The practice of using disinformation and misinformation to promote parochial agendas isn’t new. Both have been used by tyrants, demagogues, dictators, authoritarians, and manipulators of every stripe for millennia. One thing that’s new to our generation is the digital twist of Internet trolling. The effectiveness and increasing use of this tactic, highlighted in the 2016 US presidential election, justifies increased attention. An earlier Computer column encouraged such attention, and we elaborate here.

Disinformation and misinformation both involve the distribution of false information, but with differing objectives. Disinformation involves the intentional planting of false information to conceal truth or deceive the audience, especially by state actors, whereas misinformation is more generic and relaxed regarding intention, concealment, and source. For our purposes, we intend the definition of disinformation to include not just governments but also political groups, ideological movements, and other social entities. Disinformation is more pernicious, being necessarily both intentional and deceptive in its pursuit of social engineering goals. Although some trolling might be without willful deception (as in the case of mistaken “true believers”), disinformation is the more natural ally of trolling and is thus our focus.

The topic of disinformation is both complex and varied: it’s complex owing to its convoluted methods; it’s varied because of its different practitioners and contexts. It can be used to enlist support, confuse, de-legitimize, defame, intimidate, confound, escape detection or blame, avoid prosecution, and on and on. The public relations strategist uses disinformation in different ways than the tyrant owing to the latter’s assumed greater imperviousness to punishment or retribution. Similarly, the ideologue’s use of disinformation is different from that of the corrupt politician. Disinformation techniques and content vary with the purpose, targeted demographic, medium, and social networking platform.
These issues apply to trolling as well. Consequently, we’ve developed a partial taxonomy to better characterize trolling’s many manifestations. This is an appropriate time for a taxonomy, for trolling is mature enough now to reveal interesting patterns and suggest future trends and defenses.

ROOTS AND MISSING LINKS

Trolling is confirmation, in a sense, of a fundamental flaw in the notional roots of the modern Internet-enabled Web. Those roots are typified by, for example, Paul Otlet’s Mundaneum system, implemented in 1910 to collect and categorize all of the world’s important knowledge (www.mundaneum.org/en); H.G. Wells’s notion of a World Brain, outlined in a 1938 collection of essays and addresses with that title; and Vannevar Bush’s Memex system, described in his influential 1945 article “As We May Think.”

Bush envisioned a collective memory system that would advance a knowledge explosion by serving up the corpus to anyone on demand through associative indexing and browser history-like “paths” not unlike the use of hypertext to organize the Web. As was customary in the early information age, Bush was driven by the simultaneous desire for ease of information access and avoidance of information overload. He wasn’t concerned about data reliability and source authentication.

As it turns out, this overly simplistic and naive view of the information access challenge has been perpetuated ever since on the Web. To wit, subsequent work on metadata standards, including the Dublin Core elements (http://dublincore.org/documents/dces; https://tools.ietf.org/html/rfc5013), completely ignore any measure of authenticity and reliability. The closest metadata elements would include oblique terms such as “provenance,” “conforms to,” and “is referenced by.” This deficiency has been carried forward in such subsequent document type definitions as the Open Source Metadata Framework and the Resource Description Framework. To overcome this deficiency, more user control is needed—perhaps a user-driven metadata insertion tool for elements like “suspect,” “disproved,” and “content warning,” or some sort of Bayesian trigger to deal with today’s fake news and alt-facts. Otherwise, the 21st century’s spin on Bush’s vision might progressively become “As We May Deceive.”

The study of disinformation, from an information-theoretic point of view, has thus regretfully been at best occasional and informal. We have in mind, for example, contributions by David Martin and H. Michael Sweeney on disinformation and traits of disinformationists. While informative, especially with respect to the current political landscape, these works are largely anecdotal, lack examples, and aren’t directly related to trolling. Spy the Lie provides a practical guide, with examples, for detecting deception, including an analysis of behavioral cues that might betray the act. A rough equivalent for social media deceptions is sorely needed. Alas, self-published contributions on the Web, and those from the popular press, fail to do justice to the full impact of disinformation generally and trolling in particular.

TROLLING AS AN IDEOLOGICAL WEAPON

Online trolling is readily weaponized—it fits comfortably within pathocracy and kakistocracy as an effective tool of online manipulation, obfuscation, and deceit. The relationship of trolling to disinformation and politics has reached a modern zenith owing to the current US administration’s relaxation of the norms and expectations of veridical communication and the Russian government’s embrace of trolling. That said, the White House’s proneness to misinformation and even outright disinformation is a symptom of a more general social problem—namely, political emotionalism, in which facts are too often considered less of a foundation and more of a hindrance. That trend manifests itself in a tolerance of falsehoods under the guise of alt-facts, the inability to distinguish confirmable statements from beliefs and opinions, and an unreflective commitment to ideology-based and simplistic slogans, catch phrases, sound bites, formulas, and beliefs. Social scientists have developed theories of social dominance, authoritarianism, and instability that explain some these characteristics in terms of group behavior, economics, and social hierarchy.
WHY DISINFORMATION? WHY TROLLING?
Disinformation generally and trolling specifically are expedient ways to manipulate public opinion. Authoritarians of all generations understood that sound and reasoned argument isn’t sufficient to exercise control over others. Something more powerful but short of force is needed. Such machinations, to be effective, must be carefully engineered and targeted, an objective often unachievable through reasoned public debate. If politicians were to rely on logical debate, free of manipulative rhetorical devices, public consensus might be influenced by the merits of the arguments themselves when interests, often authoritarian or domineering, wish to avoid this.

Disinformation and trolling are expedient ways to manipulate public opinion. They can polarize issues to exploit a human bias toward binary choices.

Carefully crafted disinformation campaigns and trolling efforts can be instrumental in achieving the desired effect. They can artificially polarize issues to exploit a human bias toward binary choices—seeing the world in black and white, big and small, rich and poor. This is related to what Hans Rosling calls the gap instinct.15 Its appeal must follow in part from the cognitive simplicity of binary distinctions, much as we experience with true/false questions on exams. Other things being equal, cognitive effort is lower on true/false than multiple-choice questions because there’s less to think about.

Disinformationists and trolls seek to create a sense of extremes where the extreme they tout is cast in a more appealing way than the alternative. In order to force the information consumer to the desired extreme, they use lies, prevarications, untruths, alt-facts, unlikely theories, distortions, ad hominem attacks, and other rhetorical devices as part of a Machiavellian propaganda or “messaging” campaign to create the desired artificial duality in lieu of the more nuanced and reality-based presentation that would result from clear-headed analysis. Modern online disinformation and trolling campaigns functionally resemble phishing attacks in combining a modest amount of computing and networking skill to cloak the real goal and lure the target using perception management (manipulating the public into thinking they perceive something they don’t, or vice versa) and social engineering (motivating the public to do something they otherwise wouldn’t have done).

In his book Factfulness,15 Rosling describes how evolutionary traits like hard-wired fast-response brains produce simplistic world views that discourage adequate reflection and deliberation for decision making. He identifies 10 evolutionary “instincts” that no longer serve humanity well in separating truth from predatory fiction. Such instincts should be critically discussed as part of college-level general education, if not in high school. Primary education should provide practical skill in BS detection, right along with the 3 Rs. Call it the 4th R: reality checking.

A TAXONOMY OF TROLLING
Online trolling has matured to the point that we can discern some evolutionary patterns and future directions. The value proposition is obvious from the 2016 US presidential election: low-cost, potentially high-impact voter manipulation through micro-targeting. Political scientists and others continue to study the degree to which trolling influenced the vote. UK-based Cambridge Analytica executive Mark Turnbull took credit for playing a key role in Donald Trump’s win,16 and there’s now sufficient concern over the use of trolling by foreign governments to undermine US federal elections that, as part of the Mueller probe, the US Department of Justice indicted the Russian trolling factory, the Internet Research Agency, for 8 federal crimes17 as well as 13 Russians and 3 Russian companies for attempting to subvert the 2016 election.18

One thing is certain: online trolling is here to stay. Even if federal legislation were passed to outlaw it, problems like reliable cyber-attribution19—at least which that is admissible in court—will prevent trolls many avenues to circumvent whatever laws might be enacted.

So what’s the future of online trolling and its containment? We offer the following informal taxonomy as a means to focus our response.

Provocation trolling. To elicit a particular response, such as hostility, from participants of an online forum. For example, in the “Reactions” section of a Yahoo! article about a 20-year-old Guatemalan woman shot dead in Texas by a US border agent, many top comments seemed intended to spark a flame rather than shed light. For example, the first comment was “Medal of Honor!!!” (http://www.webcitation.org/7I0m5nOWF). Similarly, in an online discussion, blaming liberals or conservatives for a tragic or controversial incident will likely cause some offended readers to lunge for the bait.

Social-engineering trolling. To incite participants to activities they normally wouldn’t have undertaken—convince readers to join an organization, send a donation, observe a boycott, vote for/against a candidate, and so on.

Grooming trolling. Sending messages intended to insinuate the sender into the mind of the recipient as a slippery slope to further persuasion. Radical organizations are notorious for
using this variant of social-engineering trolling to recruit members: ISIS was widely noted for “fishing” for new members on Twitter this way, and US extremist groups are frequently noted for using this tactic.

**Partisan trolling.** To use social media surreptitiously to achieve political ends. Here’s where the heavyweights really get involved. For example, trolling has been exposed as an important component of Russia’s “firehose of falsehood” (see below) propaganda strategy, especially in the recent US presidential race.20

**Firehose trolling.** High-volume, rapid, continuous trolling without concern for consistency. Apparently a favorite of Russia, it focuses not on promoting a particular position or viewpoint but on divisiveness for its own sake. For example, according to Charles Clover, Aleksandr Dugin’s book *The Foundations of Geopolitics* is influential at the highest levels of the Russian government and “assigned as a textbook at the General Staff Academy and other military universities in Russia.”21 (A good English translation of the entire book isn’t yet available.) Clover quotes Dugin as writing, “It is especially important to introduce geopolitical disorder into internal American activity, encouraging all kinds of separatism and ethnic, social and racial conflicts, actively supporting all dissident movements—extremist, racist, and sectarian groups, thus destabilizing internal political processes in the U.S.” Trolling is certainly well suited to this activity. And it can be tough to counter. Christopher Paul22 recommends against trying “to fight the firehose of falsehood with the squirt gun of truth,” but fails to provide fully satisfying alternatives.

**Ad hominem trolling.** Defaming or discrediting individuals or groups to delegitimize their positions without engaging them on their merits. The following snippet from an exchange on an email list exemplifies this.

ML: [Controversial claim] Anybody who claims otherwise is ignorant, uninformed, or lying.

A naive respondent might be whiplashed at this point because a counterargument, reasoned or not, has already been pre-characterized as ignorant, uninformed, or a lie. The best response is probably to simply point out the rhetorical device used here, as respondent PD does next.

PD: Ooh—is this the choose-your-own-ad-hominem part of the show?

Yet even this response is hobbled because the discussion has now been diverted into a rhetorical cul-de-sac that saves ML from losing the argument.

**Partisan trolling.** To use social media surreptitiously to achieve political ends. Here’s where the heavyweights really get involved. For example, trolling has been exposed as an important component of Russia’s “firehose of falsehood” (see below) propaganda strategy, especially in the recent US presidential race.20

**Firehose trolling.** High-volume, rapid, continuous trolling without concern for consistency. Apparently a favorite of Russia, it focuses not on promoting a particular position or viewpoint but on divisiveness for its own sake. For example, according to Charles Clover, Aleksandr Dugin’s book *The Foundations of Geopolitics* is influential at the highest levels of the Russian government and “assigned as a textbook at the General Staff Academy and other military universities in Russia.”21 (A good English translation of the entire book isn’t yet available.) Clover quotes Dugin as writing, “It is especially important to introduce geopolitical disorder into internal American activity, encouraging all kinds of separatism and ethnic, social and racial conflicts, actively supporting all dissident movements—extremist, racist, and sectarian groups, thus destabilizing internal political processes in the U.S.” Trolling is certainly well suited to this activity. And it can be tough to counter. Christopher Paul22 recommends against trying “to fight the firehose of falsehood with the squirt gun of truth,” but fails to provide fully satisfying alternatives.

**Ad hominem trolling.** Defaming or discrediting individuals or groups to delegitimize their positions without engaging them on their merits. The following snippet from an exchange on an email list exemplifies this.

ML: [Controversial claim] Anybody who claims otherwise is ignorant, uninformed, or lying.

A naive respondent might be whiplashed at this point because a counterargument, reasoned or not, has already been pre-characterized as ignorant, uninformed, or a lie. The best response is probably to simply point out the rhetorical device used here, as respondent PD does next.

PD: Ooh—is this the choose-your-own-ad-hominem part of the show?

Yet even this response is hobbled because the discussion has now been diverted into a rhetorical cul-de-sac that saves ML from losing the argument.

**Jam trolling.** Disrupting a discussion or communication channel with high message volume (the trolling equivalent to a DOS attack). Technologically, automated trollbots will make this an increasing problem.

**Sport trolling.** Trolling for the self-gratification of the troll (just for the fun of it).

**Snag trolling.** Evoking responses to satisfy curiosity. One of the less toxic varieties, this nevertheless tends to divert and obscure.

**Nuisance trolling.** Derailing the thread of an online forum (blog, chatroom, and so on) for no other reason than to irritate other participants. A variant of sport trolling.

**Diversion trolling.** An insidious tactic for blocking legitimate communication by diverting a thread in a direction that’s misleading, irrelevant, false, and so on. Thus, a discussion about rising crime rates could be diverted by citing a small community that hasn’t had a murder in 20 years, or a discussion about falling crime rates could be diverted by mentioning a recent crime.

**False-flag trolling.** Pretending to be of a group or hold an opinion that the troll actually opposes, and presenting a message intended to make that group or opinion look bad. This is one of the harder forms of trolling to detect, because the writer could in theory really have the opinion claimed but not realize how his obnoxiousness is creating the opposite of the desired effect. For example, a type of robocall used in political campaigns pretends

---

**Problems like reliable cyber-attribution will provide trolls many avenues to circumvent whatever laws might be enacted against trolling.**
provoking the spirited practice the troll wants. The troll thus hones debate skills for uses like higher-stakes trolling later.

**Proxy trolling.** Using intermediary trolls to do the heavy lifting. De rigueur for large organizations, which hire people to do it. One application is astroturfing: promoting a position, product, person, and so on for which there’s little awareness or support by making it look like that entity is widely approved of. Websites and organizations set up by special interests but given names like “Citizens for X” are standard examples. Proxy trolling provides rich opportunities for all manner of resource-rich, unscrupulous actors.

**Faux-facts trolling.** Deliberate spreading of fake news, alt-facts, and other lies under the guise of truth. To fight this type of trolling, refereeing organizations, typified by the well-regarded Snopes (https://www.snopes.com/about-snopes), are a socially valuable, even essential institution. We can expect large organizational trolls to sow chaos and confusion with fake fact-checking organizations of their own.

**Insult trolling.** Insults spark responses that drain the target’s energy. They also make the target look bad and are demoralizing.

**PR trolling.** Making the troll or the views the troll is promulgating look good rather than attacking others. For example, the troll could make a claim and unverifiably cite a brother-in-law “who was there.” But the most common example is state approval of another text. It’s easy to upvote another troll’s message, or respond to a posting with “Right on!” or “Thank you for saying what so many know but are afraid to say.” This boosts persuasiveness via a bandwagon effect.

**Chaff trolling.** Sending messages that are essentially content free and thus vacuous. For example, on social media platform Quora someone claimed that a relative assigned to help guard former president Obama said that the president was “… fake as [expletive deleted].” One might well question if this relative really existed, and if he did, whether the quote was accurate. Yet consider also the word “fake”: here it carries little if any information about its subject but is an effective insult for the many unsavvy readers.

**Wheat trolling.** High-quality trolling using content that’s hard or impossible to refute—for example, a cleverly doctored photo or text incorporating seemingly well-sourced “facts.” Some lies contain their own logical inconsistencies; others smell bad only to a domain expert.

**Satire trolling.** Good satire cuts deep. It’s hard to create and even harder to generate automatically. Thus, effective as it is, satire trolling will likely remain a relatively small player in the trolling world.

**TURING TROLLBOTS**

A trollbot is simply an automated troll. Like a chatbot, it generates texts computationally. Unlike chatbot texts, trollbot output possesses markedly weaker requirements for coherence and continuity from its context. Consider, for example, a program that uses a simple bag-of-words algorithm to detect tweets or other posts critical of a particular position or public figure. It then posts replies randomly picked from a set of stock replies like “You tell’em baby!” and “That’s SO right.”

Informally, let’s refer to a trollbot that’s indistinguishable from a human troll as a Turing trollbot—one that has passed the trolling equivalent of the Turing test. A computer-controlled chatbot passes the traditional Turing test if and only if the human tester cannot distinguish the chatbot from a human. Compared to a chatbot, a trollbot has a much easier time passing—the weaker constraints on trolling make it so. Sure, there are human trolls for whom sophisticated trolling is an unsavory art form that would be hard to imitate, but a Turing trollbot need only mimic the lowest-common-denominator human troll to masquerade as a real person.

The concept of the Turing trollbot is increasingly recognized. The hardest technical aspect of primitive Turing trollbot design is sneaking through smart filters like CAPTCHA. In fact, such trollbots could soon emerge as easily downloaded freeware apps. But primitive Turing trollbots are just a start. As we were writing this article, IBM unveiled its Debater system, which successfully took on a college debate champion. This is a much greater challenge than deploying successful trollbots, which can be ever so much more efficient and economical than a paid human.

With armies of well-nigh undetectable trollbots on the horizon, what’s one to do against this threat? One approach is to simply ignore outright all controversial social media comments—that might protect individual readers. Another approach is mass immunization. The simplest way to ensure public health is for enough people to reply to suspected troll messages by shining a light on them. “Are you a troll?” might serve not just as a comment but as a warning and reminder to readers who otherwise might have overlooked the possibility. But one way or another, society must
develop strategies to reduce trolling and trollbot effectiveness.

Research is also needed to investigate the potential for automatic trolling detection software. What kinds of trolling are undetectable? What kinds have already been detected, and who are their sponsors? We also need to educate the public. An increasingly necessary goal of primary education is training people to approach social media statements with suspicion, especially when it comes to bias and misinformation.

The Internet—through social media and fake news outlets—has saddled us with the biases of those seeking to manipulate others through new forms of information corruption such as source displacement/concealment, decontextualization, and the like. Where the traditional measures of networks were in terms of value,26,27 a new and useful measure of networks is their potential for abuse.28

**POLITICAL TROLLING**

In addition to the computer and networking context, online trolling must be understood in a geopolitical context29,30 especially with respect to its utility in international competition and rivalry. For example, a measurable amount of the identified external political trolling used to influence the outcome of the 2016 US election appears to have been either sponsored or inspired by Russia. China certainly has the capability for effective political trolling as well. As time passes, more countries will inevitably engage in it as a useful and cost-effective way to project influence.

Free societies are the most susceptible to political trolling because in those countries mass opinion is a strong driver of national policy. Moreover, polarization and partisanship have been increasing for decades.31,32 Social media and networks evolve, the current deficiencies in teaching disinformation tactics widely as an important civic skill will become more apparent. Our children, like all too many adults, lack the basic skills to look upon divisive, emotive communication critically. This is a severe educational shortcoming that promises to exact a considerable toll on democratic systems.

Society needs to understand why people troll. It seems to be one of many addictive behaviors mostly afflicting alienated young males and enabled by the anonymity and easy accessibility of the Internet, much like overindulging in online porn or videogames (https://www.quora.com/Whats-it-like-to-be-an-Internet-troll). But perhaps it’s not as important to understand the psychology underlying trolling as it is to avoid being manipulated by it. As Lee Edwin Coursey24 advises,

> The next time you see a hyperbolic social media post that confirms your worst fears about people of a particular race, gender, religion, or political affiliation, your first reaction should be, “nice try, Russian troll,” rather than “OMG I MUST REPOST THIS EVERYWHERE!!!” Learn to take a breath and pause before you immediately like, retweet, or share divisive messages from obscure sources. Be especially wary of emotional manipulation. Most importantly, fact check yourself before spreading information designed to foment outrage and factionalism. Remember that the phrase “Russian disinformation campaign” does not describe some outdated method from a bygone era, but instead represents an active, effective tool being used against you right now.

The cognitive load for detection and prevention is considerable, even for a coalition of the willing to do so. There’s little cognitive load for tribalists because of illusory feelings of superiority, anosognosia (critical lack of self-awareness), and other cognitive biases. Part of the threat (and hence the value) of trolling is that so many independent-minded people don’t have the time and energy to check facts or verify claims, while tribalists and authoritarianist followers don’t feel the need.

As a consequence, trolling is convenient fodder for the gullible. It’s free, self-reinforcing propaganda that unifies true believers and confuses or obfuscates issues sufficiently to manipulate fence-sitters. The game changing potential lies with the latter (for example, the 40,000 votes in three states that effected the Electoral College outcome of the 2016 US presidential election). This is where trolls and other social media manipulators can thus be stimulated to action or opinion without benefit of the reflection that would call into question the validity of the message or stimulate thoughtful evaluation. Fact checking, introspection, and analysis work against the interests of trolls. In this way, trolling is similar to a military campaign where the goal is action without debate.

We might take a lesson from Winn Schwartau’s Time-Based Security Model...
in this regard. The model posits that a security system can be effective only when the time it takes to detect a security breach and mitigate against the threat is less than the time it takes for the security breach to achieve its objective. There’s a parallel when it comes to mitigating against the effects of abusive social media. For it to be effective, the detection time must be near zero because the reaction time required to re-tweet, forward, and so on is negligible. The parallel with trolling is that the troll is focused on achieving quick results before second thoughts might be raised.

It’s worth adding that trolling’s ability to promote division can also be used to nurture social reform and is thus a doubled-edged sword for authoritarian and totalitarian states. For that reason, such states must carefully monitor and control trolling and related digital media manipulation tools within their borders.

New though it is in the toolbox of Machiavellian kingpins and social misfits alike, the effectiveness of trolling ensures that it’ll continue to play an important role in future politics.

REFERENCES


HAL BERGHEL is an IEEE and ACM Fellow and a professor of computer science at the University of Nevada, Las Vegas. Contact him at hlb@computer.org.

DANIEL BERLEANT is a professor of information science at the University of Arkansas at Little Rock and author of the book The Human Race to the Future (4th ed., Lifeboat Foundation, 2017). Contact him at berleant@gmail.com.