Динамическая оценка рисков информационной безопасности на основе данных мониторинга

Как актуальные данные могут помочь в снижении рисков

Зуев С.В. 15-03-2021

Что такое «черный лебедь»?

«Этого же никогда не было. И вот - опять!» (В. Черномырдин)

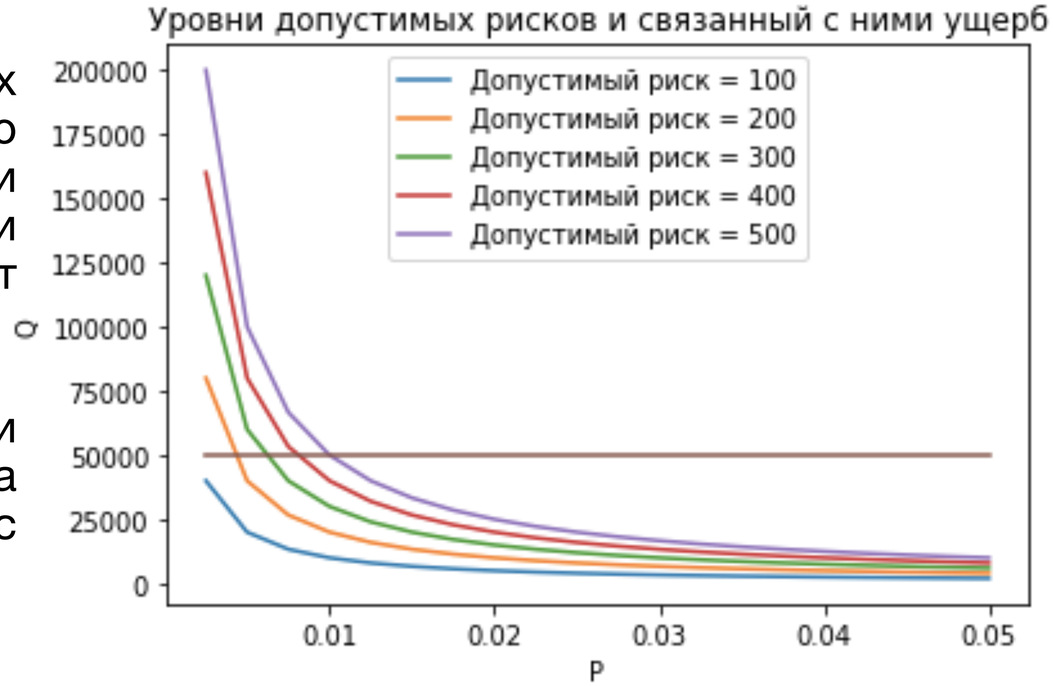
- «Черный лебедь» - это риск, связанный с редким событием, несущим большие последствия:

$$R = P \cdot Q$$

где вероятность P мала, а критичность Q велика.

- На рисунке видно, что при разных допустимых рисках действительный ущерб может составлять значительную величину: ущерб в 50000 соответствует вероятности примерно 0,07 при допустимом риске 300, а при допустимом риске 500 та же вероятность соответствует почти вдвое большему ущербу.

- Недооценка «черного лебедя» по вероятности или установка слишком высокого уровня допустимого риска может привести к потерям, не соизмеримым с сохранением системы.



Оценка вероятности «черного лебедя»

«Что мы знаем о лисе? Ничего. И то не все» (Б. Заходер)

- Вероятность с точки зрения риска - величина, характеризующая возможность события в системе в заданном временном интервале, то есть, она зависит от этого промежутка времени, от системы и от события;
- Событие описывается, как правило, его типом и интервалами значений параметров, в числе которых обязательным параметром является параметр ущерба, причиненного в результате происшедшего события всей системе или отдельным ее элементам (критичность);
- Система описывается ее признаками; системы, обладающие близкими признаковыми структурами, считаются похожими и могут объединяться в статистические ансамбли;
- Оценить вероятность означает сопоставить ей значение от 0 до 1; конечно, это невозможно без информации об уже случившихся событиях в системе или в ансамбле
- Даже при наличии информации о случившихся в ансамбле событиях, имеется возможность по-разному оценить вероятность наступления следующего подобного события - в этом заключается неопределенность;
- Обычно, чем больше информации - тем меньше неопределенность (Шеннон), но есть нюанс: только если среди имеющейся информации нет искаженной; при наличии искаженной информации, добавление любой информации может не дать снижение неопределенности или (еще хуже) привести к ложной определенности.

Оценка вероятности «черного лебедя» (продолжение)

У всякого пути есть начало,
но лишь пройдя до конца можно обрести славу (Ф. Дрейк)

- Начало пути оценки - это наиболее ранний из моментов формулировки гипотез и требований, с одной стороны, и появления данных, которым исследователь доверяет, с другой стороны;
- «Черный лебедь» - событие редкое и оценка его вероятности требует либо широкого исследования (рассмотрения ансамбля), либо глубокого (дальний ретроспективный анализ); в любом случае способ сбора информации для «черных лебедей», чаще всего, представляет собой отдельную проблему;
- Как «дойти до конца»? Для этого нужно знать «дорогу»: способ анализа данных и методику вычисления вероятности, то есть, вид вероятностного распределения;
- А кто строит «дорогу» и вдруг она ведет не туда? Критерием всегда является эксперимент; его и используем для проверки

Гипотеза

Пытаться не должна ты, делать должна (магистр Йода)
Чуть-чуть не считается (народная мудрость)

- **Любое событие обуславливается какими-то предшествующими ему событиями и, если этих обуславливающих событий недостаточно, то рассматриваемое событие не происходит**.

- Природа событий не важна, количество обуславливающих событий - не важно (главное - конечно), но важно, что все обуславливающие события должны произойти в течение какого-то времени ***с момента первого такого события***;

- Последнее очень важно для гипотезы, так как если предположить, что обуславливающие события должны происходить в течение какого-то фиксированного промежутка времени ***до рассматриваемого события***, то это приведет к совершенно иной статистике и именно это мы проверяли в эксперименте.

- Иными словами: либо начало отсчета времени для данного наблюдателя фиксировано (то есть, он когда-то начал этот отсчет и не может начать его откуда угодно), либо это начало не фиксировано и наблюдатель может наблюдать события в любых интервалах времени; оказывается, имеется принципиальная разница между этими положениями, влияющая на статистику

Немного теории … вероятностей

«Еще 10000 ведер и ключик у нас в кармане»
(Дуримар, «Приключения Буратино»)

- Вероятности «живут» в вероятностных пространствах; наше пространство состоит из бинарных серий длиной $k \in \mathbf{N}$, различающихся тем, что число положительных исходов в них меньше некоторого $a < k$ или наоборот больше или равно a; в последнем случае серию назвали a-положительной;
- В работе строго доказано, что такие серии образуют вероятностное пространство в аксиоматике Колмогорова;
- Вероятность в таком пространстве определена следующим образом:

$$P(l, a, k, p) = \frac{\sum_{\Omega} f_l^{a,k}(\omega)}{\sum_{\Omega, l} f_l^{a,k}(\omega)},$$

где $f_l^{a,k}(\omega)$ есть флаг того, что цепь k-серий $\omega \in \Omega$ имеет ровно l штук a-положительных серий; параметр p есть вероятность положительного исхода внутри k-серий.

Результат общения с теорией вероятностей

**Машина может быть любого цвета,
но автомобиль должен быть черным (Г. Форд)**

- В рамках нашей гипотезы, вероятность получить l инцидентов в первых N шагах последовательности событий (начавшейся с 1) равна

$$P(l,a,k,N,p) = \frac{\sum_{m}^{[N/k]} mC_{N-m(k-1)}^{m} C_{m}^{l} \left(\pi_{ka}\right)^{l} \left(1-\pi_{ka}\right)^{m-l}}{\sum_{m}^{[N/k]} mC_{N-m(k-1)}^{m}}$$

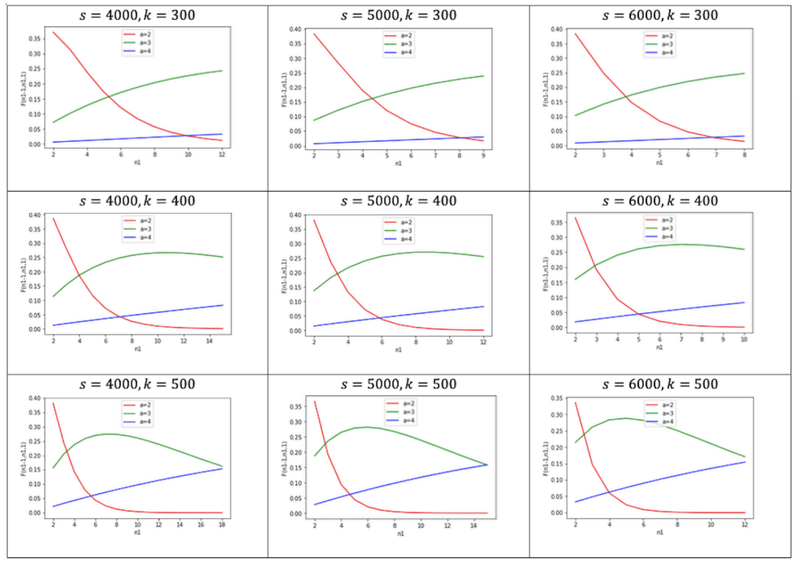
где $\pi_{ka} = \sum_{i=a}^{k} C_{k}^{i} p^{i} (1-p)^{k-i}$

- Вероятность получить первый инцидент в n-м измерении, занимающем s шагов последовательности, равна

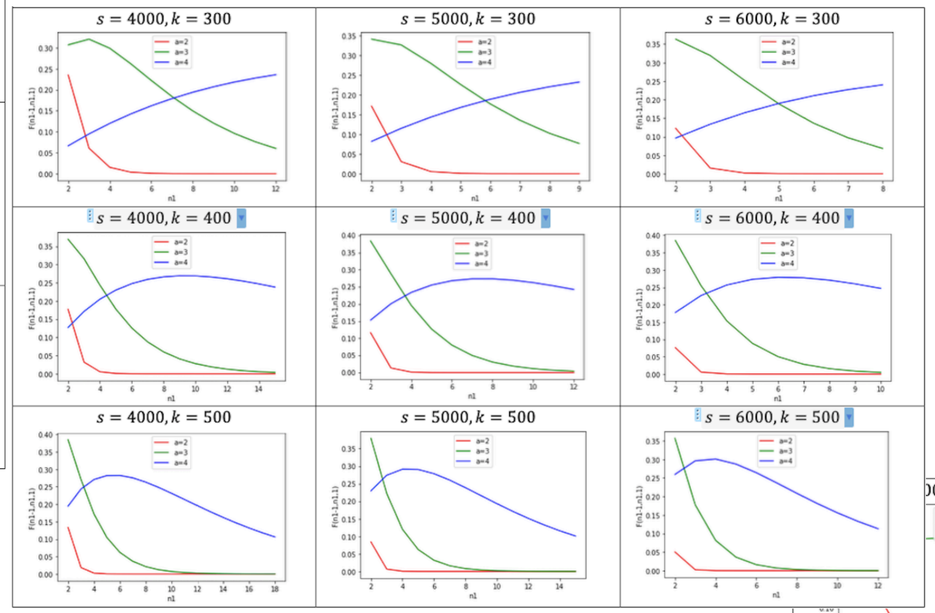
$$F_{a,s,k,p}(n) = P(0,a,k,(n-1)s,p) \cdot \left(1 - P(0,a,k,ns,p)\right)$$

Графики распределений

«Похожий» - не значит «тот самый»
(основная проблема китайских производителей начала века)

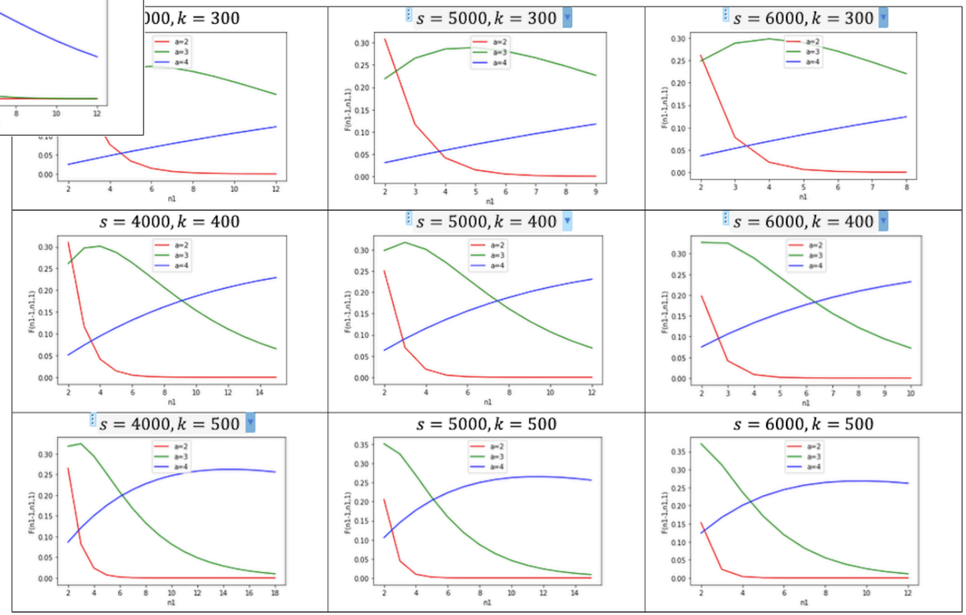


p=0.001



p=0.0015

p=0.002

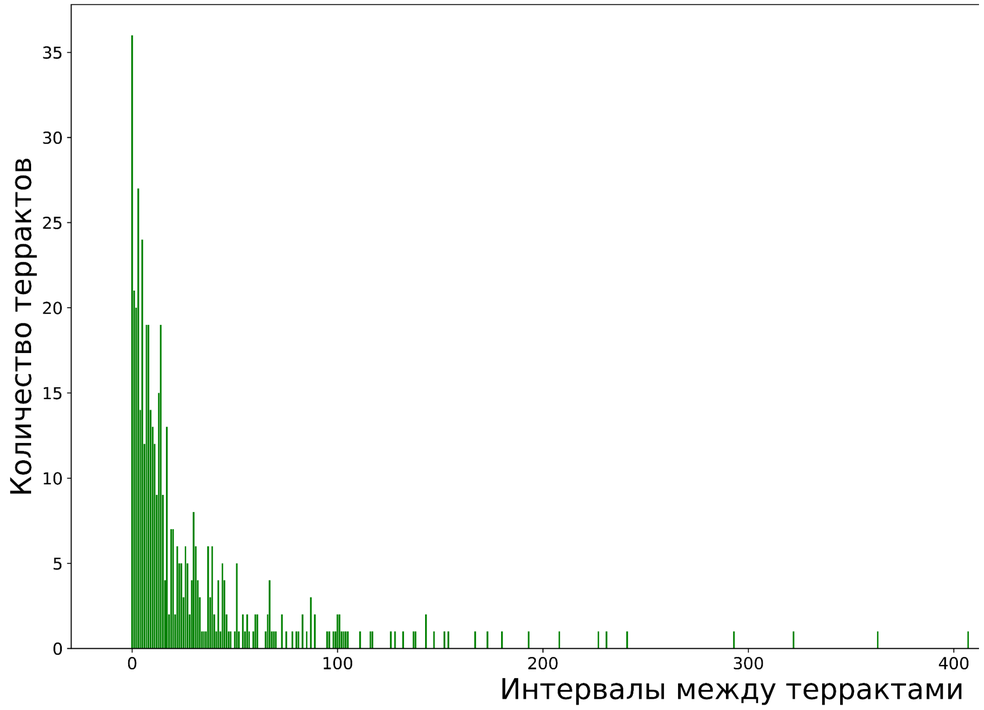


Ненормальные теракты

— А где я могу найти кого-нибудь нормального?
— Нигде, — ответил Кот, — нормальных не бывает.
Ведь все такие разные и непохожие. И это, по-моему, нормально
Льюис Кэрролл. Алиса в стране чудес

Посмотрим на анализ данных >180000 террористических атак за 1970-2017 годы (START Consortium, выложены в открытом доступе)

- Если наша гипотеза верна, то распределение вероятности повторного события любого определенного класса не будет нормальным, а будет подчиняться найденной нами статистике
- На графике - группа терактов с 51-100 жертвами. Но отличия от нормального распределения доказаны для всех групп терактов. Использовался точный тест Фишера и критерий хи-квадрат
- Для указанной группы терактов, данные согласуются с распределением на слайде 7 с $p = 0.001, \quad s = 5000, \quad k = 400, \quad a = 4$
- Можно подбирать значения параметров точнее для более полного соответствия выборке, но принципиально уже видно отличие фактической статистики от традиционной



Ну и что? Польза для информационной безопасности

**«Это какие-то неправильные пчелы.
И они дают неправильный мед» (В. Пух)**

- Пусть у нас есть ИС и в ней актив, подверженный риску R, который рассчитали на год (и собираются переоценивать через год);
- Где-то, в подобной ИС произошел инцидент, соответствующий профилю риска;
- Если риск считали как $R = P_0 \cdot C$, то теперь этот риск равен

$$R' = P_d \cdot C,$$

$$P_d = P_0 + 365\frac{1 - P(0,a,k,ds,p\sigma)}{d},$$

где σ есть масштабный фактор целевой системы в ансамбле (эксплуатируем эргодичность), а параметры a, k, s надо подбирать

- Очень легко может возникнуть ситуация, когда R был в 20 раз ниже допустимого риска, а R' в течение каких-нибудь 10 дней после инцидента в ансамбле будет в 2 раза выше допустимого риска (количественно этот случай описан в статье)

Вывод и заключение

«Не всякий вывод приводит к длительному заключению. А жаль.»
(один из хозяев ресторана на мосту Галата, Стамбул)

- Если научиться собирать данные о редких и опасных инцидентах ИБ в большом ансамбле систем, то можно снижать риски «черных лебедей», принимая защитные меры (часто дорогостоящие) на коротких промежутках времени;
- В деятельности такого типа важна скорость реакции на инцидент и, следовательно, скорость получения и обработки информации;
- Выигрыш - резкое (в разы) снижение ущерба от «черных лебедей»

СПАСИБО ЗА ВНИМАНИЕ!