All.Net Analyst Report and Newsletter

Welcome to our Analyst Report and Newsletter

Deprecating the use of social media by terrorist group for recruiting and grooming

I have been watching video presentations of the use of the Internet, and in particular social media, for recruitment and grooming of young people by terrorists for terrorist attacks. It seems to me that this is a problem that is largely solvable today. It will not be trivial or necessarily easy, but it goes to the same issues that need to be addressed by social media for election tampering and of course electioneering. Social influence is the core issue and social media is the means of mass and potentially personalized influence operations.

If everybody else jumped off a bridge would you?

This is a question my mother used to ask me whenever I asserted social proof as a basis for anything I wanted to do. Of course it was rhetorical. But I suspect the answer for most socialized young people, the answer is yes – given the proper circumstances. Of course the terrorist groups use social media to communicate with lots of folks, and only a few have to take up the cause in order to produce real harm. There is plenty of psychology research showing that this sort of thing will work and plenty of real-world experience by now showing that it does work.

Free speech to what limit?

In order to deprecate this sort of thing, we will have to start by limiting the freedom of speech. I believe that the existing legal limits can be used as is, but there is a case to be made for more / new laws in this regard.

- It may seem obvious, but clearly instructions on making bombs to kill innocents should be stopped. And indeed, some of the terrorist sites say and show precisely that. They instruct people top kill people and provide the details of how to do it.
    - *Technically, I think the existing AI capabilities can identify the vast majority of these sorts of things and that a responsible social media company should be required to check for such things before allowing release of content onto the Internet.*
        - Yes I know – we can hide this sort of thing in other content, use covert channels for messaging, and so on.
            - *Technically, we can get rid of almost all current ISIS (be example) messaging by looking for the presence of their flag in the video. Same for neo-nazi and many other similar groups, who brand their propaganda.*
                - But they will adapt to counter these tactics.
                    - *They may, but if they do they are deprecating their brand and thus reducing the effectiveness of their meme campaign.*
                        - But…
                            - *Sure – but we don't have to be perfect in order to be effective.*

This is going to be thematic here. We don't have to be perfect in order to improve things.

Technology reasonably applied

I think this is a case where extreme positions are unnecessary and harmful to the greater objective. It's pretty obvious when you watch the videos of beheadings that this is not something you want to have promoted. It should not be hard to identify snuff films, explosives going off, etc. Hint: search for "videos of explosions". But many of them are "legitimate" - or at least not terrorist recruiting and training tools. And that is also obvious. All you have to do is watch them.

- Watch them! Do you know how many videos there are? That would cost a fortune!
  - *Indeed, and what is the cost we are already paying for terrorism and defenses against it? Perhaps if we spent some of it on identifying the spread of the memes and interdicting them it would be more cost effective than many of the alternatives we already pay for.*
  - *Plus, we can augment people in these roles with technology to learn how people categorize things and automate the triage with human statistical checking*
    - But some will slip through!
      - *Yes they will, but most / almost all will not!*
        - But we will stop some legitimate things!
          - *Yes we will, but we can easily have an appeals process.*
            - But…
              - *Yes I know – there is always a but.*
              - *But we don't have to be perfect to be effective.*

Default deny and the resulting delay

Security has long had the notion of default deny. You have to prove it's allowed before it's allowed. This is usually done be people getting approvals for their user identities. But the approval process for social media would inhibit the entire process and mechanism, produce delays without limit, and likely be completely ineffective.

It is really the content to be controlled and not the people anyway. The use of default deny on content without automated processing would be disastrous for any sort of conversation. But there are approaches for using automated processing that could work reasonably well. Among these are:

- Limited transitivity (any given meme expression can only go to so many people/places)
- Limited rate of spread (increased delays with increased uses)
- Limited volume over time (each account can only say so much so fast)
- Attribution (all of your accounts and actions are traced to you as a human)
- Reputation (the attributed identity gets a reputation that can control use)

And so forth. All of these are imperfect. All of these have points and counter-points, and all of these and others can be practically used and should be. However

The equities issue

The history of intelligence services in this regard indicates that they don't want to stop every action as soon as they see it. That's because when you stop (some of) them the rest of them may know it, and then they adapt. Since you cannot find it all, it pays to watch the event sequence taking place and pull on all the threads, only stopping the bad things just before they happen, and hopefully stopping most or all of the other bad things happening along with them.

It's all about the risks and rewards.

- A reward of stopping someone as soon as you find out about them is they are stopped.
  - *But if you stop them before you can prove a crime, they may walk and try again.*
    - But the longer you wait, the more likely they will actually do the/a bad thing.
      - *So we need to determine when to act and when to watch more.*
- Another reward is that if you stop them soon enough, you prevent them from becoming the real threat you are concerned with.
  - *But then the threat actor will know of it and adapt*
    - If you detect the attempt to turn people early enough, they are not turned, and you can use them to deceive the threat actor.
      - *So we need to determine when to act and when to watch more.*

Which is to say, we need to determine when to act in what way and when to watch more.

A mixed strategy based on where they are

It seems to me that the useful approach is a mixed strategy. By example:

- Regularly stop "bad" content and people at entry (by the application), in transport and storage (by the underlying service providers), at endpoints (by applications on all the endpoints), from spreading too far too fast (by attribution and limiting), and otherwise.
- Limit the freedom of speech in the same ways as it has been limited over the long history of freedom of speech, but using technology to detect and react automatically with appeals processes for the imperfections in the system.
- Make decisions about special application or non-application of these steps based on intelligence and law enforcement needs, so that they can take over and/or allow some things to continue in order to penetrate the threat actors and their organizations and make reasoned interdiction decisions. Specifically:
  - At some thresholds stop innocents from being turned, take over for the innocent to track down the real bad actors, and then act to remove the threats.
  - At some thresholds allow actions to continue under increased surveillance and other coverage to expose more of the existing threat actors.
  - At some threshold, before physical harm is done to others, stop the threat actors by more permanent or lasting actions and take down the networks and groups using the legal system to prosecute.

The strategy won't work as well in isolated parts

For this strategy to work, the components should operate simultaneously. For example

- If we are not regularly stopping content and people at admission and subsequently, then when we stop content or people for a specific reason, the threat actors will detect the stoppage as intelligence operations acting on to them, and change their strategies and tactics before they can be caught.
- If we regularly and automatically stop content and people at various places, then when we stop content or people for a specific reason, the threat actors will detect the stoppage as normal network behavior, and use their standard approaches to continue operations, which the intelligence agencies can observe, react to, and plan for.

Intelligence and interdiction ultimately depends on the ability to effectively deceive the threat actors. Controlling the certainty and uncertainty of threat actors with regard to the environment they operate in is therefore critical to success.

The hall of mirrors

Of course if we create an irrational infrastructure with unpredictable results for normal users, we will destroy the utility and the users will go elsewhere or, worse yet, the world will lose the tremendous benefits of the Internet. For this reason, the introduction of deceptions must be highly selective and oriented only toward achieving specific goals, presumably as authorized by national command authority or the courts. Of course this applies to many nation states who might be at odds with each other, so each will apply its own sets of rules and thus we have the global intelligence challenge of fighting cyber warfare in the age of influence operations.

Who is responsible for checking and what does that mean?

If you are working for Russia, you obviously will be following their approach. Same with any other nation state. My general advice however, is that all application providers, platform providers, infrastructure providers, and other parties on the path from person to person should check at their ingest, storage, and egress before sharing, with presumptive positives going to review and, if still positive, requiring a successful appeal for correction. Liability should be removed from such actions as long as they follow legal guidelines and policies of the respective entities. Those guidelines should identify standards of practice for what to stop, and the methods for identifying undesired content should be made available to all. In this approach the technical capacity of each substantial company to detect bad things in this context cannot be kept proprietary, and must be made openly available to others through enforced information sharing. Thus the code or capability to detect Nazi, ISIS, etc. symbols; beheadings, forced drownings, etc.; and other such bad things not permitted in each jurisdiction; should be made available to all providers and updated to increase detection capacity globally. Each should also be able to augment these with some limited added capacity to keep detection of what will and won't pass uncertain around the edges.

Conclusions

Countering terrorism and state sponsored threats to the social fabric is a complex business and it requires a mixed strategy enforced by many parties. There are many implications and complex interdependencies that should be considered. But that should not stop us from making progress, starting now, and continuing into the future.

