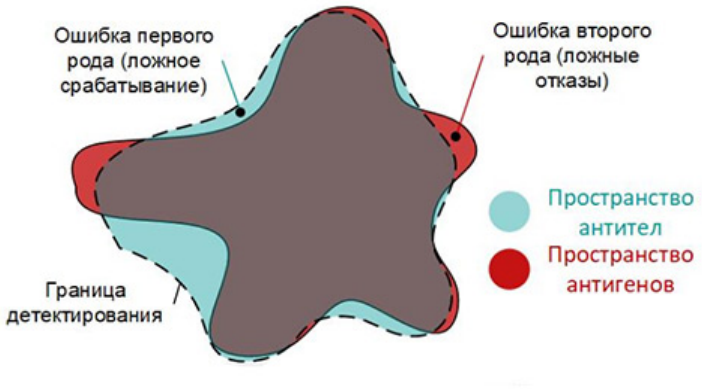
Цифровой иммунитет: учёные СФУ научат искусственный интеллект «отлавливать» утечку данных



Учёные СФУ разрабатывают самообучающуюся систему искусственного интеллекта, способную детектировать скрытую информацию, встроенную в различный медиаконтент, распространяемый в интернет-ресурсах. Такая искусственная иммунная система имитирует некоторые естественные процессы, происходящие в биологической иммунной системе человека, например, вырабатывает особые цифровые «антитела» для обнаружения «враждебных антигенов» — таких медиафайлов, в которые скрытым образом внедрена дополнительная информация. Основные результаты исследования отражены в статье.

«Стеганография — это очень древнее искусство „скрытописи“. Выбривание тайного послания на голове раба, симпатические чернила, которые становятся видимыми только в ультрафиолетовых лучах — это всё первые попытки человечества сделать неочевидное для непосвящённых сообщение „с двойным дном“. С развитием интернета и различных социальных сервисов с середины 2000-х годов в сети стали циркулировать поистине гигантские объёмы медиаданных — в том числе, изображений одного из самых популярных форматов JPEG. Как раз на способах обнаружить то, что скрыто за простой на первый взгляд картинкой, мы сосредоточились в своём исследовании», — сообщил доцент кафедры прикладной математики и компьютерной безопасности СФУ **Алексей Шниперов**.

Учёный рассказал, что отличить «чистое» изображение от наполненного скрытой информацией невозможно без специальных инструментов — только зная способ, которым туда была встроено дополнительное содержание, получится извлечь, например, карту стратегического объекта. Или рассмотреть скрытое послание экстремистского толка. Для обывателя изображение дерева или улыбающейся девушки, снабжённое секретной «начинкой», ничем не будет отличаться от миллионов подобных файлов.

«В наше время проблема пересылки потенциально опасного медиаконтента набирает обороты и касается, в том числе, крупных компаний и корпораций. Чтобы не стать жертвами информационных утечек и избежать крупных финансовых убытков, они тратят внушительные средства на разработку систем, способных фильтровать контент Всемирной паутины. Сложность работы с медиафайлами состоит в их многообразии. Для тех же изображений формата JPEG есть огромное количество параметров, — их контент, параметры цветопередачи, коэффициент сжатия и т. д. А ещё за последние десятилетия лавинообразно нарастает количество способов встроить секретное содержание в обычную картинку, включая уже готовое программное обеспечение. На сегодняшний день не существует единственного супералгоритма, который, как Супермен, мгновенно спасёт сеть от контента „с двойным дном“. Но мы в СФУ разрабатываем один из возможных инструментов борьбы с такими изображениями. Систему построена по аналогии с доказавшей свою эффективность человеческой иммунной системой и поэтому способна обнаруживать

подозрительные для неё файлы, примерно как наши клетки крови обнаруживают вирус или бактерию», — продолжил учёный.

«Иммунная сеть», создающаяся красноярскими исследователями, равно как иммунитет человека не знает наверняка, с каким противником ей придётся столкнуться. Она не зацикливается на «бесполезных» признаках — размере или предмете, изображённом на картинке (они будут абсолютно идентичными у «безопасного» и «опасного» файлов), но постоянно воспроизводит своего рода «антитела», отличающиеся друг от друга, чтобы с большей вероятностью детектировать «недоброкачественные» файлы, в которых может скрываться дополнительное содержимое.

«Разработанная система постоянно самообучается, безостановочно производя всё новые виды цифровых „антител“, каждое из которых представляет собой многомерное пространство специфических векторов-характеристик. Когда одно из них „на удачу“ вдруг цепляет какой-то медиафайл, мы, как „цифровые доктора“, смотрим, что не так с этим файлом. Если „антитело“ сработало правильно и сумело обнаружить скрытый контент в файле JPEG, например, то похожие на него цифровые „антитела“ будут продуцироваться системой в бóльших количествах. В качестве нормальных объектов в этом случае наша сеть рассматривает изображения, не нагруженные скрытым контентом. На них „цифровой иммунитет“ просто не реагирует», — отметил **Алексей Шниперов**.

Ещё одно сходство с биологическим иммунитетом можно обнаружить в «прививках», которые разработчики делают разрабатываемой системе: в неё загружают определённое количество (например, несколько сотен) «недоброкачественных» файлов — это помогает системе выработать «приобретённый цифровой иммунитет», примерно как вакцина от гриппа мобилизует защитные силы организма и защищает человека от вспышки сезонных заболеваний.

«Выборка „опасных“ изображений, которые мы используем для обучения нашей искусственной иммунной системы, — это капля в море медиафайлов, циркулирующих в сети. Конечно, система быстро учится — многократно множатся „цифровые антитела“, а также осуществляется их отрицательный и клональный отбор — системой выбраковываются те, которые по ряду признаков антителами считаться уже не могут и не подходят для обнаружения противника. Но сейчас мы подошли к пределу технических возможностей нашего оборудования: требуются более мощные процессоры, чтобы „цифровой иммунитет“ тренировался дальше. Следующим шагом станет перенос всех вычислений на тысячи специализированных процессоров, которые имеются в любом современном графическом адаптере или их объединениях» , — констатировал исследователь.

Кстати, ещё одной проблемой «цифрового иммунитета» может стать аналог аутоиммунной реакции — «цифровые антитела» пока несовершенны и порой могу реагировать на безобидные изображения — такая «цифровая аллергия», по мнению учёных СФУ, может развиться при значительном увеличении «рабочих антител» в искусственной иммунной системе после переезда её на более современное компьютерное оборудование, что потребует дополнительных исследований и корректировок модели «цифровой иммунной системы».

О формировании «цифрового иммунитета» учёными СФУ написал портал «Научная Россия».

На фото: схематичное представление функции детектирования стеганоконтейнера искусственной иммунной системой

Пресс-служба СФУ, 3 февраля 2020 г.