to force yourself to ignore the large labels for a moment and just focus on the tags (small dots). Notice something similar about the orange cluster at the top? How about those toward the bottom left? Now look at the larger labels matching those colors, and they probably tell you what you already know. All the orange tags fall under the broad category of *Compliance*. The red ones in the lower left all concern security events and/or the tactics, techniques, and procedures (TTPs) that lead to them. Looking more closely at the clustering of red tags suggests that the *Events and TTPs* category may be too broad. Those in the lower right may deserve their own category, like *Application Security*.

Take some time to cruise around Figure 10 to take it all in. (Here's a full-size version to help.) If you think about what it represents—10 years of topics at our industry's largest conference and how they interrelate—it's quite fascinating. You notice that some of the categories we've applied seem to fit the data pretty well, while others not so much. As we've said several times now, classification is an iterative process where man and machine work together to continually improve the outcome.

In fact, you can help with this. If you do happen to attend a session at RSA Conference 2018 that covers both *FERPA* and *forced browsing*, please let us know so we can fix this (and make sure to record it for posterity).



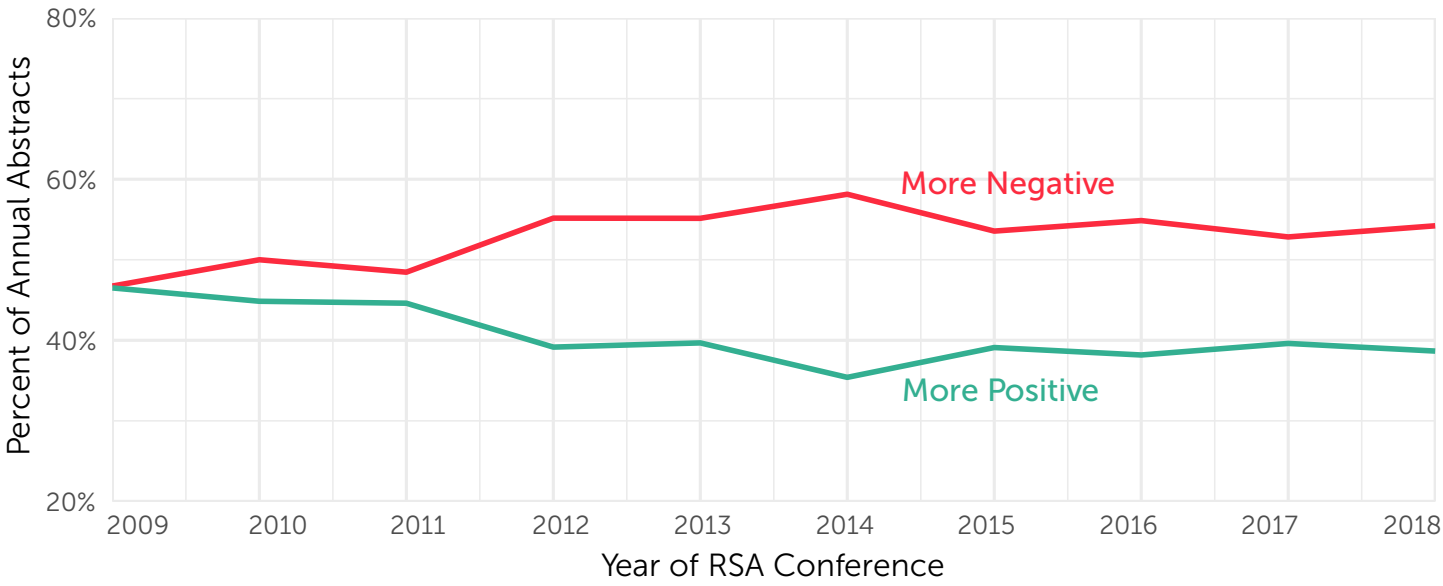
Source: Cyentia Institute with data from RSA Conference

Figure 10: Cluster Chart of Tag Associations Among RSA Conference Abstracts. Full-size version here.

Getting All Sentimental

We've looked from various angles at what we're discussing as an industry, but how are we feeling about those things? Answering that question is a perfect application of a technique called sentiment analysis. In a nutshell, sentiment analysis does exactly what the name suggests: it aims to determine the attitude, emotions, tone, polarity, etc., of a given text. We could study any number of these, but we're going to narrow it down to just positive vs negative sentiments for this section.

Our sentiment analysis on session *titles* a couple years back found more positivity than negativity, which we thought rather surprising given all the challenges security professionals deal with regularly. We were curious to see if examining longer abstracts would arrive at a different conclusion. Per Figure 11, it did.



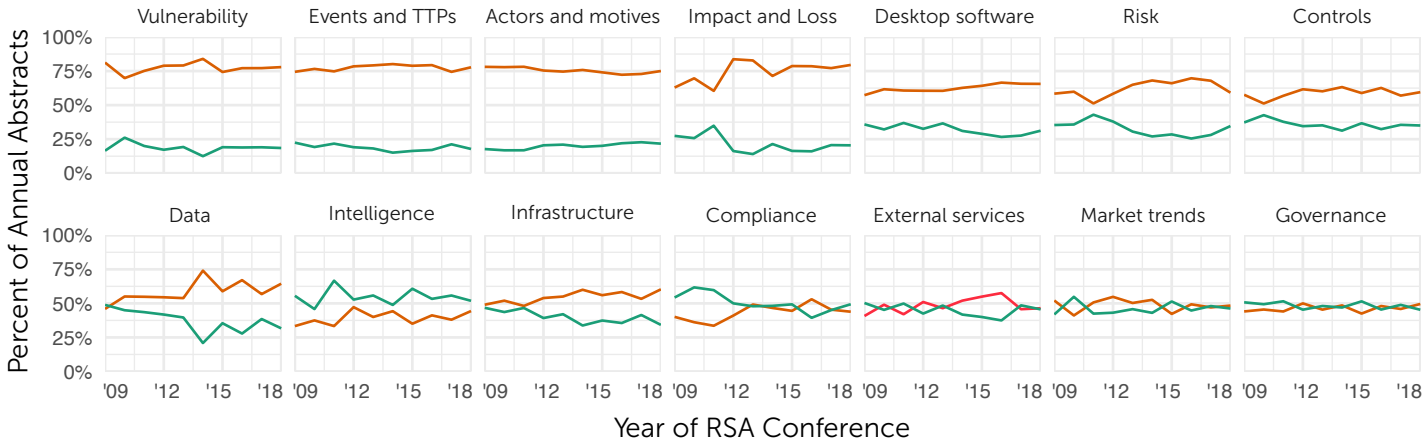
Source: Cyentia Institute with data from RSA Conference

Figure 11: Positive vs Negative Sentiments in RSA Conference Abstracts

Before we get into that, we feel the need to caveat results in this section. "Negative" here doesn't mean the author was in a bad mood or something like that (at least that we can prove). It's based on the general connotation of words in the abstract. For example, "critical," "complex," and "difficult" all have negative connotations, while "solution," "effective," and "benefit" are more positive words. But this makes sentiment analysis a little unique in our field. We might perceive "threat" as a natural and common term in a security context, but it's considered a negative term in regular English. Thus, the word "threat" was excluded from our analysis, but we undoubtedly missed some similarly-loaded security terms.

All that said, Figure 11 shows us that more abstracts leaned negative than positive. (Each abstract is given one overall sentiment rating.) The height of that disparity appears to be around 2014-2015, which happens to coincide with the peak of the *APT* theme at the Conference.

Figure 12 applies the same technique to determine whether sentiments vary across the topical categories identified back in Figure 10. The result is pretty neat. Some topics seem inherently more positive or negative, while others are more balanced. Most noteworthy from our perspective is the ordering: tactical topics (upper left) show much stronger negativity overall than more strategic topics (lower right).



Source: Cyentia Institute with data from RSA Conference

Figure 12: Positive vs Negative Sentiments Associated With Topic Categories in RSA Conference Abstracts

Also of interest is that sentiments associated with some topics bounce back and forth across the years, though this may simply suggest neutrality. There are no large switchbacks between extremely negative and extremely positive. One final observation: notice that *Intelligence* is the only category that is consistently positive. The buzz-worthiness of the associated tags *Threat intel* and *Intel sharing* might be waning, but there seems to be consensus that *Intelligence* is a positive thing to talk about.

Echoes in the Vendor Hall

There's one final thing we wanted to look into before closing out this report. Everything to this point focuses on abstracts as a window into the security community. But what about security vendors? Sure, they're part of the community too, but do they share or reflect our interests? To answer that question, we gathered descriptions supplied by all companies in the RSA Conference vendor hall from 2014 through 2017.

Figure 13 compares the most common tags from abstracts[5] in the left column with the most common tags from vendor descriptions in the right column. Lines in the middle help you find matches across columns and the grey shading denotes a tag absent from the other column. As you look this over, keep a couple things in mind: 1) abstracts are longer blocks of text and 2) vendor descriptions are often more aspirational than informational. Inspect Figure 13 in as much detail as you like, but overall, we're impressed with the level of consensus shown here.

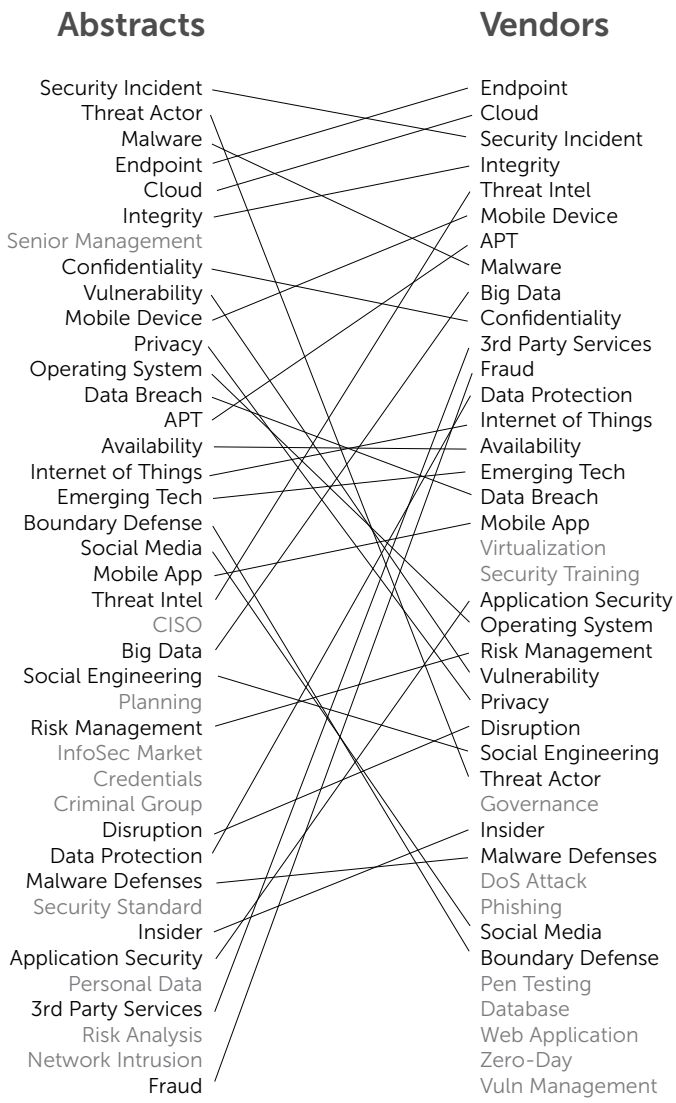


Figure 13: Most Common Tags in RSA Conference Abstracts vs Vendor Descriptions

[5]The order is slightly different from Figure 6 because this does not use the full decade of CFPs.

To test whether this apparent consensus represents actual correlation, we plotted all tags for all abstracts and vendors across all years in Figure 14. The pattern is pretty clear—topics common among Conference sessions also tend to be common among vendor descriptions. Security practitioners and vendors might have their own dialects, but at least they share the same language and culture. The question of who's influencing whom will have to wait for another time and another report.

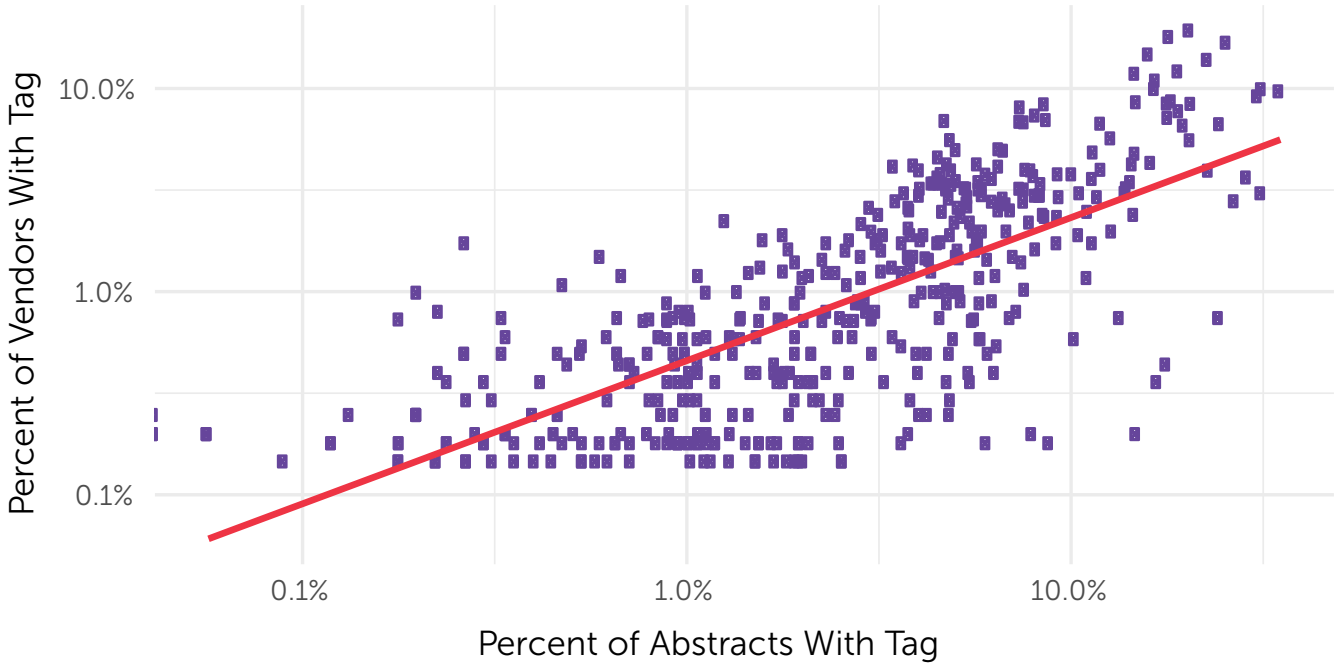


Figure 14: Correlation Between Tags in RSA Conference Abstracts vs Vendor Descriptions

Conclusion

And there you have it. After digging through nearly 15,000 RSAC Call for Paper submissions, we've unearthed some pretty powerful industry trends. Telling of both the past and current cybersecurity environment, this data offers yet another dimension to our understanding of the industry and acts as evidence of the rapid developments cybersecurity has undergone in the past decade.

But this report is far from the end game. The informational gold mine that we've struck here was only made possible through major collaborative efforts between the Cyentia Institute and RSA Conference. And as is clear from this report, there's a lot we stand to gain through continuing these conversations. Like Pink Floyd so wisely said, "All we need to do is make sure we keep talking." And what better place to do that than at RSA Conference?

Uniting industry experts, innovators, and professionals alike, RSA Conference helps keep cybersecurity on the cutting edge through international events, virtual communities, and relevant content. So join us as we tackle today's biggest challenges and continue to propel the cybersecurity conversation forward.

For more engaging content and to receive special offers on upcoming RSA conferences, visit rsaconference.com today.

Visit RSA Conference.com

Follow us on: #RSAC