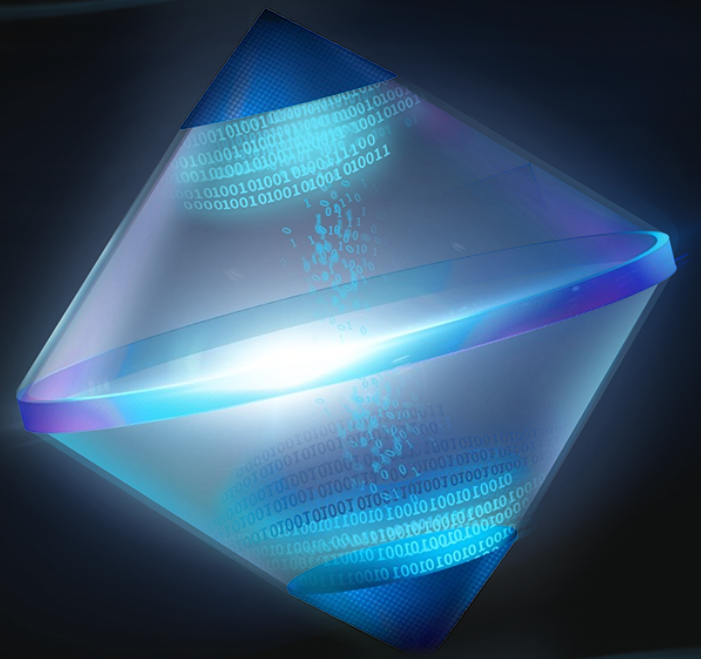
KCon

归源·智变

Returning to the source · Intelligent transformation

AI纪元:新视角下的企业蓝军



演讲人：俍梁 东北大学 软件学院

个人及团队介绍

关于个人:
东北大学软件学院信息安全专业 大二在读
寻云安全团队核心技术支撑
广州某公司实验室核心安全研究员
上海某局合作技术支撑

寻云安全团队
Cloud seeking security team

关于寻云安全团队:
自2019年成立以来,团队一直致力于信息安全多领域的研究探索，包括应用安全、系统安全、硬件安全、工控安全、二进制安全、区块链安全等领域
团队的目标是研发实用型网络安全产品，输送网络安全优质人才,构筑网络安全防线

目前已服务各行业客户:
公安部第一研究所　公安部第三研究所
广东省公安厅　广州市公安局
中国电子第六研究所　中国联通
中国电信　华为

KCon

归源·智变

Returning to the source · Intelligent transformation

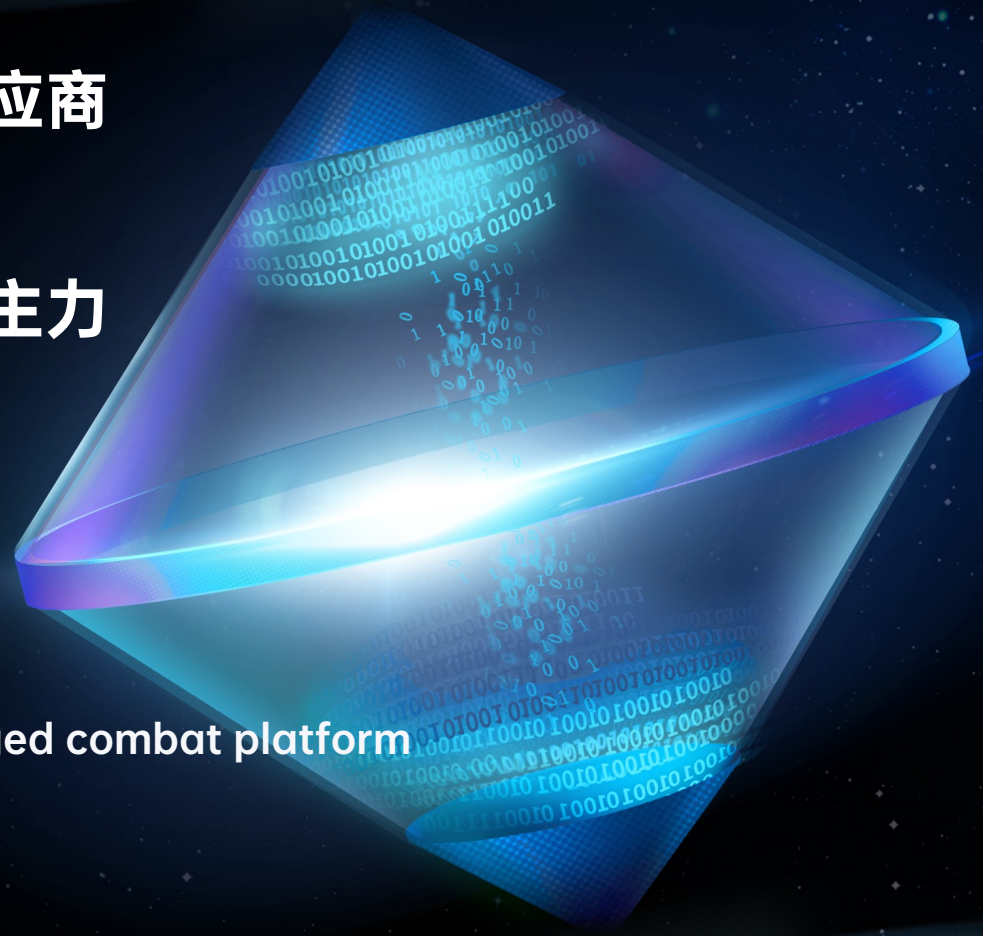
AI纪元:如何让你的AI快速成为核武器供应商

AI纪元:如何让你的AI迅速成为蓝军突破主力

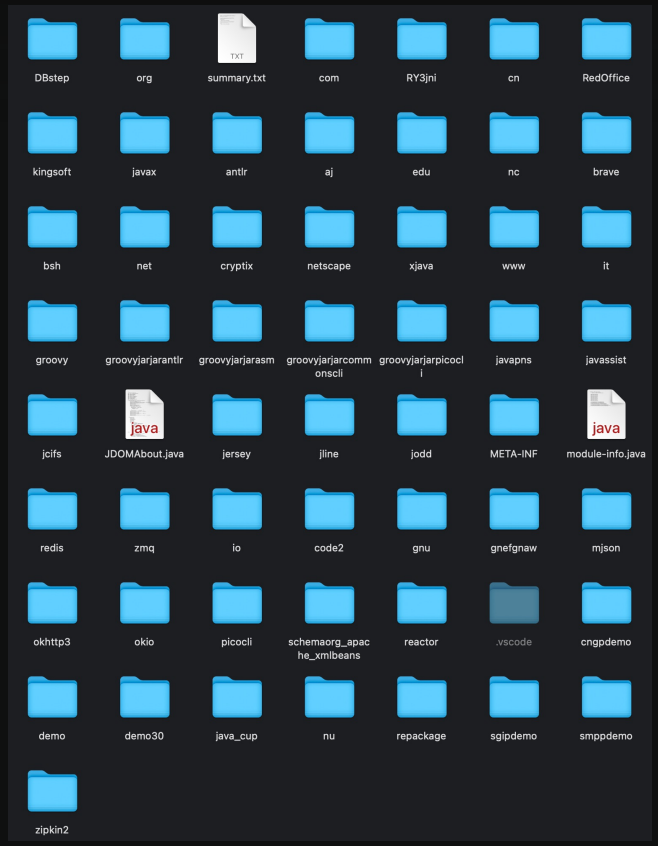
AI纪元:如何让你的AI深入敌后

AI纪元:AlphaMap-AI coordinated blue Army converged combat platform

致谢

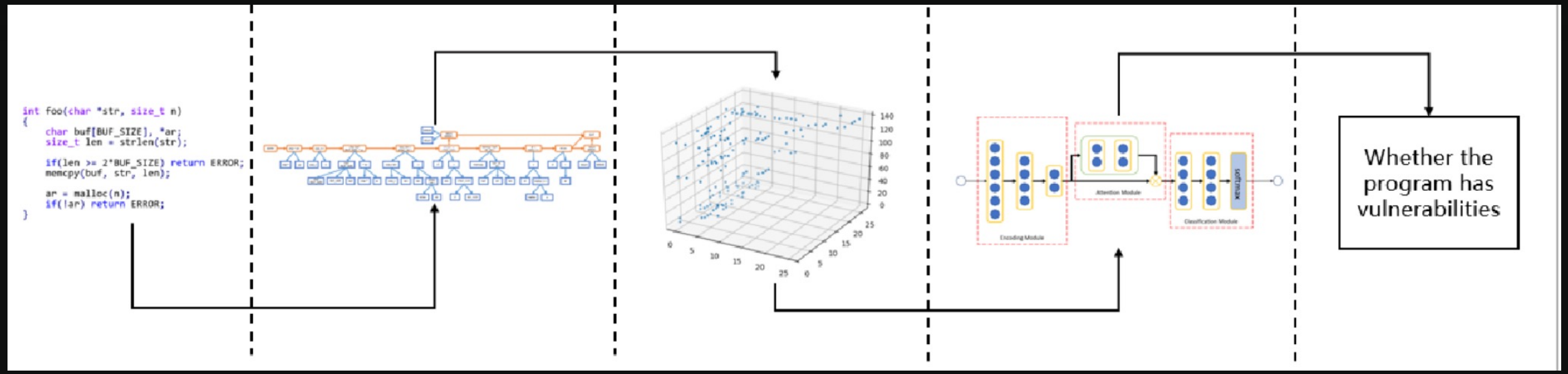


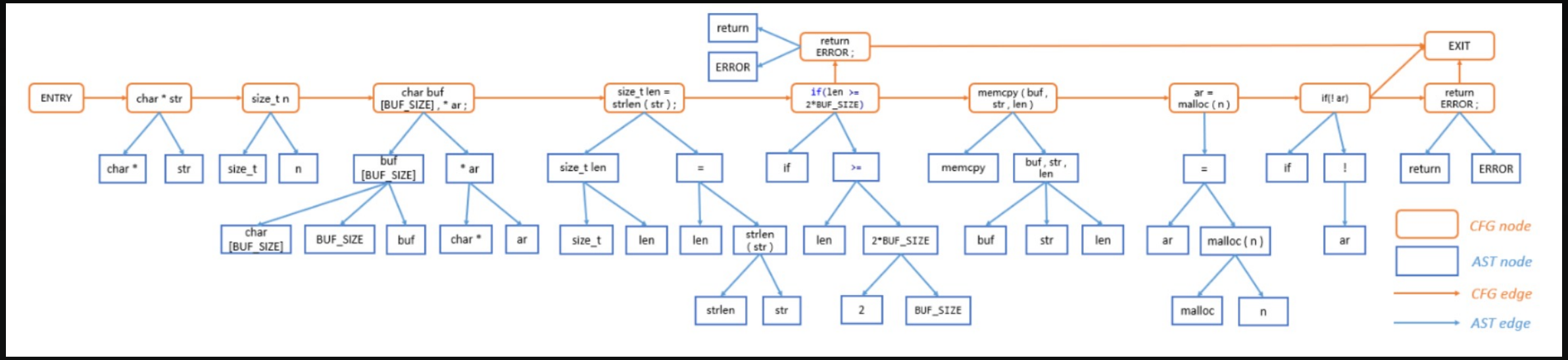
AI纪元:如何让你的AI快速成为核武器供应商



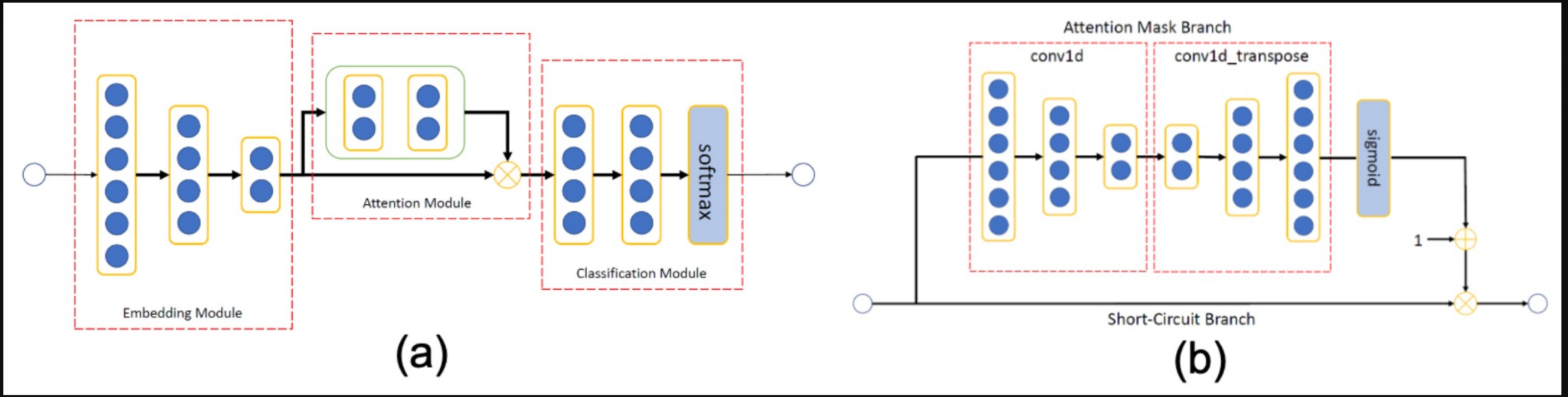
静态审计模型之
注意力模型的细粒度漏洞检测

VulSniper: Focus Your Attention to Shoot Fine-Grained Vulnerabilities. IJCAI-2019, 2019: 4665--4671





$$t_{i,j,k} = \begin{cases} 1 & is\_match(relations(v_i, v_j), f(k)) \\ 0 & otherwise \end{cases}$$


Embedding Module
Attention Module
Classification Module
softmax
(a)
Attention Mask Branch
conv1d
conv1d_transpose
sigmoid
1
Short-Circuit Branch
(b)

```python
import sklearn.decomposition as sk_decomposition
pca = sk_decomposition.PCA(n_components='mle',whiten=False,svd_solver='auto')
pca.fit(iris_X)
reduced_X = pca.transform(iris_X) #reduced_X为降维后的数据
print('PCA:')
print ('降维后的各主成分的方差值占总方差值的比例',pca.explained_variance_ratio_)
print ('降维后的各主成分的方差值',pca.explained_variance_)
print ('降维后的特征数',pca.n_components_)
```

```python
import sklearn.discriminant_analysis as sk_discriminant_analysis
lda = sk_discriminant_analysis.LinearDiscriminantAnalysis(n_components=2)
lda.fit(iris_X,iris_y)
reduced_X = lda.transform(iris_X) #reduced_X为降维后的数据
print('LDA:')
print ('LDA的数据中心点:',lda.means_) #中心点
print ('LDA做分类时的正确率:',lda.score(X_test, y_test)) #score是指分类的正确率
print ('LDA降维后特征空间的类中心:',lda.scalings_) #降维后特征空间的类中心
```

那么 如何通过细粒度检测模型进一步提高AI代码审计的准确度？

1.增加敏感函数回溯参数过程跟踪

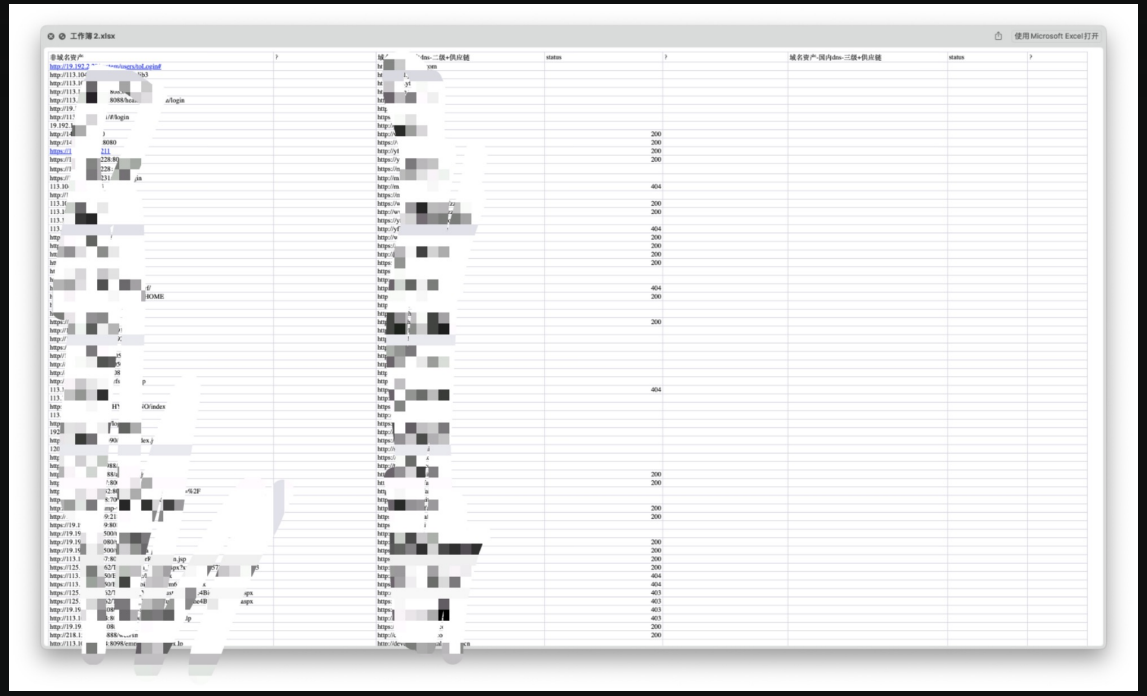
2.跟读首页文件

3.根据功能点定向审计

4.增量数据集回溯训练

AI纪元:如何让你的AI迅速成为蓝军突破主力

某地市要命的百万级级资产



模块化的半自动AI渗透测试

Environment Analyses	intranet side environment analyses such as : BloodHound
Information collection	Internet side
	Routine information collection: AP, subdomains, ip, index url, Sensitive url, Dynamic url
	Social information collection: email file information, chat tool, mail, AP, github, hvv information, media
	Intranet side: secret file, mail, AP, email file information

Lethal scan module	Internet side
	weak password Scan
	Tamplate scan Swarm tactics -- Penetration tests can be performed on any device Iday&nday Timely Internet update security-lab--0day Product core support
	Intranet side such as Fscan
AIModule	GNN-bypassWaf

email fishing	Phishing emails are automatically generated and sent
online fishing	AI-NLP such as :chat-gpt

Main function:AI-BigData Analyses

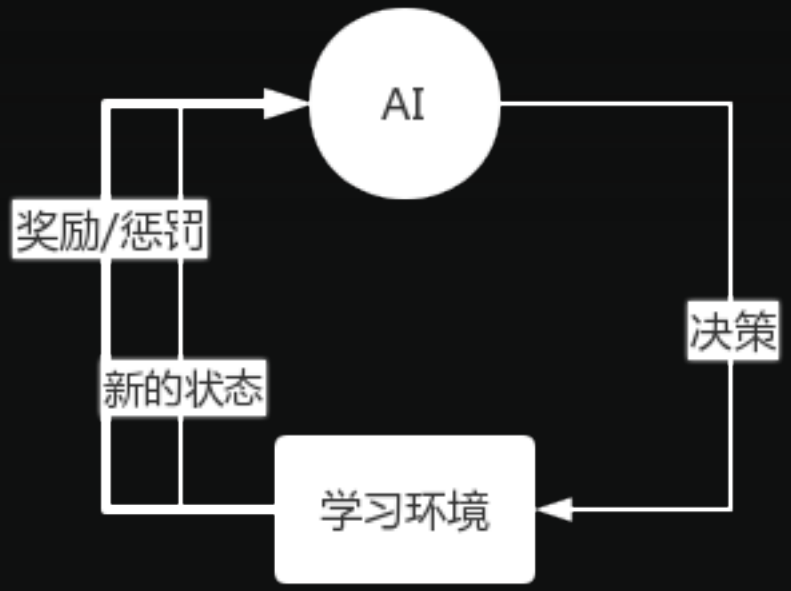
Internet side

Routine information collection	AP
	subdomains
	ip
	index url
	Sensitive url
	Dynamic url
Social information collection	email file information
	chat tool
	mail
	AP
	github
	hvv information
	media

Intranet side

secret file	
mail	
AP	
email file information	

如何解决模型的搜索效率问题：强化学习



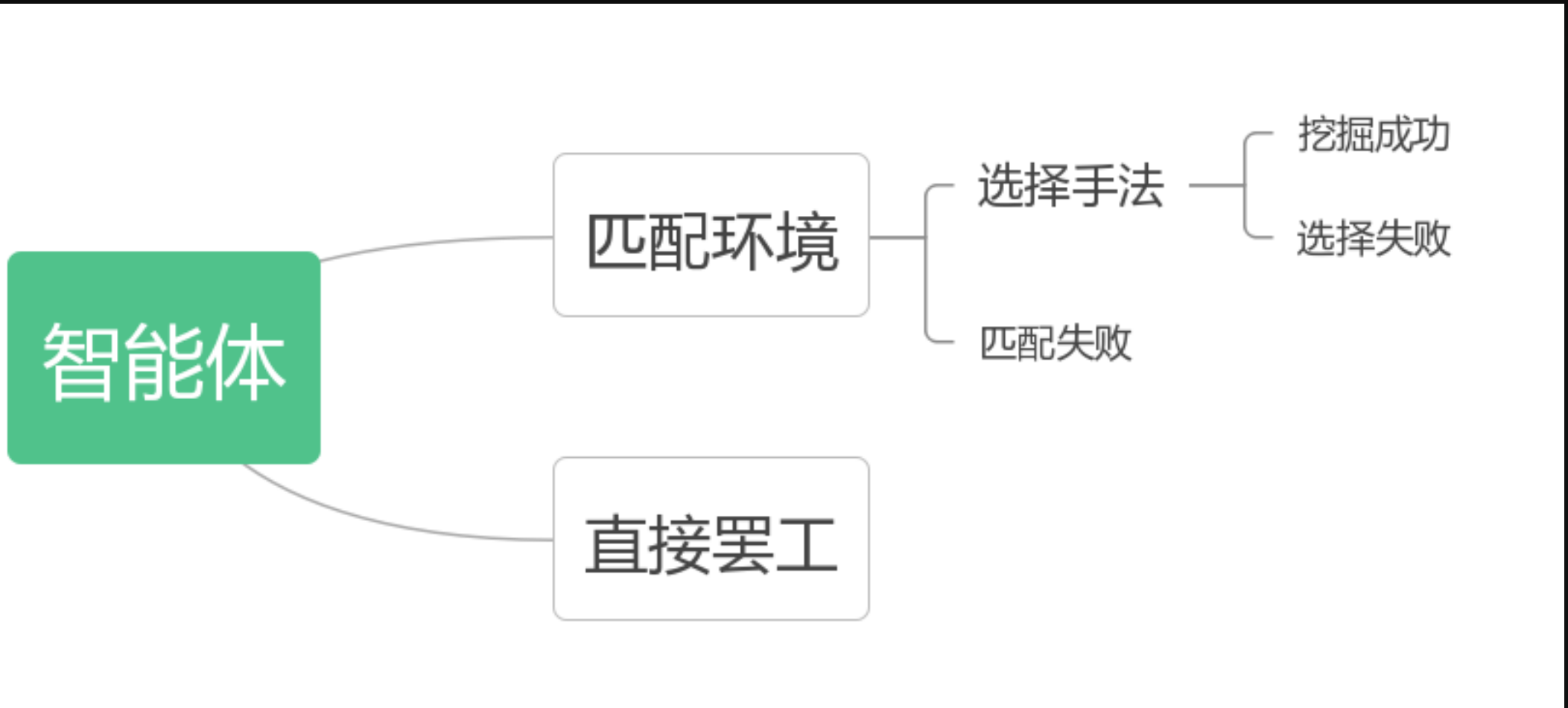
目标:提高构建的AI模型的挖掘效率

奖惩标准:从开始匹配环境到报告挖掘完成的时间长短

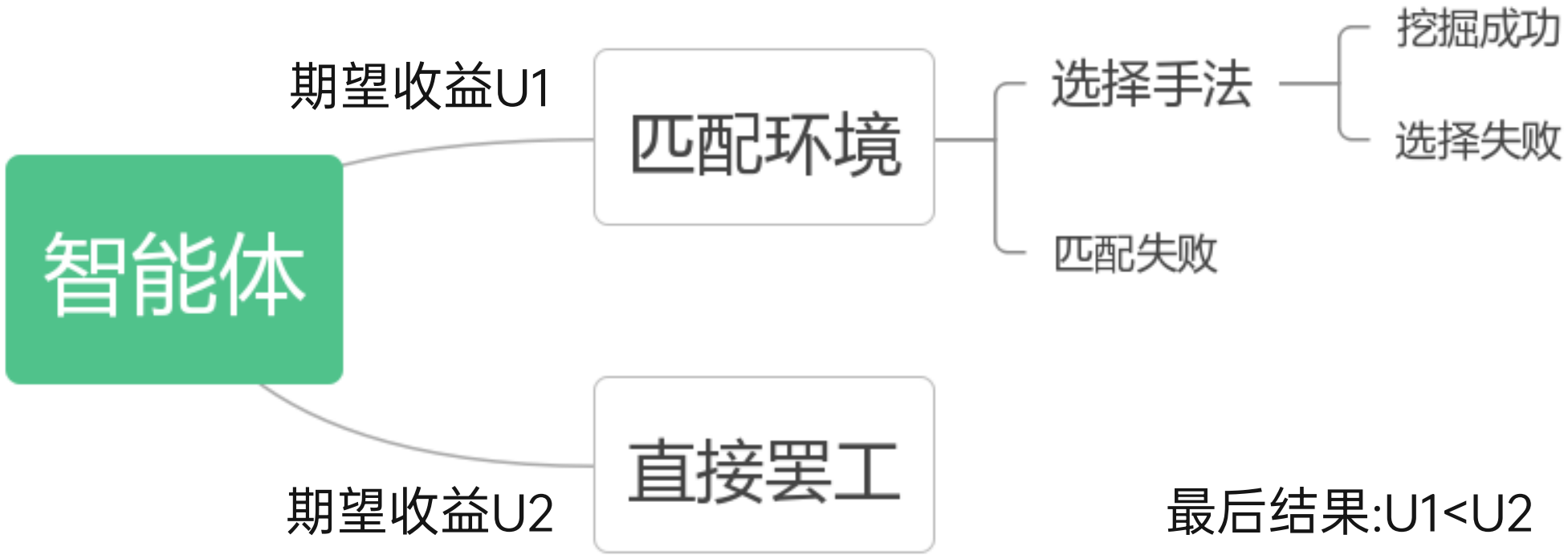
创造合适的奖励函数,使得用时越短,奖励越多

基于价值的强化学习模型

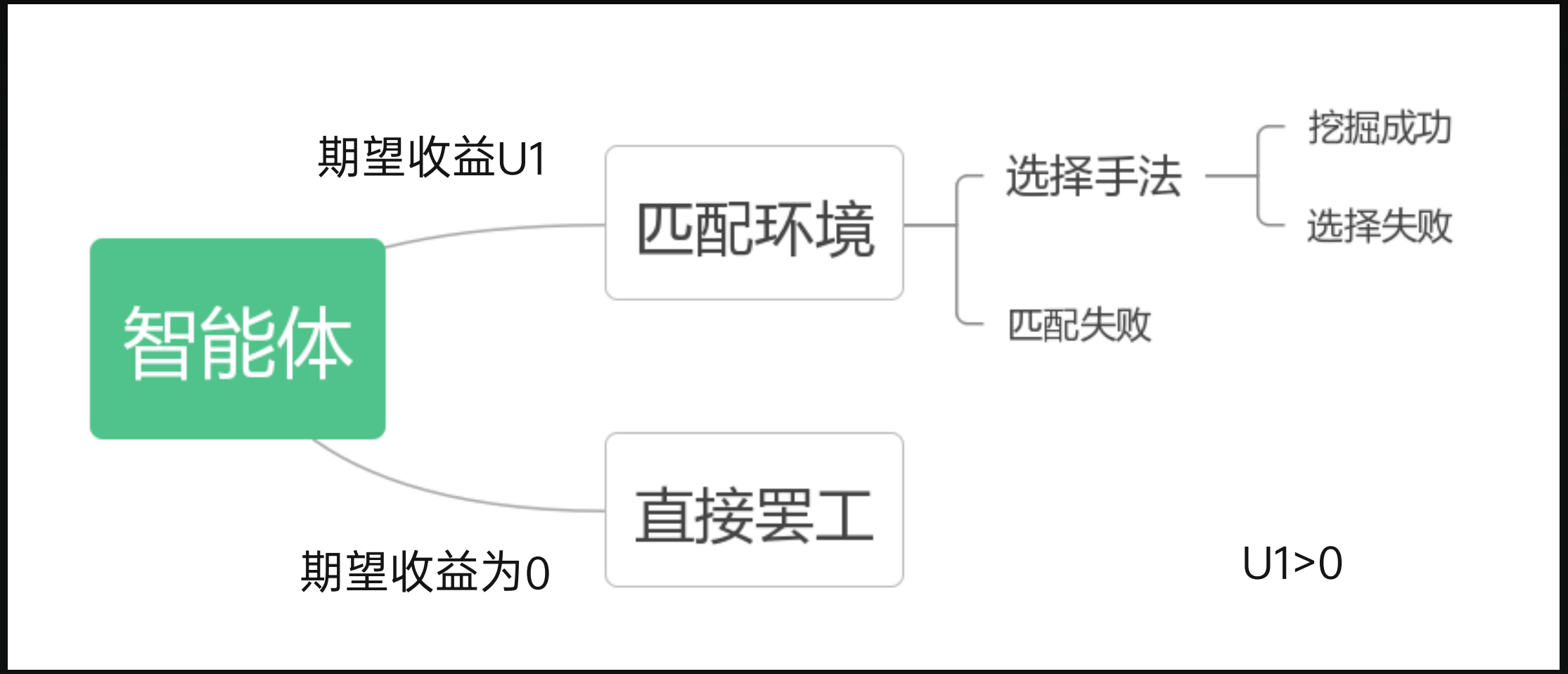
基于博弈论分析强化学习是否应当设置惩罚值



设置惩罚值

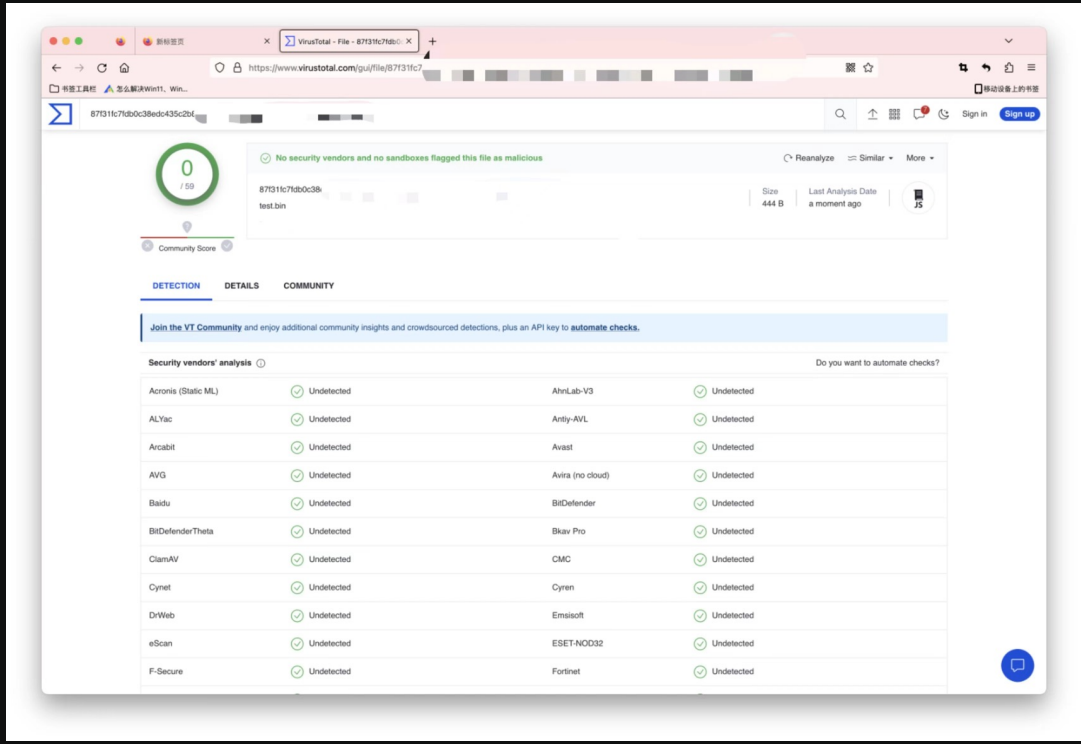


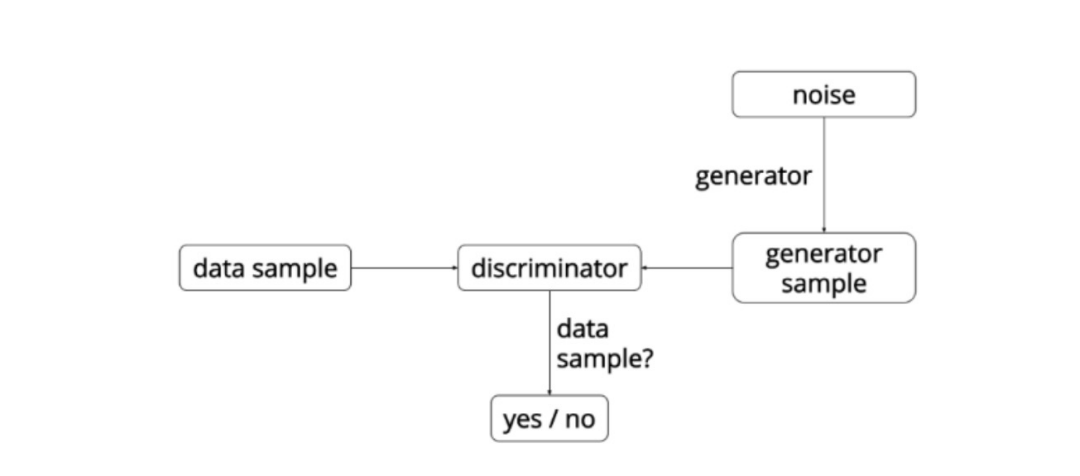
不设置惩罚值



AI纪元:如何让你的AI深入敌后

基于生成对抗的半自动化AV对抗

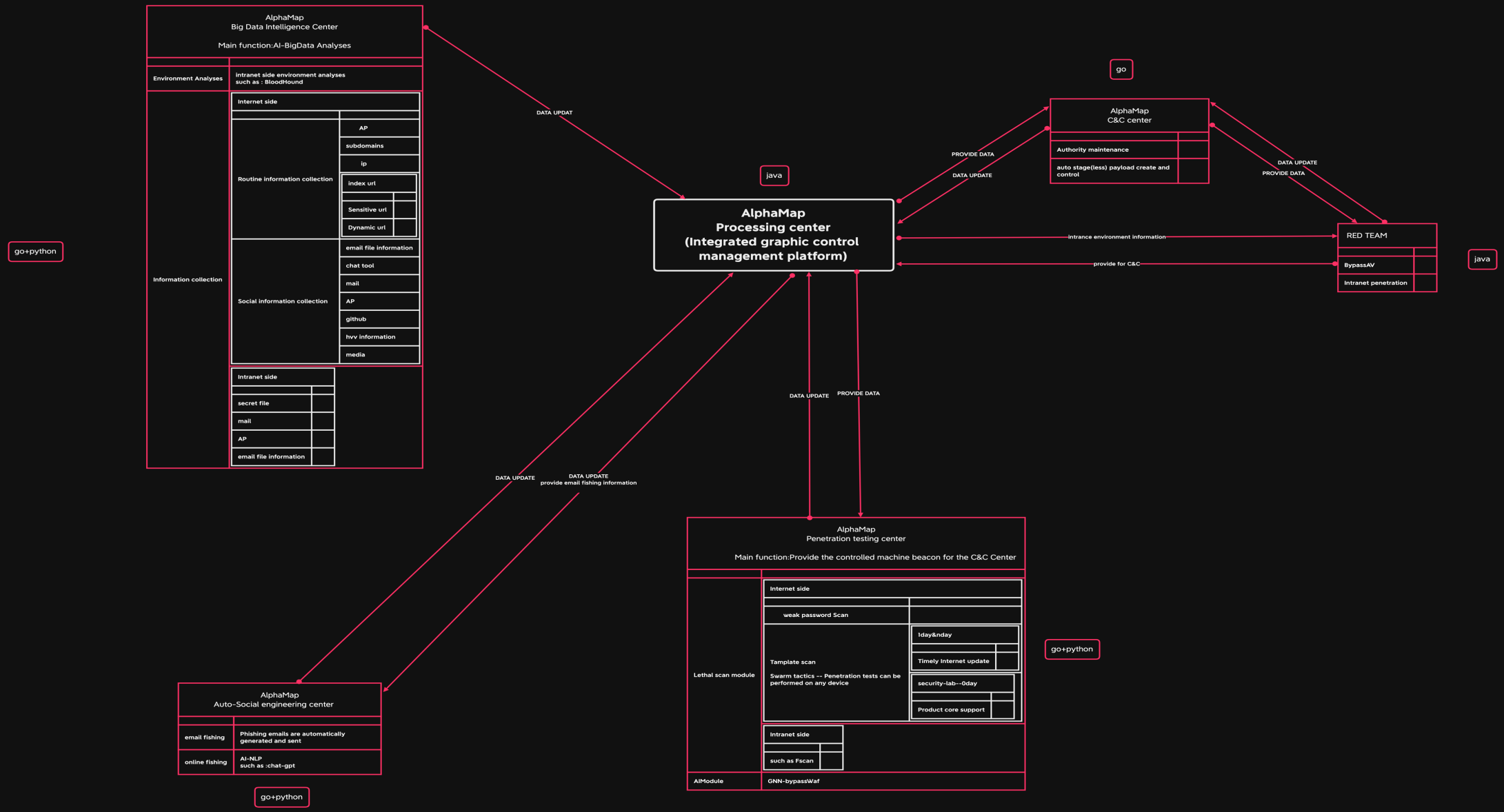




$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))].$$

•GANs是一种以半监督方式训练分类器的方法, 在没有很多带标签的训练集的时候,可以不做任何修改的直接使用代码
•G的参数更新不是直接来自数据样本,而是使用来自D的反向传播
•理论上,只要是可微分函数都可以用于构建D和G,因为能够与深度神经网络结合做深度生成式模型
•GANs可以比完全明显的信念网络(NADE,PixelRNN,WaveNet等)更快的产生样本,因为它不需要在采样序列生成不同的数据.
•模型只用到了反向传播,而不需要马尔科夫链
•相比于变分自编码器, GANs没有引入任何决定性偏置(deterministic bias),变分方法引入决定性偏置,因为他们优化对数似然的下界,而不是似然度本身,这看起来导致了VAEs生成的实例比GANs更模糊.
•相比非线性ICA(NICE, Real NVE等,),GANs不要求生成器输入的潜在变量有任何特定的维度或者要求生成器是可逆的.
•相比玻尔兹曼机和GSNs,GANs生成实例的过程只需要模型运行一次,而不是以马尔科夫链的形式迭代很多次.

AI纪元:AlphaMap-AI coordinated blue Army converged combat platform


AlphaMap
Big Data Intelligence Center
Main function:AI-BigData Analyses
Environment Analyses
Intranet side environment analyses such as : BloodHound
Information collection
Internet side
Routine information collection
AP
subdomains
ip
index url
Sensitive url
Dynamic url
Social information collection
email file information
chat tool
mail
AP
github
hvv information
media
Intranet side
secret file
mail
AP
email file information
go+python
DATA UPDAT
java
AlphaMap
Processing center
(Integrated graphic control management platform)
go
AlphaMap
C&C center
Authority maintenance
auto stage(less) payload create and control
PROVIDE DATA
DATA UPDATE
DATA UPDATE
PROVIDE DATA
intrance environment information
provide for C&C
RED TEAM
BypassAV
Intranet penetration
java
DATA UPDATE
PROVIDE DATA
DATA UPDATE
DATA UPDATE
provide email fishing information
AlphaMap
Auto-Social engineering center
email fishing
Phishing emails are automatically generated and sent
online fishing
AI-NLP
such as :chat-gpt
go+python
AlphaMap
Penetration testing center
Main function:Provide the controlled machine beacon for the C&C Center
Lethal scan module
Internet side
weak password Scan
Tamplate scan
Swarm tactics -- Penetration tests can be performed on any device
1day&nday
Timely Internet update
security-lab--0day
Product core support
Intranet side
such as Fscan
AIModule
GNN-bypassWaf
go+python

致谢

- 寻云安全团队 白羽 沉欢
- 火山信安高级攻防实验室
- 英国华威大学（网络安全中心）
  University of Warwick （BSC）

KCon

归源·智变

Returning to the source · Intelligent transformation

感谢您的观看！

THANK YOU FOR YOUR WATCHING