Supplementary Material: A Field Study of Computer-Security Perceptions Using Anti-Virus Customer-Support Chats

Mahmood Sharif [*] [†] **Kevin A. Roundy** [‡] **Matteo Dell'Amico** [‡] **Christopher Gates** [‡]
Daniel Kats [‡] **Lujo Bauer** [†] **Nicolas Christin** [†]

NOTE

In this document, we provide an in-depth look into the trend-analysis experiment that is discussed in Sec. 5 of our paper [2].

1 TREND IDENTIFICATION

We studied the potential of unsupervised methods to discover patterns which are not expressed in the labels that we developed and, more interestingly, discover new trends in the data. To this end, we applied `TaxoGen` [3] on our problem description dataset, augmented with cases up to August 2018 for a total of 210,850 unique problem descriptions. `TaxoGen` creates, in an unsupervised manner, a taxonomy of keywords from a corpus of text data, resulting in a hierarchical clustering where each cluster contains a set of documents and keywords. In Table 1, we show some statistics about the five top-level clusters returned from the algorithm, correlating them with the manually attributed labels on perceived and expert issues; more comprehensive results, including hierarchical clusters, are shown in Sec. 2.

By examining the clusters manually, we found out that cluster C0 is often related to scams—sometimes mistaken by users for successful threats, sometimes correctly recognized by them. Cluster C1 contains various questions about license handling and expiration. Cluster C2 contains requests by confused users, who are worried about unusual behavior; in many cases, we find that this is due to PUA, misunderstanding by the users, or scam. Cluster C3 has the largest percentage of actual successful threats, and—as we will see—it is the one where users report more detailed information, and many names of threats end up as keywords in this cluster. Cluster C4 is related to requests for help and instructions.

To verify that this taxonomy is useful in recognizing and categorizing trends, we isolated a set of keywords having spikes in the number of mentions—we did this by considering keywords appearing in at least 100 documents, and taking those with the lowest variance in the timestamps of documents they appear in. We ended up with 12 keywords, which were clustered sensibly by `TaxoGen`: cluster C3.4.1—a subcluster of C3, which, as discussed above, contains several threat names—contains the keywords *kotver*, *lnk*, *ink* and *bat* which are all related with the Kovter malware (and the way AV software labels it) referred to in the main paper; *krack*, *zepto*, *spectre*, *meltdown*, *locky*, *wannacry* and *knctr*, all threat names, all end up in various subclusters of cluster C3. The only exception to this is, interestingly, *equifax*. This keyword is related to queries with respect to the Equifax breach [1], with users asking for information and assurance.

We found that `TaxoGen` provides a sensible categorization for keywords with as little as 10 mentions; hence, analyzing the spikes for some keywords together with the clustering associated to the keyword appears to be a reasonable way to detect and categorize trends. This technique could also be used to route chats to agents (e.g., a keyword in the same cluster of *locky* can be used to send a chat to a ransomware expert), provide context to them (e.g., keywords similar to those appearing in the chat), or even inform other teams about abnormal situation (e.g., like in the case of Kovter, a cluster mostly made of very recent content related to the name of a piece of malware may inform of malware outbreaks and/or AV issues in dealing with them).

2 KEYWORD TAXONOMY

Here we report the results of applying `TaxoGen` on the dataset referred to in the previous section. `TaxoGen` applies local Doc2Vec embeddings (one per cluster) to refine the keyword clustering of keywords, and assigns documents to clusters based on term frequency-inverse document frequency (TF-IDF) scoring of keywords applied to it. Due to issues in running the code linked by prior work [3], we reimplemented the algorithm ourselves. As algorithm parameters, we used all the defaults included in the publication.

- **Cluster C0** (883 keywords, 13,141 docs).
  Top keywords: *wells*, *fargo*, *equifax*, *posing*.
  - **Subcluster C0.0** (143 keywords, 3,235 docs).
    Top keywords: *equifax*, *sender*, *cox*, *hotmail*, *bt*.
    - **Subsubcluster C0.0.0** (2 keywords, 289 docs).
      Top keywords: *ip*, *address*.

[*]Work partially done as an intern at Symantec Research Labs.
[†]Carnegie Mellon University, Pittsburgh, PA, USA.
{mahmoods,lbauer,nicolasc}@cmu.edu.
[‡]Symantec Research Labs, Culver City, CA, USA.
{<firstname>_<lastname>}@symantec.com.

Table 1: Top-level clusters returned by TaxoGen.

Cluster name	C0	C1	C2	C3	C4
# keywords	883	445	767	937	346
# docs	13,141	81,405	46,321	11,828	57,456
Perceived issues	attack (45%)	attack (35%)	attack (48%)	attack (63%)	attack (52%)
	scam (30%)	account (26%)	none (17%)	AV (13%)	AV (16%)
	none (15%)	AV (13%)	AV (11%)	PUA (7%)	none (12%)
	account (5%)	assurance (11%)	PUA (10%)	none (7%)	scam (9%)
	assurance (5%)	none (7%)	assurance (8%)	scam (7%)	assurance (5%)
	–	scam (6%)	scam (5%)	assurance (3%)	account (5%)
Expert issues	scam (50%)	account (25%)	none (30%)	attack (37%)	attack (31%)
	none (30%)	none (25%)	PUA (23%)	none (23%)	scam (21%)
	attack (15%)	attack (18%)	scam (23%)	scam (13%)	none (20%)
	account (5%)	scam (14%)	attack (14%)	PUA (13%)	AV (16%)
	–	AV (11%)	AV (9%)	AV (10%)	PUA (5%)

    - ∗ **Subsubcluster C0.0.1** (68 keywords, 1,485 docs). Top keywords: *amazon*, *banking*, *scammer*, *fraudulent*, *stolen*.
    - ∗ **Subsubcluster C0.0.2** (1 keywords, 6 docs). Top keywords: *robot*.
    - ∗ **Subsubcluster C0.0.3** (66 keywords, 1,443 docs). Top keywords: *equifax*, *friend*, *isp*, *cox*, *verizon*.
    - ∗ **Subsubcluster C0.0.4** (1 keywords, 5 docs). Top keywords: *sample*.
  - **Subcluster C0.1** (107 keywords, 1,729 docs). Top keywords: *advance*, *vss*, *walk*, *thx*.
    - ∗ **Subsubcluster C0.1.0** (93 keywords, 1,690 docs). Top keywords: *vss*, *assistance*, *u*, *m*, *thank*.
  - **Subcluster C0.2** (180 keywords, 2,759 docs). Top keywords: *geek*, *garmin*, *squad*, *expert*, *continuation*.
    - ∗ **Subsubcluster C0.2.0** (108 keywords, 1,340 docs). Top keywords: *geek*, *squad*, *immediate*, *llc*.
    - ∗ **Subsubcluster C0.2.1** (1 keywords, 4 docs). Top keywords: *gig*.
    - ∗ **Subsubcluster C0.2.2** (60 keywords, 1,359 docs). Top keywords: *ticket*, *ref*, *department*, *reference*, *queue*.
    - ∗ **Subsubcluster C0.2.3** (1 keywords, 4 docs). Top keywords: *adviser*.
    - ∗ **Subsubcluster C0.2.4** (2 keywords, 43 docs). Top keywords: *s*, *wasn*.
  - **Subcluster C0.3** (358 keywords, 2,858 docs). Top keywords: *gift*, *tax*, *assistant*, *verification*, *frontier*.
    - ∗ **Subsubcluster C0.3.0** (11 keywords, 75 docs). Top keywords: *invoices*, *lt*, *discount*, *emergency*, *windstream*.
    - ∗ **Subsubcluster C0.3.1** (73 keywords, 533 docs). Top keywords: *offers*, *below*, *guess*, *coputer*, *analysis*.
    - ∗ **Subsubcluster C0.3.2** (239 keywords, 2,179 docs). Top keywords: *wireless*, *request*, *concern*, *certificate*, *tax*.
    - ∗ **Subsubcluster C0.3.4** (1 keywords, 4 docs). Top keywords: *ipod*.
  - **Subcluster C0.4** (37 keywords, 2,528 docs). Top keywords: *accent*, *india*, *caller*, *pretending*, *advising*.
    - ∗ **Subsubcluster C0.4.0** (16 keywords, 836 docs). Top keywords: *apple*, *somebody*, *man*, *claim*, *pretending*.
    - ∗ **Subsubcluster C0.4.1** (1 keywords, 69 docs). Top keywords: *claiming*.
    - ∗ **Subsubcluster C0.4.2** (4 keywords, 339 docs). Top keywords: *purporting*, *telephone*, *calls*, *someone*.
    - ∗ **Subsubcluster C0.4.3** (1 keywords, 98 docs). Top keywords: *speak*.
    - ∗ **Subsubcluster C0.4.4** (15 keywords, 1,186 docs). Top keywords: *rep*, *agent*, *party*, *representative*, *group*.
- **Cluster C1** (445 keywords, 81,405 docs). Top keywords: *expires*, *expiring*, *expiration*, *november*, *visa*.
  - **Subcluster C1.0** (100 keywords, 13,300 docs). Top keywords: *debit*, *visa*, *billing*, *paypal*, *cards*.

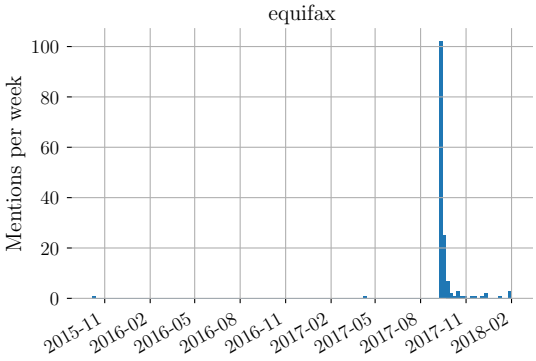
    - **Subsubcluster C1.0.0** (16 keywords, 4,047 docs). Top keywords: *cards*, *debit*, *acct*, *statement*, *visa*.
    - **Subsubcluster C1.0.1** (40 keywords, 1,019 docs). Top keywords: *storage*, *policy*, *sub*, *profile*, *knowledge*.
    - **Subsubcluster C1.0.2** (14 keywords, 2,748 docs). Top keywords: *expiration*, *expiry*, *records*, *reminders*, *tomorrow*.
    - **Subsubcluster C1.0.3** (14 keywords, 2,281 docs). Top keywords: *dollars*, *life*, *fee*, *bill*, *price*.
    - **Subsubcluster C1.0.4** (16 keywords, 3,205 docs). Top keywords: *pin*, *activation*, *digit*, *copy*, *registration*.
  - **Subcluster C1.1** (106 keywords, 10,727 docs). Top keywords: *tuesday*, *row*, *hr*, *beginning*, *monday*.
    - **Subsubcluster C1.1.0** (14 keywords, 1,421 docs). Top keywords: *ive*, *weekend*, *tuesday*, *afternoon*, *thursday*.
    - **Subsubcluster C1.1.1** (30 keywords, 2,678 docs). Top keywords: *renewing*, *april*, *december*, *march*, *dec*.
    - **Subsubcluster C1.1.2** (22 keywords, 3,027 docs). Top keywords: *hold*, *mins*, *min*, *approx*, *hrs*.
    - **Subsubcluster C1.1.3** (33 keywords, 952 docs). Top keywords: *step*, *buying*, *sometime*, *awhile*, *purchasing*.
    - **Subsubcluster C1.1.4** (7 keywords, 2,649 docs). Top keywords: *ve*, *prior*, *pm*, *session*, *line*.
  - **Subcluster C1.2** (71 keywords, 18,544 docs). Top keywords: *kindle*, *wifes*, *mom*, *depot*, *labtop*.
    - **Subsubcluster C1.2.0** (16 keywords, 1,687 docs). Top keywords: *house*, *modem*, *mother*, *kindle*, *business*.
    - **Subsubcluster C1.2.1** (8 keywords, 4,562 docs). Top keywords: *somehow*, *stating*, *think*, *got*, *saying*.
    - **Subsubcluster C1.2.2** (18 keywords, 3,870 docs). Top keywords: *pad*, *husbands*, *wifes*, *samsung*, *mom*.
    - **Subsubcluster C1.2.3** (10 keywords, 6,161 docs). Top keywords: *computor*, *comp*, *machine*, *mac*, *virus*.
    - **Subsubcluster C1.2.4** (19 keywords, 2,264 docs). Top keywords: *lenovo*, *surface*, *toshiba*, *acer*, *ten*.
  - **Subcluster C1.3** (64 keywords, 6,150 docs). Top keywords: *licences*, *seat*, *seats*, *devises*, *purposes*.
    - **Subsubcluster C1.3.0** (24 keywords, 1,609 docs). Top keywords: *definitions*, *questions*, *steps*, *risks*, *drivers*.
    - **Subsubcluster C1.3.1** (28 keywords, 1,427 docs). Top keywords: *phones*, *systems*, *mine*, *licences*, *iphones*.
    - **Subsubcluster C1.3.2** (4 keywords, 723 docs). Top keywords: *win*, *times*, *years*, *days*.
    - **Subsubcluster C1.3.3** (4 keywords, 1,017 docs). Top keywords: *laptops*, *devices*, *computers*, *one*.
    - **Subsubcluster C1.3.4** (4 keywords, 1,374 docs). Top keywords: *emails*, *threats*, *problems*, *issues*.
  - **Subcluster C1.4** (103 keywords, 32,684 docs). Top keywords: *upgrading*, *wi*, *fi*, *conflict*, *avg*.
    - **Subsubcluster C1.4.0** (14 keywords, 3,311 docs). Top keywords: *tune*, *lifetime*, *unsure*, *ready*, *suppose*.
    - **Subsubcluster C1.4.1** (26 keywords, 11,339 docs). Top keywords: *cd*, *patch*, *nis*, *nav*, *trail*.
    - **Subsubcluster C1.4.2** (16 keywords, 4,053 docs). Top keywords: *wi*, *fi*, *vpn*, *lifelock*, *impression*.
    - **Subsubcluster C1.4.3** (36 keywords, 10,322 docs). Top keywords: *family*, *ios*, *total*, *edition*, *delux*.
    - **Subsubcluster C1.4.4** (11 keywords, 3,659 docs). Top keywords: *conflict*, *difference*, *bit*, *feature*, *option*.
- **Cluster C2** (767 keywords, 46,321 docs). Top keywords: *jumping*, *movie*, *engines*, *blinking*, *bin*.
  - **Subcluster C2.0** (153 keywords, 6,831 docs). Top keywords: *yellow*, *tick*, *exclamation*, *cross*, *dialogue*.
    - **Subsubcluster C2.0.0** (46 keywords, 1,404 docs). Top keywords: *dialogue*, *image*, *progress*, *round*, *stays*.
    - **Subsubcluster C2.0.1** (65 keywords, 1,364 docs). Top keywords: *shot*, *touch*, *pending*, *size*, *indication*.
    - **Subsubcluster C2.0.2** (10 keywords, 1,172 docs). Top keywords: *hand*, *middle*, *side*, *corner*, *bottom*.
    - **Subsubcluster C2.0.3** (17 keywords, 1,754 docs). Top keywords: *beeping*, *alarm*, *repeating*, *recording*, *noise*.
    - **Subsubcluster C2.0.4** (15 keywords, 1,137 docs). Top keywords: *mark*, *tick*, *tray*, *exclamation*, *taskbar*.
  - **Subcluster C2.1** (230 keywords, 4,269 docs). Top keywords: *webcam*, *hundreds*, *puter*, *scams*, *events*.
    - **Subsubcluster C2.1.0** (43 keywords, 698 docs). Top keywords: *whats*, *compute*, *language*, *anyway*, *characters*.
    - **Subsubcluster C2.1.1** (3 keywords, 276 docs). Top keywords: *traffic*, *lot*, *messages*.
    - **Subsubcluster C2.1.2** (17 keywords, 202 docs). Top keywords: *camera*, *calendar*, *blog*, *codes*, *nature*.
    - **Subsubcluster C2.1.3** (113 keywords, 1,884 docs). Top keywords: *form*, *tv*, *difficulty*, *requests*, *attachments*.
    - **Subsubcluster C2.1.4** (36 keywords, 1,179 docs). Top keywords: *alot*, *notifications*, *frequent*, *hundreds*, *mails*.
  - **Subcluster C2.2** (129 keywords, 7,746 docs). Top keywords: *drives*, *processor*, *bin*, *recycle*, *resources*.

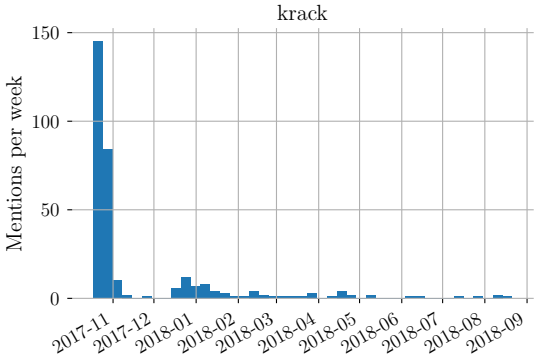
    - **Subsubcluster C2.2.0** (19 keywords, 1,354 docs). Top keywords: *processor*, *processes*, *command*, *toolbar*, *panel*.
    - **Subsubcluster C2.2.1** (6 keywords, 812 docs). Top keywords: *space*, *drives*, *usb*, *disk*, *hard*.
    - **Subsubcluster C2.2.2** (6 keywords, 551 docs). Top keywords: *reader*, *player*, *media*, *attachment*, *adobe*.
    - **Subsubcluster C2.2.3** (86 keywords, 3,771 docs). Top keywords: *extensions*, *skype*, *cookies*, *inbox*, *results*.
    - **Subsubcluster C2.2.4** (12 keywords, 1,258 docs). Top keywords: *photos*, *excel*, *documents*, *passwords*, *folders*.
  - **Subcluster C2.3** (123 keywords, 21,389 docs). Top keywords: *bing*, *redirects*, *searches*, *engine*, *crome*.
    - **Subsubcluster C2.3.0** (23 keywords, 3,776 docs). Top keywords: *adult*, *adverts*, *advertisements*, *webpages*, *sights*.
    - **Subsubcluster C2.3.1** (18 keywords, 6,497 docs). Top keywords: *sight*, *cam*, *url*, *webpage*, *https*.
    - **Subsubcluster C2.3.2** (40 keywords, 3,280 docs). Top keywords: *face*, *fb*, *post*, *movie*, *book*.
    - **Subsubcluster C2.3.3** (11 keywords, 1,877 docs). Top keywords: *sends*, *appearing*, *blocks*, *redirecting*, *pop*.
    - **Subsubcluster C2.3.4** (31 keywords, 5,959 docs). Top keywords: *crome*, *engines*, *engine*, *default*, *explore*.
  - **Subcluster C2.4** (124 keywords, 6,086 docs). Top keywords: *moves*, *acts*, *pointer*, *fan*, *funny*.
    - **Subsubcluster C2.4.0** (40 keywords, 1,132 docs). Top keywords: *stopping*, *drop*, *scripts*, *tasks*, *self*.
    - **Subsubcluster C2.4.1** (6 keywords, 1,038 docs). Top keywords: *freezing*, *sometimes*, *freezes*, *mouse*, *slow*.
    - **Subsubcluster C2.4.2** (18 keywords, 1,365 docs). Top keywords: *usage*, *funny*, *speed*, *cpu*, *become*.
    - **Subsubcluster C2.4.3** (4 keywords, 567 docs). Top keywords: *dont*, *im*, *cant*, *thing*.
    - **Subsubcluster C2.4.4** (54 keywords, 1,982 docs). Top keywords: *changing*, *shuts*, *loads*, *blinking*, *stuck*.
- **Cluster C3** (937 keywords, 11,828 docs). Top keywords: *que*, *en*, *ik*, *je*, *mi*.
  - **Subcluster C3.0** (89 keywords, 2,915 docs). Top keywords: *soft*, *locky*, *wannacry*, *ransomeware*, *zepto*.
    - **Subsubcluster C3.0.0** (3 keywords, 16 docs). Top keywords: *tests*, *gozi*, *zeuss*.
    - **Subsubcluster C3.0.1** (83 keywords, 2,881 docs). Top keywords: *ware*, *ransomware*, *spyware*, *adware*, *ransom*.
    - **Subsubcluster C3.0.2** (1 keywords, 5 docs). Top keywords: *zesus*.
    - **Subsubcluster C3.0.3** (1 keywords, 5 docs). Top keywords: *micrsoft*.
    - **Subsubcluster C3.0.4** (1 keywords, 8 docs). Top keywords: *conflicker*.
  - **Subcluster C3.1** (309 keywords, 2,853 docs). Top keywords: *failure*, *debug*, *sonar*, *urgent*, *rapport*.
    - **Subsubcluster C3.1.0** (8 keywords, 36 docs). Top keywords: *exec*, *inability*, *iexplorer*, *index*, *transit*.
    - **Subsubcluster C3.1.1** (2 keywords, 16 docs). Top keywords: *isn*, *findingdiscount*.
    - **Subsubcluster C3.1.2** (38 keywords, 343 docs). Top keywords: *trojan.gen.2*, *conduit*, *cloud*, *arcade*, *lasuperba*.
    - **Subsubcluster C3.1.3** (1 keywords, 5 docs). Top keywords: *vrus*.
    - **Subsubcluster C3.1.4** (239 keywords, 2,403 docs). Top keywords: *driver*, *attacks*, *sonar*, *redirect*, *failure*.
  - **Subcluster C3.2** (52 keywords, 1,485 docs). Top keywords: *unlocker*, *filename*, *domains*, *techsupport*, *attacking*.
    - **Subsubcluster C3.2.0** (43 keywords, 1,184 docs). Top keywords: *intrusion*, *dns*, *unlocker*, *exploit*, *attack*.
    - **Subsubcluster C3.2.1** (3 keywords, 123 docs). Top keywords: *path*, *filename*, *c*.
    - **Subsubcluster C3.2.2** (1 keywords, 5 docs). Top keywords: *tapsnake*.
    - **Subsubcluster C3.2.3** (1 keywords, 5 docs). Top keywords: *iexplore*.
    - **Subsubcluster C3.2.4** (2 keywords, 164 docs). Top keywords: *hijack*, *backdoor*.
  - **Subcluster C3.3** (265 keywords, 3,144 docs). Top keywords: *slim*, *fire*, *wan*, *wall*, *mindspark*.
    - **Subsubcluster C3.3.0** (195 keywords, 2,424 docs). Top keywords: *cleaner*, *slim*, *fire*, *reimage*, *trojans*.
    - **Subsubcluster C3.3.1** (52 keywords, 658 docs). Top keywords: *l*, *spectre*, *exe*, *slimcleaner*, *iam*.
    - **Subsubcluster C3.3.2** (1 keywords, 6 docs). Top keywords: *thnx*.
    - **Subsubcluster C3.3.3** (2 keywords, 17 docs). Top keywords: *boost*, *playthru*.
    - **Subsubcluster C3.3.4** (1 keywords, 6 docs). Top keywords: *peek*.
  - **Subcluster C3.4** (139 keywords, 1,348 docs). Top keywords: *que*, *en*, *un*, *la*, *el*.
    - **Subsubcluster C3.4.0** (3 keywords, 21 docs). Top keywords: *inf*, *tamper*, *destination*.

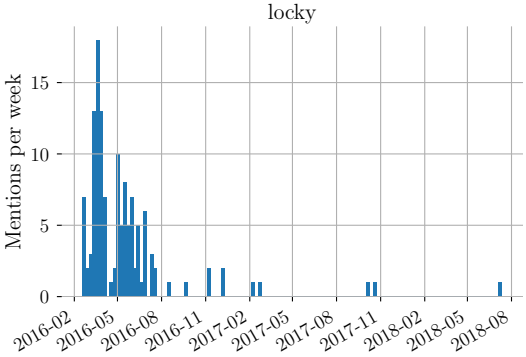
    - **Subsubcluster C3.4.1** (126 keywords, 1,263 docs). Top keywords: *que*, *o*, *kotver*, *bat*, *en*.
    - **Subsubcluster C3.4.2** (1 keywords, 39 docs). Top keywords: *amp*.
    - **Subsubcluster C3.4.3** (1 keywords, 5 docs). Top keywords: *chkdsk*.
    - **Subsubcluster C3.4.4** (1 keywords, 6 docs). Top keywords: *yontoo*.
- **Cluster C4** (346 keywords, 57,456 docs). Top keywords: *rescan*, *easer*, *nrnr*, *completes*, *scrub*.
  - **Subcluster C4.0** (50 keywords, 11,301 docs). Top keywords: *clue*, *sense*, *worry*, *wonder*, *catch*.
    - **Subsubcluster C4.0.0** (5 keywords, 1,214 docs). Top keywords: *feel*, *secure*, *wrong*, *safe*, *right*.
    - **Subsubcluster C4.0.1** (14 keywords, 4,220 docs). Top keywords: *remember*, *catch*, *anymore*, *trust*, *happen*.
    - **Subsubcluster C4.0.2** (11 keywords, 3,389 docs). Top keywords: *changes*, *wonder*, *determine*, *wondering*, *correct*.
    - **Subsubcluster C4.0.3** (7 keywords, 1,077 docs). Top keywords: *recommend*, *sense*, *change*, *pay*, *ask*.
    - **Subsubcluster C4.0.4** (13 keywords, 1,401 docs). Top keywords: *ignore*, *exclude*, *job*, *quarantine*, *worry*.
  - **Subcluster C4.1** (86 keywords, 18,818 docs). Top keywords: *factory*, *force*, *reload*, *agree*, *write*.
    - **Subsubcluster C4.1.0** (30 keywords, 7,681 docs). Top keywords: *function*, *print*, *accept*, *select*, *apply*.
    - **Subsubcluster C4.1.1** (13 keywords, 3,412 docs). Top keywords: *factory*, *force*, *wipe*, *crash*, *shutdown*.
    - **Subsubcluster C4.1.2** (16 keywords, 1,209 docs). Top keywords: *mess*, *write*, *push*, *return*, *disappear*.
    - **Subsubcluster C4.1.3** (14 keywords, 3,111 docs). Top keywords: *pick*, *hang*, *freeze*, *lock*, *hung*.
    - **Subsubcluster C4.1.4** (13 keywords, 3,405 docs). Top keywords: *move*, *stay*, *switch*, *enter*, *login*.
  - **Subcluster C4.2** (53 keywords, 6,621 docs). Top keywords: *useless*, *finishes*, *remain*, *deletes*, *functioning*.
    - **Subsubcluster C4.2.0** (4 keywords, 758 docs). Top keywords: *detecting*, *finding*, *removing*, *tells*.
    - **Subsubcluster C4.2.1** (6 keywords, 1,447 docs). Top keywords: *defender*, *installing*, *fine*, *stopped*, *working*.
    - **Subsubcluster C4.2.2** (6 keywords, 1,378 docs). Top keywords: *finds*, *happens*, *shows*, *found*, *ran*.
    - **Subsubcluster C4.2.3** (27 keywords, 1,946 docs). Top keywords: *reinstalling*, *deleting*, *caught*, *operating*, *uninstalling*.
    - **Subsubcluster C4.2.4** (10 keywords, 1,092 docs). Top keywords: *functioning*, *okay*, *savvy*, *responding*, *literate*.
  - **Subcluster C4.3** (56 keywords, 13,576 docs). Top keywords: *nbrt*, *rescan*, *easer*, *nrnr*, *powereraser*.
    - **Subsubcluster C4.3.0** (9 keywords, 3,111 docs). Top keywords: *directions*, *recovery*, *manual*, *repair*, *instructions*.
    - **Subsubcluster C4.3.1** (8 keywords, 2,380 docs). Top keywords: *autofix*, *liveupdate*, *fails*, *updates*, *update*.
    - **Subsubcluster C4.3.2** (19 keywords, 661 docs). Top keywords: *diagnostic*, *cleanup*, *rescan*, *easer*, *powereraser*.
    - **Subsubcluster C4.3.3** (13 keywords, 6,642 docs). Top keywords: *erasure*, *avail*, *luck*, *success*, *erase*.
    - **Subsubcluster C4.3.4** (7 keywords, 782 docs). Top keywords: *rescue*, *checks*, *disc*, *washer*, *mode*.
  - **Subcluster C4.4** (101 keywords, 7,140 docs). Top keywords: *include*, *reconnect*, *instal*, *compromise*, *upload*.
    - **Subsubcluster C4.4.0** (7 keywords, 991 docs). Top keywords: *refer*, *ring*, *reconnect*, *unlock*, *connection*.
    - **Subsubcluster C4.4.1** (63 keywords, 2,507 docs). Top keywords: *locate*, *visit*, *setup*, *attention*, *test*.
    - **Subsubcluster C4.4.2** (19 keywords, 1,691 docs). Top keywords: *eliminate*, *recover*, *suggestions*, *provide*, *identify*.
    - **Subsubcluster C4.4.3** (5 keywords, 815 docs). Top keywords: *transfer*, *assist*, *resolve*, *advise*, *help*.
    - **Subsubcluster C4.4.4** (7 keywords, 1,136 docs). Top keywords: *verify*, *talk*, *confirm*, *buy*, *add*.

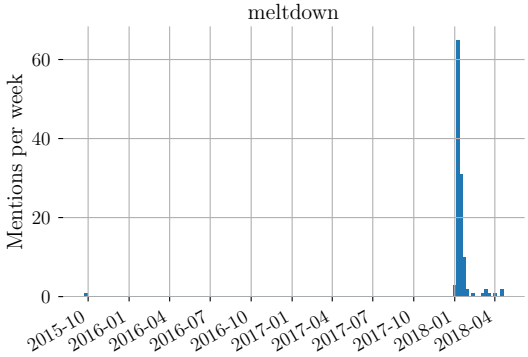
3 KEYWORDS AND TRENDS

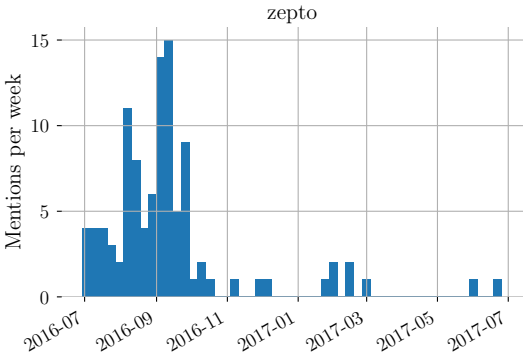
In the following graphs, we report the number of mentions per week of 12 keywords with uneven number of mentions over time, as discussed in Sec. 1. The plots refer to *equifax* (cluster C0.0.3), *locky*, *zepto* and *wannacry* (cluster C3.0.1), *krack* and *meltdown* (cluster C3.1.4), *knctr* (cluster C3.3.0), *spectre* (cluster C3.3.1), *bat*, *ink*, *kotver* and *lnk* (cluster C3.4.1—we recall that these terms are related to the AV nomenclature for Kovter). We have observed that the results of the clustering may vary slightly between different runs of the algorith—e.g., unlike the results presented here, *spectre* and *meltdown* often end up in the same cluster.

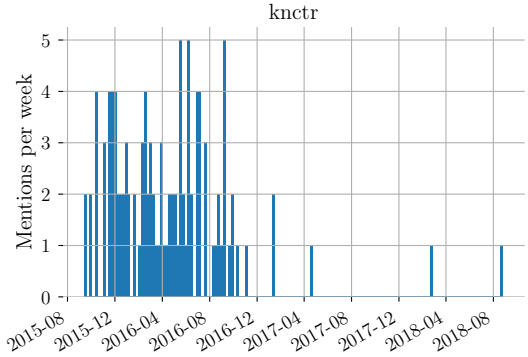

equifax
Mentions per week
0
20
40
60
80
100
2015-11
2016-02
2016-05
2016-08
2016-11
2017-02
2017-05
2017-08
2017-11
2018-02

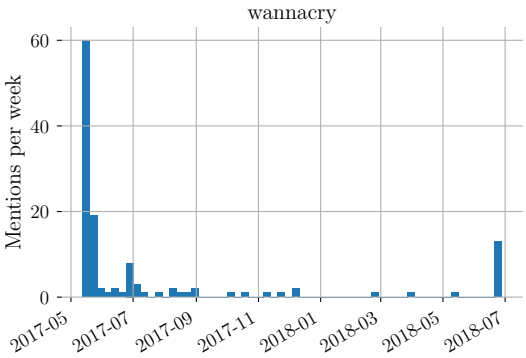

krack
Mentions per week
0
50
100
150
2017-11
2017-12
2018-01
2018-02
2018-03
2018-04
2018-05
2018-06
2018-07
2018-08
2018-09

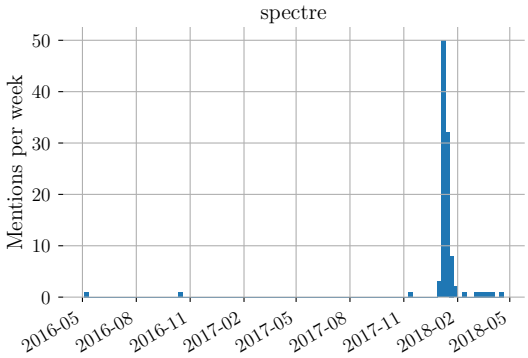

locky
Mentions per week
0
5
10
15
2016-02
2016-05
2016-08
2016-11
2017-02
2017-05
2017-08
2017-11
2018-02
2018-05
2018-08

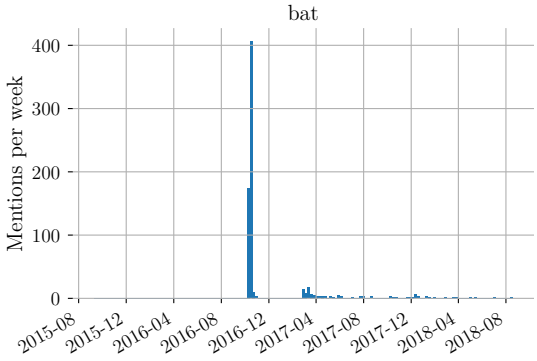

meltdown
Mentions per week
0
20
40
60
2015-10
2016-01
2016-04
2016-07
2016-10
2017-01
2017-04
2017-07
2017-10
2018-01
2018-04

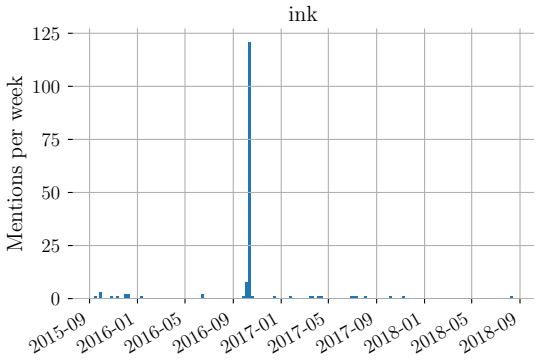

zepto
Mentions per week
0
5
10
15
2016-07
2016-09
2016-11
2017-01
2017-03
2017-05
2017-07

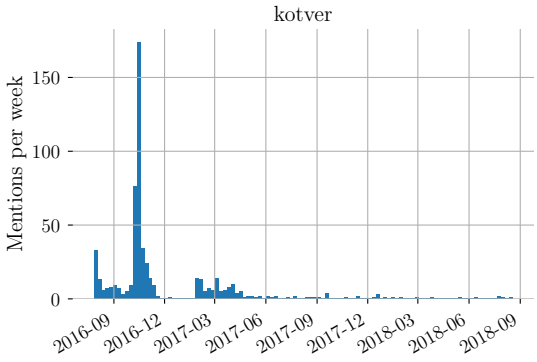

knctr
Mentions per week
0
1
2
3
4
5
2015-08
2015-12
2016-04
2016-08
2016-12
2017-04
2017-08
2017-12
2018-04
2018-08

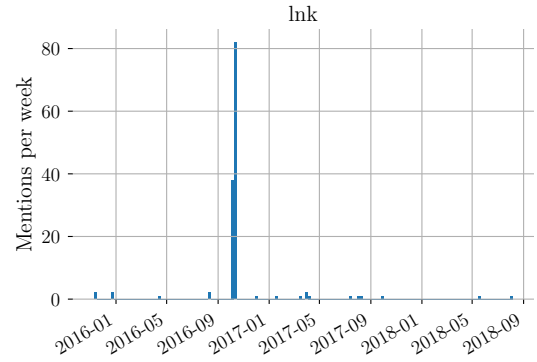

wannacry
Mentions per week
0
20
40
60
2017-05
2017-07
2017-09
2017-11
2018-01
2018-03
2018-05
2018-07


spectre
Mentions per week
0
10
20
30
40
50
2016-05
2016-08
2016-11
2017-02
2017-05
2017-08
2017-11
2018-02
2018-05









REFERENCES

[1] Federal Trade Commission. 2017. The Equifax Data Breach. https://www.ftc.gov/equifax-data-breach. Online; accessed 20 Sep 2018.

[2] Mahmood Sharif, Kevin A. Roundy, Matteo Dell'Amico, Christopher Gates, Daniel Kats, Lujo Bauer, and Nicolas Christin. 2019. A Field Study of Computer-Security Perceptions Using Anti-Virus Customer-Support Chats. In *Proc. CHI*.

[3] Chao Zhang, Fangbo Tao, Xiusi Chen, Jiaming Shen, Meng Jiang, Brian Sadler, Michelle Vanni, and Jiawei Han. 2018. TaxoGen: Constructing Topical Concept Taxonomy by Adaptive Term Embedding and Clustering. In *Proc. KDDI*.